LA-UR -91-2000

LA-UR--91-2000

DE91 014075

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

RECEIVED BY OSTI

JUN 17 1991

TITLE HUMAN RETROVIRUSES AND AIDS, 1991

AUTHOR(S) GERALD MYERS
BETTE KORBER
JAY A. BERZOFSKY
RANDALL F. SMITH
GEORGE N. PAVLAKIS

SUBMITTED TO BOOK: to be published by Theoretical Biology and Biophysics
Los Alamos National Laboratory

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy

MASTER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

FORM NO. 836 R4
ST. NO. 2629 5/81

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

HUMAN RETROVIRUSES and AIDS 1991

A COMPILATION AND ANALYSIS OF NUCLEIC ACID AND AMINO ACID SEQUENCES

Edited by

Gerald Myers
Theoretical Division
Los Alamos National Laboratory

Bette Korber
Theoretical Division
Los Alamos National Laboratory

Jay A. Berzofsky
National Cancer Institute

Randall F. Smith
MBCRR/DFCI, Harvard University

George N. Pavlakis
Frederick Cancer Research Facility
National Cancer Institute

Database and Analysis Staff

April Gifford, Jeffrey Lawrence, Rhoshel Lenroot, Kersti MacInnes
Theoretical Division
Los Alamos National Laboratory

This Publication is being funded by the Division of AIDS, National Institute of Allergy and Infectious Diseases, through an interagency agreement with the U. S. Department of Energy.

Published by Theoretical Biology and Biophysics
Group T-10, Mail Stop K710
Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA

CONTENTS

More detailed tables of contents are to be found within the various parts of the compendium. The following provides merely an overview.

ACKNOWLEDGMENTS ii

INTRODUCTION iii

GLOSSARY vi

PART I. HIV and SIV NUCLEIC ACID SEQUENCES

A. HIV-1 Sequences

Introduction and Contents I-A-1

Nucleotide Alignments and Consensus Sequences 3

Entries 70

B. HIV-2/SIV Sequences

Nucleotide Alignments and Consensus Sequences I-B-1

Entries 58

C. SIVAGM and SIVMND Sequences

Nucleotide Alignments and Consensus Sequences I-C-1

Entries 22

PART II. HIV and SIV AMINO ACID SEQUENCES

Introduction and Contents II-1

Alignments and Consensus Sequences/Pattern Catalogs 5

PART III. ANALYSIS

Introduction and Contents III-1

VPT Coding Sequences 2

Protein Information Summary 4

Analysis of the HIV-1 V3 Loop 6

Gregory HIV-2 and SIV gp120 models (update) 23

Sequence Distances and Phylogenetic Trees 25

PART IV. RELATED SEQUENCES

Introduction and Contents IV-1

Entries 2

PART V. DATABASE COMMUNICATIONS

Introduction V-1

Floppy Diskettes

ACKNOWLEDGMENTS

We thank our former editors Flossie Wong-Staal, Temple Smith and Arnold Rabson for their advice and contributions to the database project since 1986. We will also miss the services of Jeff Lawrence, a graduate research assistant who has worked on the project since 1988.

The computer-generated models of the HIV-1 principal neutralization determinant presented on the cover (also known as the V3 loop) were kindly provided by Goutam Gupta, Theoretical Division, Los Alamos National Laboratory. Chuck Calef, of Information Services, Los Alamos National Laboratory, is responsible for the typesetting of this edition.

As ever, we are grateful to the Vaccine Development Branch of the NIAID/NIH and to the Los Alamos National Laboratory for their support and guidance.

INTRODUCTION

This compendium and the accompanying floppy diskettes are the result of an effort to compile and rapidly publish all relevant molecular data concerning the human immunodeficiency viruses (HIV) and related retroviruses. The scope of the compendium and database is best summarized by the five parts that it comprises: (I) HIV and SIV Nucleotide Sequences; (II) Amino Acid Sequences; (III) Analyses; (IV) Related Sequences; and (V) Database Communications. Information within all the parts is updated at least twice in each year, which accounts for the modes of binding and pagination in the compendium.

While this publication could take the form of a review or sequence monograph, it is not so conceived. Instead, the literature from which the database is derived has simply been summarized and some elementary computational analyses have been performed upon the data. Interpretation and commentary have been avoided insofar as possible so that the reader can form his or her own judgments concerning the complex information. In addition to the general descriptions below of the parts of the compendium, the user should read the individual introductions for each part.

Part I. HIV and SIV Nucleotide Sequences. Annotated nucleic acid sequences of HIV and SIV are presented in a form close to that of the GenBank Sequence Library. Our few modifications of standard GenBank format were instituted to better serve the particular community for which this database is intended.

The LOCUS name or identifier of an entry may differ slightly from that found in the GenBank or EMBL libraries, but the ACCESSION numbers are identical for entries in all three databases. Thus each entry is universally and uniquely traceable. The SOURCE line provides some information about the infectivity or biological activity of the molecular clone from which a sequence has been derived. REFERENCES are limited to literature or personal communications having authority for the original sequence data; references that review sequence information, or that shed light light upon the function or variation of coding and regulatory sequences, are now listed in Part V.

Entries in Part I are annotated within the sequence, while their GenBank or EMBL-formatted versions on the floppy diskettes make use of FEATURES tables. The hard-copy annotation includes coding regions, regulatory structures, splice sites, and other features of functional significance. The authority for this annotation is largely invariance, the recurrence of patterns such as TATA and AATAAA. Although our practice has been to conservatively annotate, we caution the user against docility: sequence information regarding transcripts, for example, is far from satisfactory or complete at this time.

Previous releases of the database included reference sequences, or consensus sequences to replace those. Because many researchers find the nucleotide sequence alignments helpful, we are now electing to publish those in Part I. For further information about these alignments and their consensus sequences, refer to the introduction to Part I.

Part II. Amino Acid Sequences. This section contains in alignment the amino acid sequences of all known coding regions and open reading frames of HIV-1, HIV-2 and SIV's. Consensus sequences and AACC consensus-like patterns are given with their respective alignments. Protein processing sites are annotated when known. Again, the reader should consult the introduction to Part II for further explanation of the presentation and annotation of the amino acid sequences.

Part III. Analysis. This section is open-ended with the constraint that the sequence analyses and compilations be basic and of interest to the diversity of users. Currently, the analyses include new data pertaining to coding regions and transcripts not heretofore reported; summaries of protein variability; and molecular evolutionary patterns of the various regions of the HIV and SIV genomes. A large survey of protein similarities and homologies is planned for the summer release in 1990.

Part IV. Related Sequences. With the exception of the coding sequence for the human T4 surface receptor, this section contains nucleic acid sequences of lentiviruses and other retroviruses related to HIV and SIV. Entries are presented in the same format as Part I entries.

Part V. Database Communications. This part consists of a printed supplemental reference list and diskettes. The supplemental reference list contains so-called "secondary references" that review sequence information and shed light upon the function or variation of coding and regulatory sequences. With this edition, the list has been expanded to include literature pertaining to the immunology of HIV. We welcome user input to help ensure the completeness of this list.

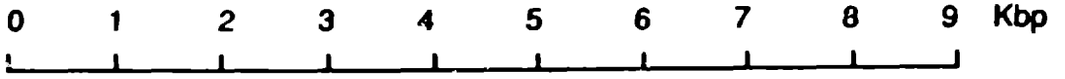
The floppy diskettes contain the nucleic acid sequences from Part I and Part IV and the translated amino acid sequences of coding regions. For the most current information regarding database files, see the READ.ME file on each diskette. Nucleotide entries are presented in GenBank format for North American users and in EMBL format for European users (unless otherwise requested). Similarly, amino acid sequences are in either PIR or Swiss-Prot format. The diskettes themselves are either 5.25" IBM-DOS format or 3.5" Macintosh format, depending upon what has been requested. If there is any trouble using these files with software designed to work with the format we have sent, please let us know the name of the program you are using and the file that it could not handle.

We are prepared to quickly enter both protein and nucleotide sequences into this database and, in the case of nucleotide sequences, oversee their entry into the GenBank and EMBL libraries. Submission of unpublished sequences is invited and encouraged. Sequence data or inquiries regarding the database should be addressed to

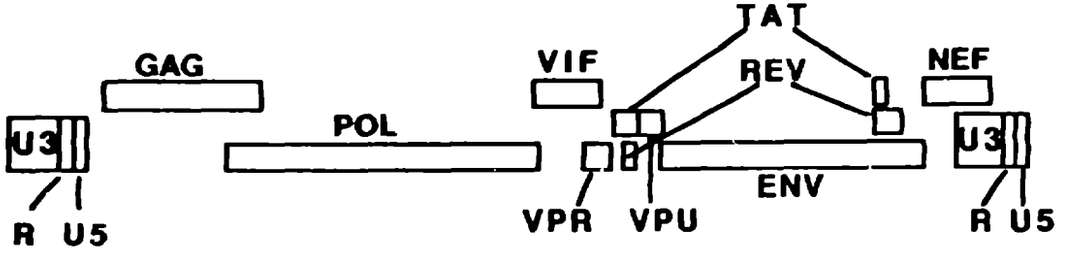
Gerald Myers
Theoretical Division
T-10, MS K710
LANL
Los Alamos, NM 87545

(505)-665-0480; fax (505)-665-3493
e-mail: glm%life@lanl.gov

A short glossary follows.

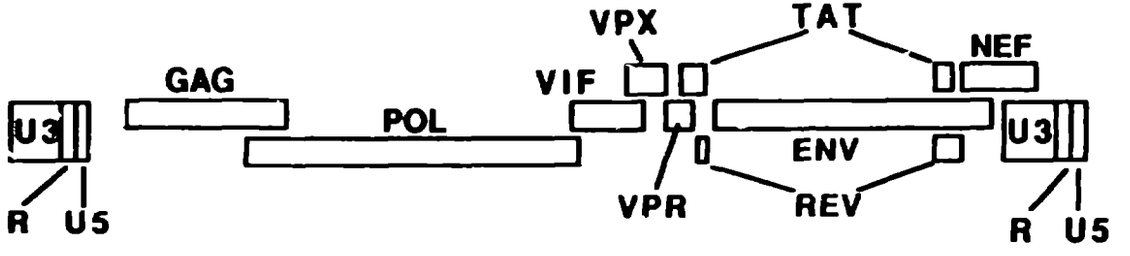


HIV-1



APR 90

HIV-2



GLOSSARY
HIV/SIV PROTEINS

Name	Typical Size	Function	Localization	Immunogenicity
GAG	p55/p17/p24/p7	structural capsid proteins	virion	H
POL	p66/p51/p32	viral enzymes	virion	H
ENV	gp120/gp41	external viral glycoproteins	plasma membrane, virion envelope	H
TAT	p16/p14	viral transactivator	primarily in nucleolus/nucleus	L
REV	p19	RNA transport, stability and utilization factor (phosphoprotein)	primarily in nucleolus/nucleus	L
VIF	p23	promotes virion infectivity	Golgi?	L
VPR	p10-15	unknown, a virion protein also called 'rap' (rapid)	virion	M
VPU	p16	promotes extra-cellular release of viral particles (phosphoprotein)	integral membrane protein only found in HIV-1	M
NEF	p27-p25	unknown (myristoylated protein)	cytoplasm, plasma membrane	H
VPX	p12-16	unknown, a virion protein	virion, only found in HIV-2 and SIV	L
TEV	p28	tripartite tat-env-rev protein (found in LAV/IIIb)	?	?

HIV GENOMIC STRUCTURAL ELEMENTS

LTR – long terminal repeat, the DNA sequence flanking the genome of integrated proviruses; contains important regulatory regions, especially those for transcriptional start and polyadenylation.

TAR – target sequence for viral transactivation, the binding site for TAT protein and possibly cellular proteins; consists of approximately the first 45 nucleotides of the viral mRNAs in HIV-1 (or the first 100 nucleotides in HIV-2 and SIV). TAR forms a hairpin stem-loop structure with a side bulge; the bulge is necessary for TAT binding and function.

RRE – (also known as CAR) REV responsive element, an RNA element encoded within the ENV region of HIV-1, consisting of approximately 200 nucleotides. The RRE is necessary for REV function; approximately seven binding sites for REV exist within the RRE RNA. Other lentiviruses (HIV-2, SIV, visna) have similar RRE elements in similar locations within ENV, and HTLVs have an analogous RNA element (RXRE) serving the same purpose within their LTR; RRE is the binding element for REV protein, and RXRE is the binding element for the Rex protein. RRE and RXRE thus form complex secondary structures important for specific protein binding.

CRS – cis-acting repressive sequences postulated to inhibit structural protein expression in the absence of REV. Their exact locations and function have not been defined; splice sites have been postulated as CRS sequences.

INS – "instability" RNA sequences found within the structural genes of HIV-1 and of other complex retroviruses. Multiple elements appear to exist within the genome; one of the best characterized elements spans nucleotides 414 to 631 in the GAG region of HIV-1.

GENES AND GENE PRODUCTS

GAG – group-specific antigens or capsid proteins; the precursor is the p55 myristoylated protein, which is processed to p17 (MAtrix) p24 (CApsid) and p7 (NucleoCapsid) proteins by the viral protease. Other small proteins are generated from the gag polyprotein.

POL – generates the viral enzymes protease, reverse transcriptase and integrase after the processing of a GAG-POL precursor polyprotein by the viral protease; GAG-POL precursor is produced by ribosome frameshifting.

ENV – viral glycoproteins produced as a precursor (gp160) and processed to the external glycoprotein gp120 and the transmembrane glycoprotein gp41. The mature proteins are held together by noncovalent interactions; as a result, a substantial amount of gp120 is released in the medium. Gp120 contains the binding site for the CD4 receptor.

TAT – transactivator of HIV gene expression; one of the two necessary viral regulatory factors (TAT and REV) for HIV gene expression. Two forms are known, TAT-1exon (minor form) of 72 amino acids, and TAT-2exon (major form) of 86 amino acids. The electrophoretic mobility of these two forms in SDS gels is anomalous; they are approximately 16 kD and 14 kD in weight. Low levels of both proteins are found in persistently infected cells. TAT is localized primarily in the nucleolus/nucleus; it acts by binding to the TAR RNA element and activating transcription from the LTR promoter. Post-transcriptional effects of TAT have been postulated.

REV – the second necessary regulatory factor for HIV expression. A 19kD phosphoprotein localized primarily in the nucleolus/nucleus, REV acts by binding to RRE and promoting the nuclear export, stabilization and utilization of the viral mRNAs containing RRE.

VIF – viral infectivity factor, typically 23 kD; required for the efficient transmission of cell-free virus in tissue culture. In the absence of VIF, the produced viral particles are defective, while the cell-to-cell transmission of virus is not affected significantly. It has been reported that the cellular localization is in the Golgi (VIF is not found in the virion).

VPR – virion-associated protein of unknown function found in HIV-1, HIV-2, SIVmac, and SIVmnd; typically 15kD. May be homologous to VPX. Also called 'RAP' for rapid.

VPU – "oUt" protein that promotes extracellular release of viral particles. Found only in HIV-1. Integral membrane phosphoprotein of 16kd; similar to M2 protein of influenza virus. It may be involved in ENV maturation. It is not found in the virion.

NEF – approximately 27 kD non-virion protein found in the cytoplasm of infected cells. Potentially myristoylated and associated with the inner plasma membrane. One of the first HIV proteins to be produced in the infected cells, it is the most immunogenic of the accessory proteins and may be used in the future for diagnosis and staging of the disease. NEF is dispensable and probably suffers counter-selection during ex vivo viral propagation. Recent evidence suggests that SIV NEF is required for viral propagation in vivo.

VPX – virion protein of 12 kD found only in HIV-2/SIV and SIVAGM; not in HIV-1 or SIVmnd. Function unknown. VPU/VPX may be used to distinguish between HIV-1 and HIV-2 infection. (VPX may have some homology with VPR.)

TEV – (also known as tnv) a tripartite 28 kD viral phosphoprotein produced very early in infection by some HIV-1 strains. Found primarily in the nucleolus/nucleus. TEV contains the first exon of TAT, a small part of ENV and the second exon of REV. It has both TAT and REV functions in the same polypeptide and can functionally replace both essential regulatory proteins of HIV-1.

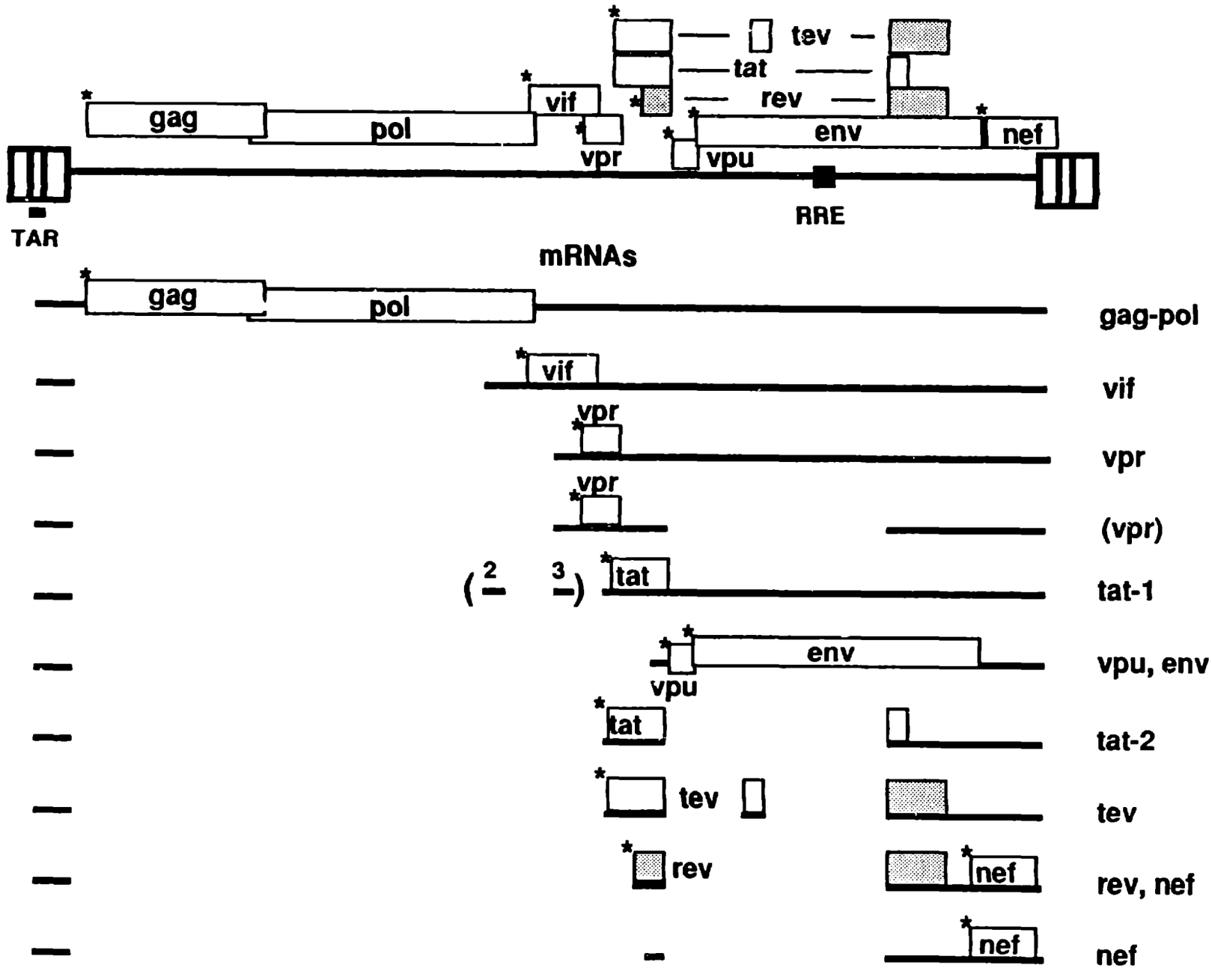
STRUCTURAL PROTEINS – the products of GAG, POL and ENV genes, which are essential components of the retroviral particle.

REGULATORY PROTEINS – TAT and REV proteins of HIV/SIV and TAX and REX proteins of HTLVs; essential for viral expression in infected cells.

ACCESSORY PROTEINS – additional (non-regulatory) virion- and non-virion-associated proteins produced by HIV/SIV retroviruses: VIF, VPR, VPU, VPX, and NEF. Although the accessory proteins are not necessary for viral propagation in tissue culture, they have been conserved in the different isolates; this conservation and experimental observations suggest that their role in vivo is very important.

COMPLEX RETROVIRUSES – retroviruses regulating their expression via viral factors and expressing additional proteins (regulatory and accessory) essential for their life cycle. The following figure shows the extent of this complexity as manifested in multiple transcripts and alternative forms of tat and rev proteins; for a fuller presentation of the information in this figure, see III-22 of SEP 90.

HIV-1


tev
tat
rev
vif
gag
pol
vpr
env
nef
vpu
RRE
TAR
mRNAs
gag
pol
gag-pol
vif
vif
vpr
vpr
vpr
(vpr)
(2 3)
tat
tat-1
env
vpu, env
vpu
tat
tat-2
tev
tev
rev
rev, nef
nef
nef

I

HIV and SIV Nucleic Acid Sequences

Nucleotide sequence alignments were generated using the PIMA program developed by Randy Smith and Temple Smith, MBCRR, Dana Farber Cancer Institute, Harvard School of Public Health, Boston, MA 02115, (617)-732-3746. By this approach, a protein pattern is created which then serves as a template for the nucleotide sequence alignment. Hand-editing, when required, was done with the MASE program (also provided by the MBCRR; see R. F. Smith and T. F. Smith, *PNAS* 87: 118–122 1990).

With a few exceptions, only full-length coding sequences were included in these alignments. Mixed case consensus sequences are used as the reference sequences for each alignment. Upper case letters indicate 100% conservation of nucleotide bases in a given position of the alignment, and lower case letters represent bases conserved in at least 50% of the sequences. To reduce the alignment uncertainty with the large HIV-1 set, sub-groups of North American and African sequences were constituted and consensus sequences were generated for each. The highly divergent SIV-cpz sequence (Huet *et al.*, *Nature* 345: 356–359 1990) was not included in the consensus sequences for the HIV-1 alignments. Large deletions, if present in only one sequence (i.e., the HIV2nihz nef coding region sequence), were also excluded from the consensus sequences. The consensus sequences in this edition take into account two unpublished sequences which were not presented in the alignments.

The authority for some of the annotation is limited largely to invariance—the recurrence of patterns such as AATAAA, for example. The reader should be cautious in drawing upon this information. In particular, the annotation of splice sites is tentative; there is a flurry of work going on now in this area (see the GLOSSARY and also III-21 and following pages of SEP 90).

Only new and revised sequences are reported as entries in this section. Otherwise, sequence entries from previous years are presented in the 1990 compendium and in the new GenBank format on the floppy diskettes.

PART I. HIV and SIV Nucleic Acid Sequences

Introduction . . . . . I - A - 1

Contents . . . . . I - A - 2

A. HIV-1 Sequences

Nucleotide alignments of HIV-1 sequences . . . . . I - A - 3

HIVLAI (K02013; M64186-190; M64194-198; M64203-223) . . . 70

HIVNL43 (M19921) 76

HIVJB01 (X57447) 80

HIVJB02 (X57450) 83

HIVJB03 (X57465) 89

HIVBAL (M68893) 92

HIVADA (M60472) 95

HIVSWB40 (M27951; M27946) 96

HIVSWB88 (M27952; M31462; M27947; M58206) 97

HIVM77, HIVV77, HIVT77 (Pending) 100

HIVM82, HIVV82, HIVT82 (Pending) 103

HIVNJA (M61028 - M61045) 117

HIVU455 (M62320) 122

HIVSF170 (M66535; M66533) 126

HIVCBL4 (M63474; M63475) 129

B. HIV-2 and Related SIV Sequences

Nucleotide Alignments of HIV-2 and SIV Sequences . . . . . I - B - 1

HIV2D194 (J04542; X52223) 58

C. SIVAGM and SIVMND Sequences

Nucleotide Alignments of AGM Sequences . . . . . I - C - 1

SIVAGM677 (M66437) 22

SIVMND121 (M62883; M62884) 26

Authority for the annotation is given on pages I-A-4 through I-A-9 of APR 90.

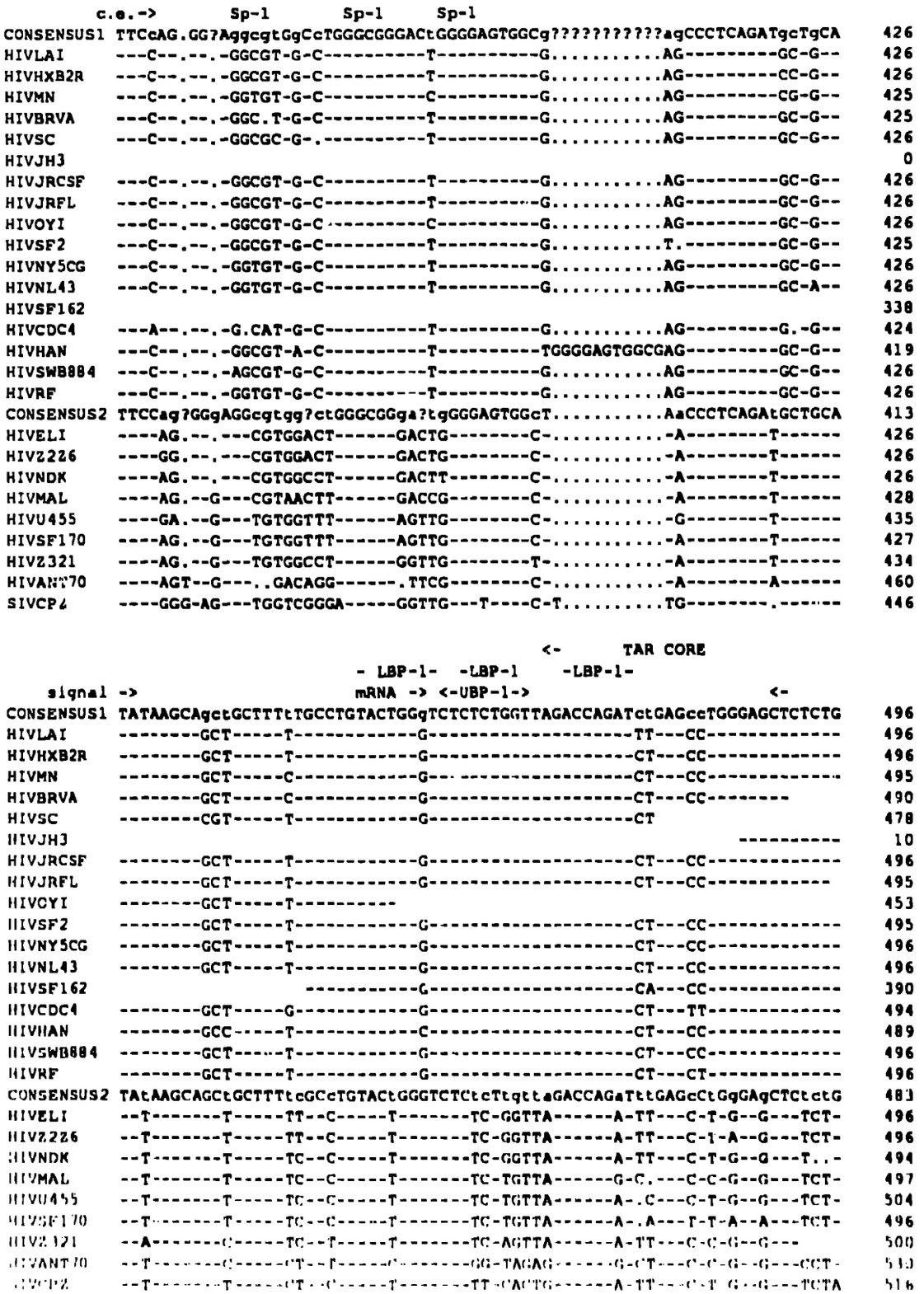
```
CONSENSUS1 TGGAaGGGcTAaTttacTCcCaaaaaAGACAaGAtATcCTTGATcTGTGGaTctAcCACACACAAGGCTA  70
HIVLAI     ----A---C--A-TCAC--C-AACGA-----A--T--C------C-----A-CT-C--------------  70
HIVHXB2R   ----A---C--A-TCAC--C-AACGA-----A--T--C------C-----A-CT-C--------------  70
HIVMN      ----T---T--A-TTAC--C-AAAAG-----A--C--C------C-----G-CT-C--------------  70
HIVBRVA    ----A---C--A-TCAC--C-AACAA-----A--T--C------T-----G-CT-T--------------  70
HIVSC      ----A---C--A-.CAC--C-AGAGA-----A--T--C------C-----A-CT-C--------------  69
HIVJH3                                                                            0
HIVBAL1    ----A---C--A-TCAC--C-AAAAA-----A--T--C------T-----G-CT-C--------------  70
HIVJRCSF   ----A---C--A-TTAC--A-AGAAA-----A--T--C------C-----A-CT-C--------------  70
HIVJRFL    ----A---C--A-TCAC--A-AGAAA-----A--T--C------C-----G-CT-C--------------  70
HIVOYI     ----A---C--A-TTAC--C-AGAAA-----A--T--T------C-----G-CT-T--------------  70
HIVSF2     ----A---C--A-TTGG--C-AAAGA-----A--G--C------C-----A-CT-C--------------  70
HIVNY5CG   ----A---C--A-TTGG--C-AAAGA-----A--T--C------C-----A-CT-C--------------  70
HIVNL43    ----A---C--A-TTGG--C-AAAAA-----A--G--C------C-----A-CT-C--------------  70
HIVSF162   ----A---T--A-TTAC--C-AAAAA-----A--T--C------C-----A-CC-C--------------  70
HIVCDC4    ----A---C--G-TTAC--C-AACAA-----A--T--C------C-----G-CT-C--------------  70
HIVSF33    ----A---C--G-TTAC--C-AAAAA-----A--T--C------C-----A-CT-C--------------  70
HIVHAN     ----A---T--A-TTAC--C-CAAAA-----A--G--C------C-----G-CT-C--------------  70
HIVSwB084  ----A---C--G-TCAC--C-AAAAA-----A--G--C------C-----G-CT-C--------------  70
HIVRF      ----T---C--G-GTTC--C-AGAAA-----A--T--C------C-----G-TT-C--------------  70
CONSENSUS2 TGGA?GGG?TAaTTt??TCcaagAaaAGAcaAGAaATCCTtGAtCT?TGGGTcTAcaAcACaCAaGGctt  65
HIVELI     ----A---C--A--TGG--CAAA-AG---CA---G-----T--T--T-----C--CA-C--A--A--CAT  70
HIVZ2Z6    ----A---C--A--TGG--AAAA-GA---CA---C-----T--T--T-----C--CA-C--A--A--CAT  70
HIVNDK     ----A---C--A--TGG--CAAG-AA---CA---G-----T--T--T-----C--CA-C--A--A--CAT  70
HIVMAL     ----T---T--G--TGG--CCCA-AA---CA---A-----T--T--G-----C--CC-C--A--A--CTA  70
HIVU455    ----T---T--A--CAC--ACAG-AA---CA---A-----T--T--G-----C--TC-C--A--A--ATT  70
HIVSF170   ----T---T--A--TAC--CAGG-GA---CA---A-----T--T--G-----C--CA-C--A--A--CTA  70
HIVZ321    ----T---C--A--TAC--CAAG-AA---CA---A-----T--C--G-----C--TC-T--A--A--CTT  70
HIVANT70   ----A---T--A--TAC--CCAT-AA---GC---A-----G--T--T-----G--TA-C--T--G--ATT  70
SIVCPZ     ----A---T--G--TAC--CAGG-GA---CA---G-----T--C--C-----C--TC-C--A--A--CTT  70

                                                 - AP-1 -      - AP-1 -
CONSENSUS1 cTTCCCTGAtTgGCAGAAcTACACAcCAGGgCCAgGgatCAGATatCC?AcTGaccttTGGATGGTGCTt 139
HIVLAI     C---------T-G------C------C----G---G-GGT-----AT--.-C--ACCTT----------A 139
HIVHXB2R   C---------T-A------C------C----G---G-GAT-----AT--.-C--ACCTT----------A 139
HIVMN      C---------T-G------C------C----G---G-GAT-----AT--.-C--ACCTT----------T 139
HIVBRVA    C---------T-G------C------C----G---G-GGT-----AT--.-C--ACCTT----------T 139
HIVSC      C---------T-G------C------C----G---G-GAT-----AT--C-C--TGTTT----------T 139
HIVJH3                                                                            0
HIVBAL1    C---------T-G------C------C----G---G-GAC-----TT--.-C--ACCTT----------T 139
HIVJRCSF   C---------T-G------C------G----A---G-GGT-----TT--.-C--ACCTT----------T 139
HIVJRFL    C---------T-G------C------C----G---G-AAT-----TT--.-T--ACCTT----------T 139
HIVOYI     T---------T-G------C------C----G---G-GAT-----AT--.-C--TGCTT----------T 139
HIVSF2     C---------T-G------T------C----G---G-GAT-----AT--.-C--ACCTT----------T 139
HIVNY5CG   C---------T-G------C------C----G---G-GAT-----AT--.-C--ACCTT----------T 139
HIVNL43    C---------T-G------C------C----G---G-GAT-----AT--.-C--ACCTT----------T 139
HIVSF162   C---------T-G------C------C----G---G-GAT-----AT--.-C--ACCTT----------T 139
HIVCDC4    C---------T-G------C------C----G---A-GAC-----TC--.-C--ACCTT----------T 139
HIVSF33    C---------T-G------C------C----G---G-GGT-----TT--.-C--ACCTT----------T 139
HIVHAN     C---------T-G------C------C----G---G-GGT-----AT--.-C--ACCTT----------T 139
HIVSWB084  C---------T-G------C------C----G---G-GAC-----GG--.-C--ACCCC----------T 139
HIVRF      C---------T-G------C------C----G---G-GAC-----AT--.-C--ACCTT----------T 139
CONSENSUS2 CTTCCCtGATTGGCAgAAcTACACACCAGGgCCAGGqAtcAGaTacCC.ACTqaccTTtGGATGGTqcTt 134
HIVELI     ------T--------A--C----------G-----G-TC--A-AT--.---AACC--T-------GC-A 139
HIVZ2Z6    ------T--------G--C----------G-----G-TC--A-AT--.---GACC--T-------GC-T 139
HIVNDK     ------C--------G--C----------G-----G-TC--A-AC--.---GACC--T-------GC-T 139
HIVMAL     ------T--------G--T----------G-----G-TT--A-TC--.---GACC--C-------GC-T 139
HIVU455    ------T--------G--T----------G-----A-TC--A-AC--.---AACA--T-------GC-A 139
HIVSF170   ------T--------G--T----------G-----G-CC--A-TC--.---AACA--T-------GC-T 139
HIVZ321    ------T--------G--C----------G-----G-CC--A-AC--.---GTGT--T-------GC-T 139
HIVANT70   ------T--------G--C----------A-----A-CC--G-TC--.---GACA--T-------TG-T 139
SIVCPZ     ------T--C-----G--C------A----A-----A-CA--A-TC--.---CTGT--T-------GC-T 139
```

HIV-1 LTR

```
                                - AP-1 -                                                <--
CONSENSUS1 CAAGcTAGTACCAgT...tgAgCCAGagaaggtagaagaggccaatgaAggAGAGAAcaaCaGctTGTtA    206
HIVLAI     ----C--------G-...TG-G----ATAAGGTAGAAGAGGCCAATAA-GG------CAC-A-CT---T-    206
HIVHXB2R   ----C--------G-...TG-G----AGAAGTTAGAAGAAGCCAACAA-GG------CAC-A-CT---T-    206
HIVMN      ----C--------G-...TG-G----AGAAGATAGAAGAGGCCAATAA-GG------CAA-T-CT---T-    206
HIVBRVA    ----C--------G-...TG-G----AGAAGATAGAAGAGGCCAATGA-GG------CAA-A-TT---T-    206
HIVSC      ----C--------G-...TA-A----AGAAGATAGAAGAAGCCAATGA-GG------CAA-A-CT---T-    206
HIVJH3                                                                                0
HIVBAL1    ----C--------G-...TG-G----AGAAGGTAGAAGAGGCCAATGA-GG------CAA-T-CT---C-    206
HIVJRCSF   ----C--------G-...TG-T----AGAAGGTAGAAGAGGCCAATGA-GG------CAA-T-CT---T-    206
HIVJRFL    ----C--------G-...TG-G----AAAAGGTAGAAGAGGCCAATGA-GG------CAA-T-CT---T-    206
HIVOYI     ----T--------A-...GG-T----ACCAGGTAGAAGAGGCCAACGA-GG------CAA-A-CT---T-    206
HIVSF2     ----C--------G-...TG-G----AGAAGGTAGAAGAGGCCAATGA-GG------CAA-A-CT---T-    206
HIVNY5CG   ----T--------G-...TG-G----GGCAGGTAGAAGAGGCCAATGA-GG------CAA-A-CT---T-    206
HIVNL43    ----T--------G-...TG-A----AGCAAGTAGAAGAGGCCAAATA-GG------GAA-A-CT---T-    206
HIVSF162   ----C--------G-...TG-T----ATTATGTAGAAGAGGCCAATGC-GG------CAA-A-CT---T-    206
HIVCDC4    ----C--------C-...TG-G----AGCAGGTAGAAGAGGCCAACGA-GG------CAA-A-CT---T-    206
HIVSF33    ----T--------G-...AG-G----AGAAAGTAGAAGAGGCCAATGA-GG------CAA-A-CT---T-    206
HIVHAN     ----T--------G-...TG-A----A.................TGA-GA------CAG-A-CC---T-    188
HIVSWB884  ----C--------G-...CG-G----AGAAGGTAGAAGAGGCCAATGA-GG------CAA-A-CT---T-    206
HIVRF      ----C--------G-...TG-G----ATAAGGTAGAAGAGGCCACTGA-GG------CAA-A-CT---T-    206
CONSENSUS2 caAgcTAGTACCAgT???tgatccAgagGaaGtAga?ga?GccActgaagGaGAGaacaac?gccTgtTa    198
HIVELI     CG-GC--------G-...TGATCC-CAG-AG-T-GAAGAA-AC-CTGAAG-A---ACCAACAGCT-GT-A    206
HIVZ2Z6    CG-GC--------G-...TGATCC-CGG-AG-T-GAAGAG-CC-CTGAAG-A---ACCAACTGCT-GT-A    206
HIVNDK     CC-GC--------G-...TGATCC-CAG-AG-T-GAAGAG-CC-CTGAAA-A---GACAACTGCT-GT-A    206
HIVMAL     TA-GT--------A-...GAGTCC-GAG-AA-T-GAGGAG-CC-ATGAAG-A---AACAACTGTC-GT-A    206
HIVU455    CA-GC--------G-...TGATCC-GCT-AA-T-GAGGAA-CT-CTGGAG-A---AACAACAGTC-A--A    206
HIVSF170   CA-GC--------G-...TGATCC-GAG-GA-T-GAGAAG-AC-CGGTGG-A---ACCAACAGCC-AT-A    206
HIVZ321    CA-GC--------G-...TGATCC-AGA-AA-T-GAGGAA-CC-ATACAG-A---AACAACTGCC-GT-A    206
HIVANT70   TA-AC--------G-GTCAGAAGA-GAG-CA-A-AGACTA-GA-ATACAT-T---AGGGCTAATC-CC-G    209
SIVCP2     CA-AC----G--CC-GACAGA...GGAGCAG-T-GAACAA-CC-ATGAAG-A--TAACAACTGCC-GT-G    206

           NFAT-1                  -->
CONSENSUS1 CAcCCtatgaGCCtGcATGGgAtgGAtGACcCgGAGaaaGAAGTgtTagtgTGGAaGTtTGACAgCcgcC    276
HIVLAI     --C--TGTGA---T-C----A-TG--T---C-T---AGA-----GT-AGAG---G--T-----G-CGC-    276
HIVHXB2R   --C--TGTGA---T-C----A-TG--T---C-G---AGA-----GT-AGAG----G--T-----G-CGC-    276
HIVMN      --C--TATGA---A-C----.-TG--T---C-G---AGA-----GT-AGTG----A--C-----G-CAC-    275
HIVBRVA    --C--TATGA---A-C----A-TG--T---C-G---AGA-----GC-GCAG----G--T-----G-CGC-    276
HIVSC      --C--GATGA---T-C----G-TG--G---C-G---AGA-----GT-AGAG----G--T-----A-CGC-    276
HIVJH3                                                                                0
HIVBAL1    --C--TATGA---T-C----G-TG--T---C-G---AAA-----GT-AGCA----A--T-----G-AGC-    276
HIVJRCSF   --C--TATGA---A-C----A-TG--C---C-A---AAG-----GT-AGTG----A--T-----G-AAG-    276
HIVJRFL    --C--TATGA---A-C----G-TA--G---C-G---AAG-----GT-AGAG----G--T-----G-AAG-    276
HIVOYI     --T--TATAA---T-C----G-TG--T---C-A---AAA-----GT-AGTG----A--T-----G-CGC-    276
HIVSF2     --C--TATGA---T-C----G-TG--G---G-G---AAA-----GT-AGTG----G--T-----G-AAA-    276
HIVNY5CG   --C--TATGA---A-C----G-TG--G---C-G---GGA-----AT-AGTG----A--T-----G-CTC-    276
HIVNL43    --C--TATGA---A-C----G-TG--G---C-G---GGA-----AT-AGTG----A--T-----G-CTC-    276
HIVSF162   --C--TATGA---A-C----G-TG--T---C-G---AAA-----AT-AGTG----G--T-----G-CGC-    276
HIVCDC4    --C--TATGA---A-G----G-TG--T---C-A---AAA-----GT-AGTG----A--T-----G-CTC-    276
HIVSF33    --C--TATGA---T-C----G-TG--G---C-G---AAA-----GT-AGTG----A--T-----G-CAC-    276
HIVHAN     --C--TGCGA---T-C----G-CA--G---A-G---AGA-----GT-AAAG----A--T-----G-CAT-    258
HIVSWB884  --C--TATGA---T-C----A-TG--G---C-G---AAA-----GT-AGTG----A--T-----G-AGC-    276
HIVRF      --C--TATAT---T-C----G-TG--T---C-A---AAA-----GT-AGTG----A--T-----G-CGC-    276
CONSENSUS2 C cCCtatatGccagCAtGGAaTgGAgGA?gcgqAgAgAgAAqtgtTaAtgTGGAagT1TgAcAGcagcC    267
HIVELI     --C--TATAT-CCAG--T---A-G--G--CCCGG-G-G-C--GTGT-A-AA----GA---A-C--CAGA-    276
HIVZ2Z6    --C--TGTAT-CCAG--T---A-G--G--CACGG-G-G-G--GTGT-A-AG----GA---A-C--CAGA-    276
HIVNDK     --C--TATGT-CCAG--A---A-G--G--CCCGG-G-G-C--GTGT-A-TG----GA---A-C--CAGA-    276
HIVMAL     --C--TATTA-CCAA--T---A-G--G--CGCAG-A-G-G--GTGC-A-AA----AG---G-C--CAGC-    276
HIVU455    --C--TATAT-CCAA--T---G-A--T--TGAAG-G-A-G--GTGT-A-TG----AG---G-C--TACC-    276
HIVSF170   --C--TATAT-CCAA--T---A-G--T--TGAGG-A-G-G--ACAT-A-TG----AG---G-C--CCGC-    276
HIVZ321    --C--CATGA-CCAG--T---A-G--T--TGATG-A-G-G--GTGT-A-TG----AG---G-C--CTCC-    276
HIVANT70   --T--AGCAT-TGCC--T---T-T--A--TACAC-T-A-G--ATAC-G-TG----AG---G-T--ATCT-    279
SIVCP2     --T--CATAT-TCAG--T---GA-G--A--TGAGG-C-A-G-GGTCT-GGTC---CGC---G-C--CAGG-    276
```

```
                  <- IL2/NRE ->
               <-    NRF      ->
CONSENSUS1 TaG.cATTtCatCAcatgGCCCGAGAgctgcATCCGGAGTacTaCAAa??.gACTG?.......CTGACA    334
HIVLAI     -A-.C---T-AT--CGTG--------GCTGC---------AC-T---G...A----.........------    334
HIVHXB2R   -A-.C---T-AT--CATG--------GCTGC---------AC-T---G...A----.........------    334
HIVMN      -A-.C---T-AG--TTAT--------GCTGC---------AC-A---G...A----.........------    333
HIVBRVA    -A-.C---T-AT--CATG--------GCTGC---------AC-A---G...A----.........------    334
HIVSC      -A-.C---T-AT--CATG--------TCTGC---------AC-A---A...G----T.......------    335
HIVJH3                                                                               0
HIVBAL1    -A-.C---C-AT--CGTG--------A                                             302
HIVJRCSF   -A-.C---G-AT--CGTG--------GCTGC---------AC-A---G...G----.........------    334
HIVJRFL    -A-.C---T-AT--CGTG--------GCTGC---------AC-A---G...G----.........------    334
HIVOYI     -A-.C---T-GT--CATG--------GGTGC---------AC-A---A...G----.........------    334
HIVSF2     -A-.C---T-AT--CATG--------GCTGC---------AC-A---A...G----.........------    334
HIVNY5CG   -A-.C---T-GT--CATG--------GCTGC---------AC-A---A...G----.........------    334
HIVNL43    -A-.C---T-GT--CATG--------GCTGC---------AC-A---A...G----.........------    334
HIVSF162   -C-.G---T-AT--CATG--------GCTGC---------AC-A---AAA.G----.........------    336
HIVCDC4    -A-.C---T-GT--CATG--------GCTGC---------AC-A---A...G----.........------    334
HIVSF33    -A-.C---T-GT--CATG--------GCTGC---------AC-A---A...G----.........-         329
HIVHAN     -A-.C---T-AT--CAAG--------GCTGC---------AC-A---A...G----.........------    316
HIVSWB884  -A-.C---T-AC--CAAG--------GCTGC---------AT-A---A...G----.........------    334
HIVRF      -C-.C---T-AT--CGTC--------GAAGC---------AC-A---A...G----.........------    334
CONSENSUS2 TaG?CAttaaaaCAca?aGCccgagag?tgCAtCCgGAGttCTacaAa???gACTg????????cTGACa    323
HIVELI     -A-.--TTTGAG--CAAG--CCGAGAGATG--T--G---TT--ACA-A...A---G........A----A    334
HIVZ2Z6    -A-.--TTTGAA--CAAG--CCGAGAGCTG--T--G---TT--ACA-A...G---G........C----A    334
HIVNDK     -A-.--CTTGAG--CAAG--CCGAGAGCTG--T--G---TT--ACA-A...G---G........C----A    334
HIVMAL     -A-.--CTAAGA--CAGA--CAGAGAACAA--T--G---TA--ACA-A...G---G........C----A    334
HIVU455    -G-.--TTAAAA--CAGA--TTATGAGCTG--T--G---TT--ATA-A...G---GAAAGACTGC----A    342
HIVSF170   -G-.--TTAAAA--CAGA--CCAAGAGATG--T--G---TT--ACA-A...G---G........C---  --A    334
HIVZ321    -A-.--CGAAAA--CCTA--CCGAGAGATG--T--G---TT--ACA-A...G---.AAAGACTGC----A    341
HIVANT70   -A-G--ACAC.C--TGTT--TATGATAACT--C--A---CT--TCC-GAAGG---.AAAAACTGC----C    347
SIVCPZ     -G-.-CTTAAGA--TATT--CAGAGAACAA--T--G---TA--ACA-A...G---GACTTTAGAC--G-G    342

                                          <- c.e. ->                    <-
CONSENSUS1 tCyAgcTttCT.aCaa.............GGGACTTTCC.........gCT........GGGGacT    369
HIVLAI     T-G-GC-TG--.A-AA.............----------.........G--........----AC-    369
HIVHXB2R   T-G-GC-TG--.A-AA.............----------.........G--........----AC-    369
HIVMN      T-G-GC-AT--.A-AA.............----------.........G--........----AC-    368
HIVBRVA    T-G-GC-AT--.G-AA.......  ....----------.........G--........----AC-    369
HIVSC      T-G-TG-GT--.A-AA.............----------.........G--........----CA-    370
HIVJH3                                                                           0
HIVJRCSF   C-G-GC-TT--.A-AA.............----------.........G--........----AC-    369
HIVJRFL    T-G-GC-TT--.A-AA.............----------.........A--........----AC-    369
HIVOYI     T-G-GC-TT--.A-AG.....  .....---------- .........G--........----AC-    369
HIVSF2     T-G-GC-TT--.A-AA.............----------.........G--........----AC-    369
HIVNY5CG   T-G-GC-TT--.A-AA.............----------.........G--........----AC-    369
HIVNL43    T-G-GC-TT--.A-AA.............----------.........G--........----AC-    369
HIVSF162   T-                                                                    338
HIVCDC4    C-T-TC-TT--.A-AA.............----------.........G--........----AC-    369
HIVHAN     C-G-GT-TT--.A-AA.............----------.........G--........----AC-    351
HIVSWB884  T-G-GT-TT--.A-GA.............----------.........G--........----AC-    369
HIVRF      T-G-GT-TT--.A-AA.............----------.........G--........----AC-    369
CONSENSUS2 cagA?gTTqCTqACA?...........????GggacTTtCc??????????qCT????????GGGGACT    357
HIVELI     CCG-GC--T--.---A.............-GGAC--T-C.........G--........-------    369
HIVZ2Z6    CCA-GT--T--.---A.............-GGAC--T-C.........G--........-------    369
HIVNDK     TCG-GT--C--.---A.............-GGAC--T-C.........G--........-------    369
HIVMAL     CAG-AG--G--G---G.............-GGAC--T-C.........G--........-------    370
HIVU455    CAG-AG--G--G---G.............-GACT--C-A.........--........-------    377
HIVSF170   CAA-AG--G--G---.............-AGAC--T-C.........G--........-------    369
HIVZ321    CAG-AG--G--G---.............-GGAC--T-C.........G--........-------    376
HIVANT70   TGA-GA--G--G---...........CTGT-GAAC--T-CAGCAAAGACTG--GACACTGC-------    404
SIVCPZ     CA.....-G-GC---AGAACTGCTGACTCTGC-GGAC--T-CA.................A------    407
```

HIV-1 LTR


c.e.-> Sp-1 Sp-1 Sp-1
CONSENSUS1 TTCcAG.GG?AggcgtGgCcTGGGCGGGACtGGGGAGTGGCg???????????agCCCTCAGATgcTgCA 426
HIVLAI ---C--.--.-GGCGT-G-C----------T----------G.....AG---------GC-G-- 426
HIVHXB2R ---C--.--.-GGCGT-G-C----------T----------G.....AG---------CC-G-- 426
HIVMN ---C--.--.-GGTGT-G-C----------C----------G.....AG---------CG-G-- 425
HIVBRVA ---C--.--.-GGC.T-G-C----------T----------G.....AG---------GC-G-- 425
HIVSC ---C--.--.-GGCGC-G-.----------T----------G.....AG---------GC-G-- 426
HIVJH3 0
HIVJRCSF ---C--.--.-GGCGT-G-C----------T----------G.....AG---------GC-G-- 426
HIVJRFL ---C--.--.-GGCGT-G-C----------T----------G.....AG---------GC-G-- 426
HIVOYI ---C--.--.-GGCGT-G-C----------C----------G.....AG---------GC-G-- 426
HIVSF2 ---C--.--.-GGCGT-G-C----------T----------G.....T.---------GC-G-- 425
HIVNY5CG ---C--.--.-GGTGT-G-C----------T----------G.....AG---------GC-G-- 426
HIVNL43 ---C--.--.-GGTGT-G-C----------T----------G.....AG---------GC-A-- 426
HIVSF162 338
HIVCDC4 ---A--.--.-G.CAT-G-C----------T----------G.....AG---------G.-G-- 424
HIVHAN ---C--.--.-GGCGT-A-C----------T----------TGGGGAGTGGCGAG---------GC-G-- 419
HIVSWB884 ---C--.--.-AGCGT-G-C----------T----------G.....AG---------GC-G-- 426
HIVRF ---C--.--.-GGTGT-G-C----------T----------G.....AG---------GC-G-- 426
CONSENSUS2 TTCCag?GGgAGGcgtgg?ctGGGCGGga?tgGGGAGTGGcT.....AaCCCTCAGAtGCTGCA 413
HIVELI ----AG.--.---CGTGGACT------GACTG--------C-.....-A--------T------ 426
HIVZ226 ----GG.--.---CGTGGACT------GACTG--------C-.....-A--------T------ 426
HIVNDK ----AG.--.---CGTGGCCT------GACTT--------C-.....-A--------T------ 426
HIVMAL ----AG.--G---CGTAACTT------GACCG--------C-.....-A--------T------ 428
HIVU455 ----GA.--G---TGTGGTTT------AGTTG--------C-.....-G--------T------ 435
HIVSF170 ----AG.--G---TGTGGTTT------AGTTG--------C-.....-A--------T------ 427
HIVZ321 ----AG.--G---TGTGGCCT------GGTTG--------T-.....-A--------T------ 434
HIVANT70 ----AGT--G---..GACAGG------.TTCG--------C-.....-A--------A------ 460
SIVCPZ ----GGG-AG---TGGTCGGGA-----GGTTG---T----C-T.....TG--------.------ 446
<- TAR CORE
- LBP-1- -LBP-1 -LBP-1-
signal -> mRNA -> <-UBP-1-> <-
CONSENSUS1 TATAAGCAgctGCTTTtTGCCTGTACTGGgTCTCTCTGGTTAGACCAGATctGAGccTGGGAGCTCTCTG 496
HIVLAI --------GCT-----T----------G----------TT---CC---------- 496
HIVHXB2R --------GCT-----T----------G----------CT---CC---------- 496
HIVMN --------GCT-----C----------G----------CT---CC---------- 495
HIVBRVA --------GCT-----C----------G----------CT---CC-------- 490
HIVSC --------CGT-----T----------G----------CT 478
HIVJH3 ---------- 10
HIVJRCSF --------GCT-----T----------G----------CT---CC---------- 496
HIVJRFL --------GCT-----T----------G----------CT---CC---------- 495
HIVOYI --------GCT-----T---------- 453
HIVSF2 --------GCT-----T----------G----------CT---CC---------- 495
HIVNY5CG --------GCT-----T----------G----------CT---CC---------- 496
HIVNL43 --------GCT-----T----------G----------CT---CC---------- 496
HIVSF162 ----------G----------CA---CC---------- 390
HIVCDC4 --------GCT-----G----------G----------CT---TT---------- 494
HIVHAN --------GCC-----T----------C----------CT---CC---------- 489
HIVSWB884 --------GCT-----T----------G----------CT---CC---------- 496
HIVRF --------GCT-----T----------G----------CT---CT---------- 496
CONSENSUS2 TAtAAGCAGCtGCTTTtcGCcTGTACtGGGTCTCTctTtgttaGACCAGaTttGAGcCtGgGAgCTCtctG 481
HIVELI --T-------T-----TT--C-----T-------TC-GGTTA------A-TT---C-T-G--G---TCT- 496
HIVZ226 --T-------T-----TT--C-----T-------TC-GGTTA------A-TT---C-T-A--G---TCT- 496
HIVNDK --T-------T-----TC--C-----T-------TC-GGTTA------A-TT---C-T-G--G---T..- 494
HIVMAL --T-------T-----TC--C-----T-------TC-TGTTA------G-C.---C-C-G--G---TCT- 497
HIVU455 --T-------T-----TC--C-----T-------TC-TGTTA------A-.C---C-T-G--G---TCT- 504
HIVSF170 --T-------T-----TC--C-----T-------TC-TGTTA------A-.A---T-T-A--A---TCT- 496
HIVZ321 --A-------C-----TC--T-----T-------TC-AGTTA------A-TT---C-C-G--G--- 500
HIVANT70 --T-------C-----CT--T-----C-------GG-TAGAG------G-CT---C-C-G--G---CCT- 513
SIVCPZ --T-------T-----CT--C-----T-------TT-CACTG------A-TT---C-T-G--G---TCTA 516

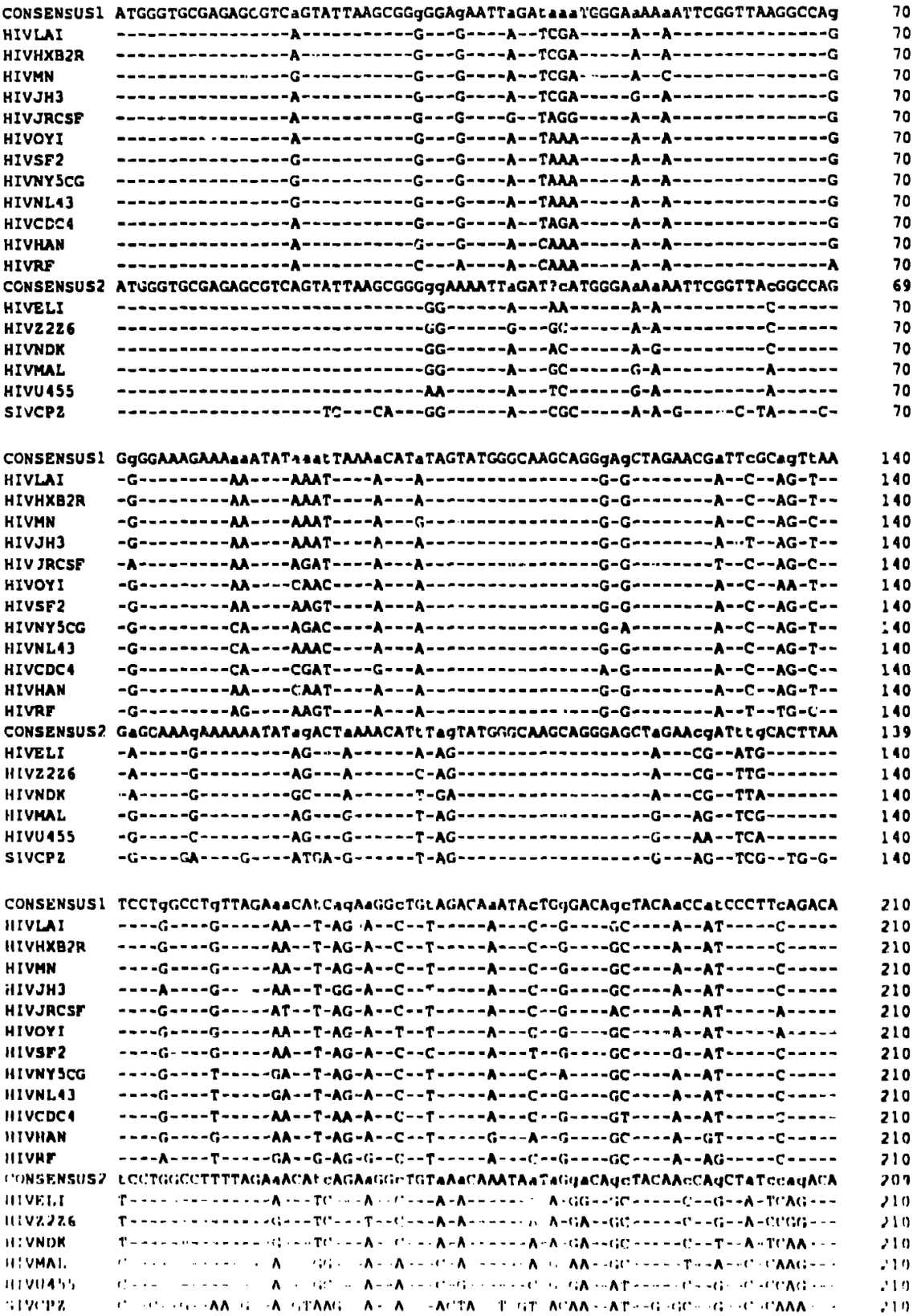
```
 TAR CORE ->
     UBP-1->                          signal ->                          <- mRNA
CONSENSUS1 G.CTa?CTAGGG.AACCCACT?GCTTAAGCCTCAATAAAGCTTGCCTTGA.GTGCTtcAAGTA.GTGTGT    560
HIVLAI     -.--AA------.--------.----------------------------.-----TC-----.------    561
HIVHXB2R   -.--AA------.--------.----------------------------.-----TC-----.------    561
HIVMN      -.--AA------.--------.----------------------------.-----TC-----.------    560
HIVJH3     -.--AA------.--------.----------------------------.-----TC-----.------     75
HIVJRCSF   -.--AG------.--------.----------------------------.-----TC-----.------    561
HIVJRFL                                                                            495
HIVOYI                                                              ----.------    463
HIVSF2     -.--AA------.--------.----------------------------.-----TC-----.------    560
HIVNY5CG   -.--AG------.--------.----------------------------.-----AC-----.------    561
HIVNL43    -.--AA------.--------.----------------------------.-----CA-----.------    561
HIVSF162   -.--AG------.--------G------------------------                            433
HIVCDC4    -.--AG------.--------.----------------------------.-----TT-----.------    559
HIVHAN     -.--AG------.--------.----------------------------.-----TC-----.------    554
HIVSWB884  -.--AG------.--------.----------------------------.-----TC-----.------    561
HIVRF      -.--AG------.--------.----------------------------.-----TC-----.------    561
CONSENSUS2 G?CTagctAggg?AaCCcaCT.gcTTAAgcCTCAATAAAGCTTGCCTTGA?GtGcttcAAGtA.GTGTGT    548
HIVELI     -.--AGCT-GGG.-A--CA--.GC----GC----------.-T-CTTC---T-.------    561
HIVZ226    -.--AGCT-GGG.-A--CA--.GC----GC----------.-T-CTTC---T-.------    561
HIVNDK     -.--AATT-GGG.-A--CA--.GC----GC----------.-C-CTTC---T-.------    559
HIVMAL     -.--AGC.-AGG.-A--CA--.GC----GC----------.-T-CCTC---C-.------    561
HIVU455    -.--AGCG-GGGA-C--.A--.GC----GC----------G-T-CTTG---T-.------    570
HIVSF170   -.--AGCG-GGGA-C--.A--.GC----GC----------                        536
HIVANT70   -C--..CT-GCTG-A--CG--.CG----CG----------.-T-AGA                 503
SIVCPZ     ..-CAGTG-AGG.-A--CA--.GC----GC----------.-T-TATTTGAGCA------    581

CONSENSUS1 GCCCgTCTGTTgTGTGACTCTGGTAaCTAGAGATCCCTCAGAcCctTTTAGtCAGtGTGGAAAA?TCTCT    629
HIVLAI     ----G------G----------A----------C-CT-----T---T--------.-----    630
HIVHXB2R   ----G------G----------A----------C-CT-----T---T--------.-----    630
HIVMN      ----G------A----------G----------T-CT-----G---T--------.-----    629
HIVJH3     ----G------G----------A----------C-AT-----T---C--------.-----    144
HIVJRCSF   ----G------C----------A----------C-CT-----T---T--------.-----    630
HIVJRFL                                                                    495
HIVOYI     ----G------G----------A----------C-CT-----T---T--------.-----    532
HIVSF2     ----G------G----------A----------C-CT-----T---T--------A-----    630
HIVNY5CG   ----G------G----------A----------C-CT-----T---T--------.-----    630
HIVNL43    ----G------G----------A----------C-CT-----T---T--------.-----    630
HIVCDC4    ----G------G----------A----------C-CT-----T---T--------A-----    629
HIVHAN     ----G------G----------G----------T-CT-----T---T--------.-----    623
HIVSWB884  -                                                               562
HIVRF      ----G------G----------A----------C-TC-----T---T--------.-----    630
CONSENSUS2 GCCCgTCTGTTgTgTGACTCTGGTAACTAGAGATCCCTCAGAcCccTtTAGtC?GaGTg?AAAAaTCTCT    616
HIVELI     ----G------G-G----------C-CC-T---T-A-A--GG----.-----    630
HIVZ226    ----G------G-G----------C-CC-T---T-A-A--GG----.-----    630
HIVNDK     ----G------G-G----------T-CG-T---T-C-A--G.---A-----    628
HIVMAL     ----A------G-G----------C-AC-C---A-G-T--..----A-----    629
HIVU455    ----G------A-T----------C-AC-A---A-T-T--..----A-----    638
SIVCPZ     ---TA---A--CAGAC----------TTAAAT--TAGTCAA.GG----.-----    649
```

HIV-1 LTR

```
                <-
                  <- Lys-tRNA pbs
CONSENSUS1 AGCAGTGGCGCCCGAACAGGgACttgAaAgcGAAAGtaa.....................aaCCaGAGgag    679
HIVLAI     --------------------G--TTG-A-GC-----GGA.....................AA--A---GAG    680
HIVHXB2R   --------------------G--CTG-A-GC-----GGA.....................AA--A---...    677
HIVMN      --------------------G--T G-A-GC-----AAA.....................AA--A---...    676
HIVJH3     --------------------G--CGG-A-GC-----AGA.....................AA--A---GAG    194
HIVJRCSF   --------------------G--CGG-A-GC-----AGA.....................AA--A---GAG    680
HIVJRFL                                                                              495
HIVOYI     ----------------------G--CTT-A-GT-----TGG.....................AA--A---GAG    582
HIVSF2     --------------------G--GCG-A-GC-----TAG.....................AA--A---GAG    680
HIVNY5CG   --------------------G--TTG-G-GC-----TAA.....................AG--A---GAG    680
HIVNL43    --------------------G--TTG-A-GC-----TAA.....................AG--A---GAG    680
HIVCDC4    --------------------.--TTG-A-GC-----TGA.....................GA--A---AAG    678
HIVHAN     --------------------G--TTG-A-GC-----TAG.....................AA--A---GAG    673
HIVRF      --------------------G--CAG-A-GC-----TAG.....................AA--A---GAG    680
CONSENSUS2 AGCAGTGGCGCCCGAACAGGGACctgAAAGcGAAAGTag??????????????????????aacCAGAGaAG    666
HIVELI     ------------------------CTG----C------AG.....................AAC-----G--    680
HIVZ226    ------------------------CTG----C------AG.....................AAC-----A--    680
HIVNDK     ------------------------CTG----C------AG.....................AAC-----A--    678
HIVMAL     ------------------------TTT----T------AACAGGGACTCGAAAGCGGAAGTTC-----A--    699
HIVU455    ------------------------TCG----C------TC.....................-----A--    685
SIVCPZ     ------------------------CTG-.--T------AA.....................CCGG--GCTGA    698

CONSENSUS1 cTCTCTCGACGCA.GGACTCGGCTTGCTgaAGCGC??GCaCgGCAAGAGGCGAGGGGcGGCGactGGTGA    746
HIVLAI     C-----------.-------------GA-----..--A-G--------------A----ACT-----    747
HIVHXB2R   C-----------.-------------GA-----CC--A-G--------------C----ACT-----    746
HIVMN      C-----------.-------------GA-----..--A-G--------------C----ACT-----    743
HIVJH3     A-----------.-------------GA-----..--A-A--------------C----GCT-----    261
HIVJRCSF   A-----------.-------------GA-----..--A-A--------------C----ACT-----    747
HIVJRFL          -------.-------------GA-----..--A-G--------------C----ACT-----    556
HIVOYI     C-----------.-------------T.-----..--A-G--------------C----ACT-----    648
HIVSF2     C-----------.-------------GA-----..--A-A--------------C----ACT-----    747
HIVNY5CG   A-----------.-------------GA-----..--A-G--------------C----ACT-----    747
HIVNL43    A-----------.-------------GA-----..--A-G--------------C----ACT-----    747
HIVCDC4    C-----------.-------------GA-----..--A-G--------------C----ACT-----    745
HIVHAN     A-----------.-------------GA-----..--A-G--------------C----ACT-----    740
HIVRF      A-----------.-------------GA-----..--G-G--------------C----AAC-----    747
CONSENSUS2 ?TCTCTCGACGCA.GGACTCGGCTTGCTGAagcGcgCACa??GCAAGAGGCGAGgggCaGCGACtGGTGA    732
HIVELI     C-----------.----------------AGC-CG---G..-----------GGG-A-----T-----    747
HIVZ226    C-----------.----------------AGC-CG---G..-----------GGG-A-----C-----    747
HIVNDK     A-----------.----------------AGC-CG---A..-----------GGG-A-----T-----    745
HIVMAL     T-----------.----------------GGT-CA---A..-----------A.G-G-----T-----    765
HIVU455    T-----------.----------------G..-TG---ACA-----------AG.-G-----T-----    751
SIVCPZ     A----C-------C----------------...-TG---ACA-----------G..-G-----T-----    763
```

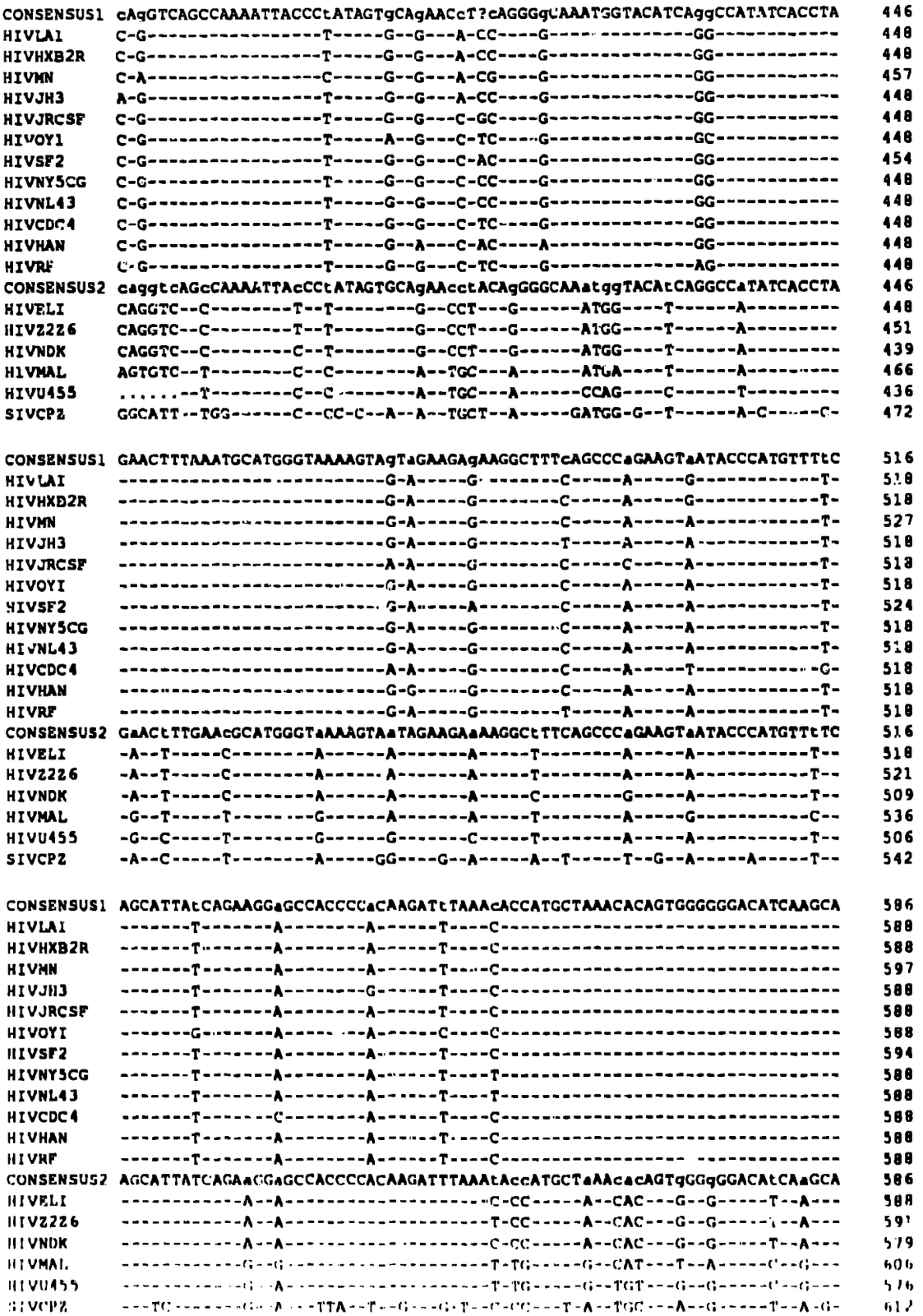
```
          \/ major 5'sj                              gag cds  ->
CONSENSUS1 GTACGCCaaaaaTTtt?.GACTAGCGGAGGCTAGAAGGAGAGAGAG      788
HIVLAI     -------AAAAA--TT..-------------------------         789
HIVHXB2R   -------AAAAA--TT..-------------------------         788
HIVMN      -------AAAAA--CTT.-------------------------         786
HIVJH3     -------AAAAA--TT..-------------------------         303
HIVJRCSF   -------GAAAT--T...-------------------------         788
HIVJRFL    -------AAAAA--TT..-------------------------         598
HIVOYI     -------AAAAA--TTT.--------------------------        692
HIVSF2     -------...AA--TTT.---------------------------       789
HIVNY5CG   -------AAAAA--TT..-------------------------         789
HIVNL43    -------AAAAA--TT..-------------------------         789
HIVCDC4    -------AAAAA--TT..-------------------------         787
HIVHAN     -------.AAAA--TT..--------------------------        782
HIVRF      -------GAAAA--TT..-------------------------         789
CONSENSUS2 GTACGCt?aaAaTTTTt?GACTAGCGGAGGctAGAAGGAGAGAgA      775
HIVELI     ------TAAA-T----..-----------CT----------G          789
HIVZ2Z6    ------TAAA-A----..-----------CT----------G          789
HIVNDK     ------TG..A-A----T.-----------..----------G-        786
HIVMAL     ------C...-A----T.-----------CT----------G          805
HIVU455    ------CTAA-A----TT-----------CT----------A-         796
SIVCPZ     ------....-A----T..G-------T--CT---CCT----GA- >     826
```

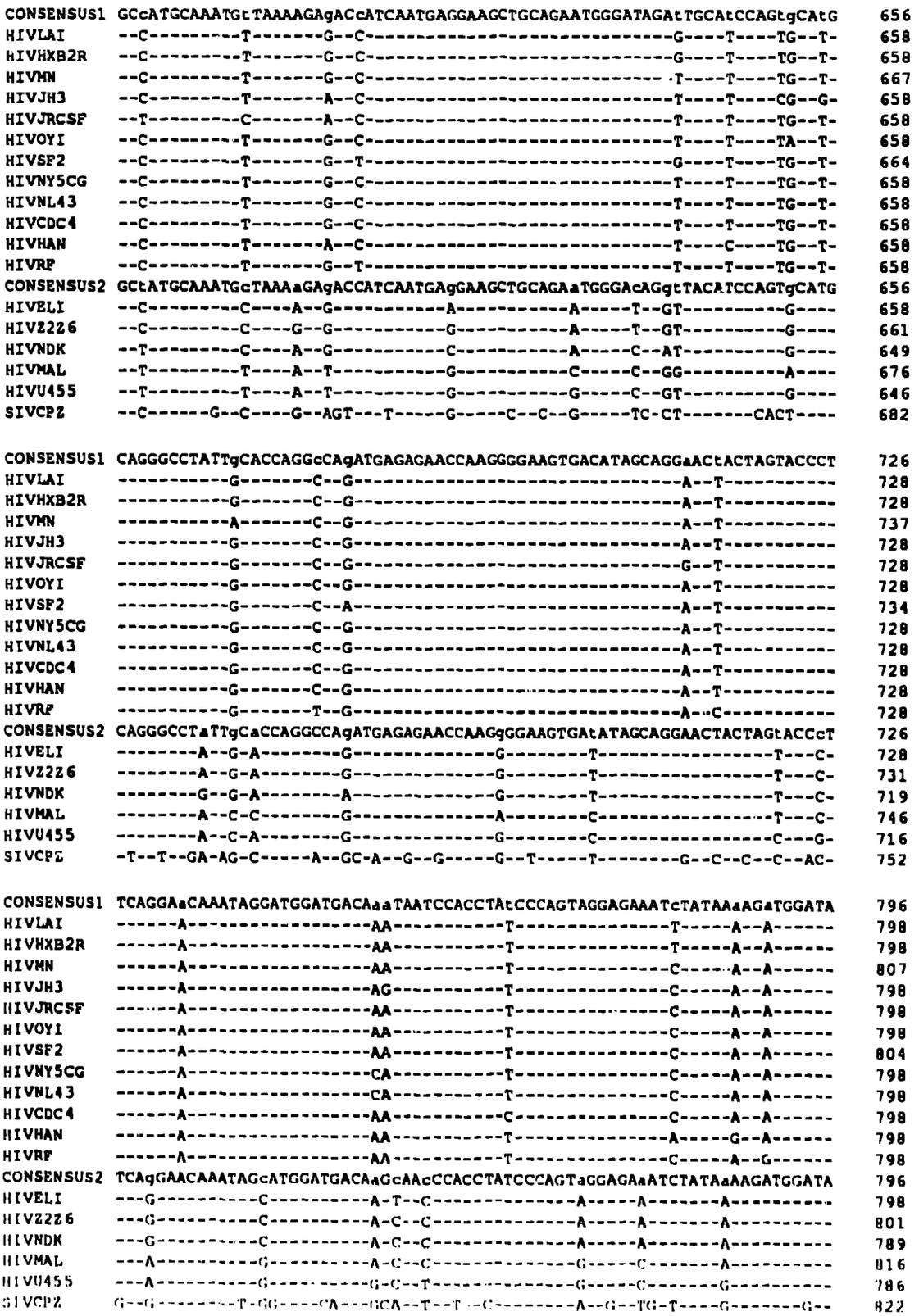
HIV-1 GAG

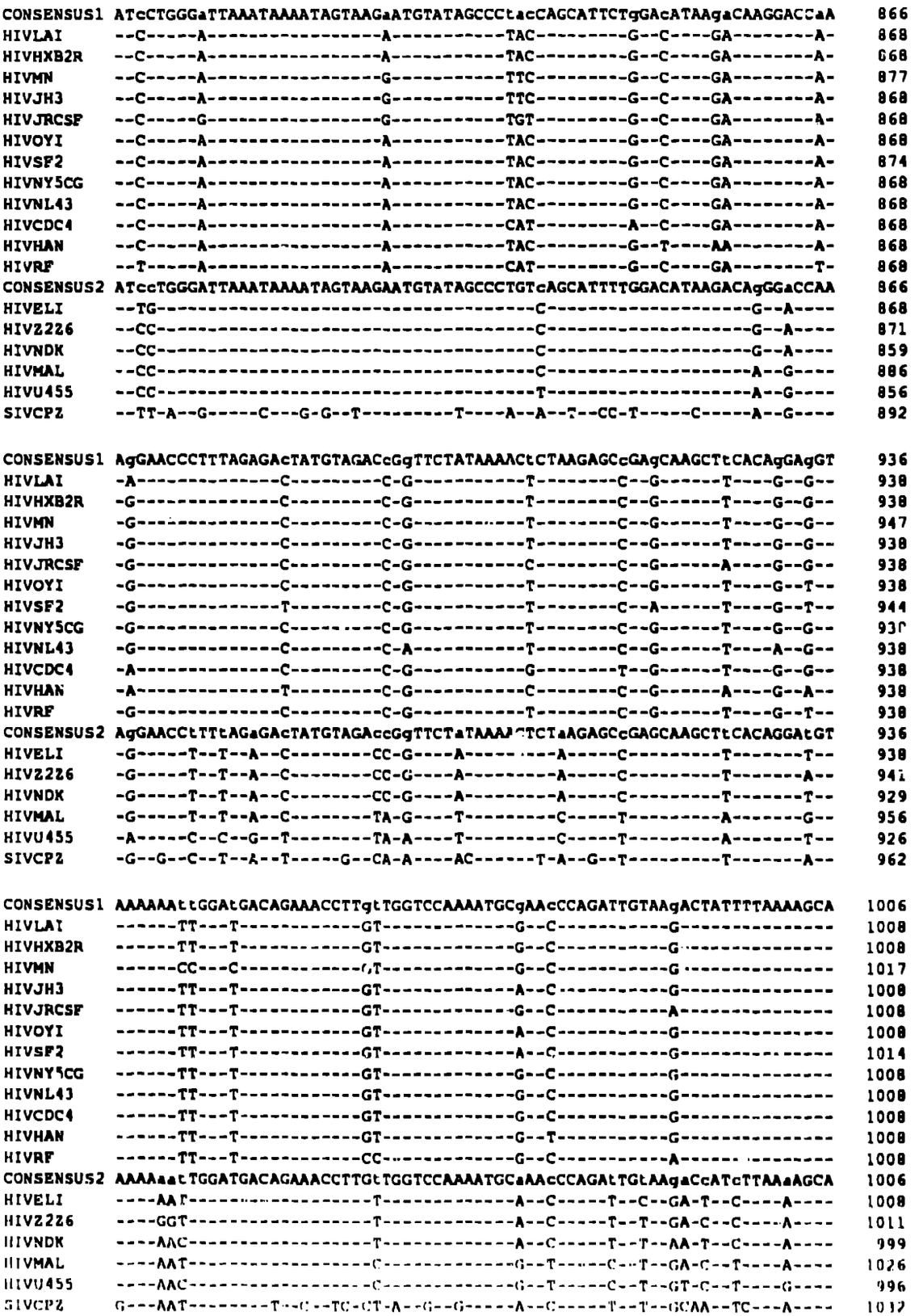


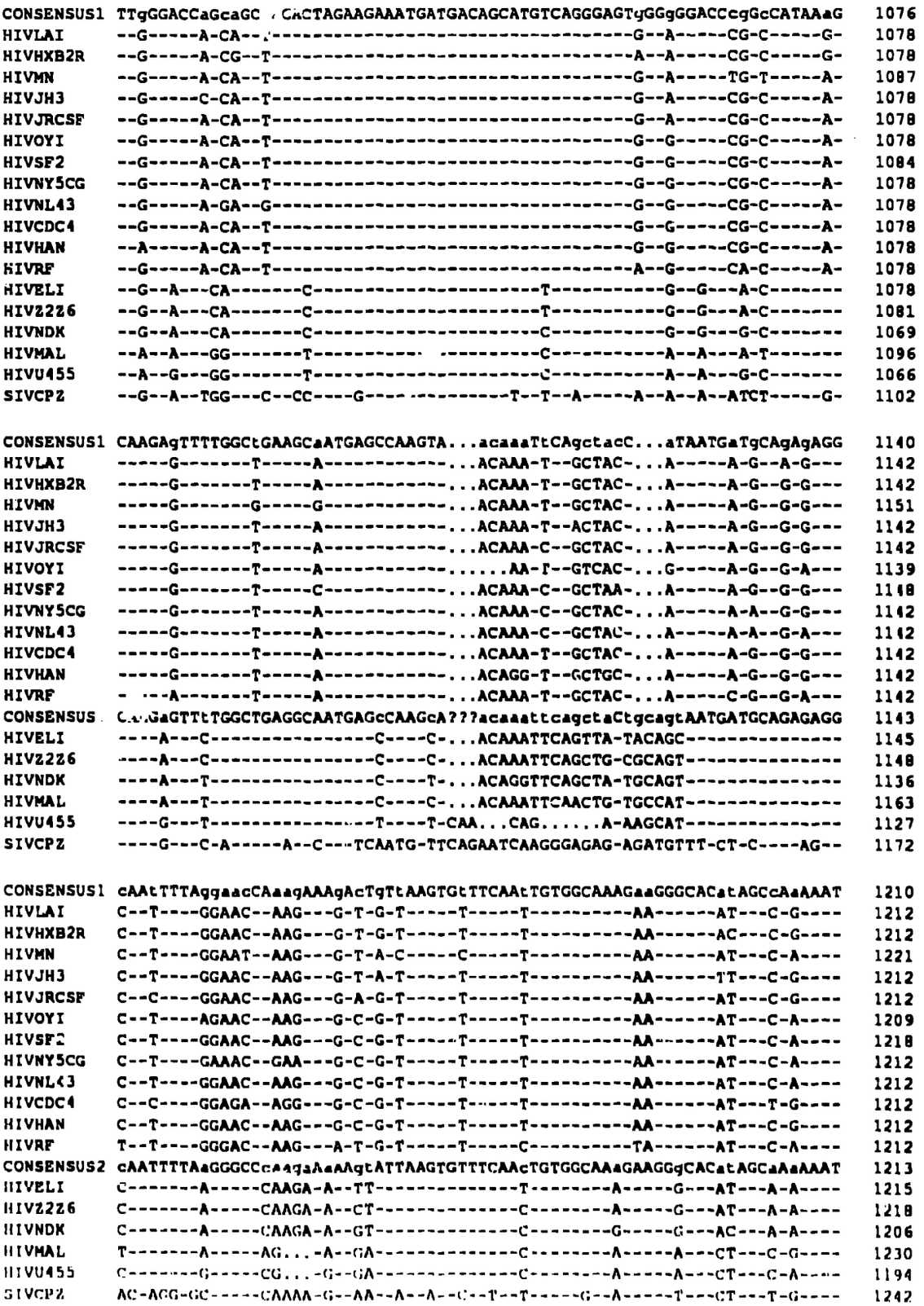
```
CONSENSUS1 GGATCAGAAGAActTAgATCATTATaTAATaCAgTAGCAaccCTcTATTGTGTgCATCAAAggATAGAgg  280
HIVLAI     ----------CT--G--------A----A--G-----ACC--C--------G-------GG-----GA  280
HIVHXB2R   ----------CT--G--------A----A--G-----ACC--C--------G-------GG-----GA  280
HIVMN      ----------CT--A--------A----A--G-----ACC--C--------G-------AG-----GA  280
HIVJH3     ----------CT--A--------T----A--G-----ACC--C--------G-------GA-----GG  280
HIVJRCSF   ----------CT--C--------A----A--G-----ACC--C--------A-------GG-----GA  280
HIVOYI     ----------AT--G--------A----A--G-----ACT--T--------G-------AG-----GG  280
HIVSF2     ----------CT--G--------A----A--G-----ACC--C--------A-------GG-----TG  280
HIVNY5CG   ----------CG--G--------T----A--G-----GTC--C--------G-------GG-----TG  280
HIVNL43    ----------CT--G--------A----A--A-----GTC--C--------G-------GG-----TG  280
HIVCDC4    ----------CT--G--------A----A--G-----ACC--C--------G-------GG-----GG  280
HIVHAN     ----------CT--G--------A----A--G-----ACC--C--------G-------AG-----GG  280
HIVRF      ----------CT--A--------A----G--G-----ACC--C--------A-------AT-----GG  280
CONSENSUS2 GGAtCAGAAGAAcTTAgATCATTATaTAATACAGTAGCAaccCCTCTATTGTGTACAT?AAaGgATAGAtG  278
HIVELI     ---A--------C---G--------A----------AC---------------A--G-A-----T-  280
HIVZ2Z6    ---T--------C---G--------T----------AC---------------G--A-G-----G-  280
HIVNDK     ---T--------A---G--------A----------AC---------------G--A-G-----G-  280
HIVMAL     ---T--------A---A--------A----------AC---------------C--A-G-----T-  280
HIVU455    ---A--------C---G--------A----------GT---------------C--A-G-----T-  280
SIVCPZ     --CT------G-C-GCGG--C--G-T---C--TC-G---GTA--G-GG--CA-----AGTGAC--CACT-  280

CONSENSUS1 TAAaAGACACCAAgGAAGCtTTAGA?AAgaTAGAGGAAGAgCAAAaCAAaAGT...AaGAAAAAaGCACA  346
HIVLAI     ---A---------G-----T-----C--GA----------G----A---A---...-A------A-----  347
HIVHXB2R   ---A---------G-----T-----C--GA----------G----A---A---...-A------A-----  347
HIVMN      ---A---------G-----T-----G--AA----------G----A---A---...-A------A-----  347
HIVJH3     ---A---------G-----T-----G--GA----------G----A---A---...-A------A-----  347
HIVJRCSF   ---A---------G-----T-----A--GA----------G----C---A---...-T------G-----  347
HIVOYI     ---A---------G-----T-----T--GA----------G----A---A---...-A------A-----  347
HIVSF2     ---A---------G-----T-----G--GA----------G----A---A---...-A------G-----  347
HIVNY5CG   ---A---------G-----C-----T--GA----------G----A---A---...-A------G-----  347
HIVNL43    ---A---------G-----C-----T--GA----------G----A---A---...-A------G-----  347
HIVCDC4    ---G---------A-----T-----C--GA----------A----A---A---...-A------A-----  347
HIVHAN     ---A---------G-----T-----C--GG----------G----A---C---...-A------G-----  347
HIVRF      ---G---------G-----T-----C--GA----------G----A---A---...-A------A-----  347
CONSENSUS2 TAAAAGACACCAAgGAAGCttTAgAaAAgATgGAGGAAgaaCAAAAcAAaAgt???AagaAAAaGaCACA  345
HIVELI     ----------G-----TT--G-A--G--G------GAG-----C--A-GT...-AGA---A-G----  347
HIVZ2Z6    ----------G-----TT--G-A--G--G------GAA-----C--A-GTAAG-ACA---A-G----  350
HIVNDK     ----------A-----TG--G-A--G--G------GAA-----C--A-GT...-AGA---A-A----  347
HIVMAL     ----------G-----GC--G-T--A--A------ATA-----T--G-GC...-GGC---A-A----  347
HIVU455    ----------G-----TT--A-T--A--A------ATG-----T--G-AC...-AGC---G-A----  347
SIVCPZ     --G-------AC-GA----TC--..............GAA--GCTA--GCGG...CATC-TGGAGA---  332

CONSENSUS1 GCAA??????.....................gcaGCAGCTGaCaCA?????????GGAaACagcAgc  377
HIVLAI     ----.............................GCA-------A-A--.........---C--AGC-GC  378
HIVHXB2R   ----.............................GCA-------A-A--.........---C--AGC-AT  378
HIVMN      ----.............................GCA-------A-A--GGAAACAGA---A--AGC-GC  387
HIVJH3     ----.............................GCA-------A-A--.........---A--AGC-GC  378
HIVJRCSF   ----.............................GCA-------A-A--.........---A--AGC-GC  378
HIVOYI     ----.............................ACA-------A-A--.........---A--AGC-GC  378
HIVSF2     ----GCAGCA.......................GCT-------G-A--.........---A--AGC-GC  394
HIVNY5CG   ----.............................GCA-------A-A--.........---A--AGC-GC  378
HIVNL43    ----.............................GCA-------A-A--.........---A--AAC-GC  378
HIVCDC4    ----.............................GCA-------A-A--.........---A--AGC-GC  378
HIVHAN     ----.............................GAA-------A-G--.........---A--AGA-AC  378
HIVRF      ----.............................GCA-------A-A--.........---A--GGC-GC  378
CONSENSUS2 GCAa??????????????????.....GCAGCAGCTgacaca.....gaaaacAgCagc  376
HIVELI     ---A.....................------- -GACACA.....GGAAAC-A-AGC  378
HIVZ2Z6    ---A.....................---------GACGCA.....GGGAAC-A-AGC  381
HIVNDK     ---A.....................---------GAT.....-G-AGC  369
HIVMAL     ---GGCAGCAGCTGCACAGCAG.....---------GCCACA.....AAAAAC-G-AGC  396
HIVU455    ---G.....................---------AACACA.....GGA...-G-...  372
SIVCPZ     A--GAGCAAAACTGAAAGTAACTCAGGAACCCGT-A--GG--GAGCCAGTCAAGGCGCTAGTGCCTCTGCT  402
```

HIV-1 GAG







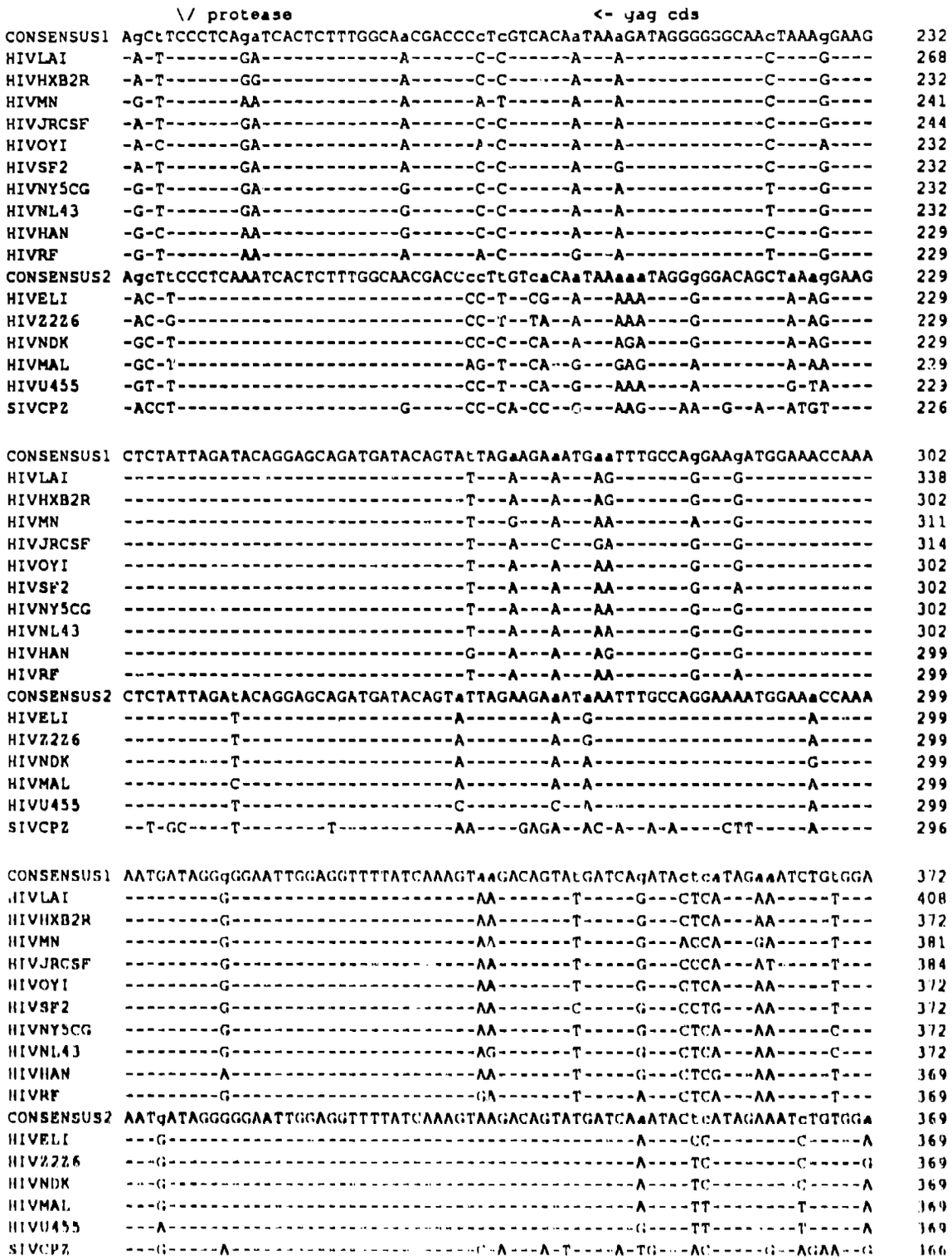


```
CONSENSUS1 TGCAgGGCcCCTAGGAAAAaGGGCTGTTGGAaATGTGGAAaGGAAGGACAcCAAATGAAaGAtTGtActg   1280
HIVLAI     ----G---C----------A-----------A--------A---------C--------A--T--T-CTG   1282
HIVHXB2R   ----G---C----------A-----------A--------A---------C--------A--T--T-CTG   1282
HIVMN      ----G---C----------G-----------A--------A---------C--------A--T--T-CTG   1291
HIVJH3     ----G---C----------A-----------A--------A---------C--------A--T--T-ATG   1282
HIVJRCSF   ----G---C----------A-----------A--------A---------C--------A--G--T-CTG   1282
HIVOYI     ----G---T----------A-----------A--------G---------C--------A--T--T-CTG   1279
HIVSF2     ----G---C----------A-----------G--------G---------C--------A--T--C-CTG   1288
HIVNY5CG   ----G---C----------A-----------A--------A---------C--------A--T--T-CTG   1282
HIVNL43    ----G---C----------A-----------A--------A---------C--------A--T--T-CTG   1282
HIVCDC4    ----A---C----------A-----------A--------G---------C--------A--T--T-CTG   1282
HIVHAN     ----G---C----------A-----------A--------A---------T--------G--T--C-CAG   1282
HIVRF      ----G---C----------A-----------A--------A---------C--------A--T--C-CTA   1282
CONSENSUS2 TGcAGGGCCCCTAGgAaAAAGGGCTGTTGGAaATGtGGaAagGAAGGACACCAAaTgAAaGAtTGCACTG   1283
HIVELI     --C-----------G-A--------------G---T--A-AG------------C-A--A--T-------   1285
HIVZ2Z6    --C-----------G-G--------------A---T--A-AG------------C-G--G--T-------   1288
HIVNDK     --C-----------A-A--------------A---C--A-GG------------A-G--A--T-------   1276
HIVMAL     --C-----------G-A--------------A---T--G-AG------------A-G--A--C-------   1300
HIVU455    --T-----------G-A--------------A---C--G-AA------------A-G--A--C-------   1264
SIVCPZ     --T-AA--A--A--A-G---A---------CGG--T--GCAA--G---------A-G--A--C-----A-   1312

                      pol cds ->
CONSENSUS1 Ag??????aGACAGGCTAATTTTTTAGGGAAgATCTGGCCTTCCcaCAAGGGaAGGccagGgAAt?????   1339
HIVLAI     -G......A----------------------G------------TA------A---CCAG-G--TTTTCT   1346
HIVHXB2R   -G......A----------------------G------------TA------A---CCAG-G--T.....   1341
HIVMN      -G......A----------------------G------------TG------A---...C-G--T.....   1347
HIVJH3     -A......A----------------------G------------TC------A---CCAG-G--T.....   1341
HIVJRCSF   -G......A----------------------G------------TA------A---CCAG-G--T.....   1341
HIVOYI     -G......A----------------------G------------CA------A---CCAG-G--T.....   1338
HIVSF2     -G......A----------------------G------------TA------A---CCAG-G--T.....   1347
HIVNY5CG   -G......A----------------------G------------CA------A---CCAG-G--T.....   1341
HIVNL43    -G......A----------------------G------------CA------A---CCAG-G--T.....   1341
HIVCDC4    -G......A----------------------G------------CA------A---CCAG-G--T.....   1341
HIVHAN     -A......A----------------------G------------CA------G---CCAG-A--T.....   1341
HIVRF      -TGAGGGAC----------------------A------------CA------A---CCAG-G--C.....   1347
CONSENSUS2 Ag......AGACAGGCTAATTTTTTAGGGAaaATTTGGCCTTCCcACAAGGGaAGGCCgGGGAAt.....   1342
HIVELI     -G......----------------------GA------------C-------A-----G-----C.....   1344
HIVZ2Z6    -A......----------------------AG------------C-------A-----G-----C.....   1347
HIVNDK     -A......----------------------AG------------C-------A-----G-----T.....   1335
HIVMAL     -G......----------------------AA------------C-------A-----A-----T.....   1359
HIVU455    -G......----------------------AA------------A-------G-----A-----T.....   1323
SIVCPZ     GA......-------TG-------------AAGG----------CG--GC--G-----A--C--C.....   1371

CONSENSUS1 ????????????????????????????????TTtCtTCAGAgCAGAcCAGAGCCAACAGCCCCaCCAGaa   1378
HIVLAI     TCAGAGCAGACCAGAGCCAACAGCCCCACCA--T-T-----G----C----------------A----AA   1416
HIVHXB2R   ..............................--T-T-----G----C----------------A----AA   1380
HIVMN      ..............................--T-C-----G----A----------------A----AA   1386
HIVJH3     ..............................--T-T-----G----C----------------T----AG   1380
HIVJRCSF   ..............................--C-T-----G----C----------------A----AA   1380
HIVOYI     ..............................--T-T-----A----C----------------A----CA   1377
HIVSF2     ..............................--T-T-----G----C----------------A----AA   1386
HIVNY5CG   ..............................--T-T-----G----C----------------A----AA   1380
HIVNL43    ..............................--T-T-----G----C----------------A----AA   1380
HIVCDC4    ..............................--T-T-----G----C----------------A----AG   1380
HIVHAN     ..............................--T-T-----G----C----------------A----AA   1380
HIVRF      ..............................--T-T-----G----C----------------A----AA   1386
CONSENSUS2 ..............................TTtCttCAgAGCAGACCAGAGCCAACAGCCCCACCAGCA   1381
HIVELI     ..............................--T-TC--A------------------------------   1393
HIVZ2Z6    ..............................--T-TT--G------------------------------   1386
HIVNDK     ..............................--T-TT--G------------------------------   1374
HIVMAL     ..............................--C-TT--G------------------------------   1398
HIVU455    ..............................--T-CT--G------------------------------   1362
SIVCPZ     ..............................--TGTG--G-A----A----A-----G--------GAT-   1410
```

```
CONSENSUS1 GAGAGCTTCaGGTtTGGGgaaGAgacAacaaCtCCCtcTCAGAAGCAGGAG????????????ccGAtAG   1436
HIVLAI     ---------A---C----GTA--GAC-ACAA-T---TC------------.............CC--T--   1474
HIVHXB2R   ---------A---C----GTA--GAC-ACAA-T---CC------------.............CC--T--   1438
HIVMN      ---------A---T----GAA--GAC-ACAA-T---TA------------AAGAAGCAGGAGAC--T--   1456
HIVJH3     ---------A---T----GAA--GAC-ACAA-C---TC------------.............CC--G--   1438
HIVJRCSF   ---------A---T----GAA--GAC-GCAA-T---TC------------CAGAAGCAGGAGCC--T--   1450
HIVOYI     ---------G---T----GAA--GAC-ACAA-T---CC------------.............CC--T--   1435
HIVSF2     ---------A---T----GAG--GAA-ACAA-T---TC------------.............CC--T--   1444
HIVNY5CG   ---------A---T----GAA--GAC-ACAA-T---TC------------.............CC--T--   1438
HIVNL43    ---------A---T----GAA--GAC-ACAA-T---TC------------.............CC--T--   1438
HIVCDC4    ---------A---T----GAC--GAC-ACAA-T---TC------------.............CC--G--   1438
HIVHAN     ---------A---T----...--AGC-ACAG-T---TC------------.............CC--T--   1435
HIVRF      ---------A---T----GAA--GAC-...A-T---TC------------.............AA--T--   1441
CONSENSUS2 GAgAgCTTcGGGtTtGGGGAagAGATa...AcccCCtCTcagAAACAGGAG.............CaGAAAG   1436
HIVELI     --G-G---C---T-T-----AG----A...-CCC--T--CAA---------.............-A-----   1438
HIVZ2Z6    --G-G---C---T-T-----AG----A...-CCC--T--CAG---------.............-A-----   1441
HIVNDK     --G-G---C---T-T-----GG----A...-CCC--T--CAG---------.............-A-----   1429
HIVMAL     --G-G---C---T-T-----GG----A...-AAC--T--CAG---------.............-A-----   1453
HIVU455    --A-T---T---A-G-----AA----G...-CCT--C--GCG---------.............-T-----   1417
SIVCPZ     --G-GT-AT---TACCA---GG-AGAG...-AGAG-CAGGAG--GA-A---.............GGA...-   1462

CONSENSUS1 ACAAGGaaaTGTATCCtTTAgCtTCCCTCAgaTCACTCTTTGGCAaCGACCCcTcGTCACaa***        1501
HIVLAI     ------AAC-------T---A-T-------GA-------------A------C-C-----AA---        1539
HIVHXB2R   ------AAC-------T---A-T-------GG-------------A------C-C-----AA---        1503
HIVMN      ------ACC-------T---G-T-------AA-------------A------A-T-----AA---        1521
HIVJH3     ------AAC-------T---G-T-------GA-------------A------C-C-----CC---        1503
HIVJRCSF   ------AAT-------T---A-T-------GA-------------A------C-C-----AA---        1515
HIVOYI     ------GAC-------T---A-C-------GA-------------A------A-C-----AA---        1500
HIVSF2     ------AAC-------T---A-T-------GA-------------A------C-C-----AA---        1509
HIVNY5CG   ------AAC-------T---G-T-------GA-------------G------C-C-----AA---        1503
HIVNL43    ------AAC-------T---G-T-------GA-------------G------C-C-----AA---        1503
HIVCDC4    --- --AAC-------T---G-T-------GA-------------A------A-C-----AA---        1503
HIVHAN     ------AAC-------C---G-C-------AA-------------G------C-C-----AA---        1500
HIVRF      ------AAC-------T---G-T-------AA-------------A------A-C-----AG---        1506
CONSENSUS2 ACAaGGAActGtaTCCTTtAgcTtCCCTCAAATCACTCTTTGGCAACGACCccTtGTcaCAa---        1501
HIVELI     ---A----CT-TA-----T-AC-T-------------------------CC-T--CG--A---        1503
HIVZ2Z6    ---A----CT-TA-----C-AC-G-------------------------CC-T--TA--A---        1506
HIVNDK     ---A----CT-TA-----T-GC-T-------------------------CC-C--CA--A---        1494
HIVMAL     ---A----TT-TA-----T-GC-T-------------------------AG-T--CA--G---        1518
HIVU455    ---G----CA-AC-----T-GT-T-------------------------CC-T--CA--G---        1482
SIVCPZ     -G-GCAGTCTATA---ACC-ACCT--------------------G-----CC-CA-CC--G---        1527
```

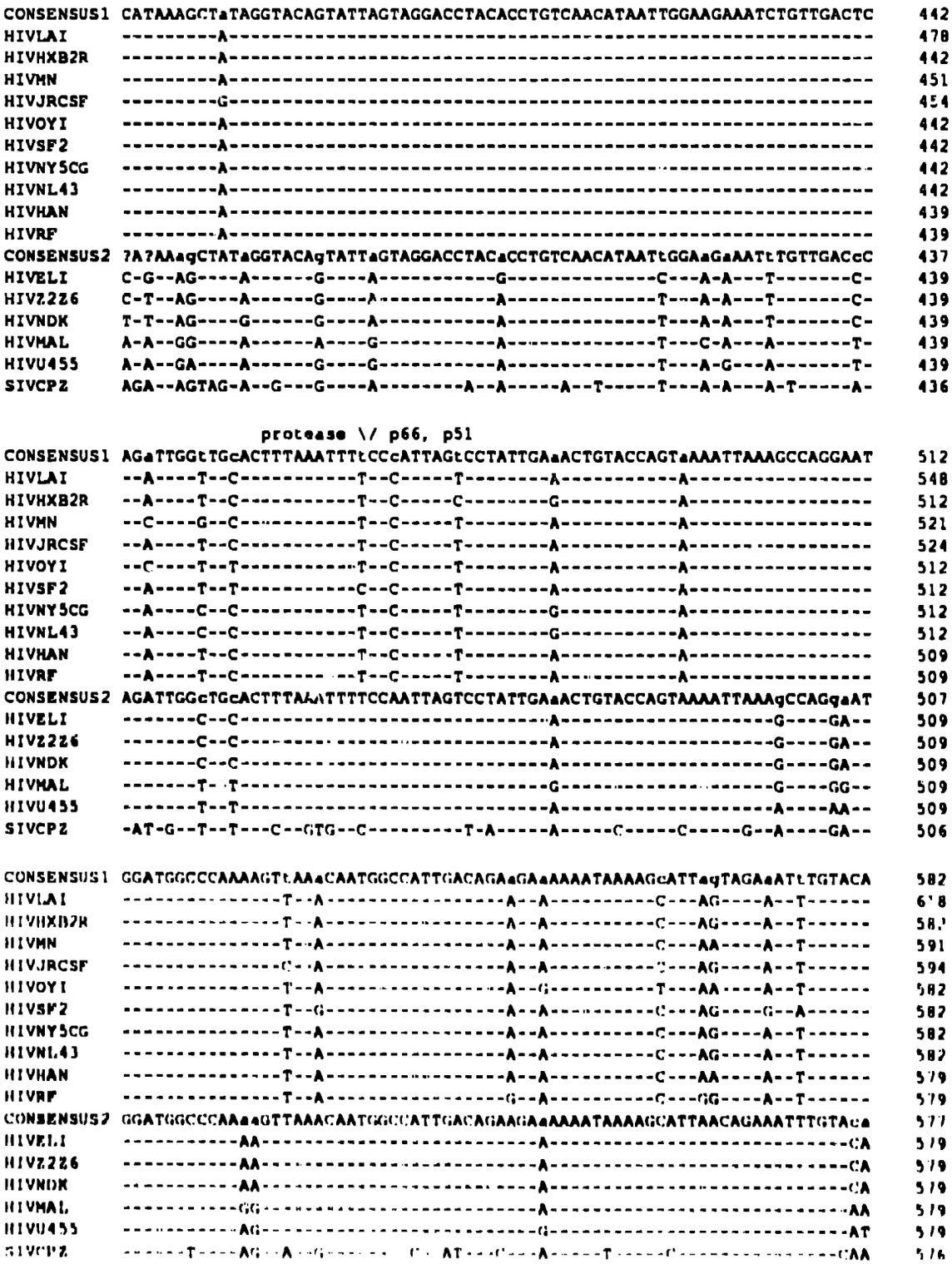
HIV-1 POL

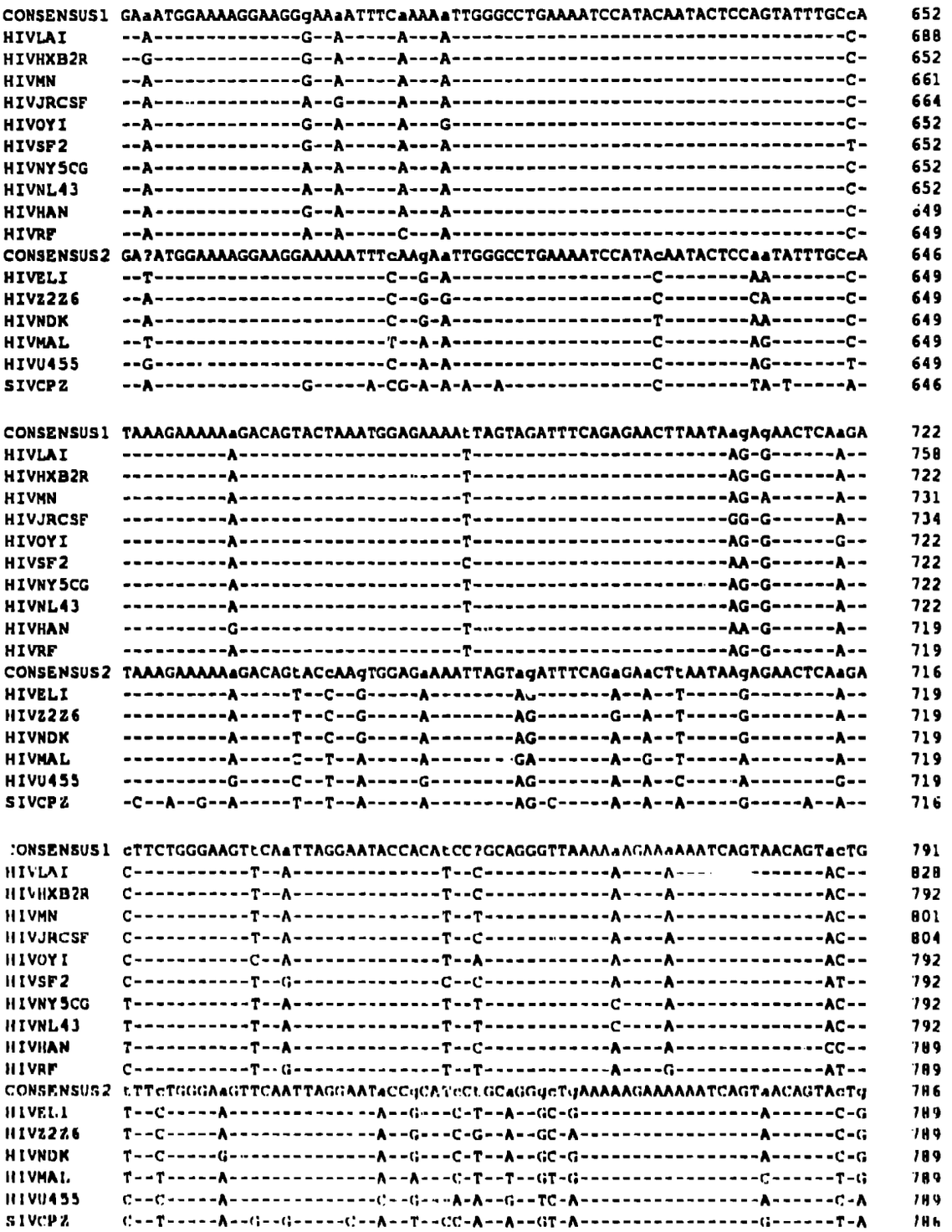
```
 gag ->
CONSENSUS1 TTTTTTAGGGAAgATCTGGCCTTCCcaCAAGGGaAGGCcagggAAtTT??????????????????????  48
HIVLAI     ----------G----------TA------A----CAGGG--T--TCTTCAGAGCAGACCAGAGCCA  70
HIVHXB2R   ----------G----------TA------A----CAGGG--T--.....  48
HIVMN      ----------G----------TG------A----G...G--T--.....  45
HIVJRCSF   ----------G----------TA------A----CAGGG--T--.....  48
HIVOYI     ----------G----------CA------A----CAGGG--T--.....  48
HIVSF2     ----------G----------TA------A----CAGGG--T--.....  48
HIVNY5CG   ----------G----------CA------A----CAGGG--T--.....  48
HIVNL43    ----------G----------CA------A----CAGGG--T--.....  48
HIVHAN     ----------G----------CA------G----CAGGA--T--.....  48
HIVRF      ----------A----------CA------A----CAGGG--C--.....  48
CONSENSUS2 TTTTTTAGGGAaaATTTGGCCTTCCcACAAGGGaAGGCCgGGGAAtTT.....  48
HIVELI     ----------GA----------C-------A-----G-----C--.....  48
HIVZ2Z6    ----------AG----------C-------A-----G-----C--.....  48
HIVNDK     ----------AG----------C-------A-----G-----T--.....  48
HIVMAL     ----------AA----------C-------A-----A-----T--.....  48
HIVU455    ----------AA----------A-------G-----A-----T--.....  48
SIVCPZ     ----------AAGG----------CG--GC--G-----A--C--C--.....  48

CONSENSUS1 ??????????????tCtTCAGAgCAGAcCAGAGCCAACAGCCCCACCAGaAGAGAGCTTCaGGTtTGGGG  104
HIVLAI     ACAGCCCCACCATTT-T-----G----C----------A----------A---C-----  140
HIVHXB2R   ..............T-T-----G----C----------A----------A---C-----  104
HIVMN      ..............T-C-----G----A----------A----------A---T-----  101
HIVJRCSF   ..............C-T-----G----C----------A----------A---T-----  104
HIVOYI     ..............T-T-----A----C----------C----------G---T-----  104
HIVSF2     ..............T-T-----G----C----------A----------A---T-----  104
HIVNY5CG   ..............T-T-----G----C----------A----------A---T-----  104
HIVNL43    ..............T-T-----G----C----------A----------A---T-----  104
HIVHAN     ..............T-T-----G----C----------A----------A---T-----  104
HIVRF      ..............T-T-----G----C----------A----------A---T-----  104
CONSENSUS2 ..............tCttCAgAGCAGACCAGAGCCAACAGCCCCACCAGCAGAgAgCTTcgggtttgggGG  104
HIVELI     ..............T-TC--A----------G-G---CGGGTTTGG--  104
HIVZ2Z6    ..............T-TT--G----------G-G---CGGGTTTGG--  104
HIVNDK     ..............T-TT--G----------G-G---CGGGTTTGG--  104
HIVMAL     ..............C-TT--G----------G-G---CGGGTTTGG--  104
HIVU455    ..............T-CT--G----------A-T---...TGGGAT--  101
SIVCPZ     ..............TGTG--G-A----A----A-----G--------GAT---G-GT-ATGGGTACCA--  104

CONSENSUS1 aaGagAcAaCaacTCCCtcTCA????????????GAAGCAGGAGccGATAGACAAGGaacTGTATCCtTT  162
HIVLAI     TA-AG-C-A-AAC----TC---.....----------CC----------AAC-------T--  198
HIVHXB2R   TA-AG-C-A-AAC----CC---.....----------CC----------AAC-------T--  162
HIVMN      AA-AG-C-A-AAC----TA---GAAGCAGGAGAA----------AC----------ACC-------T--  171
HIVJRCSF   AA-AG-C-G-AAC----TC---GAAGCAGGAGCA----------CC----------AAT-------T--  174
HIVOYI     AA-AG-C-A-AAC----CC---.....----------CC----------GAC-------T--  162
HIVSF2     AG-AG-A-A-AAC----TC---.....----------CC----------AAC-------T--  162
HIVNY5CG   AA-AG-C-A-AAC----TC---.....----------CC----------AAC-------T--  162
HIVNL43    AA-AG-C-A-AAC----TC---.....----------CC----------AAC-------T--  162
HIVHAN     AA-CA-C-G-...----TC---.....----------CC----------AAC-------C--  159
HIVRF      AA-AG-C-A-...----TC---.....----------AA----------AAC-------T--  159
CONSENSUS2 agGAgAtaAc...ccCCTCtCa.....???qAAACAGGAGCaGAAAGACAaGGAActGtaTCCTTt  159
HIVELI     AA--G-TA-C...CC----T-A.....A----------A--------A----CT-TA-----T  159
HIVZ2Z6    AA--G-TA-C...CC----T-A.....G----------A--------A----CT-TA-----C  159
HIVNDK     AG--G-TA-C...CC----T-A.....G----------A--------A----CT-TA-----T  159
HIVMAL     AG--G-TA-A...AC----T-A.....G----------A--------A----TT-TA-----T  159
HIVU455    GG--A-AG-T...GA----C-C.....TGCG----------T--------G----CA-AC-----T  159
SIVCPZ     AG--AGA......GAAGAGC-A.....GGAG--GA-A--GGAG-GAG--G...TCTATA---ACC  156
```


\/ protease
<- gag cds
CONSENSUS1 AgCtTCCCTCAgaTCACTCTTTGGCAaCGACCCcTcGTCACAaTAAaGATAGGGGGGCAAcTAAAgGAAG 232
HIVLAI -A-T-------GA-----------A------C-C------A---A-----------C----G---- 268
HIVHXB2R -A-T-------GG----------A------C-C------A---A-----------C----G---- 232
HIVMN -G-T-------AA----------A------A-T------A---A-----------C----G---- 241
HIVJRCSF -A-T-------GA----------A------C-C------A---A-----------C----G---- 244
HIVOYI -A-C-------GA----------A------A-C------A---A-----------C----A---- 232
HIVSF2 -A-T-------GA----------A------C-C------A---G-----------C----G---- 232
HIVNY5CG -G-T-------GA----------G------C-C------A---A-----------T----G---- 232
HIVNL43 -G-T-------GA----------G------C-C------A---A-----------T----G---- 232
HIVHAN -G-C-------AA----------G------C-C------A---A-----------C----G---- 229
HIVRF -G-T-------AA----------A------A-C------G---A-----------T----G---- 229
CONSENSUS2 AgcTtCCCTCAAATCACTCTTTGGCAACGACCccTtGTcaCAaTAAaaaTAGGgGGACAGCTaAagGAAG 229
HIVELI -AC-T----------CC-T--CG--A---AAA----G---------A-AG---- 229
HIVZ226 -AC-G----------CC-T--TA--A---AAA----G---------A-AG---- 229
HIVNDK -GC-T----------CC-C--CA--A---AGA----G---------A-AG---- 229
HIVMAL -GC-T----------AG-T--CA--G---GAG----A---------A-AA---- 229
HIVU455 -GT-T----------CC-T--CA--G---AAA----A---------G-TA---- 229
SIVCPZ -ACCT----------G-----CC-CA-CC--G---AAG---AA--G--A--ATGT---- 226

CONSENSUS1 CTCTATTAGATACAGGAGCAGATGATACAGTAtTAGaAGAaATGaaTTTGCCAgGAAgATGGAAACCAAA 302
HIVLAI ----------T---A---A---AG-------G---G---------- 338
HIVHXB2R ----------T---A---A---AG-------G---G---------- 302
HIVMN ----------T---G---A---AA-------A---G---------- 311
HIVJRCSF ----------T---A---C---GA-------G---G---------- 314
HIVOYI ----------T---A---A---AA-------G---G---------- 302
HIVSF2 ----------T---A---A---AA-------G---A---------- 302
HIVNY5CG ----------T---A---A---AA-------G---G---------- 302
HIVNL43 ----------T---A---A---AA-------G---G---------- 302
HIVHAN ----------G---A---A---AG-------G---G---------- 299
HIVRF ----------T---A---A---AA-------G---A---------- 299
CONSENSUS2 CTCTATTAGAtACAGGAGCAGATGATACAGTaTTAGAAGAaATaAATTTGCCAGGAAAATGGAAaCCAAA 299
HIVELI ----------T----------A---------A--G----------A----- 299
HIVZ226 ----------T----------A---------A--G----------A----- 299
HIVNDK ----------T----------A---------A--A----------G----- 299
HIVMAL ----------C----------A---------A--A----------A----- 299
HIVU455 ----------T----------C---------C--A----------A----- 299
SIVCPZ --T-GC----T---------T-----------AA----GAGA--AC-A--A-A----CTT-----A----- 296

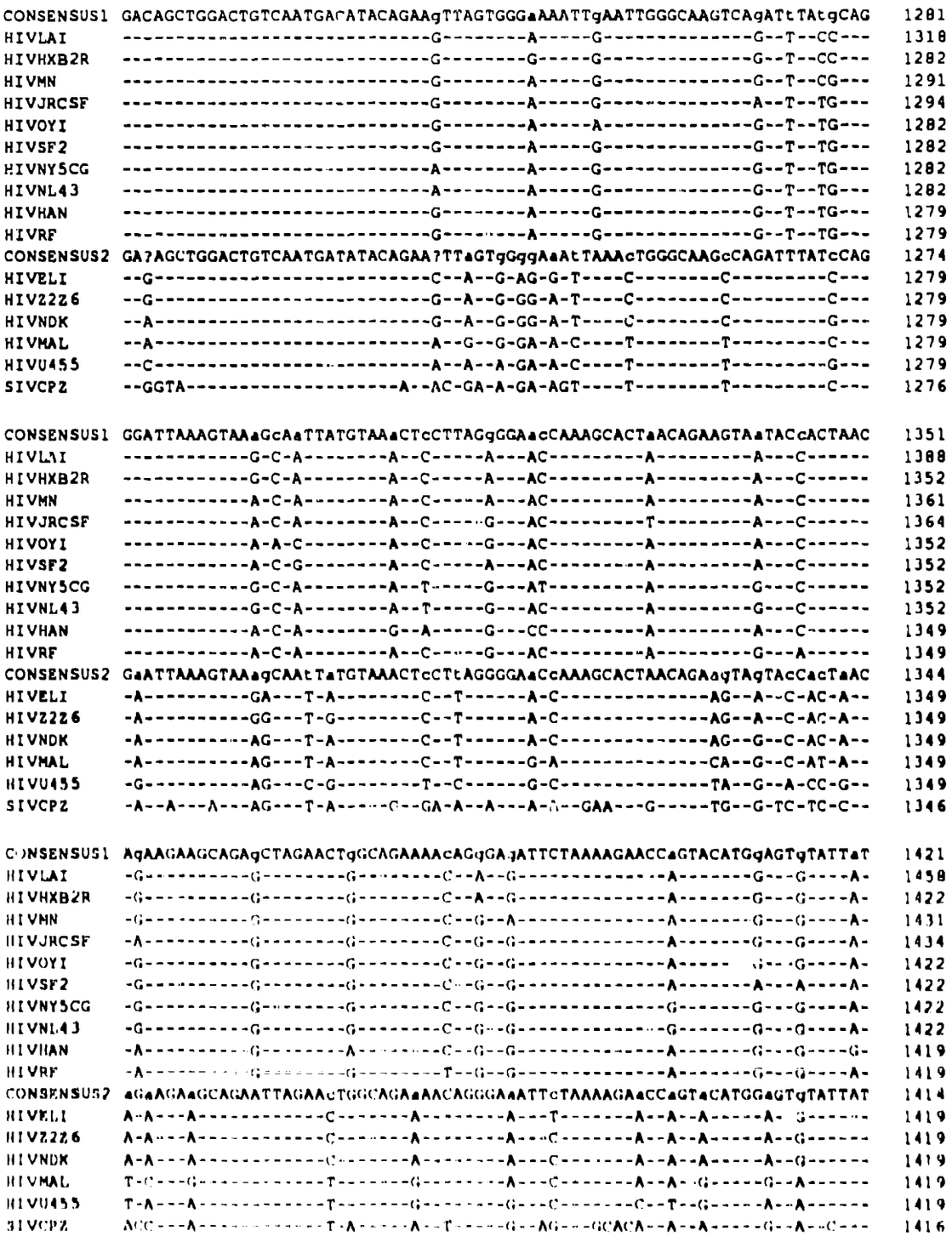
CONSENSUS1 AATGATAGGgGGAATTGGAGGTTTTATCAAAGTaaGACAGTAtGATCAgATActcaTAGaaATCTGtGGA 372
HIVLAI ---------G----------AA-------T-----G---CTCA---AA-----T--- 408
HIVHXB2R ---------G----------AA-------T-----G---CTCA---AA-----T--- 372
HIVMN ---------G----------AA-------T-----G---ACCA---GA-----T--- 381
HIVJRCSF ---------G----------AA-------T-----G---CCCA---AT-----T--- 384
HIVOYI ---------G----------AA-------T-----G---CTCA---AA-----T--- 372
HIVSF2 ---------G----------AA-------C-----G---CCTG---AA-----T--- 372
HIVNY5CG ---------G----------AA-------T-----G---CTCA---AA-----C--- 372
HIVNL43 ---------G----------AG-------T-----G---CTCA---AA-----C--- 372
HIVHAN ---------A----------AA-------T-----G---CTCG---AA-----T--- 369
HIVRF ---------G----------GA-------T-----A---CTCA---AA-----T--- 369
CONSENSUS2 AATgATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAaATACtcATAGAAATcTGTGGa 369
HIVELI ---G----------A----CC---------C-----A 369
HIVZ226 ---G----------A----TC---------C-----G 369
HIVNDK ---G----------A----TC---------C-----A 369
HIVMAL ---G----------A----TT---------T-----A 369
HIVU455 ---A----------G----TT---------T-----A 369
SIVCPZ ---G-----A----------A---A-T----A-TG---AC-----G--AGAA--G 366


protease \/ p66, p51



```
CONSENSUS1 GATGTGGGtGATGCATAtTTTTCAGTTCCCtTAGATaAAGAcTTcAGgAAgTATACTGCATTTACCATAC  861
HIVLAI     --------T--------T-----------T-----G----C--C--G--G------------------  898
HIVHXB2R   --------T--------T-----------T-----G----C--C--G--G------------------  862
HIVMN      --------T--------T-----------T-----A----C--C--G--G------------------  871
HIVJRCSF   --------T--------T-----------T-----A----C--C--G--G------------------  874
HIVOYI     --------T--------C-----------T-----A----C--C--A--G------------------  862
HIVSF2     --------T--------C-----------T-----A----C--T--A--G------------------  862
HIVNY5CG   --------T--------T-----------T-----A----C--C--G--G------------------  862
HIVNL43    --------C--------T-----------T-----A----C--C--G--G------------------  862
HIVHAN     --------T--------T-----------C-----A----C--C--A--A------------------  859
HIVRF      --------T--------T-----------T-----A----G--C--G--G------------------  859
CONSENSUS2 GATGTGGGtGAtGCATATTTtTCAGTtCCcTTAGATgAAgatTTtAGgAAaTATACtGCaTTtACcATAc  856
HIVELI     --------T--T--------T-----T--C------G--GAT--T--G--A-----C--C--T--C---T  859
HIVZ2Z6    --------T--T--------T-----T--C------A--GAC--T--G--A-----T--A--T--C---C  859
HIVNDK     --------T--T--------C-----T--C------G--GAT--T--G--A-----C--A--T--C---C  859
HIVMAL     --------G--T--------T-----C--T------G--GAT--C--G--G-----T--A--C--T---C  859
HIVU455    --------G--C--------T-----T--T------G--AGC--T--A--G-----T--G--C--C---C  859
SIVCPZ     -----A--A--T--C--C--C--TTGT--CC-G---A--GAT--C--G--A-----T--A--T--A--TC  856

CONSENSUS1 CTAGTAtAAACAATGAgACACCAgGGATTAGATATCAGTACAATGTGCTtCCACAgGGaTGGAAAGGATC  931
HIVLAI     ------T---------G------G------------------------T-----G--A-----------  968
HIVHXB2R   ------T---------G------G------------------------T-----G--A-----------  932
HIVMN      ------T---------A------G------------------------T-----G--A-----------  941
HIVJRCSF   ------T---------G------G------------------------T-----G--A-----------  944
HIVOYI     ------T---------G------G------------------------T-----G--A-----------  932
HIVSF2     ------T---------G------G------------------------G-----G--A-----------  932
HIVNY5CG   ------T---------G------G------------------------T-----G--A-----------  932
HIVNL43    ------T---------G------G------------------------T-----G--A-----------  932
HIVHAN     ------T---------G------G------------------------T-----G--A-----------  929
HIVRF      ------T---------A------C------------------------T-----A--G-----------  929
CONSENSUS2 CtAGTATaAAcAATGAGACACCAGGgaTtAGaTATCAGTACAATGTGCTtCCaCAGGGATGGAAAGGATC  926
HIVELI     -T-----A--C--------------GA-T--A-----------------T--A-----------------  929
HIVZ2Z6    -T-----A--T--------------GA-T--A-----------------T--A-----------------  929
HIVNDK     -T-----A--C--------------GA-T--A-----------------C--A-----------------  929
HIVMAL     -C-----T--T--------------GA-T--A-----------------A--A-----------------  929
HIVU455    -T-----A--C--------------AG-C--G-----------------T--G-----------------  929
SIVCPZ     -C-----A--T--------C-----AG-T--A-----A--T-----TT-G--A--A-----------G--  926

CONSENSUS1 ACCAGCaATATTCCAaaGTAGcATGACAAAAATCTTaGAGCCTTTTAgAAAACAaAATCCAGAcATAgTt 1001
HIVLAI     ------A--------AA----C--------------A----------G------A--------C---G-T 1038
HIVHXB2R   ------A--------AA----C--------------A----------G------A--------C---G-T 1002
HIVMN      ------A--------AA----C--------------A----------G------A--------C---G-T 1011
HIVJRCSF   ------A--------AA----C--------------A----------G------A--------C---A-T 1014
HIVOYI     ------G--------AA----T--------------A----------G------A--------C---G-T 1002
HIVSF2     ------A--------AA----C--------------A----------G------G--------C---G-T 1002
HIVNY5CG   ------A--------GT----C--------------G----------G------A--------C---G-C 1002
HIVNL43    ------A--------GT----C--------------A----------G------A--------C---G-C 1002
HIVHAN     ------A--------AA----C--------------A----------G------A--------C---G-T  999
HIVRF      ------A--------AA----T--------------A----------A------A--------A---G-T  999
CONSENSUS2 ACCqgCAATATTCCAaAGTAGCATGACAAAAATCTTAGAqCCCTTTAGAaaAcAAaATCCAGAaATaGTt  996
HIVELI     ---GG----------A----------------------G----------AA-C--A-------A--G-T  999
HIVZ2Z6    ---GG----------A----------------------G----------AA-C--A-------A--A-T  999
HIVNDK     ---GG----------A----------------------G----------AA-C--A-------A--A-T  999
HIVMAL     ---AG----------G----------------------A----------AC-A--A-------A--A-C  999
HIVU455    ---AT----------G----------------------G----------TC-C--C-------C--A-T  999
SIVCPZ     T--AT-T--C-----A--C--T-----------TC----A-----C---GA-A-GA----T--C--TACT  996
```

```
CONSENSUS1 ATCTAtCAATACATGGATGATTTGTATGTAGGATCTGAcTTAGAAATAGgGCAGCATAGAacAAAAATAG  1071
HIVLAI     ----------------------------------------C----------G----------AC--------  1108
HIVHXB2R   ----------------------------------------C----------G----------AC--------  1072
HIVMN      ----------------------------------------C----------G----------GC--------  1081
HIVJRCSF   ----------------------------------------C----------G----------AC--------  1084
HIVOYI     ----------------------------------------C----------G----------AC--------  1072
HIVSF2     ----------------------------------------C----------G----------AC--------  1072
HIVNY5CG   ----------------------------------------C----------G----------AC--------  1072
HIVNL43    ----------------------------------------C----------G----------AC--------  1072
HIVHAN     ----------------------------------------C----------A----------AC--------  1069
HIVRF      ----------------------------------------T----------G----------AT--------  1069
CONSENSUS2 ATcTAtCAATACATGGATGAtTTGTATGTaGGaTCTGAcTTAGAAATAGGgCAgCATAGaaCAAAAATAG  1066
HIVELI     --C--T--------------T--------A--A-----C----------G--G-----GA----------  1069
HIVZ2Z6    --C--T--------------T--------A--A-----C----------G--G-----AA----------  1069
HIVNDK     --C--T--------------T--------A--A-----C----------G--G-----AA----------  1069
HIVMAL     --A--C--------------T--------A--G-----T----------A--A-----AA----------  1069
HIVU455    --C--T--------------C--------G--A-----T----------G--A-----AG----------  1069
SIVCPZ     --T--C--G-----------CC-A-----G--G-----TC-T-----T-AT--A-----AAA---GG-G-  1066

CONSENSUS1 AGGAaCTGAGAcaaCATCTGTTgAgGTGGGGatTTACCACACCaGACAAaAAACATCAGAAAGAACCTCC  1141
HIVLAI     ----G------CAA--------G-G------AC----------A-----A--------------------  1178
HIVHXB2R   ----G------CAA--------G-G------AC----------A-----A--------------------  1142
HIVMN      ----A------CGA--------G-G------AT----------A-----A--------------------  1151
HIVJRCSF   ----A------CAA--------G-A------AT----------A-----A--------------------  1154
HIVOYI     ----A------CAA--------G-G------AT----------A-----A--------------------  1142
HIVSF2     ----A------CAG--------G-G------AT----------A-----A--------------------  1142
HIVNY5CG   ----A------CAA--------G-G------AT----------A-----A--------------------  1142
HIVNL43    ----A------CAA--------G-G------AT----------A-----A--------------------  1142
HIVHAN     ----A------CAG--------G-A------AT----------A-----A--------------------  1139
HIVRF      ----A------GAA--------A-A------GT----------G-----G--------------------  1139
CONSENSUS2 AGgAAtTAAGAGaaCATCTATTgAggTGGGGATTtAccAcaCCAGAtAAAAAaCATCAGAAAGAACCcCC  1136
HIVELI     --A--T------AA--------G-GG--------T-CC-GA-----T-----A--------------C--  1139
HIVZ2Z6    --G--T------AA--------A-GG--------T-CC-CA-----T-----A--------------C--  1139
HIVNDK     --G--T------AA--------G-GG--------T-CC-CA-----T-----A--------------T--  1139
HIVMAL     --G--C------AA--------G-AA--------T-CC-CA-----C-----G--------------C--  1139
HIVU455    --G--T------CT--------G-GC--------C-TT-CC-----C-----G--------------T--  1139
SIVCPZ     -AG--C-----CAA---T-GC-T-AA-----G--C-CA-CC-----C-----A-----A--G--G--A--  1136

CONSENSUS1 ATTcCTTTGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAqCCTATAqTGCTqCCAGAAAAa  1211
HIVLAI     ---C----------------------------------------------G------G----G--------A  1248
HIVHXB2R   ---C----------------------------------------------G------G----G--------A  1212
HIVMN      ---C----------------------------------------------G------G----A--------A  1221
HIVJRCSF   ---C----------------------------------------------G------G----G--------A  1224
HIVOYI     ---C----------------------------------------------G------A----G--------A  1212
HIVSF2     ---C----------------------------------------------G------A----G--------A  1212
HIVNY5CG   ---C----------------------------------------------G------G----G--------G  1212
HIVNL43    ---C----------------------------------------------G------G----G--------G  1212
HIVHAN     ---C----------------------------------------------G------G----G--------A  1209
HIVRF      ---T----------------------------------------------G------G----G--------A  1209
CONSENSUS2 ATTTCTTTGGATGGGqTATGAACTcCATCCTGAtAAATGGACAGTaCAGcCTATAaAacTGCCAGAaAAq  1206
HIVELI     ----------------T--------C--T-----T----------A---T-----A-AC-------A--G  1209
HIVZ2Z6    ----------------G--------C--T-----T----------A---T-----A-AT-------A--G  1209
HIVNDK     ----------------T--------C--T-----T----------A---C-----A-CC-------A--A  1209
HIVMAL     ----------------G--------C--C-----C----------G---C-----C-AC-------C--G  1209
HIVU455    ----------------G--------T--T-----C----------T---C-----C-GC-------A--A  1209
SIVCPZ     C---T-A--------A-----G--C--C--A--C----------C---C-----TC-AT-A-----A--A  1206
```

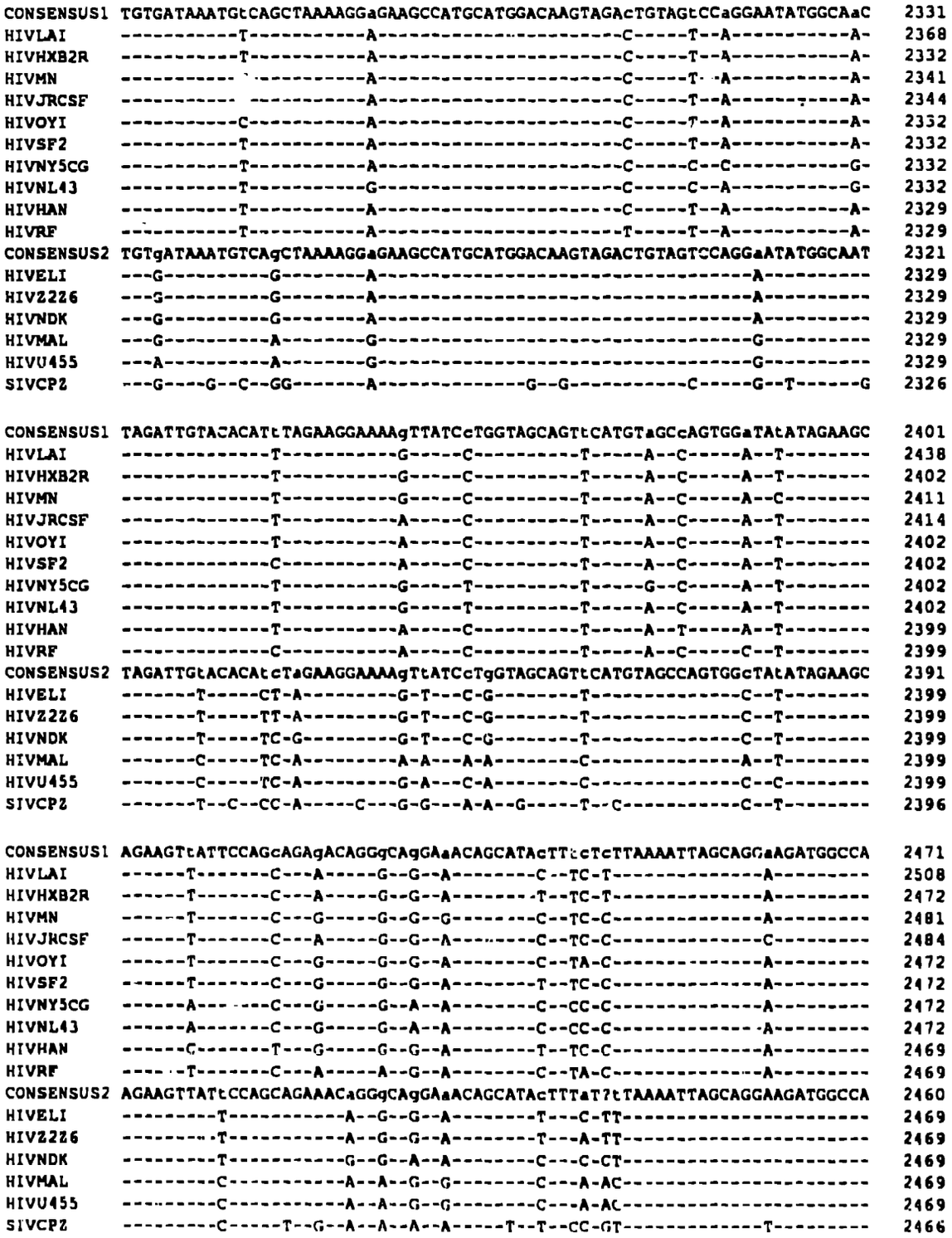


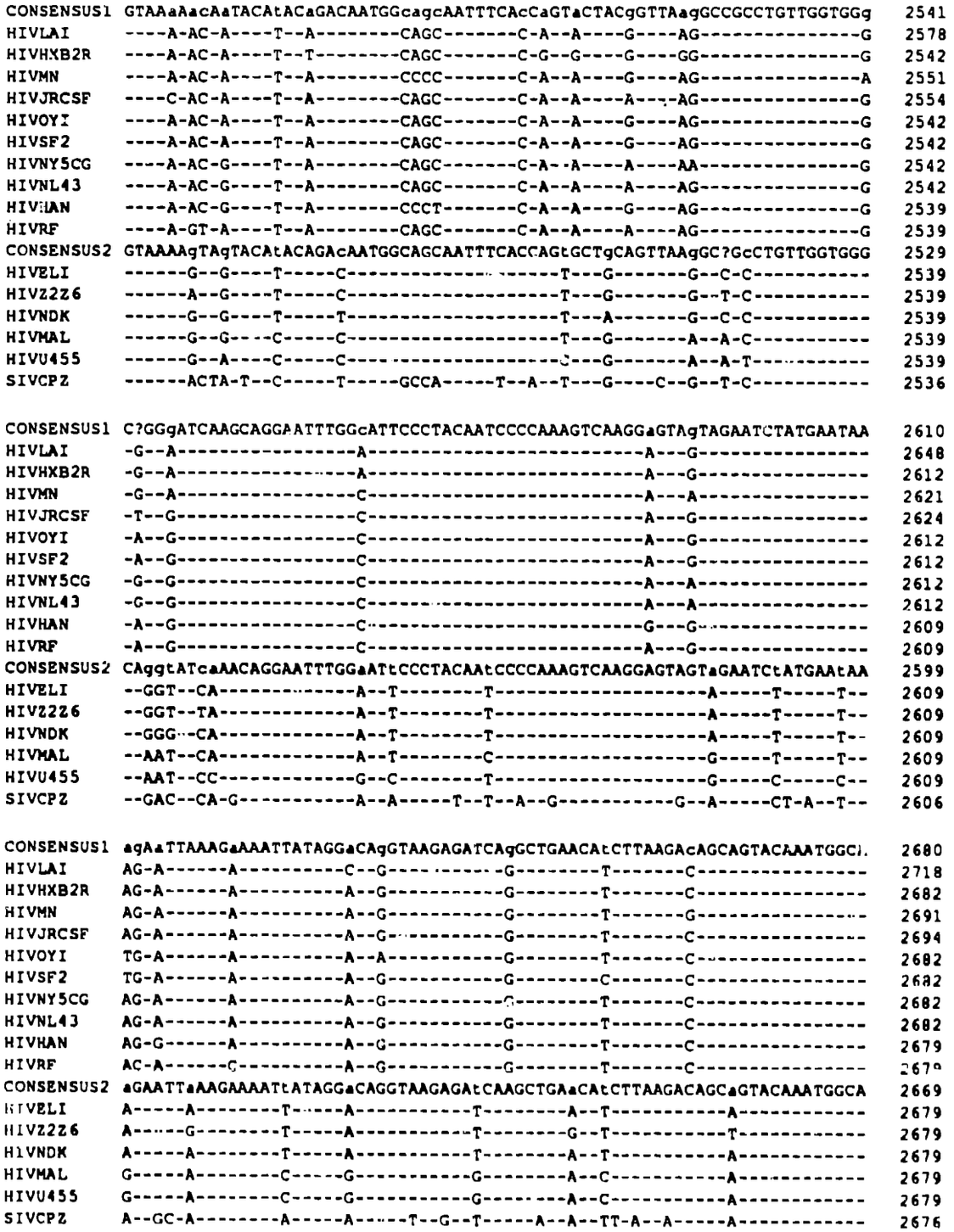
```
CONSENSUS1 GACCCATCAAAAGACTTAaTAGcAGAAaTACAGAAgCAGGGgcAAGGCCAATGGACATAtCAAATTTaTC  1491
HIVLAI     ----------------A---C----A-------G-----GC----------------T-------A--  1528
HIVHXB2R   ----------------A---C----A-------G-----GC----------------T-------A--  1492
HIVMN      ----------------A---C----G-------G-----GC----------------T-------A--  1501
HIVJRCSF   ----------------A---T----A-------G-----GC----------------T-------T--  1504
HIVOYI     ----------------G---C----T-------A-----AC----------------T-------A--  1492
HIVSF2     ----------------G---C----A-------G-----GC----------------T-------A--  1492
HIVNY5CG   ----------------A---C----A-------G-----GC----------------T-------A--  1492
HIVNL43    ----------------A---C----A-------G-----GC----------------T-------A--  1492
HIVHAN     ----------------G---C----A-------G-----GG----------------T-------A--  1489
HIVRF      ----------------A---C----A-------G-----GC----------------C-------A--  1489
CONSENSUS2 GACCCATCAAAAGACtTAaTAGCAGAAaTACAGAAaCAaGGGcAcGgcCAATGGACATAcCAAATtTAtC  1484
HIVELI     ---------------T--A--------A-------A--A---C-C-GC----------C-----T--T-  1489
HIVZ2Z6    ---------------T--A--------A-------A--A---C-C-GC----------C-----T--T-  1489
HIVNDK     ---------------T--A--------C-------A--A---G-C-GC----------C-----T--T-  1489
HIVMAL     ---------------T--A--------A-------G--G---C-A-GT----------T-----A--C-  1489
HIVU455    ---------------C--G--------A-------A--A---C-A-AC----------T-----T--T-  1489
SIVCPZ     --T---GAC-----GC-TA--------A-------A--G--CA-CTGC--G-----T--T--G--A-TT-  1486

CONSENSUS1 AAGAgCCATTTAAaAAtCTGAAAACAGGaAAgTATGCAAgaAtGAgGGGTgCCCACACTAATGATgTaAA  1561
HIVLAI     ----G--------A--T----------A--A-------GA-C--G----G--------------G-A--  1598
HIVHXB2R   ----G--------A--T----------A--A-------GA-T--G----G--------------G-A--  1562
HIVMN      ----G--------A--T----------C--A-------GA-T--G----G--------------G-A--  1571
HIVJRCSF   ----G--------A--T----------A--A-------GA-C--G----G--------------G-A--  1574
HIVOYI     ----G--------A--T----------A--G-------GA-T--G----G--------------G-A--  1562
HIVSF2     ----G--------A--T----------A--G-------GG-T--G----G--------------G-A--  1562
HIVNY5CG   ----G--------A--T----------A--G-------GA-T--A----G--------------G-G--  1562
HIVNL43    ----G--------A--T----------A--A-------GA-T--A----G--------------G-G--  1562
HIVHAN     ----A--------G--T----------A--G-------GA-T--G----G--------------A-A--  1559
HIVRF      ----G--------A--C----------A--G-------GA-T--G----G--------------G-A--  1559
CONSENSUS2 AAGAaCcATtTAAAAATCTgAAAACAGGaAAgTATGCAAGAAtgAgGggTGCcCACACTAATGATGTAAA  1554
HIVELI     ----A-C--T---------G--------A--G----------TG-G-GG---C----------------  1559
HIVZ2Z6    ----A-C--T---------G--------A--G----------TG-G-GG---C----------------  1559
HIVNDK     ----A-C--T---------A--------A--G----------CG-G-GG---C----------------  1559
HIVMAL     ----G-A--A---------G--------G--G----------TA-A-TC---C----------------  1559
HIVU455    ----G-C--T---------G--------A--A----------AA-G-TC---T----------------  1559
SIVCPZ     -G--A-C-CA---G---T-G--G-----G--A-----C---CAA-G-TCA--A-----A-----CA-C-G  1556

CONSENSUS1 ACAaTTAACAGAGGCAGTgCAAAAAaTAgCCacaGAAaGCATAGTAATATGGGGAAAGAcTCCTAAATTT  1631
HIVLAI     ---A----------G------A--A--ACA---A----------C----------               1668
HIVHXB2R   ---A------  ------G------A--A--ACA---A----------C----------               1632
HIVMN      ---A----------G------A--G--ACA---A----------C----------               1641
HIVJRCSF   ---A----------G------A--G--AAT---A----------T----------               1644
HIVOYI     ---G----------G------A--A--CAA-- A----------C----------               1632
HIVSF2     ---G----------G------G--T--ACA---A----------T----------               1632
HIVNY5CG   ---A----------A------A--G--ACA---A----------C----------               1632
HIVNL43    ---A----------A------A--G--ACA---A----------C----------               1632
HIVHAN     ---G----------G------A--G--ACA---G----------C----------               1629
HIVRF      ---A----------A------G--G--ACA---A----------C----------               1629
CONSENSUS2 aCAATTAaCAGAgGcaGTGCAAAaaaTatCCacAGAAAGCATAGTgATATGGGGAAagActCCTAAATTT  1624
HIVELI     G------G----G-CA-------GAA-AT--AC----------G----------GG-CT---------  1629
HIVZ2Z6    A------G----G-TA-------AAA-AT--AC----------G----------AG-CT---------  1629
HIVNDK     A------A----G-CA-------AAA-AG--AC----------G----------AG-CT---------  1629
HIVMAL     A------A----A-CA-------AGA-AG--CA----------A----------AA-CT---------  1629
HIVU455    A------A----A-TG-------AAG-GT--AC----------A----------AG-TC---------  1629
SIVCPZ     G---C-GG-G--A-CA--A----AAA-TG-TACT----------A--T-------AA-CA---------  1626
```

```
CONSENSUS1 AaAcTaCCCATACAAAAgGAAACATGGGAagCaTGGTGGAcaGAgTATtgGcAAGCcACCTGGATTCCTG  1701
HIVLAI     -A-C-A----------G----------AA-A-------CA--G---TG-C----C-------------  1738
HIVHXB2R   -A-C-G----------G----------AA-A-------CA--G---TG-C----C-------------  1702
HIVMN      -G-C-A----------A----------AA-A-------CA--G---AC-T----C-------------  1711
HIVJRCSF   -A-T-A----------A----------AA-A-------CA--G---TG-C----C-------------  1714
HIVOYI     -A-C-A----------A----------AG-A-------CG--G---TG-C----C-------------  1702
HIVSF2     -A-C-A----------G----------AG-A-------TG--G---TG-C----T-------------  1702
HIVNY5CG   -A-T-A----------G----------AG-G-------CG--G---TG-C----C-------------  1702
HIVNL43    -A-T-A----------G----------AG-A-------CA--G---TG-C----C-------------  1702
HIVHAN     -G-C-A----------G----------AG-A-------CG--G---TG-C----C-------------  1699
HIVRF      -A-C-A----------A----------GG-A-------CA--G---TG-C----C-------------  1699
CONSENSUS2 AgAcTACCCATACAAAAgGAAACATGGGAaaCATGGTGGatAGAgTATTGGCAaGCcACCcTGGATtCCTG  1694
HIVELI     -G-C----------G----------AA--------GCA--G---------A--C--T-----T----  1699
HIVZ2Z6    -G-T----------G----------AA--------GTA--G---------A--C--T-----T----  1699
HIVNDK     -A-C----------G----------AA--------ATA--G---------A--C--C-----T----  1699
HIVMAL     -G-C----------A----------GG--------ACA--A---------A--C--C-----C----  1699
HIVU455    -G-C----------G----------AG--------ATG--G---------G--T--C-----T----  1699
SIVCPZ     -GGT----AG----G--A----GT-----AG--------GCA--A--C-----G--A--C-----C----  1696

CONSENSUS1 AgTGGCAGtTTGTcAATACCCCTCCcTTAGTgAAaTTaTGGTACCAGTTAGAgAAaGAaCCCATAgTAGG  1771
HIVLAI     -G------T----C----------T-----G--A--A----------G--A--A------G----  1808
HIVHXB2R   -G------T----T----------C-----G--A--A----------G--A--A------G----  1772
HIVMN      -G------G----C----------C-----G--A--A----------G--A--A------G----  1781
HIVJRCSF   -G------T----C----------C-----G--A--A----------A--A--A------G----  1784
HIVOYI     -A------T----C----------C-----A--A--A----------G--A--C------G----  1772
HIVSF2     -G------T----C----------C-----G--A--A----------G--A--A------G----  1772
HIVNY5CG   -A------T----C----------C-----G--G--A----------A--G--A------A----  1772
HIVNL43    -G------T----C----------C-----G--G--A----------G--A--A------A----  1772
HIVHAN     -G------T----C----------C-----G--A--A----------G--A--A------G----  1769
HIVRF      -G------T----C----------C-----A--A--G----------A--A--A------A----  1769
CONSENSUS2 AgTGGGAaTTTGTCAATACcCCTCCttTAGTAAAAtTATGGTACCAGTTAGAgAagGAaCCCATAatAGG  1764
HIVELI     -G-----A----------C-----TT--------T----------G-AG--A------AT---  1769
HIVZ2Z6    -G-----A----------C-----TT--------T----------G-AG--A------AT---  1769
HIVNDK     -G-----A----------C-----TT--------T----------G-AG--A------AT---  1769
HIVMAL     -A-----G----------T-----CC--------C----------A-CA--A------GT---  1769
HIVU455    -A-----G----------C-----CC--------T----------G-AA--C------GC---  1769
SIVCPZ     -A-----A---A------C--A--CT----C---T--------AGT-----G-CA--A--T---CC-AC  1766

CONSENSUS1 aGCAGAAACtTTCTATGTAGATGGqGCAGCtAatAGgGAgACTAaAttAGGAAAAGCAGGaTATGTtACT  1841
HIVLAI     A--------G----------G-----T-GC--G--G----A-TT----------T---  1878
HIVHXB2R   A--------C----------G-----T-AC--G--G----A-TT----------T---  1842
HIVMN      T--------T----------G-----T-AC--G--G----A-AA----------T---  1851
HIVJRCSF   A--------T----------G-----T-AC--G--G----A-TT----------T---  1854
HIVOYI     A--------T----------G-----T-AT--G--G----A-TT----------T---  1842
HIVSF2     A--------T----------G-----T-AT--G--G----A-TT----------T---  1842
HIVNY5CG   A--------T----------A-----C-AT--G--A----A-TT----------A---  1842
HIVNL43    A--------T----------G-----C-AT--G--A----A-TT----------A---  1842
HIVHAN     A--------T----------G-----T-AT--G--G----A-TT----------T---  1839
HIVRF      A--------T----------G-----T-AT--A--G----A-TT----------T---  1839
CONSENSUS2 AGCAGAaACtTTCTATGTAGATGGGGCAGCTAATAGaGAgACTAAa.taGGAAAaGCAGGaTATGTtACT  1834
HIVELI     ------A--T----------A--G-----ATTA-----A-----A-----T---  1839
HIVZ2Z6    ------A--T----------A--G-----ATTA-----G-----A-----T---  1839
HIVNDK     ------A--T----------A--G-----ATTA-----A-----A-----T---  1839
HIVMAL     ------A--T----------A--A-----AAAG-----A-----A-----T---  1839
HIVU455    ------G--A----------G--A-----GCTA-----A-----G-----C---  1839
SIVCPZ     CA----T--T-AT----------A-----A-----G--A--A--AACT--G--A-----A-----A--A  1836
```

```
CONSENSUS1 gacAGaGGaAGACAaAAaGTTGTctCCcTAactGACACAACAAATCAGAAgACTGAgTTACAAGCAATTc  1911
HIVLAI     AAT--A--A-----A--A-----CA--C--ACT----------------G-----G----------C  1948
HIVHXB2R   AAT--A--A-----A--A-----CA--C--ACT----------------G-----G----------T  1912
HIVMN      AAC--A--A-----A--G-----CT--C--ACT----------------G-----G----------C  1921
HIVJRCSF   AGC--A--A-----A--A-----CT--C--ACA----------------A-----G----------C  1924
HIVOYI     GAC--A--A-----G--A-----CT--C--ACT----------------G-----A----------C  1912
HIVSF2     GAC--A--A-----A--A-----CT--A--GCT----------------G-----A----------C  1912
HIVNY5CG   GAC--A--T-----A--A-----CC--T--ACG----------------G-----A----------C  1912
HIVNL43    GAC--A--A-----A--A-----CC--C--ACG----------------G-----G----------C  1912
HIVHAN     GAC--G--A-----A--A-----CT--C--AAT----------------G-----G----------C  1909
HIVRF      GAC--A--A-----A--A-----CT--C--ACT----------------G-----G----------C  1909
CONSENSUS2 GACAGAGGAAGACAgAAaGTTGTccCttT?ACTGA?ACaACAAATCAgAAGACTGAgTTACAaGCAATta  1902
HIVELI     ----------G--A-----CC-TT-G-----C--G--------G--------G-----A-----TA  1909
HIVZ2Z6    ----------G--A-----CC-TT-T-----T--A--------G--------G-----A-----TA  1909
HIVNDK     ----------G--A-----CC-TT-C-----C--G--------G--------G-----A-----TA  1909
HIVMAL     ----------A--G-----CT-CT-A-----A--A--------G--------A-----A-----CC  1909
HIVU455    ----------A--G-----TT-CC-A-----G--A--------A--------A-----T-----CC  1909
SIVCPZ     ----A---G-A---A--AA-CA-TAGCT-AGAAA-T--C--T-----GC-AG-A--A---A-G--TT-GC  1906

CONSENSUS1 AtCTAGCTtTGCAgGATTCgGGatTaGAAGTAAAcATAGTaACAGACTCACAATATGCATTaGGAATCAT  1981
HIVLAI     -T------T----G-----G--AT-A--------T-----A--------------------A--------  2018
HIVHXB2R   -T------T----G-----G--AT-A--------C-----A--------------------A--------  1982
HIVMN      -T------T----A-----A--GT-A--------C-----A--------------------A--------  1991
HIVJRCSF   -C------T----G-----A--AT-A--------C-----A--------------------A--------  1994
HIVOYI     -T------C----G-----G--AT-A--------C-----A--------------------A--------  1982
HIVSF2     -T------T----G-----G--AT-A--------C-----A--------------------A--------  1982
HIVNY5CG   -T------T----G-----G--AT-A--------C-----G--------------------G--------  1982
HIVNL43    -T------T----G-----G--AT-A--------C-----G--------------------G--------  1982
HIVHAN     -T------T----G-----G--AC-A--------C-----A--------------------A--------  1979
HIVRF      -T------T----G-----G--AT-A--------C-----A--------------------G--------  1975
CONSENSUS2 AtcTAGCtTTgCAGGATTCgGGATtAGAAGTAAAcATAGTAACAGAtTCACAaTATGCAtTAGGaATcAT  1972
HIVELI     -TC----C--G--------G----T---------C----------T-----A------T----A--C--  1979
HIVZ2Z6    -TT----T--G--------G----T---------C----------T-----A------T----A--C--  1979
HIVNDK     -TC----T--A--------G----T---------C----------T-----A------C----A--C--  1979
HIVMAL     -CT----T--A--------A----C---------C----------C-----G------T----G--T--  1979
HIVU455    -TC----C--G--------A----C---------T----------C-----G------T----G--C--  1979
SIVCPZ     TTC-G--C--G--------A-ATCA-C-G--T--C-----G--T--C-----A----TGT----G--T--  1976

CONSENSUS1 TCAAGCACAACCAGATaAaAGTGAATCAGAGTTAGTCAgTCAaATAATAGAgcAGtTAATAAAAAAGGAA  2051
HIVLAI     ----------------A A----------------A---A--------GC--T----------------  2088
HIVHXB2R   ----------------C-A----------------A---A--------GC--T----------------  2052
HIVMN      ----------------A-A----------------G---A--------GC--T----------------  2061
HIVJRCSF   ----------------A-A----------------G---A--------AC--C----------------  2064
HIVOYI     ----------------A-A----------------G---A--------GC--T----------------  2052
HIVSF2     ----------------A-G----------------G---A--------GC--T----------------  2052
HIVNY5CG   ----------------A-G----------------G---A--------GC--T----------------  2052
HIVNL43    ----------------A-G----------------G---A--------GC--T----------------  2052
HIVHAN     ----------------A-A----------------G---A--------GC--T----------------  2049
HIVRF      ----------------A-A----------------G---G--------GC--T----------------  2049
CONSENSUS2 TCAaGCACAACCAGAtAaqAGTGAATCAGAgtTaGTcAaTCAAATAATAGAGcAgtTAATAaAaAAGGAa  2042
HIVELI     ---A----------T-AG----------GT-A--C-A-------------C-GT-----A-A-----A  2049
HIVZ2Z6    ---A----------T-AG----------GT-A--C-G-------------C-GT-----A-A-----A  2049
HIVNDK     ---A----------T-AG----------GT-A--C-G-------------C-GC-----A-A-----A  2049
HIVMAL     ---A----------T-AA----------GA-T--T-A-------------C-AT-----C-G-----C  2049
HIVU455    ---G----------C-GC----------AA-A--C-A-------------A-GC-----G-A-----A  2049
SIVCPZ     ---GT---G-----TGAC----------AT-A--C-A---G--------AG-GT----TA-A--A--A  2046
```

```
CONSENSUS1 AAgGTCTAcCTGGCATGGGTACCAGCACAcAAAGGaATTGGAGGAAATGAACAaGTAGATaaaTTaGTCA  2121
HIVLAI     --G-----T--------------------C-----A----------------A------AAA--A----  2158
HIVHXB2R   --G-----T--------------------C-----A----------------A------AAA--A----  2122
HIVMN      --G-----T--------------------C-----A----------------A------AAA--A----  2131
HIVJRCSF   --A-----C--------------------C-----A----------------G------AAA--A----  2134
HIVOYI     --G-----C--------------------C-----A----------------A------AAA--A----  2122
HIVSF2     --G-----C--------------------C-----A----------------A------AAA--A----  2122
HIVNY5CG   --A-----C--------------------T-----A----------------A------AAA--A----  2122
HIVNL43    --A-----C--------------------C-----A----------------A------GGG--G----  2122
HIVHAN     --G-----C--------------------C-----A----------------A------AAA--A----  2119
HIVRF      --G-----C--------------------C-----G----------------A------AGA--A----  2119
CONSENSUS2 AAgGTtTACCTGgCATGGGTACCAGCaCAcAAAGGaATTGGAGGAAATGAaCAAGTAGATAAATTAGTCA  2112
HIVELI     --G--T------G-------------A--C-----A--------------A------------------  2119
HIVZ2Z6    --G--T------G-------------A--T-----A--------------A------------------  2119
HIVNDK     --G--T------G-------------A--C-----A--------------A------------------  2119
HIVMAL     --G--C------T-------------A--C-----G--------------A------------------  2119
HIVU455    --A--C------T-------------G--C-----G--------------G------------------  2119
SIVCPZ     --AA-C-----CT-C-----------A--C-----T--A----------G-----G-------------  2116

CONSENSUS1 GTgCTGGAATCAGgAAAGTaCTATTTTTagATGGAATAGATAAGGCCCAAGAaGAaCATGAgAAATATCA  2191
HIVLAI     --G----------G-----A--------AG----------------------T--A-----G--------  2228
HIVHXB2R   --G----------G-----A--------AG----------------------T--A-----G--------  2192
HIVMN      --G----------G-----A--------AG----------------------A--C-----G--------  2201
HIVJRCSF   --G----------G-----G--------AG----------------------A--T-----A--------  2204
HIVOYI     --G----------G-----A--------GG----------------------G--A-----G--------  2192
HIVSF2     --G----------G-----A--------GA----------------------A--A-----G--------  2192
HIVNY5CG   --G----------G-----A--------AG----------------------A--A-----G--------  2192
HIVNL43    --G----------G-----A--------AG----------------------A--A-----G--------  2192
HIVHAN     --G----------A-----A--------AG----------------------A--C-----G--------  2189
HIVRF      --A----------G-----A--------GG----------------------T--A-----G--------  2189
CONSENSUS2 Gtca?GGAATCAGgAAaGTaCTATTTTTgGATGGaATAGATAAgGCTCAaGAaGAaCATGAgAAATATCA  2181
HIVELI     -TCAA--------G--A--A--------G-----A--------G-----A--A--A-----G--------  2189
HIVZ2Z6    -TCAG--------G--A--A--------G-----A--------A-----A--A--A-----G--------  2189
HIVNDK     -TCAG--------G--A--A--------G-----A--------G-----G--A--A-----G--------  2189
HIVMAL     -CAGT--------A--G--A--------A-----G--------G-----A--A--A-----A--------  2189
HIVU455    -TTCT--------G--G--G--------A-----G--------G-----A--G--C-----A--------  2189
SIVCPZ     -TGCA--------G--A--G-----CC-A-----A-----C-GG--C--A--A--A-----A-GG-----  2186

CONSENSUS1 CAGTAAtTGGAGAGCAATGGCTAgTGAtTTTAACCTgCCACCTgTAGTAGCAAAAGAAATAGTAGCCAGC  2261
HIVLAI     ------T----------------G---T--------G------G------------------------  2298
HIVHXB2R   ------T----------------G---T--------G------G------------------------  2262
HIVMN      ------T----------------G---C--------A------A------------------------  2271
HIVJRCSF   ------T----------------G---T--------G------A------------------------  2274
HIVOYI     ------C----------------G---T--------G------G------------------------  2262
HIVSF2     ------T----------------G---T--------G------G------------------------  2262
HIVNY5CG   ------T----------------G---T--------A------G------------------------  2262
HIVNL43    ------T----------------G---T--------A------G------------------------  2262
HIVHAN     ------T----------------A---T--------G------G------------------------  2259
HIVRF      ------T----------------G---T--------G------G------------------------  2259
CONSENSUS2 CaaCAAtTGGAGAGCAATGGCTAGTGAtTTTAAcCTaCCACCtgTgGTAGCgAAaGAAATAGTAGCtAGC  2251
HIVELI     -AA---T----------------T-----C--A-----CG-G-----A--A----------T---  2259
HIVZ2Z6    -AA---T----------------T-----C--A-----TG-G-----A--A----------T---  2259
HIVNDK     -AA---T----------------T-----C--A-----TG-G-----G--A----------T---  2259
HIVMAL     -AG---T----------------C-----T--A-----TA-A-----G--G----------C---  2259
HIVU455    -TG---C----------------T-----T--G-----TG-G-----G--G----------C---  2259
SIVCPZ     TAGT--T----A---T----------T-----TT-A-----CA-A-----A--A--------G--CCAT  2256
```

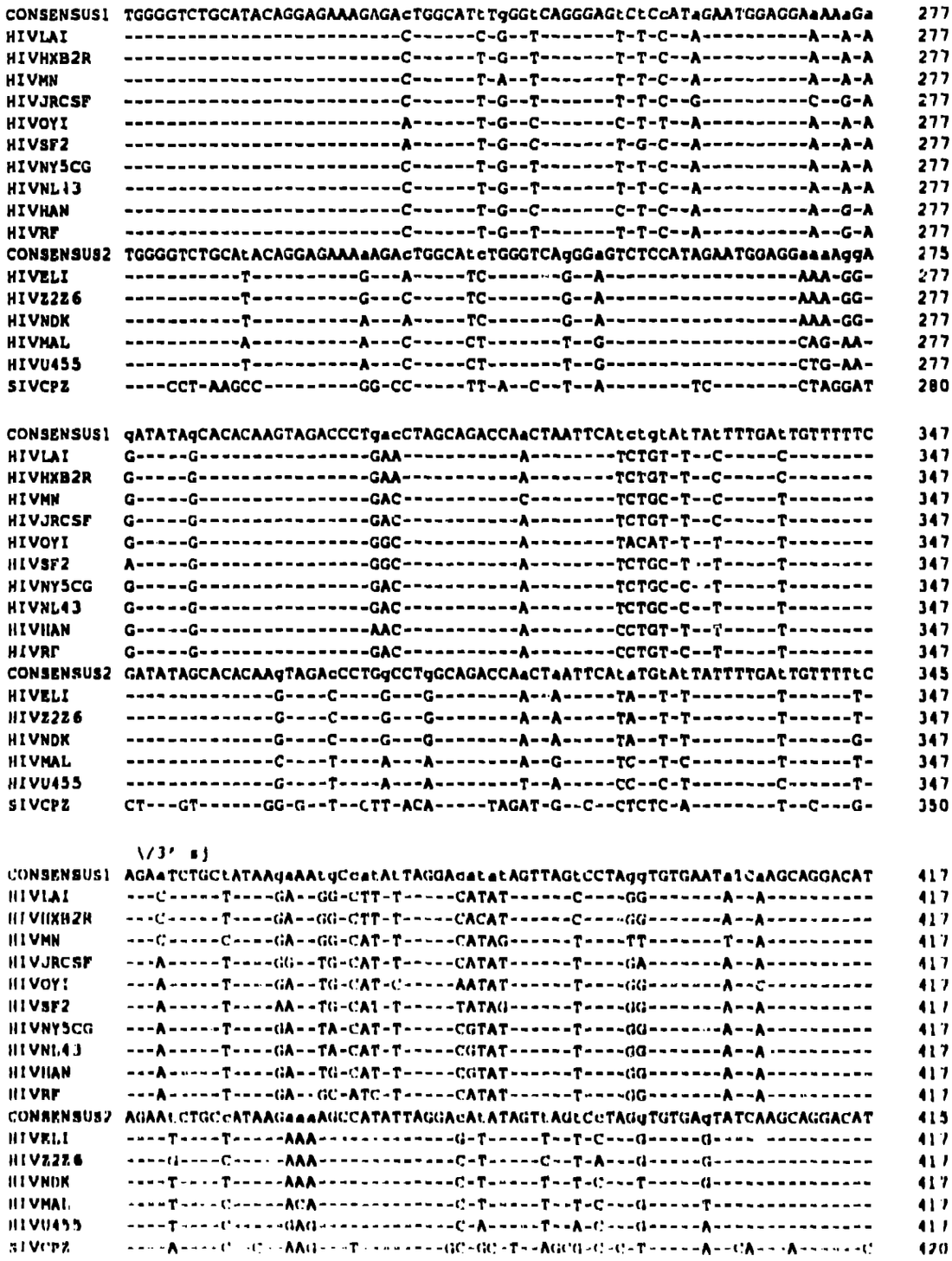

CONSENSUS1 TGTGATAAATGtCAGCTAAAAGGaGAAGCCATGCATGGACAAGTAGAcTGTAGtCCaGGAATATGGCAaC 2331
HIVLAI ----------T----------A----------C-----T--A----------A- 2368
HIVHXB2R ----------T----------A----------C-----T--A----------A- 2332
HIVMN ----------T----------A----------C-----T--A----------A- 2341
HIVJRCSF ----------T----------A----------C-----T--A----------A- 2344
HIVOYI ----------C----------A----------C-----T--A----------A- 2352
HIVSF2 ----------T----------A----------C-----T--A----------A- 2332
HIVNY5CG ----------T----------A----------C-----C--C----------G- 2332
HIVNL43 ----------T----------G----------C-----C--A----------G- 2332
HIVHAN ----------T----------A----------C-----T--A----------A- 2329
HIVRF ----------T----------A----------T-----T--A----------A- 2329
CONSENSUS2 TGTgATAAATGTCAgCTAAAAGGaGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGaATATGGCAAT 2321
HIVELI ---G----------G--------A---------- 2329
HIVZ2Z6 ---G----------G--------A---------- 2329
HIVNDK ---G----------G--------A---------- 2329
HIVMAL ---G----------A--------G---------- 2329
HIVU455 ---A----------A--------G---------- 2329
SIVCPZ ---G----G--C--GG-------A----------G--G----------C-----G--T------G 2326
CONSENSUS1 TAGATTGTACACATtTAGAAGGAAAAgTTATCcTGGTAGCAGTtCATGTaGCcAGTGGaTAtATAGAAGC 2401
HIVLAI ----------T----------G-----C----------T-----A--C-----A--T-------- 2438
HIVHXB2R ----------T----------G-----C----------T-----A--C-----A--T-------- 2402
HIVMN ----------T----------G-----C----------T-----A--C-----A--C-------- 2411
HIVJRCSF ----------T----------A-----C----------T-----A--C-----A--T-------- 2414
HIVOYI ----------T----------A-----C----------T-----A--C-----A--T-------- 2402
HIVSF2 ----------C----------A-----C----------T-----A--C-----A--T-------- 2402
HIVNY5CG ----------T----------G-----T----------T-----G--C-----A--T-------- 2402
HIVNL43 ----------T----------G-----T----------T-----A--C-----A--T-------- 2402
HIVHAN ----------T----------A-----C----------T-----A--T-----A--T-------- 2399
HIVRF ----------C----------A-----C----------T-----A--C-----C--T-------- 2399
CONSENSUS2 TAGATTGtACACAtcTaGAAGGAAAAgTtATCcTgGTAGCAGTtCATGTAGCCAGTGGcTAtATAGAAGC 2391
HIVELI -------T-----CT-A---------G-T---C-G--------T----------C--T-------- 2399
HIVZ2Z6 -------T-----TT-A---------G-T---C-G--------T----------C--T-------- 2399
HIVNDK -------T-----TC-G---------G-T---C-G--------T----------C--T-------- 2399
HIVMAL -------C-----TC-A---------A-A---A-A--------C----------A--T-------- 2399
HIVU455 -------C-----TC-A---------G-A---C-A--------C----------C--C-------- 2399
SIVCPZ -------T--C--CC-A-----C---G-G---A-A--G-----T--C----------C--T-------- 2396
CONSENSUS1 AGAAGTtATTCCAGcAGAgACAGGgCAgGAaACAGCATAcTTtcTcTTAAAATTAGCAGGaAGATGGCCA 2471
HIVLAI ------T-------C---A-----G--G--A--------C--TC-T----------A--------- 2508
HIVHXB2R ------T-------C---A-----G--G--A--------T--TC-T----------A--------- 2472
HIVMN ------T-------C---G-----G--G--G--------C--TC-C----------A--------- 2481
HIVJRCSF ------T-------C---A-----G--G--A--------C--TC-C----------C--------- 2484
HIVOYI ------T-------C---G-----G--G--A--------C--TA-C----------A--------- 2472
HIVSF2 ------T-------C---G-----G--G--A--------T--TC-C----------A--------- 2472
HIVNY5CG ------A-------C---G-----G--A--A--------C--CC-C----------A--------- 2472
HIVNL43 ------A-------C---G-----G--A--A--------C--CC-C----------A--------- 2472
HIVHAN ------G-------T---G-----G--G--A--------T--TC-C----------A--------- 2469
HIVRF ------T-------C---A-----A--G--A--------C--TA-C----------A--------- 2469
CONSENSUS2 AGAAGTTATtCCAGCAGAAACaGGgCAgGAaACAGCATAcTTTaT?tTAAAATTAGCAGGAAGATGGCCA 2460
HIVELI ---------T----------A--G--G--A--------T---C-TT---------- 2469
HIVZ2Z6 ---------T----------A--G--G--A--------T---A-TT---------- 2469
HIVNDK ---------T----------G--G--A--A--------C---C-CT---------- 2469
HIVMAL ---------C----------A--A--G--G--------C---A-AC---------- 2469
HIVU455 ---------C----------A--A--G--G--------C---A-AC---------- 2469
SIVCPZ ---------C-----T--G--A--A--A--A-----T--T--CC-GT----------T--------- 2466



```
CONSENSUS1 GTATTcATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGgTACAGTGCAGGGGAAAGAATAgTAGacA  2750
HIVLAI     -----C-----------------------------------G--------------------G---AC-  2788
HIVHXB2R   -----C-----------------------------------G--------------------G---AC-  2752
HIVMN      -----C-----------------------------------G--------------------G---GC-  2761
HIVJRCSF   -----C-----------------------------------G--------------------A---AC-  2764
HIVOYI     -----T-----------------------------------G--------------------G---AT-  2752
HIVSF2     -----C-----------------------------------A--------------------G---AC-  2752
HIVNY5CG   -----C-----------------------------------G--------------------G---AC-  2752
HIVNL43    -----C-----------------------------------G--------------------G---AC-  2752
HIVHAN     -----C-----------------------------------G--------------------G---AC-  2749
HIVRF      -----C-----------------------------------G--------------------G---AC-  2749
CONSENSUS2 GTaTTcATcCACAATTTTAAAAGAAaAgGGGGGATTGGGGGaTACAGTGCAGGGGAaAGAATAATAGACA  2739
HIVELI     --A--C--C----------------G-A-------------A--------------A-------------  2749
HIVZ2Z6    --A--C--C----------------A-G-------------A--------------G-------------  2749
HIVNDK     --A--T--C----------------A-G-------------A--------------A-------------  2749
HIVMAL     --G--C--T----------------A-G-------------G--------------A-------------  2749
HIVU455    --A--C--T----------------A-G-------------A--------------A-------------  2749
SIVCPZ     --G--C--T----------------A-G-------------G----C---------A-------------  2746

CONSENSUS1 TAATAGCaACAGAcATACAAAcTAAAGAAtTACAAAAaCAAATTACAAAaATTCAAAATTTTCGGGTTTA  2820
HIVLAI     -------A-----C-------C-------T-------A-----------A--------------------  2058
HIVHXB2R   -------A-----C-------C-------T-------A-----------A--------------------  2822
HIVMN      -------A-----C-------C-------C-------A-----------A--------------------  2831
HIVJRCSF   -------A-----C-------C-------T-------A-----------A--------------------  2834
HIVOYI     -------T-----C-------C-------C-------A-----------A--------------------  2822
HIVSF2     -------A-----C-------C-------C-------G-----------A--------------------  2822
HIVNY5CG   -------A-----C-------T-------T-------A-----------A--------------------  2822
HIVNL43    -------A-----C-------C-------T-------A-----------A--------------------  2822
HIVHAN     -------A-----C-------C-------T-------A-----------A--------------------  2819
HIVRF      -------A-----C-------C-------C-------A-----------A--------------------  2819
CONSENSUS2 TaATAGCAACAGACATACAAACTAaAGAATTACAAAAACAAATtacAAAAATTCAAAATTTTCGGGTTTA  2809
HIVELI     -A----------------------A-----------------TAT----------------------  2819
HIVZ2Z6    -A----------------------A-----------------CAC----------------------  2819
HIVNDK     -A----------------------G-----------------CAT----------------------  2819
HIVMAL     -G----------------------A-----------------TAC----------------------  2819
HIVU455    -A----------------------A-----------------TTC----------------------  2819
SIVCPZ     -A--------------------A-GC----------------TTT----G-------A----------  2816

                \/3' sj                                          5' sj\/
CONSENSUS1 TTACAGGGACAgCAgAgAtCCacTTTGGAAAGGACcaGCAAAGCTtCTcTGGAAAGGTGAAGGGGCAGTA  2890
HIVLAI     ----------G--G-G-T--AC-----------CA--------C--C--------------------  2928
HIVHXB2R   ----------G--G-A-T--AC-----------CA--------C--C--------------------  2892
HIVMN      ----------G--G-G-T--AC-----------CA--------T--C--------------------  2901
HIVJRCSF   ----------A--G-G-T--AA-----------CA--------T--C--------------------  2904
HIVOYI     ----------G--G-G-A--AC-----------CA--------T--T--------------------  2892
HIVSF2     ----------A--A-G-T--CC-----------CA--------T--C--------------------  2892
HIVNY5CG   ----------G--G-G-T--AG-----------CA--------C--C--------------------  2892
HIVNL43    ----------G--G-G-T--AG-----------CA--------C--C--------------------  2892
HIVHAN     ----------G--G-G-A--AT-----------CA--------T--C--------------------  2889
HIVRF      ----------G--G-G-T--AC-----------AC--------T--C--------------------  2889
CONSENSUS2 TTACAGgGACAgCAGAGAtCCaATTTGGAAAGGACCAGCAAAgCT?CTCTGGAAAGGTGAAGGGGCAGTA  2878
HIVELI     ------A----G------T--A--------------------G--C--------------------  2889
HIVZ2Z6    ------G----G------T--A--------------------G--C--------------------  2889
HIVNDK     ------G----G------T--A--------------------G--T--------------------  2889
HIVMAL     ------G----A------C--A--------------------A--A--------------------  2889
HIVU455    ------G----G------C--C--------------------A--A--------------------  2889
SIVCPZ     ------A----G------C--A-------------G----CCT-A--G--------------------  2886
```

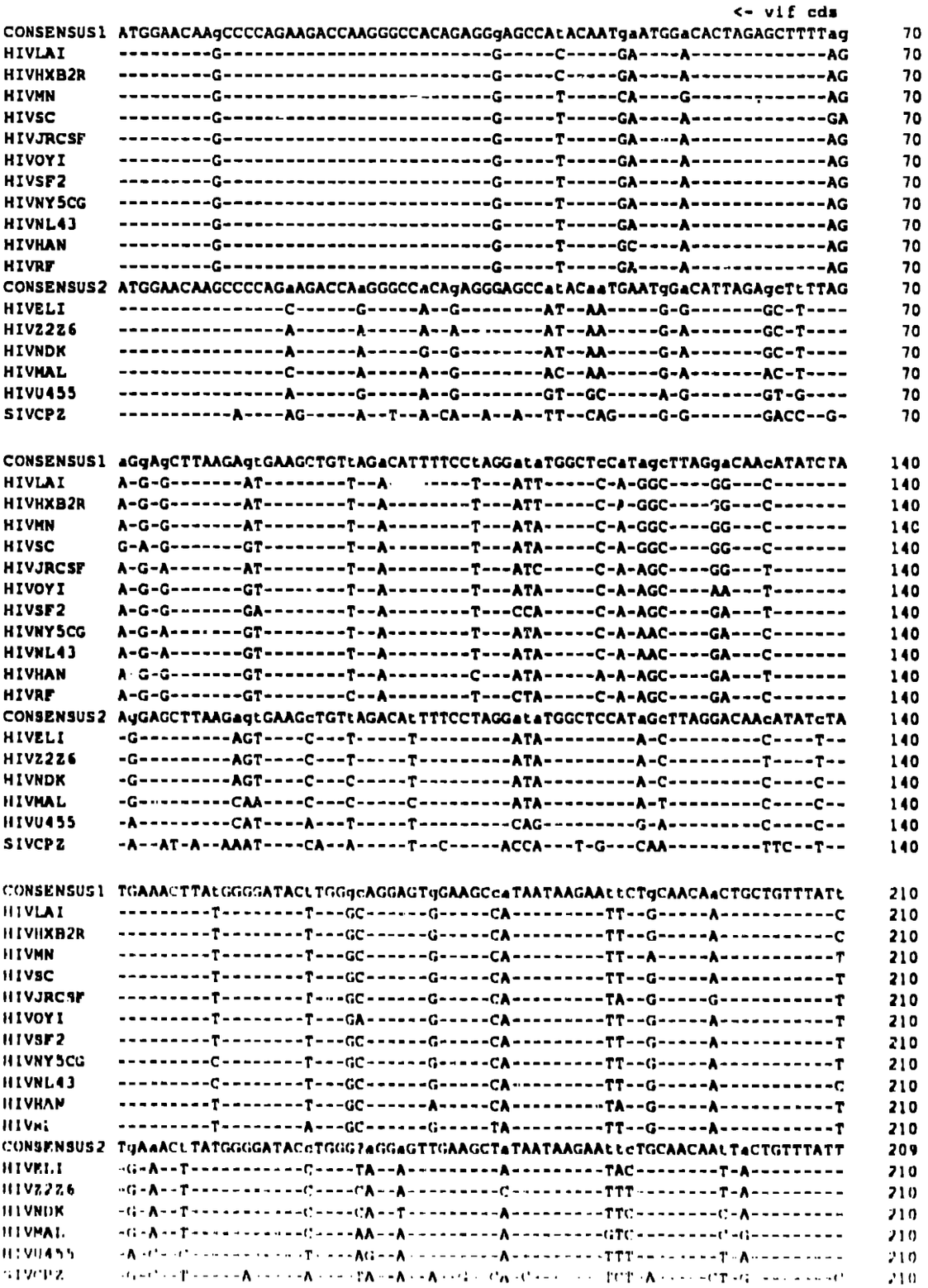
```
                                                                                vif cds ->
CONSENSUS1 GTAATACAAGATAATAgTGACATAAAAGTAGTGCCAAGAAGaAAAGcAAAgaTCATtAGGGATTATGGAA   2960
HIVLAI     ----------------G------------------------A----C---GA----T-------------   2998
HIVHXB2R   ----------------G------------------------A----C---GA----T-------------   2962
HIVMN      ----------------A------------------------A----C---GG----T-------------   2971
HIVJRCSF   ----------------G------------------------A----T---AA----T-------------   2974
HIVOYI     ----------------G------------------------G----C---GA----T-------------   2962
HIVSF2     ----------------G------------------------A----C---AA----T-------------   2962
HIVNY5CG   ----------------G------------------------A----C---GA----C-------------   2962
HIVNL43    ----------------G------------------------A----C---GA----C-------------   2962
HIVHAN     ----------------G------------------------A----C---GA----T-------------   2959
HIVRF      ----------------G------------------------A----C---GA----T-------------   2959
CONSENSUS2 GTAATACAaGACAAtAGTGAcATAAAGGTAGTACCAAGAAGAAAAGtAAAgATcATtAGGGATTATGGAA   2948
HIVELI     --------A-----G-----C-------------------------T---G--T--T-------------   2959
HIVZ2Z6    --------A-----T-----C-------------------------T---G--T--C-------------   2959
HIVNDK     --------A-----T-----C-------------------------T---G--C--T-------------   2959
HIVMAL     --------G-----T-----T-------------------------C---A--C--T-------------   2959
HIVU455    --------A-----T-----T-------------------------C---G--C--T-------------   2959
SIVCPZ     --G--C--G--TC-AG-G--AC-------------------G----C---G--C--T--A----------   2956

CONSENSUS1 AACAGAtGGcagGTGATGATTGTGTGGCAAGtAGACAGGATGAGGAT***   3010
HIVLAI     ------T--CAG-------------------T------------------   3048
HIVHXB2R   ------T--CAG-------------------T------------------   3012
HIVMN      ------C--CAG-------------------C------------------   3021
HIVJRCSF   ------T--CAG-------------------T------------------   3024
HIVOYI     ------T--CAG-------------------T------------------   3012
HIVSF2     ------T--CAG-------------------T------------------   3012
HIVNY5CG   ------T--CAG-------------------T------------------   3012
HIVNL43    ------T--CAG-------------------T------------------   3012
HIVHAN     ------T--GCA-------------------T------------------   3009
HIVRF      ------T--CAG-------------------T------------------   3009
CONSENSUS2 AACAGATGGCAGGTGATGATTGTgTGGCAaGTaGACAGGATGAgGAT---   2998
HIVELI     -----------------------G-----A--A----------G------   3009
HIVZ2Z6    -----------------------G-----A--A----------G------   3009
HIVNDK     -----------------------G-----A--A----------G------   3009
HIVMAL     -----------------------G-----G--G----------G------   3009
HIVU455    -----------------------A-----G--A----------A------   3009
SIVCPZ     -----------------------G-----A--A-----A----G------   3006
```

```
                                                                  <- pol cds
CONSENSUS1 ATGGAAAACAGAtGGcagGTGATGATTGTGTGGCAAGtAGACAGGATGAGGATTAgaACATGGAAaAGtt   70
HIVLAI     ------------T--CAG--------------------T-----------------GA--------A--TT   70
HIVHXB2R   ------------T--CAG--------------------T-----------------GA--------A--TT   70
HIVMN      ------------C--CAG--------------------C-----------------GA--------A--TT   70
HIVJRCSF   ------------T--CAG--------------------T-----------------GA--------C--TT   70
HIVOYI     ------------T--CAG--------------------T-----------------GA--------A--TT   70
HIVSF2     ------------T--CAG--------------------T-----------------GA--------A--TT   70
HIVNY5CG   ------------T--CAG--------------------T-----------------AC--------A--AT   70
HIVNL43    ------------T--CAG--------------------T-----------------AC--------A--AT   70
HIVHAN     ------------T--GCA--------------------T-----------------GA--------C--TT   70
HIVRF      ------------T--CAG--------------------T-----------------GA--------A--TT   70
CONSENSUS2 ATGGAAAACAGATGGCAGGTGATGATTGTqTGGCAaGTaGACAGGATGAgGATTAgaACATGGaAaAGTT   70
HIVELI     ----------------------------G-----A--A----------G-----AA------A-A----   70
HIVZ2Z6    ----------------------------G-----A--A----------G-----GA------A-A----   70
HIVNDK     ----------------------------G-----A--A----------G-----AC------A-A----   70
HIVMAL     ----------------------------G-----G--G----------G-----GA------C-C----   70
HIVU455    ----------------------------A-----G--A----------A-----GA------A-C----   70
SIVCPZ     ----------------------------G-----A--A-----A----G-----AG--T---A-T----   70

CONSENSUS1 TAGTaAAACACCATATqTATatTTCAagGAAACCTAaqqgatGGttTTATAgACAtCACTATGAAAGcaC  140
HIVLAI     ----A----------G---GT----GG---------GGGGAT--TT-----G---T----------CC-  140
HIVHXB2R   ----A----------G---GT----GG---------GGGGAT--TT-----G---T----------CC-  140
HIVMN      ----A----------G---AT----AA---------AAGGAC--TT-----G---T----------CA-  140
HIVJRCSF   ----A----------G---AT----GG---------AGGGAT--AT-----A---T----------CA-  140
HIVOYI     ----A----------G---GT----AA---------AGGGAT--TT-----G---T----------CA-  140
HIVSF2     ----A----------G---AT----AA---------AAGGAT--TT-----G---T----------TA-  140
HIVNY5CG   ----A----------G---AT----AG---------AGGACT--TT-----G---T----------TA-  140
HIVNL43    ----A----------G---AT----AG---------AGGACT--TT-----G---T----------TA-  140
HIVHAN     ----A----------A---TG----CG---------AGAACT--GT-----G---T----------CA-  140
HIVRF      ----A----------G---AT----AG---------AGGGAT--TT-----G---T----------CA-  140
CONSENSUS2 TAGTAAAAcAcCAtATGTATGTtTCaAAGAAAGCTaacaqaTGGTTTTATAGACATCACTATGAaAGcc?  139
HIVELI     --------C-C--T--------T--A---------AACAGA----------------------A--CCC  140
HIVZ2Z6    --------C-C--T--------T--A---------AGCAGA----------------------C--CCC  140
HIVNDK     --------T-C--T--------T--A---------AACAGA----------------------C--CCA  140
HIVMAL     --------C-T--T--------C--A---------AAAAAT----------------------A--CAG  140
HIVU455    --------C-T--C--------C--C---------CAAGGT----------------------A--TAG  140
SIVCPZ     ----T---T-T--T--T---AGG--A--A--G--AAGAGGT--------C--------T-----TCACCC  140

CONSENSUS1 TcATCCAAqAaTAAGTTCAGAaGTACAcATCCCaCtAGGGGAt...GctAqaTTqGTAaTAACAACATAT  207
HIVLAI     -C------G-A----------A-----C-----A-T------T...-CT-GA--G---A----------  207
HIVHXB2R   -C------G-A----------A-----C-----A-T------T...-CT-GA--G---A----------  207
HIVMN      -C------G-A----------A-----C-----A-T------T...-CT-GA--G---A----------  207
HIVJRCSF   -A------G-G----------A-----A-----A-T------T...-CT-GA--G---A----------  207
HIVOYI     -C------G-A----------A-----C-----A-T------T...-CT-CC--G---G----------  207
HIVSF2     -C------G-G----------A-----C-----C-T------T...-CT-AA--G---A----------  207
HIVNY5CG   -A------A-A----------A-----C-----A-T------T...-CT-AA--A---A--------  --  207
HIVNL43    -A------A-A----------A-----C-----A-T------T...-CT-AA--A---A----------  207
HIVHAN     -A------G-C----------A-----C-----A-T------A...-CT-AA--G---A----------  207
HIVRF      -C------G-A----------G-----C-----A-C------T...-AA-GG--G---A----------  207
CONSENSUS2 cCAccCAAaAaTAAGTTCAGAAGTACACATcCCACTAGGaGAa...GCTAqacTqGTAqTAA?AACATAT  205
HIVELI     C--CC---A-A----------C--------A--A...----GAC-G-  -A---A--------  207
HIVZ2Z6    C--CC---A-A----------T--------A--A...----TGC-G---G---A--------  207
HIVNDK     C--CC---A-A----------C--------A--A...----GAC-G---G---C--------  207
HIVMAL     G--TC---A-G----------C--------G--T...----GAT-A---G---G--------  207
HIVU455    A--TT---G-G----------C--------G--A...----GAT-A---G---G--------  207
SIVCPZ     TA-TC---A-G-TGC---T--A--------A--TT-CA-A--TTATT---AAT AA--G-G-C---C---  210
```



```
          5' sj\/
CONSENSUS1 AaCAAGGTAGGAtcCTCTACAgTACttGGCACTAgCAGCATTaAtAacACCAAAAAaAGAtAAAGCCACCTT   487
HIVLAI     -A----------T-------A---T--------G-------A-T-AC-------A---T----------   487
HIVHXB2R   -A----------T-------A---T--------G-------A-T-AC-------A---T----------   487
HIVMN      -A----------C-------G---T--------A-------A-T-AC-------A---T----------   487
HIVJRCSF   -G----------T-------G---T--------A-------A-T-AA-------A---T----------   487
HIVOYI     -A----------T-------A---T--------G-------G-T-AA-------A---T----------   487
HIVSF2     -A----------T-------A---T--------G-------A-T-AC-------A---C----------   487
HIVNY5CG   -A----------T-------G---T--------G-------A-T-AA-------C---T----------   487
HIVNL43    -A----------T-------G---T--------G-------A-T-AA-------C---T----------   487
HIVHAN     -G----------T-------A---T--------G-------A-C-AC-------A---T----------   487
HIVRF      -A----------T-------G---C--------G-------A-C-AC-------A---T----------   487
CONSENSUS2 AaCAAGGTAGGATCctTACAgTATTTGGCaCTaacAGCATTAaTAgCaCCAAaAaaGatAAaGCCaCCTT   485
HIVELI     -A------------CC----G--------A--AAC-------A--G-A----A-CA-AT--A---A----   487
HIVZ2Z6    -G------------CT----G--------A--AAC-------G--G-A----A-AA-AT--A---G----   487
HIVNDK     -A------------CT----G--------A--AGC-------A--G-A----A-AA-AT--A---A----   487
HIVMAL     -A------------TT----G--------A--AAC-------A--G-A----A-AA-AC--G---A----   487
HIVU455    -A------------TC----A--------T--GAA-------G--A-C----C-AG-GC--A---A----   487
SIVCPZ     -GGC----------CT----G-TC-----C--AAA---T---A--T-AGAG-G-AG-CAT-GA--A----   490

                                    vpR cds ->
CONSENSUS1 TGCCTAgTGTTA?gAAACTgACAGAgGAtAGATGGAACAAgCCCCAGAAGACCAAGGGCCACAGAGGgAG   556
HIVLAI     ------G-----CG-----G-----G--T----------G-------------------------G--   557
HIVHXB2R   ------G-----CG-----G-----G--T----------G-------------------------G--   557
HIVMN      ------G-----AG-----G-----G--T----------G-------------------------G--   557
HIVJRCSF   ------G-----AG-----A-----G--T----------G-------------------------G--   557
HIVOYI     ------G-----CG-----A-----G--T----------G-------------------------G--   557
HIVSF2     ------G-----AG-----G-----G--T----------G-------------------------G--   557
HIVNY5CG   ------G-----GG-----G-----G--C----------G-------------------------G--   557
HIVNL43    ------G-----GG-----G-----G--C----------G-------------------------G--   557
HIVHAN     ------T-----CA-----G-----A--T----------G-------------------------G--   557
HIVRF      ------G-----AG-----G-----G--T----------G-------------------------G--   557
CONSENSUS2 TGCCTAGTGTTAgGAAgcTAACAGAaGATAGATGGAACAAGCCCCAGaAGACCAaGGGCCaCAgAGGGAG   555
HIVELI     ------------G---GC-------A--------------------C------G-----A--G------   557
HIVZ2Z6    ------------G---GT-------A--------------------A------A-----A--A------   557
HIVNDK     ------------G---GC-------A--------------------A------A-----G--G------   557
HIVMAL     ------------G---GC-------A--------------------C------A-----A--G------   557
HIVU455    ------------A---AT-------G--------------------A------G-----A--G------   557
SIVCPZ     -A--------CGCC--AT-------G---------------- A----AG-----A--T--A-CA--A--A   560

CONSENSUS1 CCAtACAATqaATGGaCAC***    578
HIVLAI     ---C-----GA----A------    579
HIVHXB2R   ---C-----GA----A------    579
HIVMN      ---T-----CA----G------    579
HIVJRCSF   ---T-----GA----A------    579
HIVOYI     ---T-----GA----A------    579
HIVSF2     ---T-----GA----A------    579
HIVNY5CG   ---T-----GA----A------    579
HIVNL43    ---T-----GA----A------    579
HIVHAN     ---T-----GC----A------    579
HIVRF      ---T-----GA----A------    579
CONSENSUS2 CCatACaaTGAATq'aCAT***    577
HIVELI     --AT--AA-----G-G------    579
HIVZ2Z6    --AT--AA-----G-A------    579
HIVNDK     --AT--AA-----G-A------    579
HIVMAL     --AC--AA-----G-A------    579
HIVU455    --GT--GC-----A-G------    579
SIVCPZ     --TT--CA-----G-G------    582
```

HIV-1 VPR

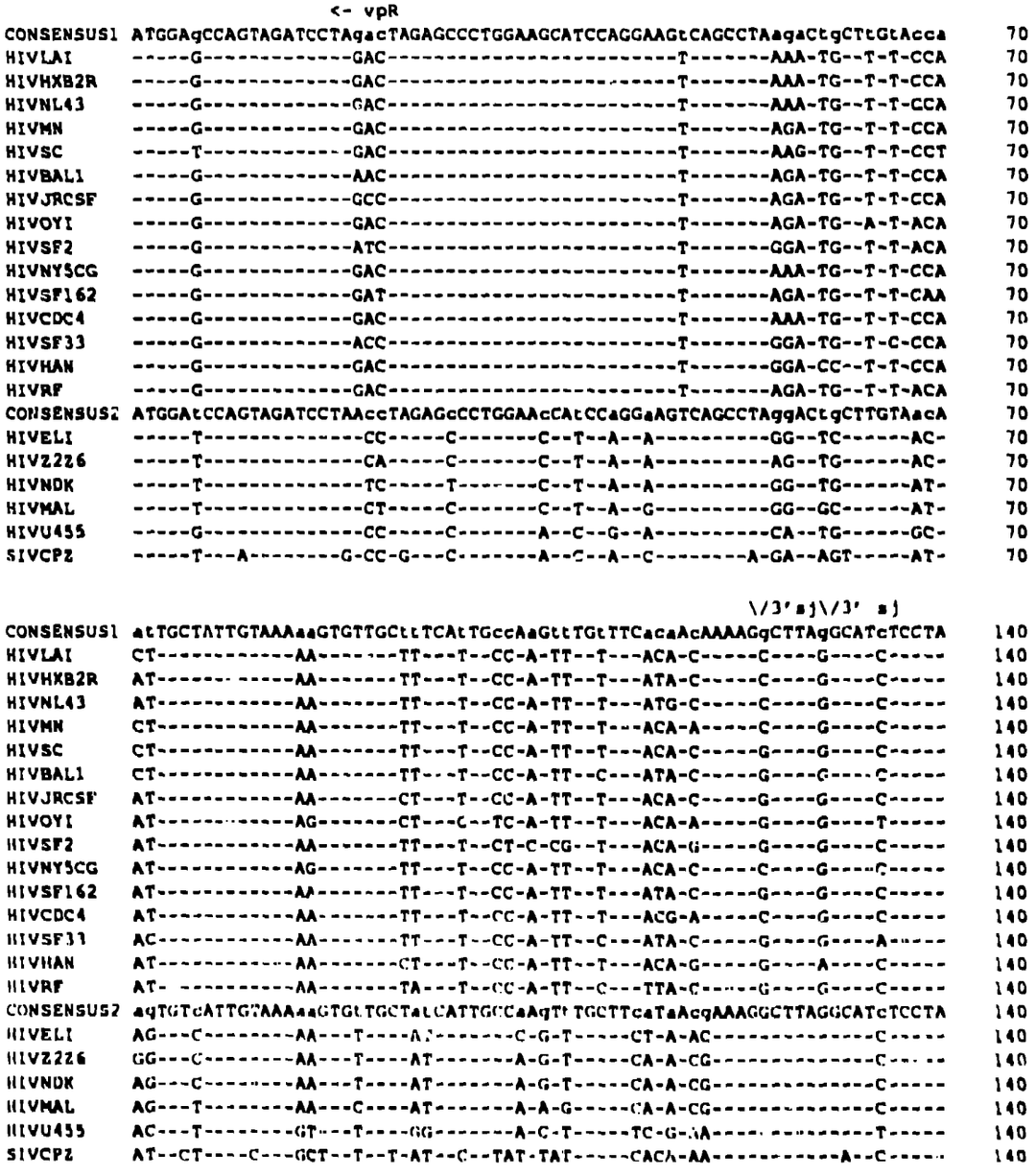

<- vif cds
CONSENSUS1 ATGGAACAAgCCCCAGAAGACCAAGGGCCACAGAGGgAGCCAtACAATgaATGGaCACTAGAGCTTTTag 70
HIVLAI ---------G-------------------------G-----C-----GA----A-------------AG 70
HIVHXB2R ---------G-------------------------G-----C-----GA----A-------------AG 70
HIVMN ---------G-------------------------G-----T-----CA----G-------------AG 70
HIVSC ---------G-------------------------G-----T-----GA----A-------------GA 70
HIVJRCSF ---------G-------------------------G-----T-----GA----A-------------AG 70
HIVOYI ---------G-------------------------G-----T-----GA----A-------------AG 70
HIVSF2 ---------G-------------------------G-----T-----GA----A-------------AG 70
HIVNY5CG ---------G-------------------------G-----T-----GA----A-------------AG 70
HIVNL43 ---------G-------------------------G-----T-----GA----A-------------AG 70
HIVHAN ---------G-------------------------G-----T-----GC----A-------------AG 70
HIVRF ---------G-------------------------G-----T-----GA----A-------------AG 70
CONSENSUS2 ATGGAACAAGCCCCAGaAGACCAaGGGCCaCAgAGGGAGCCatACaaTGAATgGaCATTAGAgcTtTTAG 70
HIVELI ----------------C------G-----A--G--------AT--AA-----G-G--------GC-T---- 70
HIVZ2Z6 ----------------A------A-----A--A--------AT--AA-----G-A--------GC-T---- 70
HIVNDK ----------------A------A-----G--G--------AT--AA-----G-A--------GC-T---- 70
HIVMAL ----------------C------A-----A--G--------AC--AA-----G-A--------AC-T---- 70
HIVU455 ----------------A------G-----A--G--------GT--GC-----A-G--------GT-G---- 70
SIVCPZ ----------A----AG-----A--T--A-CA--A--A--TT--CAG----G-G--------GACC--G- 70

CONSENSUS1 aGgAgCTTAAGAgtGAAGCTGTtAGaCATTTTCCtAGGataTGGCTcCaTagcTTAGgaCAAcATATCtA 140
HIVLAI A-G-G--------AT---------T--A-----T---ATT-----C-A-GGC----GG---C-------- 140
HIVHXB2R A-G-G--------AT---------T--A---------T---ATT-----C-A-GGC----GG---C-------- 140
HIVMN A-G-G--------AT---------T--A---------T---ATA-----C-A-GGC----GG---C-------- 140
HIVSC G-A-G--------GT---------T--A---------T---ATA-----C-A-GGC----GG---C-------- 140
HIVJRCSF A-G-A--------AT---------T--A---------T---ATC-----C-A-AGC----GG---T-------- 140
HIVOYI A-G-G--------GT---------T--A---------T---ATA-----C-A-AGC----AA---T-------- 140
HIVSF2 A-G-G--------GA---------T--A---------T---CCA-----C-A-AGC----GA---T-------- 140
HIVNY5CG A-G-A--------GT---------T--A---------T---ATA-----C-A-AAC----GA---C-------- 140
HIVNL43 A-G-A--------GT---------T--A---------T---ATA-----C-A-AAC----GA---C-------- 140
HIVHAN A-G-G--------GT---------T--A---------C---ATA-----A-A-AGC----GA---T-------- 140
HIVRF A-G-G--------GT---------C--A---------T---CTA-----C-A-AGC----GA---C-------- 140
CONSENSUS2 AgGAGCTTAAGagtGAAGcTGTtAGACAtTTTCCTAGGataTGGCTCCATaGcTTAGGACAAcATATcTA 140
HIVELI -G----------AGT----C---T-----T----------ATA----------A-C----------C----T-- 140
HIVZ2Z6 -G----------AGT----C---T-----T----------ATA----------A-C----------T----T-- 140
HIVNDK -G----------AGT----C---C-----T----------ATA----------A-C----------C----C-- 140
HIVMAL -G----------CAA----C---C-----C----------ATA----------A-T----------C----C-- 140
HIVU455 -A----------CAT----A---T-----T----------CAG----------G-A----------C----C-- 140
SIVCPZ -A--AT-A--AAAT----CA--A-----T--C-----ACCA---T-G---CAA----------TTC--T-- 140

CONSENSUS1 TGAAACTTAtGGGGATACtTGGgcAGGAGTgGAAGCcaTAATAAGAAttCTgCAACAaCTGCTGTTTATt 210
HIVLAI ---------T--------T---GC------G-----CA----------TT--G-----A----------C 210
HIVHXB2R ---------T--------T---GC------G-----CA----------TT--G-----A----------C 210
HIVMN ---------T--------T---GC------G-----CA----------TT--A-----A----------T 210
HIVSC ---------T--------T---GC------G-----CA----------TT--G-----A----------T 210
HIVJRCSF ---------T--------T---GC------G-----CA----------TA--G-----G----------T 210
HIVOYI ---------T--------T---GA------G-----CA----------TT--G-----A----------T 210
HIVSF2 ---------T--------T---GC------G-----CA----------TT--G-----A----------T 210
HIVNY5CG ---------C--------T---GC------G-----CA----------TT--G-----A----------T 210
HIVNL43 ---------C--------T---GC------G-----CA----------TT--G-----A----------C 210
HIVHAN ---------T--------T---GC------A-----CA----------TA--G-----A----------T 210
HIVRF ---------T--------A---GC------G-----TA----------TT--G-----A----------T 210
CONSENSUS2 TgAaACtTATGGGGATACcTGGGtaGGaGTTGAAGCTaTAATAAGAAttcTGCAACAAtTaCTGTTTATT 209
HIVELI -G-A--T----------G----TA--A---------A---------TAC--------T-A---------- 210
HIVZ2Z6 -G-A--T----------C----CA--A---------C---------TTT--------T-A---------- 210
HIVNDK -G-A--T----------C----CA--T---------A---------TTC--------C-A---------- 210
HIVMAL -G-A--T----------C----AA--A---------A---------GTC--------C-G---------- 210
HIVU455 -A-G-A-T----------T---AG--A---------A---------TTT--------T-A---------- 210
SIVCPZ -G-G-A-T-----A----A---TA--A--A---G--CA-C----TCT-A----CT-G---------C 210

```
                \/3' sj                                                tat cds ->
CONSENSUS1 CATTT?CAGAATTGGgTGtCgaCAtAGCAGAATAGGCaTTAcTCgACAgAGgAGAgCA???AGaAATGGA   276
HIVLAI     -----.---------G--T-GA--T----------G---C--A---G--G---G--...--A------   276
HIVHXB2R   -----T---------G--T-GA--***                                           236
HIVMN      -----.---------G--T-GA--T----------A---T--G---G--G---G--...--A------   276
HIVSC      -----.---------G--T-GA--T----------A---T--A---A--G---T--...--A------   276
HIVJRCSF   -----.---------G--T-GA--T----------A---C--G---G--G---G--...--A------   276
HIVOYI     -----.---------G--T-AA--T----------A---C--G---G--A---G--...--A------   276
HIVSF2     -----.---------G--T-AA--T----------A---T--A---G--G---G--AGA--A------   279
HIVNY5CG   -----.---------A--T-AA--T----------A---T--G---G--G---G--...--G------   276
HIVNL43    -----.---------G--T-GA--T----------G---C--G---G--G---G--...--A------   276
HIVHAN     -----.---------G--T-GC--T----------A---C--C---G--G---G--...--A------   276
HIVRF      -----.---------G--T-AA--T----------A---C--G---A--A---G--...--A------   276
CONSENSUS2 CATTT.CAGAATTGGGTGtCAACATAGCAGAATAgGcATTAcTCgAcaGAGAAGAGca...AGaAATGGA   275
HIVELI     -----.----------T--------------G-C----T--G-CA--------CA...--A------   276
HIVZ226    -----.----------T--------------G-T----C--G-CA--------CA...--A------   276
HIVNDK     -----.----------T--------------A-T----C--G-CA--------CA...--A------   276
HIVMAL     -----.----------T--------------G-C----C--G-CA--------CA...--A------   276
HIVU455    -----.----------C--------------G-C----T--C-GG--------GC...--G------   276
SIVCPZ     -----.T---C-------C--G----------G-C--CCTC-C-CAA------AGA...TCC------   276

CONSENSUS1 gCCAGTAGATCC***    291
HIVLAI     G--------------    291
HIVMN      G--------------    291
HIVSC      T--------------    291
HIVJRCSF   G--------------    291
HIVOYI     G--------------    291
HIVSF2     G--------------    294
HIVNY5CG   G--------------    291
HIVNL43    G--------------    291
HIVHAN     G--------------    291
HIVRF      G--------------    291
CONSENSUS2 tCCAGTAGATCC***    290
HIVELI     T--------------    291
HIVZ226    T--------------    291
HIVNDK     T--------------    291
HIVMAL     T--------------    291
HIVU455    G--------------    291
SIVCPZ     T---A----------    291
```

HIV-1 TAT


<- vpR
CONSENSUS1 ATGGAgCCAGTAGATCCTAgacTAGAGCCCTGGAAGCATCCAGGAAGtCAGCCTAagaCtgCTtGtAcca 70
HIVLAI -----G----------GAC----------T-------AAA-TG--T-T-CCA 70
HIVHXB2R -----G----------GAC----------T-------AAA-TG--T-T-CCA 70
HIVNL43 -----G----------GAC----------T-------AAA-TG--T-T-CCA 70
HIVMN -----G----------GAC----------T-------AGA-TG--T-T-CCA 70
HIVSC -----T----------GAC----------T-------AAG-TG--T-T-CCT 70
HIVBAL1 -----G----------AAC----------T-------AGA-TG--T-T-CCA 70
HIVJRCSF -----G----------GCC----------T-------AGA-TG--T-T-CCA 70
HIVOYI -----G----------GAC----------T-------AGA-TG--A-T-ACA 70
HIVSF2 -----G----------ATC----------T-------GGA-TG--T-T-ACA 70
HIVNY5CG -----G----------GAC----------T-------AAA-TG--T-T-CCA 70
HIVSF162 -----G----------GAT----------T-------AGA-TG--T-T-CAA 70
HIVCDC4 -----G----------GAC----------T-------AAA-TG--T-T-CCA 70
HIVSF33 -----G----------ACC----------T-------GGA-TG--T-C-CCA 70
HIVHAN -----G----------GAC----------T-------GGA-CC--T-T-CCA 70
HIVRF -----G----------GAC----------T-------AGA-TG--T-T-ACA 70
CONSENSUS2 ATGGAtCCAGTAGATCCTAAccTAGAGcCCTGGAAcCAtCCaGGaAGTCAGCCTAggACtgCTTGTAacA 70
HIVELI -----T----------CC-----C-------C--T--A--A----------GG--TC------AC- 70
HIVZ226 -----T----------CA-----C-------C--T--A--A----------AG--TG------AC- 70
HIVNDK -----T----------TC-----T-------C--T--A--A----------GG--TG------AT- 70
HIVMAL -----T----------CT-----C-------C--T--A--G----------GG--GC------AT- 70
HIVU455 -----G----------CC-----C-------A--C--G--A----------CA--TG------GC- 70
SIVCPZ -----T---A--------G-CC-G---C-------A--C--A--C--------A-GA--AGT-----AT- 70
\/3'aj\/3' aj
CONSENSUS1 atTGCTATTGTAAAaaGTGTTGCttTCAtTGccAaGttTGtTTCacaAcAAAAGgCTTAgGCATcTCCTA 140
HIVLAI CT----------AA-------TT---T--CC-A-TT--T---ACA-C-----C----G----C----- 140
HIVHXB2R AT-------------AA-------TT---T--CC-A-TT--T---ATA-C-----C----G----C----- 140
HIVNL43 AT----------AA-------TT---T--CC-A-TT--T---ATG-C-----C----G----C----- 140
HIVMN CT----------AA-------TT---T--CC-A-TT--T---ACA-A-----C----G----C----- 140
HIVSC CT----------AA-------TT---T--CC-A-TT--T---ACA-C-----G----G----C----- 140
HIVBAL1 CT----------AA-------TT---T--CC-A-TT--C---ATA-C-----G----G----C----- 140
HIVJRCSF AT----------AA-------CT---T--CC-A-TT--T---ACA-C-----G----G----C----- 140
HIVOYI AT----------AG-------CT---C--TC-A-TT--T---ACA-A-----G----G----T----- 140
HIVSF2 AT----------AA-------TT---T--CT-C-CG--T---ACA-G-----G----G----C----- 140
HIVNY5CG AT----------AG-------TT---T--CC-A-TT--T---ACA-C-----C----G----C----- 140
HIVSF162 AT----------AA-------TT---T--CC-A-TT--T---ATA-C-----G----G----C----- 140
HIVCDC4 AT----------AA-------TT---T--CC-A-TT--T---ACG-A-----C----G----C----- 140
HIVSF33 AC----------AA-------TT---T--CC-A-TT--C---ATA-C-----G----G----A----- 140
HIVHAN AT----------AA-------CT---T--CC-A-TT--T---ACA-G-----G----A----C----- 140
HIVRF AT----------AA-------TA---T--CC-A-TT--C---TTA-C-----G----G----C----- 140
CONSENSUS2 aqTGTcATTGTAAAaaGTGtTGCTatCATTGCCaAqTt TGCTTcaTaAcqAAAGGCTTAGGCATcTCCTA 140
HIVELI AG---C---------AA---T----A\--------C-G-T-----CT-A-AC--------------C----- 140
HIVZ226 GG---C---------AA---T----AT--------A-G-T-----CA-A-CG--------------C----- 140
HIVNDK AG---C---------AA---T----AT--------A-G-T-----CA-A-CG--------------C----- 140
HIVMAL AG---T---------AA---C----AT--------A-A-G-----CA-A-CG--------------C----- 140
HIVU455 AC---T---------GT---T----GG--------A-C-T-----TC-G-AA----------T----- 140
SIVCPZ AT--CT----C---GCT--T--T-AT--C--TAT-TAT-----CACA-AA----------A--C-----

```
   rev cds ->   \/3' sj
CONSENSUS1 TGGCAGGAAGAAGcGGAGaCAG...CGAcgaaGAgCTCcTcaagaCAGTcAGacTCATCAAGttTCTcTA   207
HIVLAI     ----------C----A---...---CGAA--C---C-CAAGG----C--AC--------TT---C--   207
HIVHXB2R   ----------C----A---...---CGAA--G---A-CAGAA----C--AC--------CT---C--   207
HIVNL43    ----------C----A---...---CGAA--G---A-CAGAA----C--AC--------CT---C--   207
HIVMN      ----------C----A---...---CGAA--G---C-GAAGA----C--AC--------TT---C--   207
HIVSC      ----------C----A---...---AGAA--G---C-CAAGA----C--AC--------TT---C--   207
HIVBAL1    ----------C----A---...---CGAA--G---C-CAAGA----G--AC--------TT---C--   207
HIVJRCSF   ----------C----A---...---CGAA--C---C-CAAGA----C--AC--------TT---C--   207
HIVOYI     ----------C----A---...---CGAA--G---C-CAAGA----A--AC--------TT---C--   207
HIVSF2     ----------C----A---...---CGAA--G---C-CAGGA----C--AC--------CT---C--   207
HIVNY5CG   ----------C----A---...---CGAA--G---C-CAAGA----G--AC--------TT---C--   207
HIVSF162   ----------C----A---...---CGAA--G---C-CCAGA----G--GT--------TT---C--   207
HIVCDC4    ----------C----A---...---CGAA--G---A-CAAGA----C--AA--------CA---C--   207
HIVSF33    ----------C----G---...---CAGA--G---C-GACAG----C--AA--------AT---C--   207
HIVHAN     ----------C----A---...---CGAA--G---C-CAAGA----C--AC--------TT---C--   207
HIVRF      ----------A----A---...---CGAG--C---C-CAAGG----C--AC--------TC---T--   207
CONSENSUS2 TGGCAgGAAGAAGCGGAgACag...cGACGAagACCTcCTCAaGGCggTcAGgcTCATCAAg?TCcTaTA   206
HIVELI     -----G----------G--AG...C-----GG----C----A---GG-C--GC-------GT--C-A--   207
HIVZ2Z6    -----G----------G--AG...A-----AG----T----A---GG-C--AC-------GA--C-A--   207
HIVNDK     -----G----------G--AG...C-----AA----C----A---GA-C--GC-------GT--C-A--   207
HIVMAL     -----G----------G--AG...C-----AG----C----G---AA-C--GC-------GA--C-C--   207
HIVU455    -----A----------A--CC...C-----GG----C----A---AG-A--GA-------AC--T-A--   207
SIVCPZ     -----G---------A-CCACAAGAC----CACGG--C--GCA---AG-A-AAA-A-----GA-TC-A--   210

                  /- env -> (for more information, see the HIV-1 ENV CDS that follows)
              exon\/exon
CONSENSUS1 tCAAAGCAACCC?CctCCCAgccccgAGGGGAC...CCgacagGCCCGaaGgAAtCGAAGaAGaAGGTGG   273
HIVLAI     T----------A-CT----ACCCCG-------...--GACAG-----AA-G--***               261
HIVHXB2R   T----------A-CT----ACCCCG-------...--GACAG-----AA-G--***               261
HIVNL43    T----------A-CT----ATCCCG-------...--GACAG-----AA-G--***               261
HIVMN      C----------G-CC----GTTCCG-------...--GACAG-----AA-G--T-----A--A------   274
HIVSC      C----------G-CT----AGCCAG-------...--GACAG-----AA-G--T-----A--A------   274
HIVBAL1    T----------A-CT----TCCTCG-------.. --GACAG-----GA-G--T-----A--A------   274
HIVJRCSF   C----------T-CT----GCAACG-------...--GACAG-----AA-G--T-----A--A------   274
HIVOYI     T----------G-CT----ACCCAG-------...--GACAG-----AA-G--T-----A--A------   274
HIVSF2     T----------G-CT----GTCCCG-------...--GACAG-----AC-G--T-----A--A------   274
HIVNY5CG   T----------G-CT----GCCCAG-------...--GACAG-----AA-G--T-----A--A------   274
HIVSF162   C----------G-TT----GCCCCA-------...--GACAG-----AA-G--T-----A--A------   274
HIVCDC4    T----------T-CT----AACCCG-------...--GACAG-----AA-G--C-----A--G------   274
HIVSF33    T----------T-AT----GCCCAG-------...--GACAG-----AA-G--T-----A--G------   274
HIVHAN     C----------T-CT----GCAACG-------...--GGACA-----AA-A--T-----A--A------   274
HIVRF      T----------A-CT----GCCCCG-------...--GACAG-----AA-G--T-----G--A------   274
CONSENSUS2 CGAaAGGAACCCTcctCCCAgccCCgagGGGaC???cCGaCAGGCCcGaAggAA??GAAGAAGAAGGTGG   271
HIVELI     ---A---------CCT----GCC--GAG---A-...C--A------C-A-GG--CA--------------   274
HIVZ2Z6    ---A---------CCT----GCC--GAG---A-...C--A------C-A-GG--***               261
HIVNDK     ---G---------CCT----GTC--GAG---A-...C--A------C-A-GA--***               261
HIVMAL     ---G---------CCT----ACA--GAG---A-CACC--A------C-A-GG--***               264
HIVU455    ---A---------TGC----ATC--AGA---T-...T--G------A-G-AG--TC--------------   274
SIVCPZ     ---A---------TAT----GTC-AGAG--AA-...AA-GA----TC-G-GA--TC--C----G------   277
```

HIV-1 TAT

```
CONSENSUS1 AGaGAGAGaCAGAGaCagATCCggtcgat???      307
HIVMN      --A-----A-----A-AC----GGTCGAT***      306
HIVSC      --A-----A-----A-AG----GGTCGAT***      306
HIVBAL1    --A-----A-----A-AG----GGTCCAT***      306
HIVJRCSF   --A-----A-----A-AG----GGACAAT***      306
HIVOYI     --A-----A-----A-AG----GGAAGAT***      306
HIVSF2     --A-----A-----A-AG----GTTCGAT***      306
HIVNY5CG   --A-----A-----A-AG----GGACCAT***      306
HIVSF162   --A-----A-----A-AG----AGTCCAT***      306
HIVCDC4    --A-----G-----A-GG----ACTCGAT***      306
HIVSF33    --A-----A-----A-AG----ACTCGAT***      306
HIVHAN     --A-----A-----G-AG----GTTCGAT***      306
HIVRF      --A-----A-----A-AG----GGCGGTGCAG***   309
CONSENSUS2 AGAGC?AGGCA?AGACAGATC?????            298
HIVELI     -----G-----G---------CG***            300
HIVU455    -----A-----A---------GATTCGCT***      306
SIVCPZ     C----A--A--G-AG------AG***            303
```

```
     tat ->      \/3' sj
CONSENSUS1 ATGGCAGGAAGAAGcGGAGaCAGCGAcgaa...GAgCTccTcaagaCaGTcAGacTcATCAAGttTCTcT     67
HIVLAI     -------------C----A------CGAA...--C--CC-CAAGG-A--C--AC-C------TT---C-     67
HIVHXB2R   -------------C----A------CGAA...--G--CA-CAGAA-A--C--AC-C------CT---C-     67
HIVNL43    -------------C----A------CGAA...--G--CA-CAGAA-A--C--AC-C------CT---C-     67
HIVMN      -------------C----A------CGAA...--G--CC-GAAGA-A--C--AC-C------TT---C-     67
HIVBRVA                                    --G--CC-GAAGA-A--C--AG-C------TT---C-     37
HIVSC      -------------C----A------AGAA...--G--CC-CAAGA-A--C--AC-C------TT---C-     67
HIVJH3           ------C----A------CGAA...--G--GC-CAGGC-.--C--GC-A------TT---C-     56
HIVBAL1    -------------C----A------CGAA...--G--CC-CAAGA-A--G--AC-C------TT---C-     67
HIVJRCSF   -------------C----A------CGAA...--C--CC-CAAGA-A--C--AC-C------TT---C-     67
HIVOYI     -------------C----A------CGAA...--G--CC-CAAGA-A--A--AC-C------TT---C-     67
HIVSF2     -------------C----A------CGAA...--G--CC-CAGGA-A--C--AC-C------CT---C-     67
HIVNY5CG   -------------C----A------CGAA...--G--CC-CAAGA-A--G--AC-C------TT---C-     67
HIVSF162   -------------C----A------CGAA...--G--CC-CCAGA-A--G--GT-C------TT---C-     67
HIVCDC4    -------------C----A------CGAA...--G--CA-CAAGA-A--C--AA-C------CA---C-     67
HIVSF33    -------------C----G------CAGA...--G--CC-GACAG-A--C--AA-C------AT---C-     67
HIVHAN     -------------C----A------CGAA...--G--CC-CAAGA-A--C--AC-C------TT---C-     67
HIVRF      -------------A----A------CGAG...--C--CC-CAAGG-A--C--AC-C------TC---T-     67
CONSENSUS2 ATGGCAgGAAGAAGCGGAgACagcGACGAa...gACCTcCTCAaGGCggTcAGgcTCATCAAg?TCcTaT     66
HIVELI     ------G----------G--AGC-----G...C----C-  --A---GG-C--GC-------GT--C-A-     67
HIVZ2Z6    ------G----------G--AGA-----A...G----T----A---GG-C--AC-------GA--C-A-     67
HIVNDK     ------G----------G--AGC-----A...A----C----A---GA-C--GC-------GT--C-A-     67
HIVMAL     ------G----------G--AGC-----A...G----C----G---AA-C--GC-------GA--C-C-     67
HIVU455    ------A----------A--CCC-----G...G----C----A---AG-A--GA-------AC--T-A-     67
SIVCPZ     ------G---------A-CCACAA-----CGCACGG--C--GCA---AG-A-AAA-A-----GA-TC-A-     70

                    /- env -> (for more information, see the HIV-1 ENV CDS that follows)
               exon\/exon
CONSENSUS1 AtCAAAGCAACCCgCctCCCAgccccgAGGGG. .ACCCgacagGCCCGaaGgAAtcGAAGaAGaAGGTG    134
HIVLAI     -T----------A-CT----ACCCCG-----...----GACAG-----AA-G--TA----A--A-----    134
HIVHXB2R   -T----------A-CT----ACCCCG-----...----GACAG-----AA-G--TA----A--A-----    134
HIVNL43    -T----------A-CT----ATCCCG-----...----GACAG-----AA-G--TA----A--A-----    134
HIVMN      -C----------G-CC----GTTCCG-----...----GACAG-----AA-G--TC----A--A-----    134
HIVBRVA    -T----------G-CT----GGCCGG-----...----GACAG-----AA-G--TC----A--A-----    104
HIVSC      -C----------G-CT----AGCCAG-----...----GACAG-----AA-G--TC----A--A-----    134
HIVJH3     -C----------G-CT----GCCCCG-----...----GACAG-----AA-G--TC----A--A-----    125
HIVBAL1    -T----------A-CT----TCCTCG-----...----GACAG-----GA-G--TC----A--A-----    134
HIVJRCSF   -C----------T-CT----GCAACG-----...----GACAG-----AA-G--TC----A--A-----    134
HIVOYI     -T----------G-CT----ACCCAG-----...----GACAG-----AA-G--TC----A--A-----    134
HIVSF2     -T----------G-CT----GTCCCG-----...----GACAG-----AC-G--TC----A--A-----    134
HIVNY5CG   -T----------G-CT----GCCCAG-----...----GACAG-----AA-G--TC----A--A-----    134
HIVSF162    C----------G-TT----GCCCCA-----...----GACAG-----AA-G--TC----A--A-----    134
HIVCDC4    -T----------T-CT----AACCCG-----...----GACAG-----AA-G--CC----A--G-----    134
HIVSF33    -T----------T-AT----GCCCAG-----...----GACAG-----AA-G--TC----A--G-----    134
HIVHAN     -C----------T-CT----GCAACG-----...----CGACA-----AA-A--TC----A--A-----    134
HIVRF      -T----------A-CT----GCCCCG-----...----GACAG-----AA-G--TC----G--A-----    134
CONSENSUS2 ACCAaAGCAACCCTcctCCCAgccCCgagGGG???aCcCGaCAGGCCcGaAggAAtaGAAGAAGAAGGTG    133
HIVELI     ----A---------CCT----GCC--GAG---...A-C--A------C-A-GG--CA-------------    134
HIVZ2Z6    ----A---------CCT----GCC--GAG---...A-C--A------C-A-GG--TA-------------    134
HIVNDK     ----G---------CCT----GTC--GAG---...A-C--A------C-A-GA--TA-------------    134
HIVMAL     ----G---------CCT----ACA--GAG---ACCA-C--A------C-A-GG--TA-------------    137
HIVU455    ----A---------TGC----ATC--AGA---...T-T--G------A-G-AG--TC-------------    134
SIVCPZ     ----A---------TAT----GTC-AGAG--A...A-AA-GA----TC-G-GA--TC--C----G-----    137
```

HIV-1 REV

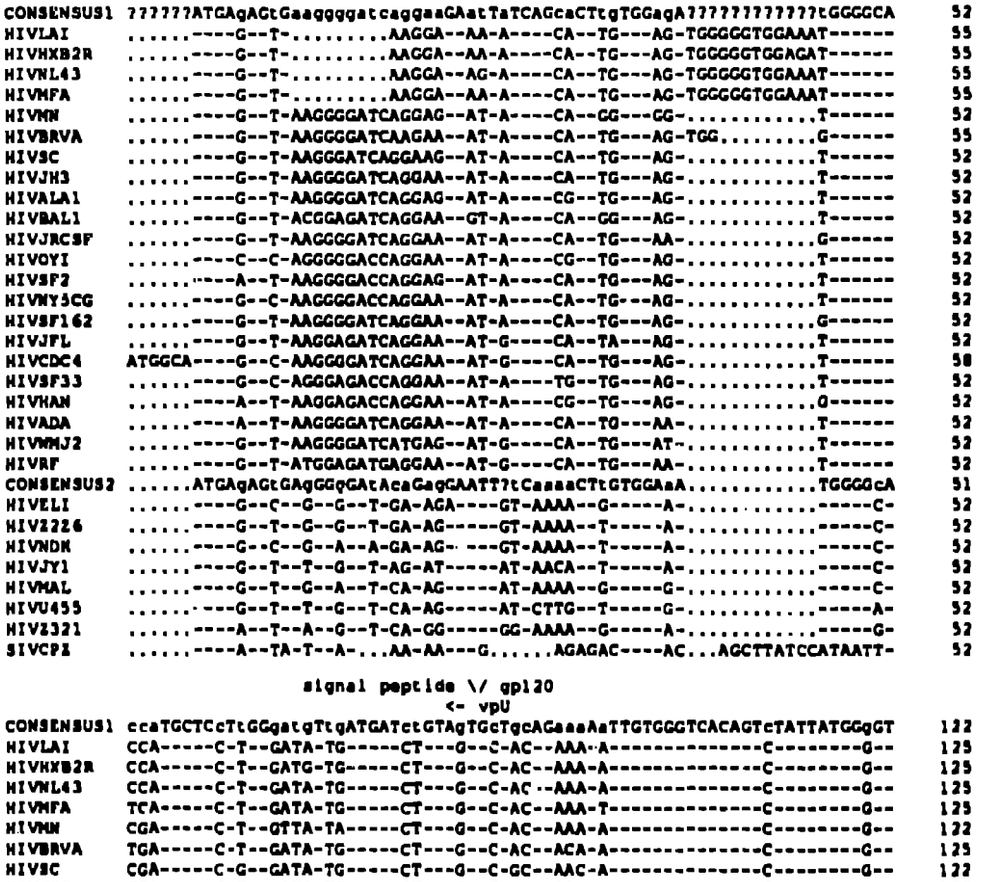
```
CONSENSUS1 GAGagaGAGaCAGAGaCagATCCggtcgattAGTgaatGGaTtcTTagCacTtaTCTGGGacGAtctgcG  204
HIVLAI     ---AGA---A-----A-AG----ATTCGATT---GAAC--A-CC--AG-AC-TA------AC--TCTGC-  204
HIVHXB2R   ---AGA---A-----A-AG----ATTCGATT---GAAC--A-CC--GG-AC-TA------AC--TCTGC-  204
HIVNL43    ---AGA---A-----A-AG----ATTCGATT---GAAC--A-CC--AG-AC-TA------AC--TCTGC-  204
HIVMN      ---AGA---A-----A-AC----GGTCGATT---GCAT--A-TC--AG-AA-TA------TC--CCTGC-  204
HIVBRVA    ---AGC---A-----A-AG----AGTCCCTT---GGAT--A-TC--AG-AC-TT------TC--TCTGC-  174
HIVSC      ---AGA---A-----A-AG----GGTCGATT---GGAT--A-TC--AG-AA-TA------TC--CTAGC-  204
HIVJH3     ---ACA---A-----A-AG----GGTCAATT---GGAT--A-TA--AG-AA-TA------TC--CCTGC-  195
HIVBAL1    ---AGA---A-----A-AG----GGTCCATT---GAAC--A-TC--GG-GC-TA------TC--TCTGC-  204
HIVJRCSF   ---AGA---A-----A-AG----GGACAATT---GAAC--A-TC--AG-AC-TA------TC--CCTGC-  204
HIVOYI     ---AGA---A-----A-AG----GGAAGATT---GGAT--A-TC--AG-AC-TA------AC--TCTGC-  204
HIVSF2     ---AGA---A-----A-AG----GTTCGATT---GGAT--A-TC--AG-AC-TA------AA--TCTGC-  204
HIVNY5CG   ---AGA---A-----A-AG----GGACCATT---GAAT--A-TC--AG-AC-CA------TC--TCTGC-  204
HIVSF162   ---AGA---A-----A-AG----AGTCCATT---GCAT--A-TA--AG-AC-CA------AC--TCTAC-  204
HIVCDC4    ---AGA---G-----A-GG----ACTCGATT---GCAT--C-TC--AG-AC-TG------AC--TCTGC-  204
HIVSF33    ---AGA---A-----A-AG----ACTCGATT---GAAC--A-TC--AG-AC-GT------AC--TCTTC-  204
HIVHAN     ---AGA---A-----G-AG----GTTCGATT---GAGC--A-TC--AG-AC-TT------AC--CCTGC-  204
HIVRF      ---AGA---A-----A-AG----GGCGGTGC---GAAT--A-TC--GA-AC-TA------AC--TCTGT-  204
CONSENSUS2 GAGAGCgAGGCAgAGaCAGATCcaTtcGaTTggTGAaCGGATTcTcaGCacTTatCTGGGACGAcCTGaG  203
HIVELI     ------G-----G--A------CG-GA-A--GC---A------C-CG--AC--AT---------C---C-  204
HIVZ2Z6    ------G-----G--A------CA-TC-A--GG---A------C-CA--AC--AT---------T---A-  204
HIVNDK     ------G-----G--A------CA-TC-A--GG---A------A-TT--AC--TT---------C---A-  204
HIVMAL     ------A-----G--G------AA-TC-A--GG---A------C-CA--AC--AT---------C---A-  207
HIVU455    ------A-----A--A------GA-TC-C--AG---G------C-TA--GA--GC---------C---C-  204
SIVCPZ     -C----A--A--G-AG------AG-GA-A--AG--GAA--G-GC-TGC-AC--AT---------C--CC-  207

CONSENSUS1 GAgCCTGTgCctCTTCAGCTACCaCCgCttgAgAGACTTAcTCTTGAttGtagcgAgGAtTGTGGAActT  274
HIVLAI     --G-----G-CT----------A--G-TTG-G-------C------TT-TAACG-G--T-------CT-  274
HIVHXB2R   --G-----G-CT----------A--G-TTG-G-------C------TT-TAACG-G--T-------CT-  274
HIVNL43    --G-----G-CT----------A--G-TTG-G-------C------TT-TAACG-G--T-------CT-  274
HIVMN      --G-----T-CT----------A--A-...-G-------C------TT-CAGCG-G--T-------CT-  271
HIVBRVA    --G-----T-CT----------A--G-TTG-G-------C------TT-TAACG-G--T-------CT-  244
HIVSC      --G-----G-CT----------A--G-TTG-G-------C------TT-TAACG-G--T-------CT-  274
HIVJH3     --G-----G-CT----------A--G-TTG-G-------C------TT-TAACG-G--T-------CT-  265
HIVBAL1    --G-----T-CT----------A--G-TTG-G-------C------TT-TGATG-G--T-------CT-  274
HIVJRCSF   --G-----T-CT----------A--G-TTG-G-------C------CT-TAACG-G--T-------CT-  274
HIVOYI     --G-----G-CT----------A--G-TTG-G-------A------TT-TAGCG-G--T-------CT-  274
HIVSF2     --G-----G-CT----------G--G-TTG-G-------C------TT-CAGCG-G--C-------AT-  274
HIVNY5CG   --G-----G-CT----------A--G-TTG-G-------C------TT-TAGCG-G--T-------CT-  274
HIVSF162   --G-----G-CT----------A--G-TTG-G-------A------TT-CAGCG-G--T-------CT-  274
HIVCDC4    --G-----G-CT----------A--G-TTG-G-------C------TT-TAGCG-G--T-------CT-  274
HIVSF33    --G-----G-CT----------A--G-TTG-C-------C------TT-TAGCG-G--T-------CT-  274
HIVHAN     --G-----G-CT----------G--G-TTG-G-------C------TT-TAGCG-A--T-------AC-  274
HIVRF      --C-----G-AG----------A--G-TTG-G-------C------TA-TAGTG-G--T-------CT-  274
CONSENSUS2 GAaCCTGTGCCTCTTCAGgTACCACCGcTTGAGAGACTtaatCTTaATTGcAgcGAGgatTGTgGAaCTT  273
HIVELI     --G---------------C--------C----------TAAT---A----C-GT---GAT---A--A---  274
HIVZ2Z6    --A---------------C--------C----------TAAT---A----C-GC---GAT---G--G---  274
HIVNDK     --A---------------C--------C----------CAAT---A----C-GC---GAT---G--A---  274
HIVMAL     --A---------------T--------C----------TACT---A----C-AC---GAT---G--A---  277
HIVU455    --A---------------C--------A----------TCGC---G----T-GC---AGC---G--A---  274
SIVCPZ     A-A----G-GA-T-GG--C-----GA-C----C-AG--TAGC--GC-A--TGTG---ACAAC-CA-GACG  277
```

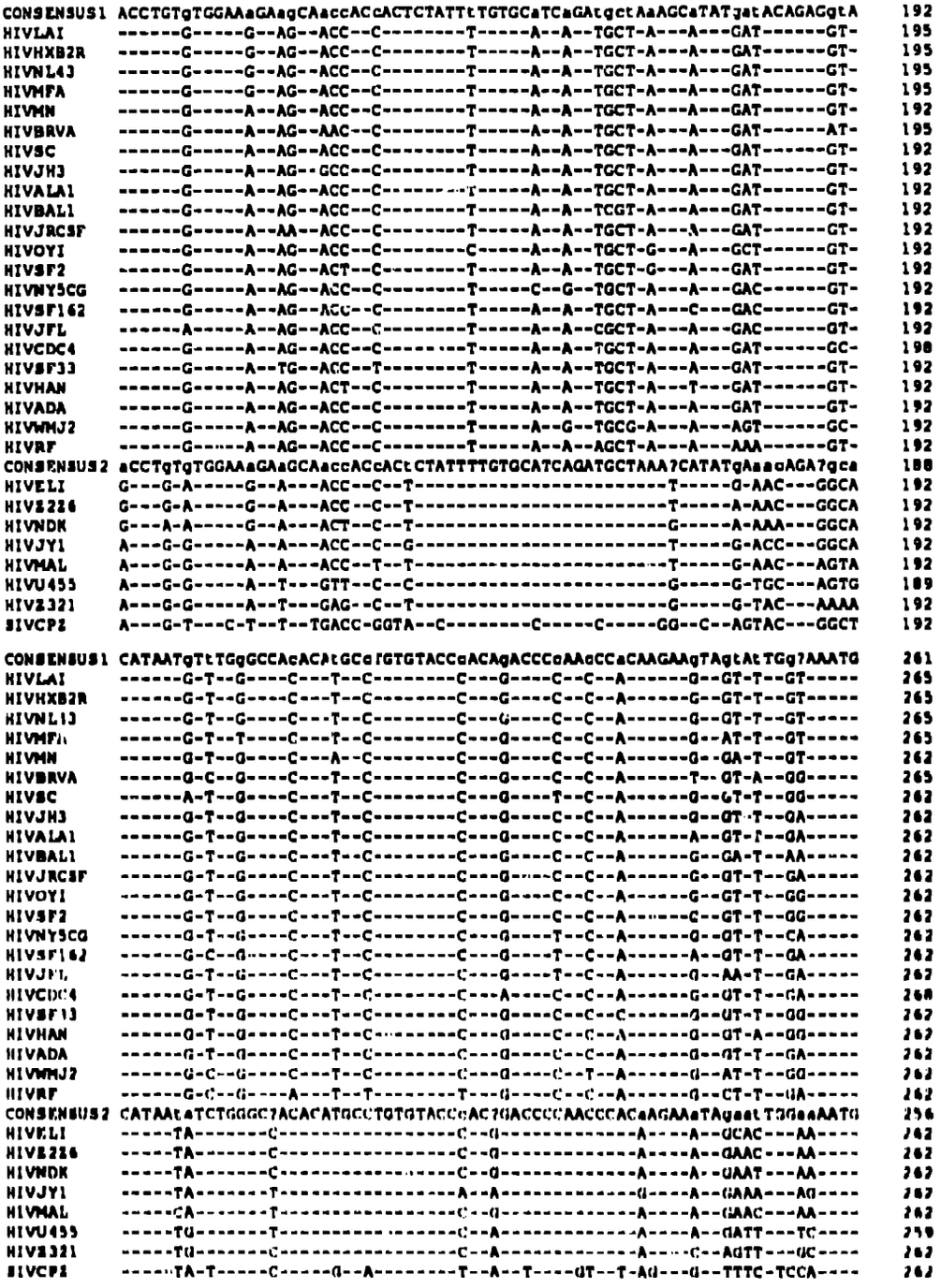
```
CONSENSUS1 CTGGGaCgcag.................???GGGGTGGGAAgcCCTcAAaTAtTGgtGCAaTCTCCT?C    322
HIVLAI     -----A-GCAG.....................----------GC---C--A--T--GT---A------A-    323
HIVHXB2R   -----A-GCAG.....................----------GC---C--A--T--GT---A------A-    323
HIVNL43    -----A-GCAG.....................----------GC---C--A--T--GT---A------A-    323
HIVMN      -----A-GCAG.....................----------GT---C--A--T--GT---A------A-    320
HIVBRVA    -----A-GCAG.....................----------GT---C--A--T--GT---A------A-    293
HIVSC      -----A-GCAG.....................----------GC---C--A--T--GT---A------A-    323
HIVJH3     -----A-GCAG.....................----------GC---C--A--T--GT---A------A-    314
HIVBAL1    -----A-TAGC.................AGG----------GT---C--A--T--GT---A------A-    326
HIVJRCSF   -----A-GCAG.....................----------AT---G--A--C--GT---A------A-    323
HIVOYI     -----A-GCAG.....................----------GT---G--A--T--GT---A------G-    323
HIVSF2     -----G-ACAG.....................----------GC---C--A--T--GT---G------G-    323
HIVNY5CG   -----A-GCAG.....................----------GC---C--A--T--TT---A------G-    323
HIVSF162   -----A-GCAG.....................----------GC---C--G--T--GG---A------G-    323
HIVCDC4    -----A-GCAG.....................----------GT---C--A--T--GT---A------G-    323
HIVSF33    -----A-GCAG.....................----------GT---C--A--T--GT---A------G-    323
HIVHAN     -----A-GCAG.....................----------GT---C--G--T--GT---A------C-    323
HIVRF      -----A-GCAG.....................----------GC---C--G--T--GT---A------G-    323
CONSENSUS2 CTGGGACGCAG?????????????????????GGGGTGGGAagcCCTCAAATAtCTGTGGAAtCTcCTgC    322
HIVELI     ----------.....................----------CAT---------T---------T--C--A-    323
HIVZ2Z6    ----------.....................----------AGC---------T---------T--C--A-    323
HIVNDK     ----------.....................----------AGC---------C---------C--C--G-    323
HIVMAL     ----------.....................----------AGC---------T---------T--C--G-    326
HIVU455    ----------CAGCCTCAAGGGACTGAGACT---------GGG---------T---------T--T--G-    344
SIVCPZ     T------ATCTAATACTTCACAGCCTCAGACT-CTAC----GAGA--GTGCCTG---G--GGAA-TA-T-    347

CONSENSUS1 agTATTGGagtCAGGAaCTAAaGAa***          350
HIVLAI     AG------AGT-----A----A--A---          351
HIVHXB2R   AG------AGT-----A----A--A---          351
HIVNL43    AG------AGT-----A----A--A---          351
HIVMN      AG------AGT-----A----A--G---          348
HIVBRVA    AG------AGT-----A----A--A---          321
HIVSC      AG------AGT-----A----G--A---          351
HIVJH3     AA------AGT-----A----A--A---          342
HIVBAL1    AG------AGT-----A----A--A---          354
HIVJRCSF   AG------AGT-----A----A--A---          351
HIVOYI     AG------AGC-----A----A--A---          351
HIVSF2     AG------ATT-----A----A  A---          351
HIVNY5CG   AG------GGT-----A----A--A---          351
HIVSF162   AG------ATT-----A----A--A---          351
HIVCDC4    AG------AGT-----A----A--A---          351
HIVSF33    TG------AGT-----A----A--A---          351
HIVHAN     AG------AGC-----A----A--A---          351
HIVRF      AG------AGT-----G----A--A---          351
CONSENSUS2 agTATTGGaGTCaGGAACTgAagAa?AGTGC***    355
HIVELI     AG------A---A------G-GG-AC--------    357
HIVZ2Z6    AG------A---G------G-AG-AC--------    357
HIVNDK     AG------A---A------G-GG-A***          351
HIVMAL     AA------G---A------G-AG-A***          354
HIVU455    TG------G---G------A-AA-T***          372
SIVCPZ     AA------G-AAA---G--A-AA-T***          375
```

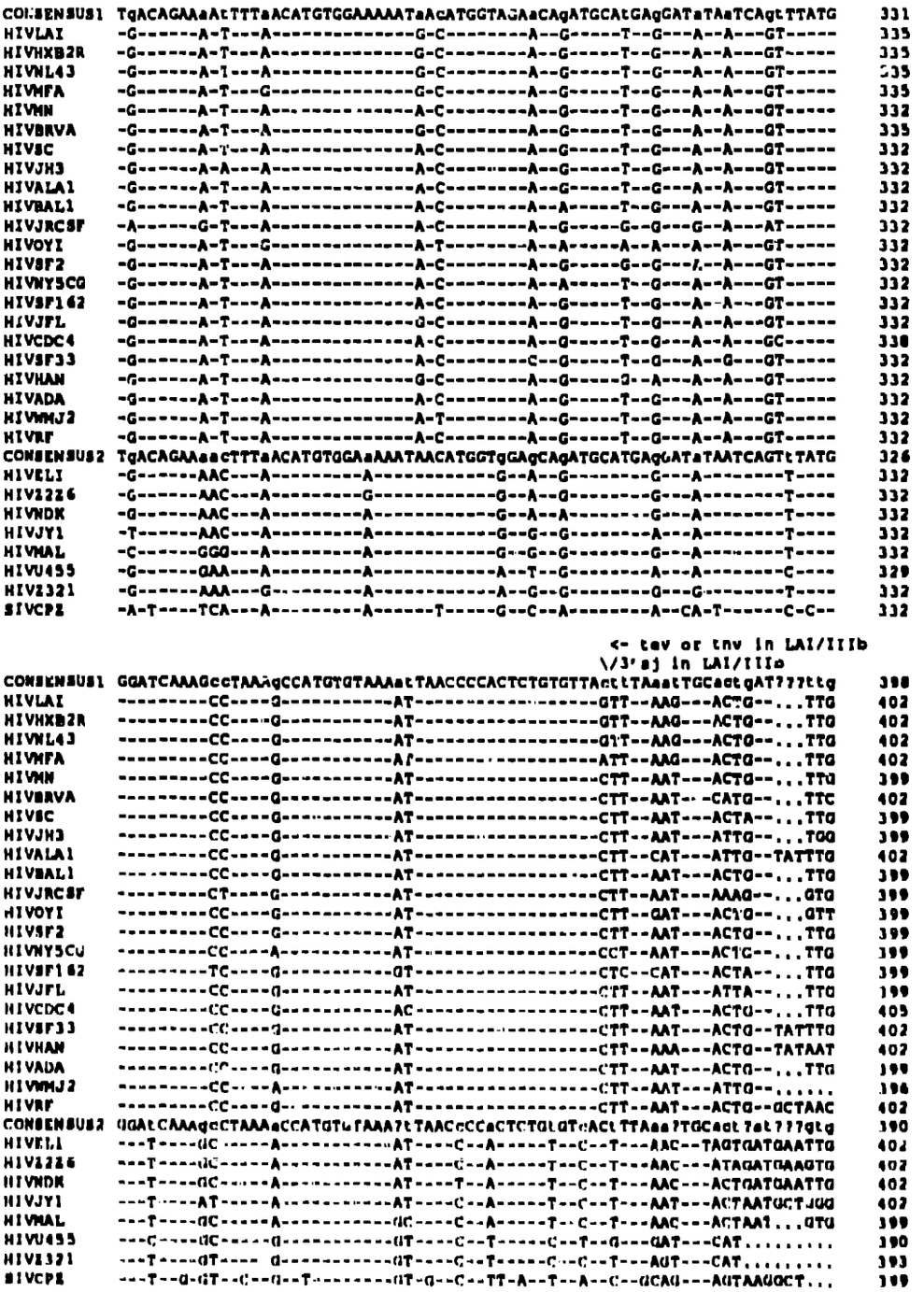
HIV-1 VPU

```
CONSENSUS1 ATGcaacCTtTacaAaTAt?agCaATAGtAgcATTAGTAGTAGcAgcAATAaTAGcAATAgTTGTgTGg.   68
HIVLAI     ---CAAC--A-ACA-A--...G-A----C-GC----------C-AT----A---C----G----G--G.   66
HIVNL43    ---CAAC--A-AAT-G--...G-A----T-GC----------C-AT----A---C----G----G--G.   66
HIVMN      ---CAAC--T-AGT-A--GCAG-A----T-GC----------C-GG----A---C----G----G--A.   69
HIVBRVA    ---CAAC--T-ACA-A--TCAG-A----T-GC----------C-GC----A---C----G----T--G.   69
HIVSC      ---CAAC--T-ACA-A--TTAT-A----T-GC----------C-GC----A---C----G----G--G.   69
HIVJH3     ---CAAT--T-ATA-A--GCAG-A----T-GC----------T-GC----A---C----G----G--G.   69
HIVBAL1    ---CAAG--T-ACA-A--TCAG-T----T-GG----------C-GC----A---C----G----G--G.   69
HIVSF2     ---CAAT--T-ACA-A--TTAG-A----T-TC----------T-GC----A---C----G----G--G.   69
HIVJRCSF   ---CAAC--T-ACA-A--TTAG-A----T-GC----------C-GG----A---C----A----G--G.   69
HIVNY5CG   ---CTTT--T-.....................................C----G----G--G.   25
HIVSF162   ---CAAC--C-TCA-A--TTAG-A----T-GC----------C-GC----A---C----G----G--G.   69
HIVHAN     ---TTAT--T-ACA-A--GCAG-A----T-GC----------C-GC----A---T----G----G--G.   69
HIVRF      ---CAAT--T-AGA-A--TTAG-A----T-GC----------C-GC----C---C----G----G--G.   69
CONSENSUS2 ATgCAACCTTtAgtGATAaTAGCAATAGcAGCATTAGTAGTAgCaatAATAaTAGCAATAGTTGTGTGG.   69
HIVELI     --G-------T-GG----A---------C-------------G-AAT----C----------------.   69
HIVZ2Z6    --G-------C-CA----A---------C-------------G-AGC----A----------------.   69
HIVNDK     --G-------T-GT----A---------C-------------G-AAT----A----------------.   69
HIVMAL     --A-------T-GT----T---------T-------------A-GCT----A----------------.   69
SIVCPZ     --GACT-TG-T-GTCGGCT--...G-TCTCAT--------GCTTAATTGC-TGGAAC--TTG-A-A---G   67

CONSENSUS1 ..?CcATAGtattcaTA?gAAtatAGgAAAatAttAAGaCAAAga.......??aAAaTAgAcAGgtTAA  125
HIVLAI     ..T-C----TAATCA--.G--TAT--G---AT-TT---A----GA.........A--A--G-C--GT---  124
HIVNL43    ..T-C----TAATCA--.G--TAT--G---AT-TT---A----GA.........A--A--G-C--GT---  124
HIVMN      ..T-C----TATTCA--.G--TAT--G---AT-AG---A----GA.........A--A--G-C--GT---  127
HIVBRVA    ..T-C----CATTAC--.G--TAT--A---TT-TT---A----GG.........A--A--G-C--GT---  127
HIVSC      ..T-T----TATTCA--TT--TAT--G---AT-TT---A----GA.........A--A--G-C--GT---  128
HIVJH3     ..T-C----TATTCA--.G--TAT--G---AT-TT---A----GA.........A--A--G-C--GT---  127
HIVBAL1    ..A-C----TATTCA--.G--TAT--G---AT-TT---G----GA.........A--A--G-C--GT---  127
HIVSF2     ..A-C----TACTCA--.G--TAT--G---AT-TT---A----GA.........--A--G-C--AT---  126
HIVJRCSF   ..T-C----TACTCA--.G--TAT--G---AT-TT---A----GA.........A--A--G-T--GT---  127
HIVNY5CG   ..A-C----TGTTCA--.G--TAC--G---GT-TT---A----GA.........A--G--G-C--GA---   83
HIVSF162   ..A-C----TGTACA--.G--TAC--G---AT-TT---A----GA.........A--A--G-T--GT---  127
HIVHAN     ..A-C----TAGTCA--.G--TAT--G---TT-TT---A----GA.........A--A--A-C--GT---  127
HIVRF      ..A-C----TTGGCA--.G--ATT--G---AC-TT---G----AA.......AAA--A--G-C--GT---  129
CONSENSUS2 ..ACCATAGTATTcATA.GAAtaTAGgAggATAAaAAggCAAAG?.........AaAATAGACtGttTaa  126
HIVELI     ..----------TC----.---TA---A-GG----A--AG-----G.........-G-------T-TT-AC  127
HIVZ2Z6    ..----------TC---.---TA---G-GG----A--GG-----A.........-A-------T-TA-AA  127
HIVNDK     ..----------AC---.---TA---G-GG----A--GG-----A.........-A-------T-TT-AA  127
HIVMAL     ..----------TT---.---AT---G-AA----G--GA-----G.........-A-------A-GT-GA  127
SIVCPZ     GATA---TA-TAAATGG.-G-TA---A-GGTAT-A--GA--T--GCTTGAGACAGA---T--GA-AC-TA  136

                                    env cds ->
CONSENSUS1 TtGATAgAaTAAqaGAaAqAGCAGAaGACAGTGGCAATGAgAGtGAaGGggatcaggaaGAa...tTaTC  192
HIVLAI     -T----G-C---TA--A-G------A--------------G--T--A--A.........--A...A-A--  182
HIVNL43    -T----G-C---TA--A-G------A--------------G--T--A--A.........--A...G-A--  182
HIVMN      -T----G-A---GC--A-G------A--------------G--T--A--GGATCAGGAG--A...T-A--  194
HIVBRVA    -A----G-A---GA--A-G------A--------------G--T--A--GGATCAAGAA--A...T-A--  194
HIVSC      -T----G-A---GA--A-G------A--------------G--T--A--GATCAGGAAG--A...T-A--  194
HIVJH3     -T----G-A---GA--A-G------A--------------G--T--A--GGATCAGGAA--A...T-A--  194
HIVBAL1    -T----G-A---CA--A-G------A--------------G--T--C--AGATCAGGAA--G...T-A--  194
HIVSF2     -T----G-A---GA--A-A------A--------------A--T--A--GGACCAGGAG--A...T-A--  193
HIVJRCSF   -T----A-A---GA--G-G------A--------------G--T--A--GGATCAGGAA--A...T-A--  194
HIVNY5CG   -T----G-A---GA--A-G------A--------------G--C--A--GGACCAGGAA--A...T-A--  150
HIVSF162   -T----G-A---CA--A-G------A--------------G--T--A--GGATCAGGAA--A...T-A--  194
HIVHAN     -T----G-C---CA--A-G------A--------------A--T--A--AGACCAGGAA--A...T-A--  194
HIVRF      -T----G-A---GA--A-G------A--------------G--T--T--AGATGAGGAA--A...T-G--  196
CONSENSUS2 TTGATAGAATAAqaGAAAGAGCAGAAGAcAGTGGCAATGAGAG?GAGGG?GAtAqAGAGqAA...TTqTC  191
HIVELI     ----------C---------------C--------------C-----G--T-G----A--...--G--  194
HIVZ2Z6    ----------G---------------C--------------T-----G--T-G----G--...--G--  194
HIVNDK     ----------G---------------C--------------C-----A--A-G----G--...--G--  194
HIVMAL     ----------G---------------T--------------T-----A--T-C----G--...--A--  194
SIVCPZ     ACCTA-T-T-G-G---------------T-----A-----A--TA-T--A--AGA---AG-GAGAC-GGA  206
```

```
CONSENSUS1 AGCaCTt????????????gTGGagAtG???GGGCAccaTGCT...CCTtGGGataTTgATGATCtG***   244
HIVLAI     ---A--TGTGGAGATGGGGG---AA-T-...-----CCA----...---T---ATA--G------T----   246
HIVNL43    ---A--TGTGGAGATGGGGG---AA-T-...-----CCA----...---T---ATA--G------T----   246
HIVMN      ---A--G...........G---GG-T-...-----CGA----...---T---TTA--A------T----   246
HIVBRVA    ---A--T...........G---AG-T-GGG-----TGA----...---T---ATA--G------T----   249
HIVSC      ---A--T...........G---AG-T-...-----CGA----...---G---ATA--G------T----   246
HIVJH3     ---A--T...........G---AG-T-...-----CGA----...---T---ATA--G------T----   246
HIVBAL1    ---A--G...........G---AG-T-...-----TCA----...---T---ATA--A------T----   246
HIVSF2     ---A--T...........G---AG-T-...-----CCT----...---T---ATG--G------T----   245
HIVJRCSF   ---A--T...........G---AA-G-...-----TCT----...---T---ACA--A------T----   246
HIVNY5CG   ---A--T...........G---AG-T-...-----CGA----...---T---ATG--A------T----   202
HIVSF162   ---A--T...........G---AG-G-...-----CCT----...---T---ATG--G------T----   246
HIVHAN     ---G--T...........G---AG-G-...-----CCT----...---T---ATG--A------A----   246
HIVRF      ---A--T...........G---AA-T-...-----CCA----...---T---ATG--G------T----   248
CONSENSUS2 AAAACT?...........GTGGAaATG...GGGCATcATGCT...CCTTGGGAT?TTGATGACCTG***   241
HIVELI     ------G...........-----A---...------C-----...---------A----------      246
HIVZ2Z6    ------T...........-----A---...------C-----...---------A----------      246
HIVNDK     ------T...........-----A---...------C-----...---------G----------      246
HIVMAL     ------G...........-----G---...------G-----...---------G----------      246
SIVCPZ     -C-G--TATCCAT...AATTACA-TCATAACAAT---TT----AAC--C....--G--..------C---   267
```

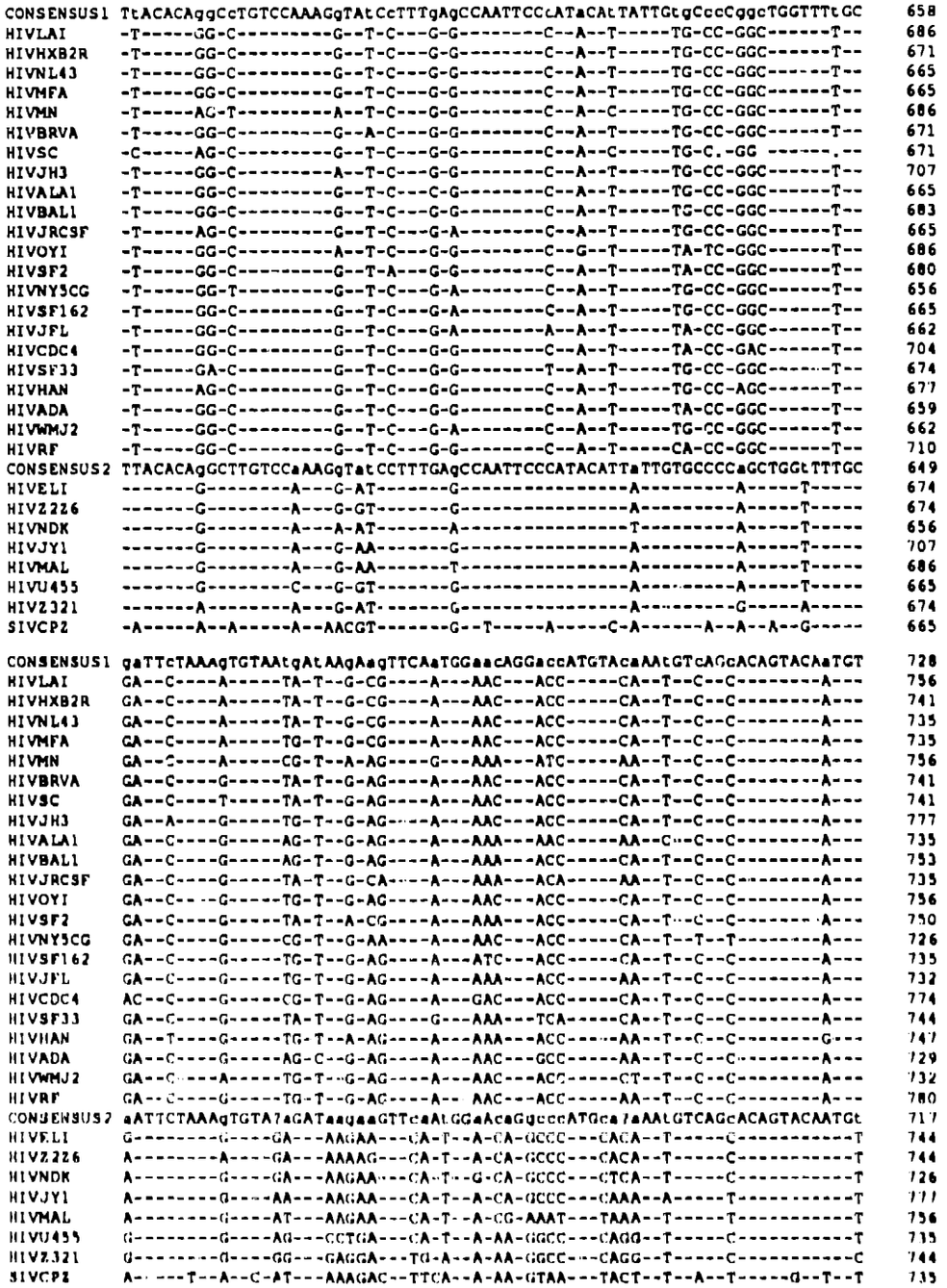

signal peptide \/ gp120
<- vpU

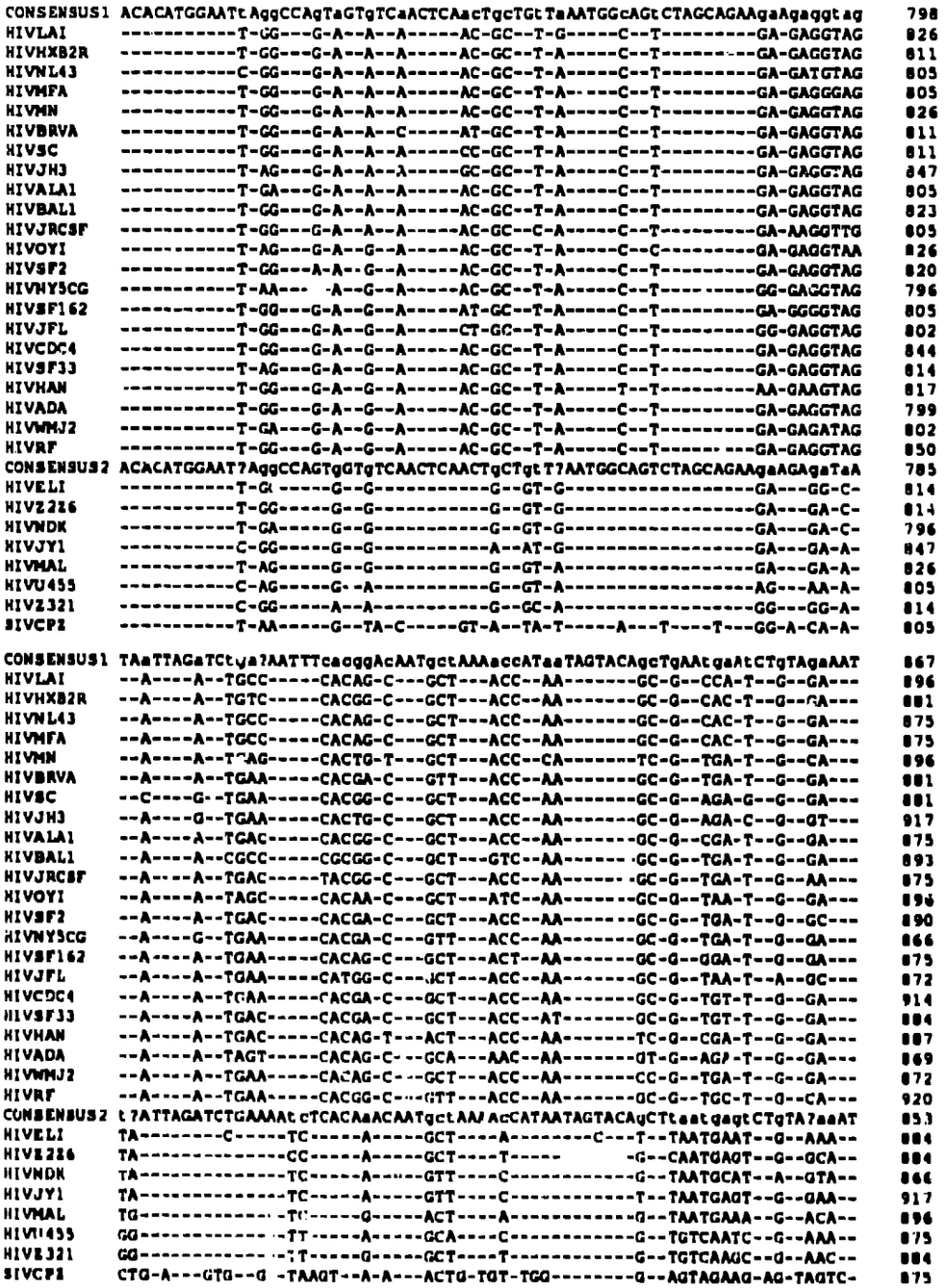


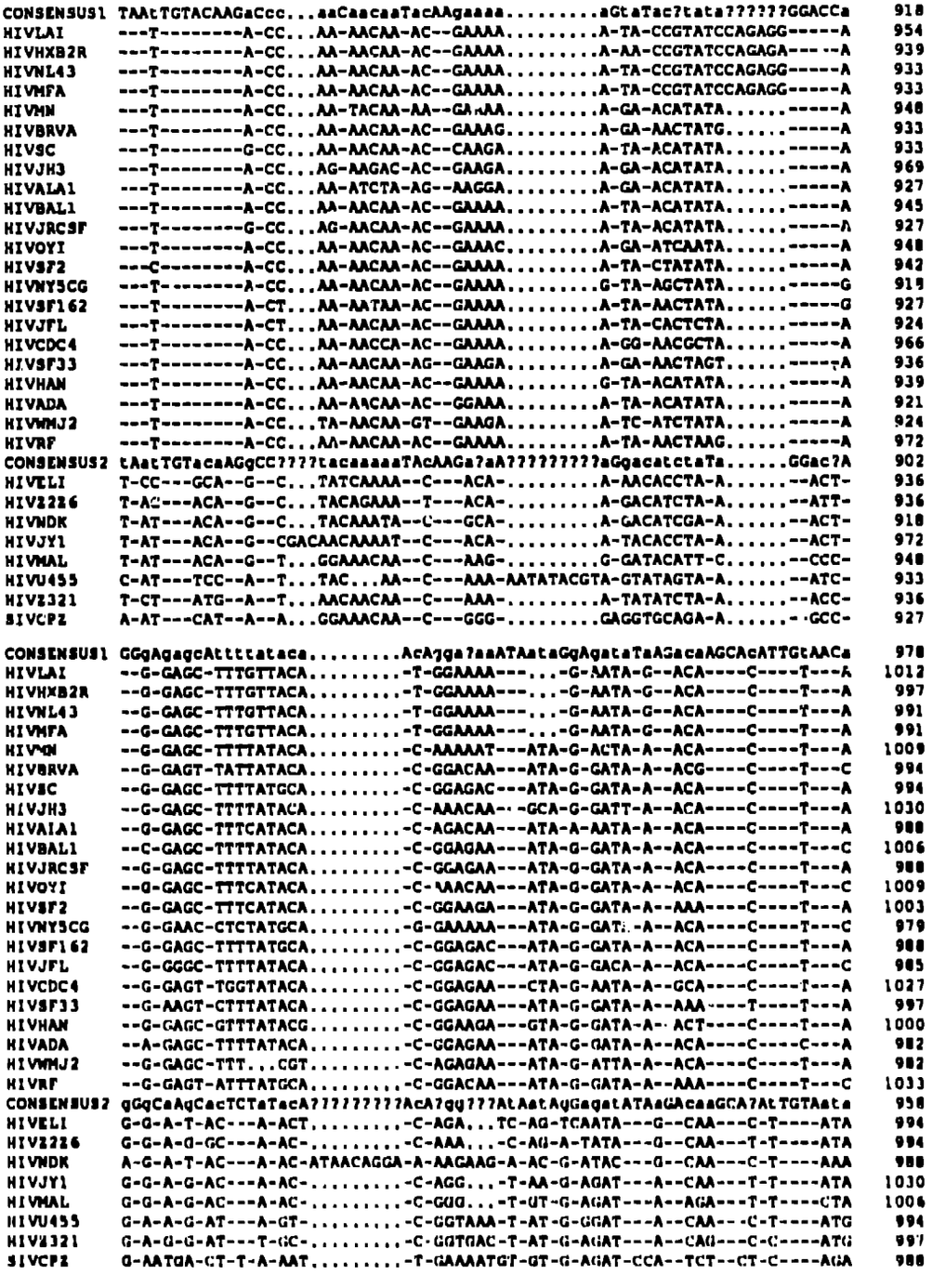

<- tev or tnv in LAI/IIIb
\/3'sj in LAI/IIIa

```
CONSENSUS1 a?g???AAtgctACT??????????????????????aataccA?tagtagtag???????????ggg   428
HIVLAI     GGG...--TGCT---AATACCAATAGTAGT.........AATACC-ATAGTAGTAGC.........GGGG   451
HIVHXB2R   AAG...--TGAT---.......................AATACC-ATAGTAGTAGC.........GGGA   436
HIVNL43    AAG...--TGAT---.......................AATACC-ATAGTAGTAGC.........GGGA   436
HIVMFA     AAG...--TGAT---.......................AATACC-ATAGTAGTAAC.........GGGA   436
HIVMN      AGG...--TACT---AATACCAATAATAGT...ACTGCTAATAAC-ATAGTAATAGC.........GAGG   454
HIVBRVA    ......--TGCT---.......................AATGCC-CTAGTAATAGC.........GGGA   433
HIVSC      AGG...--TGAT---AGCACCAATGCTACT.........AATACC-CTAGTAGTAAT.........CGGG   448
HIVJH3     GGG...--TGAT---AGCCCCAATGCTACT.........AATACC-CTAGTAGTGGC.........GGGG   448
HIVALA1    ......--TGCT---.......................AATACC-CTAGTAGTAAC.........GGGA   433
HIVBAL1    AGG...--TGCT---AATGGGAATGACACT.........AATACC-CTAGTAGTAGC.........AGGG   448
HIVJRCSF   ......--TGCT---.......................AATACC-CTAGTAGTAGT.........GAGG   430
HIVOYI     ......--TACC---AGTAGTAGTTTGAGGAATGCTACTAATACC-CAAGTAGTAGT.........TGGG   454
HIVSF2     GGG...--GGCT---.......................AATACC-ATAGTAGTAAT.........TGGA   433
HIVNY5CG   ACG...--TGCT---...TAT.............GCC...AATGGT-GTAGT.....................  429
HIVSF162   AAG...--TGCT---.......................AATACC-AGAGTAGTAAT.........TGGA   433
HIVJFL     AAG...--TGTT---..............AATATC...AATAAT-GT.....................TGGG   430
HIVCDC4    AATACT--TAAT---ACT...AATACTACT.......GAACTATC-ATAATAGTAGT.........TTGGG   457
HIVSF33    GGG...--TGCT---.......................AATACC-ACAATAGTAGT.........GGGG   436
HIVHAN     ......--TGCC---.............................-ACAGTAGT............TGGG   424
HIVADA     AGG...--TGTT---..............AATATC...AATAAT-GTAGTGAG..............G   433
HIVWMJ2    AAG...--CATT---GATTGG..................AAGAATACA-CA.....................  426
HIVRF      TTG...--TGGT---.......................AATGTC-CTAGTAGTAGC.........GGGG   436
CONSENSUS2 ??????AA?atcAcc???at??????????aatg??aatAAt?ccaat??????????????????ga?g   418
HIVELI     AGGAAC--TGGC-CT...ATG.............GGGAAC--TGTCACTACA.....................  438
HIVZ2Z6    ATGGAG--TGTC-CA...ATG.........AAGAATAAT--TGTCACA.........................  438
HIVNDK     AGG...--C...-GC.........................--GGGCAATGGG.................AAGG   427
HIVJY1     GGG...--TAAG-CT.............ACTAATGGGAAT--TACTACTAATCAG.............GAGG   445
HIVMAL     ......--TGGG-CTGCTGTG.........AATGGGACT--TGCTGGG...AGTAATAGGACTAATGCAG   451
HIVU455    ......--CATC-CC...ATTAACAACACCAAT...AAC--CACCAATATCACA.............GATG   436
HIVZ321    ......--CATC-CC...ATC.........AAGGACAAT--CACCAATGTAGAT.................  429
SIVCPZ     ......--CTTT-GC......CAGGCAAAAAACCTAACA--CCAGACATCTTCT.................  441

                                                    tev or tnv in LAI/IIIb ->
                                                                        5'sj\/
CONSENSUS1 ga???a?gatggagaaaGgAGAaaTaaaAAAcTGcTCTTTCaAtatcAcCAcAagcAtAAgAgaTAAGgt   494
HIVLAI     AAATGATGATGGAGAAA-G---GA-AAA---C--C------A-TATC-G--C-AGC-T--G-GG----GT   521
HIVHXB2R   GAATGATAATGGAGAAA-G---GA-AAA---C--C------A-TATC-G--C-AGC-T--G-GG----GT   506
HIVNL43    GAATGATAATGGAGAAA-G---GA-AAA---C--C------A-TATC-G--C-AGC-T--G-GA----GT   506
HIVMFA     GAATGATAATGGAGAAA-G---GA-AAA---C--C------A-TATC-G--C-AGC-T--G-AA----GT   506
HIVMN      GA...ACAATAAAGGGA-G---AA-GAA---C--C------A-TATC-C--C-AGC-T--G-GA----AT   521
HIVBRVA    AA...ATGATGGAGGGA-G---AA-GAA---C--C------A-TATC-C--C-AGC-T--G-GA----AT   500
HIVSC      GA...AAGATGGAGGGA-G---AA-GAC---C--C------A-TATC-C--C-AGC-T--G-AG----GT   515
HIVJH3     AA...AAGATGGAGAAA-G---AA-GAA---C--C------A-TATC-C--C-AGC-T--G-GA----GT   515
HIVALA1    CA...AAGATGGAGAGA-G---AA-AAA---C--C------A-TATC-C--C-AGC-T--G-GA----AT   500
HIVBAL1    GA...ATGGTGGGGGGA-G---AA-GAA---T--C------A-TATC-C--C-AAC-T--G-GG----GT   515
HIVJRCSF   GA...ATGATGGAGAGA-G---AA-AAA---C--C------A-TATC-C--A-AGC-T--G-GA----GT   497
HIVOYI     AA...ACGATGGAGAAA-G---AT-AAA---C--C------A-TACC-C--C-AGC-T--G-GA----AT   521
HIVSF2     AAGAAGAAATA...AAA-G---AA-AAA---C--C------A-TATC-C--C-AGC-T--G-GA----AT   500
HIVNY5CG   ........GAGGAGAGA-G---AA-AAG---T--C------A-TGTC-C--C-ATC-T--G-AA----AT   491
HIVSF162   AAGAG...ATGGACACA-G---AA-AAA---T--C------A-GGTC-C--C-AGC-T--G-AA----AT   500
HIVJFL     GAAAG...ATGGAGGAA-G---AA-AAA---C--C------A-TACC-C--C-AGC-T--A-AA----AT   497
HIVCDC4    AACAA...CGGGGTAAA-G---AA-GAG---C--T------A-TATC-C--C-AGC-T--G-GA----GT   524
HIVSF33    GA...ACGGTGGAGAAA-A---AA-AAA---C--C------A-TATC-C--C-GGC-T--G-GA----GT   501
HIVHAN     GA...AGGATGGAGAAA-G---AA-ACA---C--C------A-GGTC-C--C-AAC-T--G-GA----GT   491
HIVADA     GA...ATG......AGA-G---AA-AAA---C--C------A-TATC-C--C-AGC-T--G-GA----GT   494
HIVWMJ2    .....ATAATAGGAGGA-G---AG-AAA---C--C------A-TATC-C--C-AGC-G--G-GA----GT   491
HIVRF      GAACAATGATGGAGAAC-G---AA-AAA---C--C------C-AGTT-C--C-AGT-G--G-GA----AC   506
CONSENSUS2 ?a???a??gaggga?AaagggagaaatgaaaaacggcTCTTtCAATaTaaccacagta?LaAGaGATAAgaa   480
HIVELI     ........GAGGAG-AAGGA...ATG-----C--C----T----G-AACCACAGTACTA-AA-----GAA   497
HIVZ2Z6    ........GAGGAA-TAAGA...ATG-----C--C----T----A-AACTACAGTAGTA--A-----GA?   497
HIVNDK     TAGAAGAGGAGGAA-AAAGG.....----C--C----T----G-A.............--G-----GAG   479
HIVJY1     AACAGATGATGGAG-AAGGAGAAATG-----C--C----T----A-AACCACAGTGATA--C-----AAA   515
HIVMAL     AATTGAAAATGGAA-TTGGAGAAGTG-----C--C----T----A-AACCCCAGTAGGA--T-----A,,   519
HIVU455    GA.........GTG-GGGAAGAAATG-----T--C----T----A-GACCACAGAACTA--A-----GAA   497
HIVZ321    .....ACTGAAATG-AAGAAGAAATA-----T--T----A----A-GACCACAGAATTA--A-----GCA   494
SIVCPZ     ..........GCGCCTCTCGAAATG-----C--T--A-TT---G-GACAACAGAATTA--A-----GAA   500
```

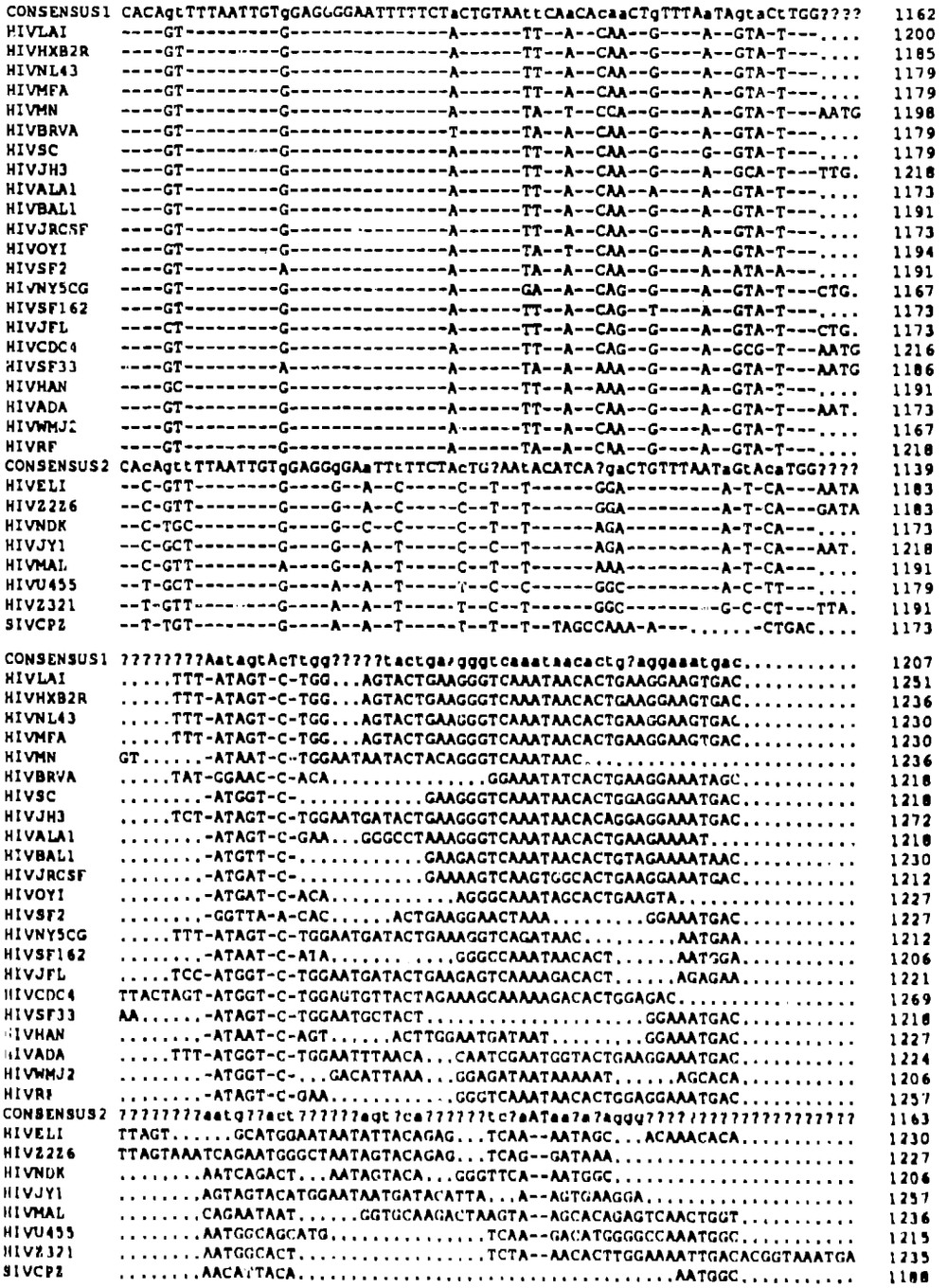
```
CONSENSUS1 gcagaaagaataTGCactttTTtaTAaacttGATgTagtAcCAATA??????gaTAAt???????????   546
HIVLAI     GCAGAAAGAATA---ATTTT--TA--AACTT---A-AAT-C-----.....GA---T.............  573
HIVHXB2R   GCAGAAAGAATA---ATTTT--TA--AACTT---A-AAT-C-----.....GA---T.............  558
HIVNL43    GCAGAAAGAATA---ATTCT--TA--AACTT---A-AGT-C-----.....GA---T.............  558
HIVMFA     GCAGAAAGAATA---ATTCT--TA--AACTT---A-AAG-C-----.....GA---T.............  558
HIVMN      GCAGAAAGAATA---ACTTC--TA--AACTT---A-AGT-T-----.....GA---T.............  573
HIVBRVA    GCAGAAAGAATA---ACTTT--TA--AACTT---A-AGT-C-----.....GA---TGAT..........  555
HIVSC      ACAGAAAGAATA---ACTTT--TA--AACTC---G-AGT-C-----.....GA---T.............  567
HIVJH3     GCAGAAAGAACA---ACTTT--TA--AACAT---G-AGT-C-----.....AA---TAGTACCAAGGAT  579
HIVALA1    GCAGAAAGAATC---GCTTT--CA--AACTT---G-AGT-C-----.....GA---T.............  552
HIVBAL1    GCAGAAAGAATA---ACTTT--TA--AACTT---A-AGC-C-----.....GA---T.............  567
HIVJRCSF   GCAGAAAGAATA---TCTTT--TA--AACTG---G-AGT-C-----.....GA---T.............  549
HIVOYI     GCAGGAACAATA---ACTTT--TA--AACTT---G-ATT-C-----.....GA---G.............  573
HIVSF2     TCAGAAAGAAAA---ACTTT--CG--ACCTT---G-AGT-C-----.....GA---TGCTAGTACT...  561
HIVNY5CG   ACAAAAAGAATA---ACTTT--TA--GACTT---A-AGT-C-----.....GA---G.............  543
HIVSF162   GCAGAAAGAATA---ACTTT--TA--AACTT---G-AGT-C-----.....GA---T.............  552
HIVJFL     GCAGAGAGAATA---ACTTT--TA--AACTT---G-AGT-C-----.....GA---G.............  549
HIVCDC4    GCAGAGAGAATA---ATTGT--TA--AACTT---G-AGA-C-----...GATGA---T.............  579
HIVSF33    ACAGAAGGCATA---ATATT--TA--AACTT---G-AGT-C-----GATGATGA---T.............  561
HIVHAN     ACAGAAAGAATC---ACTTT--TA--AAACT---G-AGT-C-----.....GA---TAATAAGACT...  552
HIVADA     GAAGAAAGACTA---ACTTT--TA--GACTT---G-AGT-C-----.....GA---T.............  546
HIVMMJ2    GCATAAAGAATA---ACTTT--TA--AACTT---G-AGT-C-----AAGGGTGA---T.............  549
HIVRF      GCAGAAAAAATA---ACTTT--TA--AACTT---G-GGT-C-----GAGAAGGG---TATTAGCCCTAAG  576
CONSENSUS2 aaagcAAgtAtATgCActtTTTTATAgaCTTGATaTAGTaCcaATa...???ga?aATaataat???...  537
HIVELI     GCAGC--GT-T--G--CTT-------GA------A----A-CA--A......GACA--GATAGTAGT...  558
HIVZ226    AAAAC--GT-C--G--CTT-------GA------A----A-CC--A......GACA--GATAATAGT...  558
HIVNDK     GGAGC--GT-T--G--CTT-------AA------A----G-CA--A......GACA--AATAATAGG...  540
HIVJY1     GAAGC--GT-C--G--CTT-------GA------G----A-CA--A......GATG--GATAATAGT...  576
HIVMAL     .AGGC--GA-T--G--ACT-------AC------C----A-AA--A......GATG--AGT.........  573
HIVU455    ACAGA--GT-T--T--CTT-------GA------A----A-AA--C.........A--AAAACT......  552
HIVZ321    AAGGA--AT-T--T--CTG-------GA------A----A-CA--T...GGGGGAA--AGT.........  552
SIVCPZ     GAAAC-GGTCT--T--CTA------GTGGAG---G-G--AAACC-A......GGGA--GAG.........  555

CONSENSUS1 ????????????????????ataata?tAcc??????a??TaTAggTTgAtAaaTTGTaAcaccTCagtCA  588
HIVLAI     ..................GAT...ACT-CC......AGC-A--CG--G-C-AG----A-CACC--AGT--  616
HIVHXB2R   ..................GAT...ACT-CC......AGC-A--AG--G-C-AG----A-CACC--AGT--  601
HIVNL43    ........................-CC......AGC-A--GG--G-T-AG----A-CACC--AGT--  595
HIVMFA     ........................-CC......ACC-A--GG--G-T-AG----A-CACC--AGT--  595
HIVMN      ..................GAT...AGT-CC......AGC-A--GG--G-T-AG----A-TACC--AGT--  616
HIVBRVA    ..................AAAACTAAT-CT......AGA-A--GG--G-T-AG----A-TACC--AGT--  601
HIVSC      ........................-CG......AGC-A--CG--G-T-AA----A-CACC--AGT--  604
HIVJH3     AATATAAAAAAT......GATAATAGT-CC......AGA-A--GA--A-T-AG----A-CACC--AGT--  637
HIVALA1    ..................AATAAT-CC......AGC-A--TA--G-T-AG----A-CACC--AGT--  595
HIVBAL1    ..................AATAGTAAT-AT......AGA-A--GG--G-T-AG----A-CACC--AGT--  613
HIVJRCSF   ..................AAGAATAAT-CC......AAA-A--GG--A-T-AG----A-CACC--AGT--  595
HIVOYI     ....................AATGAT-CT......AAA-T--GG--A-T-CA----A-CACC--AAC--  616
HIVSF2     ..............ACTACCAACTAT-CC......AAC-A--GG--G-T-CA----A-CAGA--AGT--  610
HIVNY5CG   ..................GATAATACA-GC.........-A--CA--G-T-AA----G-CACC--AGT--  586
HIVSF162   ..................GATAATACA-GC.........-A--AA--G-T-AA----A-CACC--AGT--  595
HIVJFL     ..................GATAATACA-GC.........-A--CA--G-T-AA----A-CACC--AAC--  592
HIVCDC4    .........AAAAATACTACCAACAAC-CC......AAA-A--GG--G-T-AA----A-CACC--AGT--  634
HIVSF33    ..................ACTAATACC-GC.........-A--GG--G-T-CA----A-TTCC--AGT--  604
HIVHAN     .........AGTAATAGGGATAATACT-CC......AGC-A--TG--G-T-CA----A-TAGA--TGT--  607
HIVADA     ..................GATAATACT-GC.........-A--GG--G-T-AA----A-TACC--AAC--  589
HIVMMJ2    ....................AACAGT-GT......AGA-A--GA--G-T-AA----A-TACC--AGT--  592
HIVRF      AATAATACTAGC......AATAATACT-GCTATGGTAAC-A--CA--G-T-CA----A-TTCC--AGT--  640
CONSENSUS2 ????????????????????aataatact......AatTATAGa?TAATAAATTGtaATACcTCAgccA  579
HIVELI     ..................ACCAATAGTACC......-AT-----GT----------TA----C---GCC-  604
HIVZ226    ..................ACCAATAGTACC......-AT-----AT----------TA----C---GCC-  604
HIVNDK     ..................ACCAATAGTACT......-AT-----GT----------TG----C---ACC-  586
HIVJY1     GCTAATACCAGTAAT...ACCAATTATACC......-AT-----AT----------CA----C---GCC-  637
HIVMAL     ..................GATAAT...AGT......-GT-----GC----------TA----C---GTA-  616
HIVU455    ..................GATAACAAT.........-GT-----AC----------TA----C---ACC-  595
HIVZ321    .........AGTAATGGTGAT...AGTAGT......-AA-----AC----------TA----T---GCC-  604
SIVCPZ     ....................AACAAC.........-CA-----GA----T-----CA----TA--GCC-  595
```

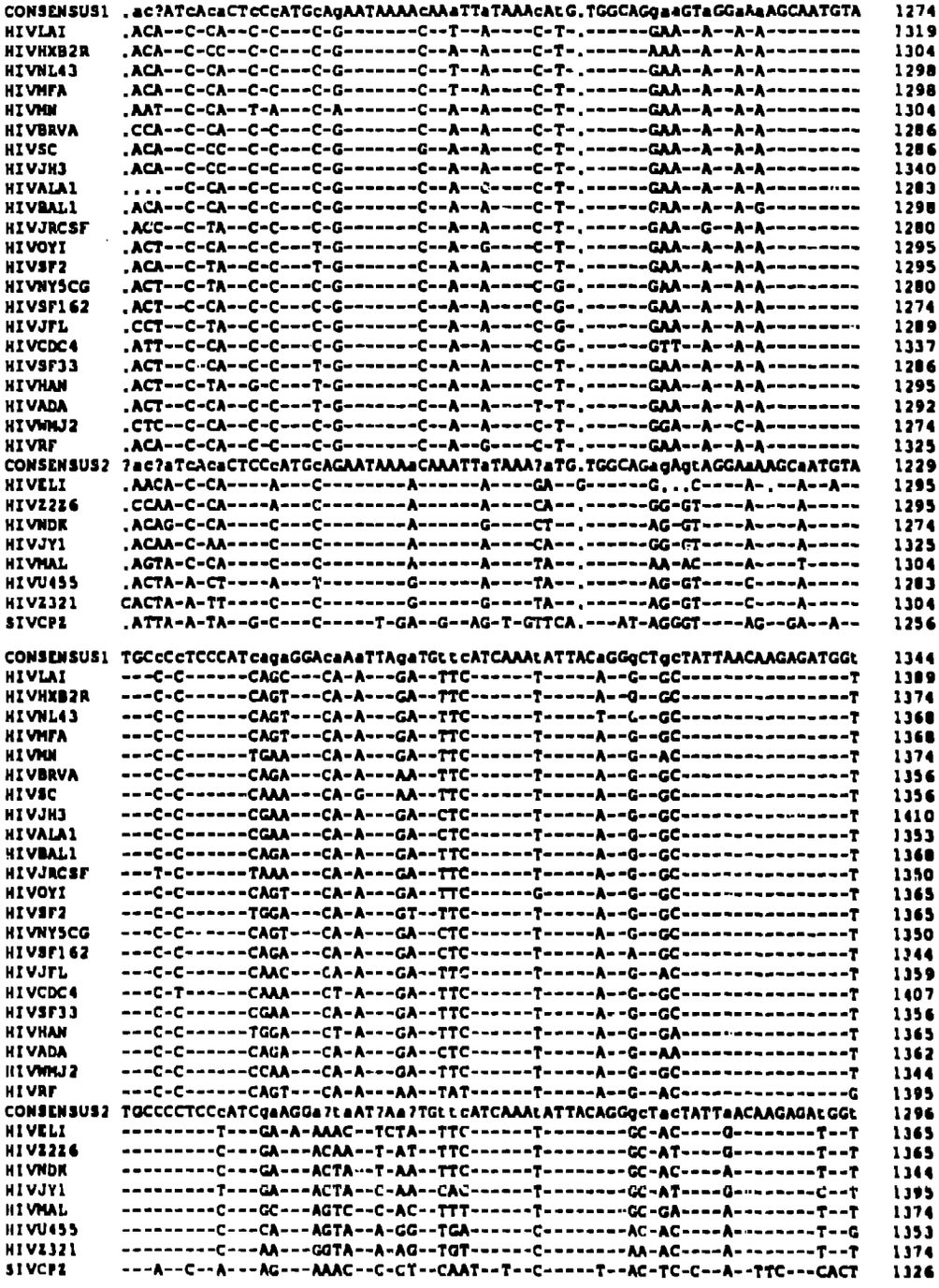

CONSENSUS1 TtACACAggCcTGTCCAAAGgTAtCcTTTgAgCCAATTCCcATaCAtTATTGtgCccCggcTGGTTTtGC 658
HIVLAI -T-----GG-C----------G--T-C---G-G--------C--A--T-----TG-CC-GGC------T-- 686
HIVHXB2R -T-----GG-C----------G--T-C---G-G--------C--A--T-----TG-CC-GGC------T-- 671
HIVNL43 -T-----GG-C----------G--T-C---G-G--------C--A--T-----TG-CC-GGC------T-- 665
HIVMFA -T-----GG-C----------G--T-C---G-G--------C--A--T-----TG-CC-GGC------T-- 665
HIVMN -T-----AG-T----------A--T-C---G-G--------C--A--C-----TG-CC-GGC------T-- 686
HIVBRVA -T-----GG-C----------G--A-C---G-G--------C--A--T-----TG-CC-GGC------T-- 671
HIVSC -C-----AG-C----------G--T-C---G-G--------C--A--C-----TG-C.-GG ------.-- 671
HIVJH3 -T-----GG-C----------A--T-C---G-G--------C--A--T-----TG-CC-GGC------T-- 707
HIVALA1 -T-----GG-C----------G--T-C---C-G--------C--A--T-----TG-CC-GGC------T-- 665
HIVBAL1 -T-----GG-C----------G--T-C---G-G--------C--A--T-----TG-CC-GGC------T-- 683
HIVJRCSF -T-----AG-C----------G--T-C---G-A--------C--A--T-----TG-CC-GGC------T-- 665
HIVOYI -T-----GG-C----------A--T-C---G-G--------C--G--T-----TA-TC-GGC------T-- 686
HIVSF2 -T-----GG-C----------G--T-A---G-G--------C--A--T-----TA-CC-GGC------T-- 680
HIVNY5CG -T-----GG-T----------G--T-C---G-A--------C--A--T-----TG-CC-GGC------T-- 656
HIVSF162 -T-----GG-C----------G--T-C---G-A--------C--A--T-----TG-CC-GGC------T-- 665
HIVJFL -T-----GG-C----------G--T-C---G-A--------A--A--T-----TA-CC-GGC------T-- 662
HIVCDC4 -T-----GG-C----------G--T-C---G-G--------C--A--T-----TA-CC-GAC------T-- 704
HIVSF33 -T-----GA-C----------G--T-C---G-G--------T--A--T-----TG-CC-GGC------T-- 674
HIVHAN -T-----AG-C----------G--T-C---G-G--------C--A--T-----TG-CC-AGC------T-- 677
HIVADA -T-----GG-C----------G--T-C---G-G--------C--A--T-----TA-CC-GGC------T-- 659
HIVWMJ2 -T-----GG-C----------G--T-C---G-A--------C--A--T-----TG-CC-GGC------T-- 662
HIVRF -T-----GG-C----------G--T-C---G-G--------C--A--T-----CA-CC-GGC------T-- 710
CONSENSUS2 TTACACAgGCTTGTCCaAAGgTatCCTTTGAgCCAATTCCCATACATTaTTGTGCCCCaGCTGGtTTTGC 649
HIVELI -------G--------A---G-AT-------G----------A---------A-----T----- 674
HIVZ2Z6 -------G--------A---G-GT-------G----------A---------A-----T----- 674
HIVNDK -------G--------A---A-AT-------A----------T---------A-----T----- 656
HIVJY1 -------G--------A---G-AA-------G----------A---------A-----T----- 707
HIVMAL -------G--------A---G-AA-------T----------A---------A-----T----- 686
HIVU455 -------G--------C---G-GT-------G----------A---------A-----T----- 665
HIVZ321 -------A--------A---G-AT-------G----------A---------G-----A----- 674
SIVCPZ -A-----A--A-----A--AACGT-------G--T-----A-----C-A------A--A--A--G----- 665
CONSENSUS1 gaTTcTAAAgTGTAAtgAtAAgAagTTCAaTGGaacAGGaccATGTAcaAAtGTcAGcACAGTACAaTGT 728
HIVLAI GA--C----A-----TA-T--G-CG----A---AAC---ACC-----CA--T--C--C--------A--- 756
HIVHXB2R GA--C----A-----TA-T--G-CG----A---AAC---ACC-----CA--T--C--C--------A--- 741
HIVNL43 GA--C----A-----TA-T--G-CG----A---AAC---ACC-----CA--T--C--C--------A--- 735
HIVMFA GA--C----A-----TG-T--G-CG----A---AAC---ACC-----CA--T--C--C--------A--- 735
HIVMN GA--C----A-----CG-T--A-AG----G---AAA---ATC-----AA--T--C--C--------A--- 756
HIVBRVA GA--C----G-----TA-T--G-AG----A---AAC---ACC-----CA--T--C--C--------A--- 741
HIVSC GA--C----T-----TA-T--G-AG----A---AAC---ACC-----CA--T--C--C--------A--- 741
HIVJH3 GA--A----G-----TG-T--G-AG----A---AAC---ACC-----CA--T--C--C--------A--- 777
HIVALA1 GA--C----G-----AG-T--G-AG----A---AAA---AAC-----AA--C--C--C--------A--- 735
HIVBAL1 GA--C----G-----AG-T--G-AG----A---AAA---ACC-----CA--T--C--C--------A--- 753
HIVJRCSF GA--C----G-----TA-T--G-CA----A---AAA---ACA-----AA--T--C--C--------A--- 735
HIVOYI GA--C----G-----TG-T--G-AG----A---AAC---ACC-----CA--T--C--C--------A--- 756
HIVSF2 GA--C----G-----TA-T--A-CG----A---AAA---ACC-----CA--T--C--C--------A--- 750
HIVNY5CG GA--C----G-----CG-T--G-AA----A---AAC---ACC-----CA--T--T--T--------A--- 726
HIVSF162 GA--C----G-----TG-T--G-AG----A---ATC---ACC-----CA--T--C--C--------A--- 735
HIVJFL GA--C----G-----TG-T--G-AG----A---AAA---ACC-----AA--T--C--C--------A--- 732
HIVCDC4 AC--C----G-----CG-T--G-AG----A---GAC---ACC-----CA--T--C--C--------A--- 774
HIVSF33 GA--C----G-----TA-T--G-AG----G---AAA---TCA-----CA--T--C--C--------A--- 744
HIVHAN GA--T----G-----TG-T--A-AG----A---AAA---ACC-----AA--T--C--C--------G--- 747
HIVADA GA--C----G-----AG-C--G-AG----A---AAC---GCC-----AA--T--C--C--------A--- 729
HIVWMJ2 GA--C----A-----TG-T--G-AG----A---AAC---ACC-----CT--T--C--C--------A--- 732
HIVRF GA--C----G-----TG-T--G-AG----A--- AAC---ACC-----AA--T--C--C--------A--- 780
CONSENSUS2 aATTCTAAAgTGTA?aGATaagaaGTTcaAtGGaAcaGgccATGca?aAAtGTCAGcACAGTACAATGt 717
HIVELI G--------G----GA---AAGAA---CA-T--A-CA-GCCC---CACA--T-----C----------T 744
HIVZ2Z6 A--------A----GA---AAAAG---CA-T--A-CA-GCCC---CACA--T-----C----------T 744
HIVNDK A--------G----GA---AAGAA---CA-T--G-CA-GCCC---CTCA--T-----C----------T 726
HIVJY1 A--------G----AA---AAGAA---CA-T--A-CA-GCCC---CAAA--A-----T----------T 777
HIVMAL A--------G----AT---AAGAA---CA-T--A-CG-AAAT---TAAA--T-----T----------T 756
HIVU455 G--------G----AG---CCTGA---CA-T--A-AA-GGCC---CAGG--T-----C----------T 735
HIVZ321 G--------G----GG---GAGGA---TG-A--A-AA-GGCC---CAGG--T-----C----------C 744
SIVCPZ A-----T--A--C-AT---AAAGAC--TTCA--A-AA-GTAA---TACT--T--A--T-----G--T--T 735

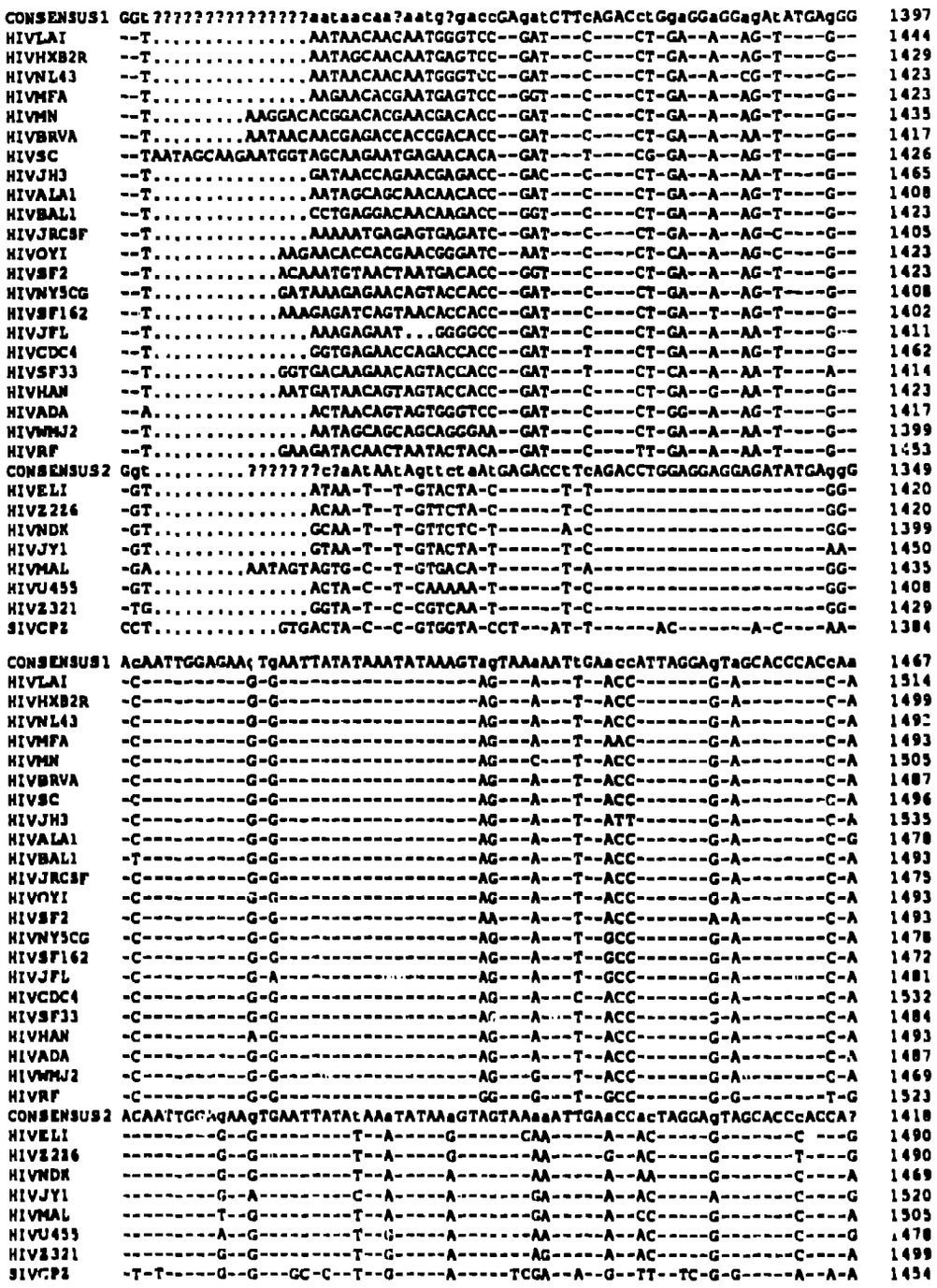


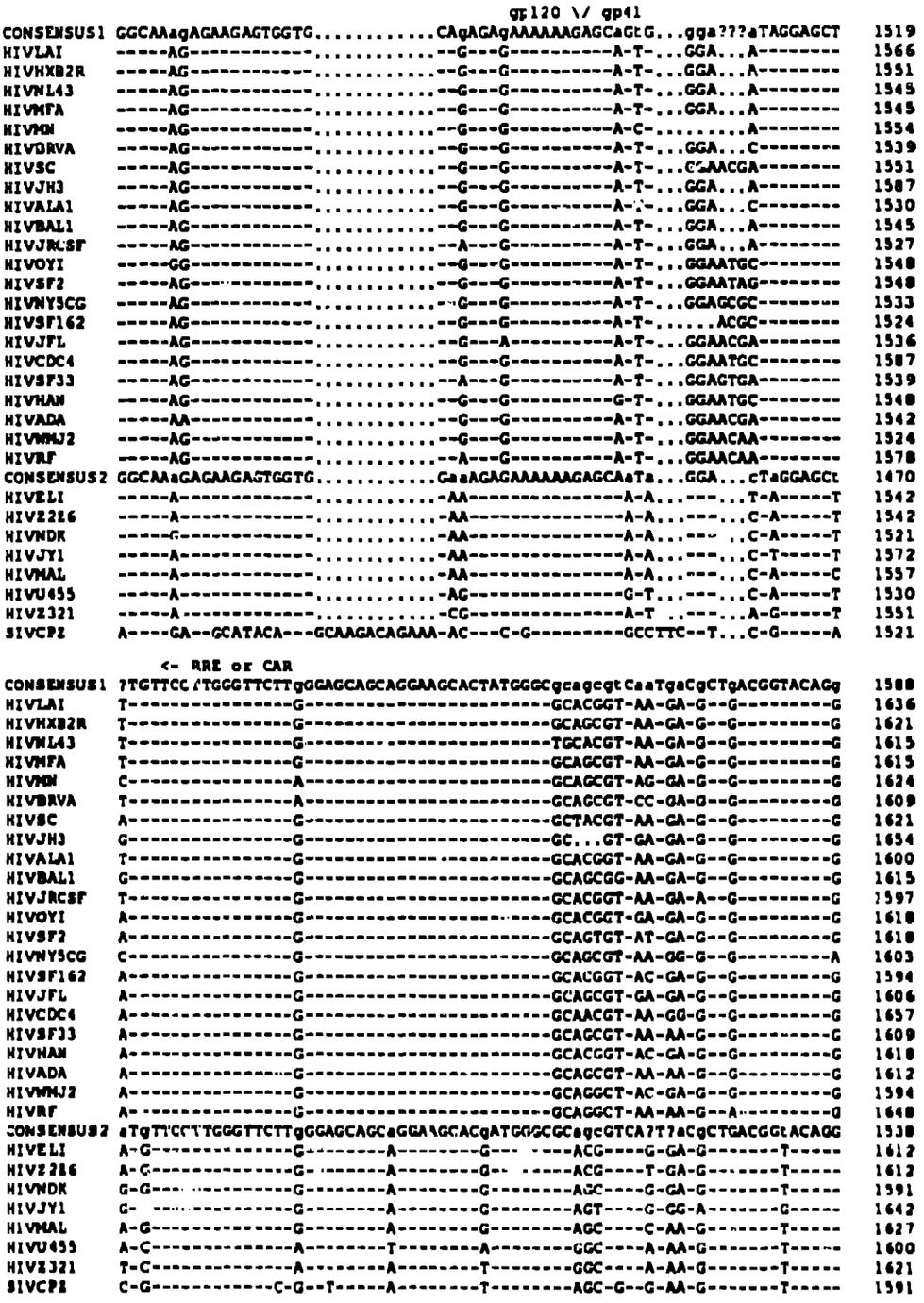


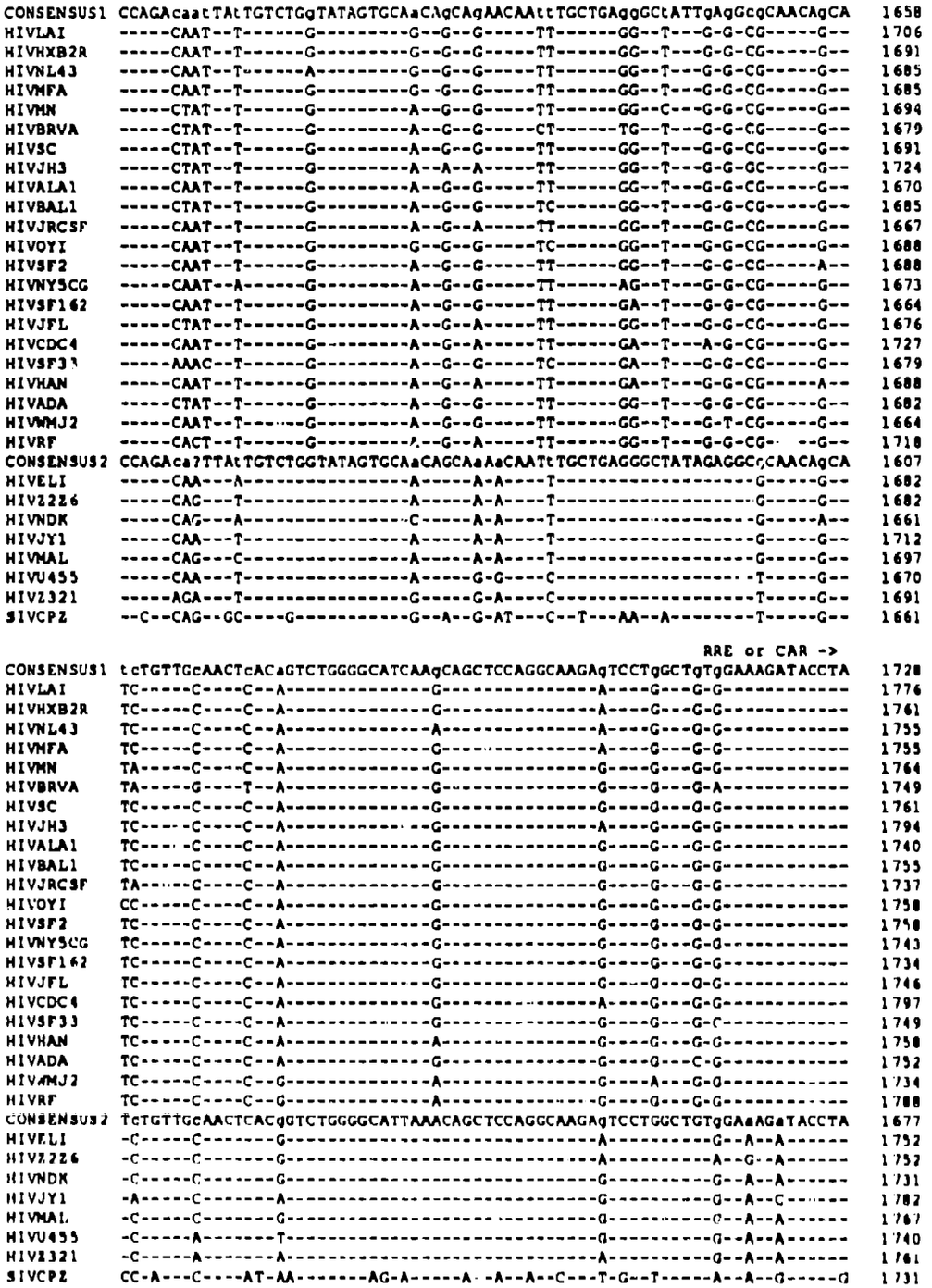
```
CONSENSUS1 TTAgTaGagcAaaATGGaataacACTtTAaaacAgaTAgtta?aAaaTTAagAgaa??????CAaTTTgg  1041
HIVLAI     ---G-A-AGC-AA----AATGCC---T--AAAC-GA--GCTAGC-AA---AG-GAA......--A---GG  1076
HIVHXB2R   ---G-A-AGC-AA----AATAAC---T--AAAC-GA--GCTAGC-AA---AG-GAA......--A---GG  1061
HIVNL43    ---G-A-AGC-AA----AATGCC---T--AAAC-GA--GCTAGC-AA---AG-GAA......--A---GG  1055
HIVMFA     ---G-A-AGC-AA----ATGTCC---T--AAAC-GA--GCAAGC-AA---AG-GAA......--A---GG  1055
HIVMN      ---G-A-AGC-AA----AATGAC---T--AGAC-GA--GTTAGC-AA---AA-GAA......--A---..  1071
HIVBRVA    ---G-A-ATC-AA----GAGAAC---T--AAAC-GA--GTTACA-AA---AG-GTA......--A---..  1056
HIVSC      ---G-A-AGC-AA----AATAAC---T--AAAC-GA--GTTATA-AA---AG-GAC......--A---GA  1058
HIVJH3     ---A-A-AGC-AG----AATGCC---T--AAAC-GA--GTTGGA-AA---AG-AAA......--A---GT  1094
HIVALA1    ---G-A-AGC-AA----GAGGAC---T--AAAC-GA--GTTGAA-AA---AG-GAA......--A---..  1050
HIVBAL1    ---G-A-AGC-AA----AATGAC---T--AATA-GA--GTTATA-AA---AG-GAA......--A---GG  1070
HIVJRCSF   ---G-A-AGC-CA----AATAAC---T--AAAC-GA--GTTGAA-AA---AG-GAA......--A---..  1050
HIVOYI     ---G-A-AGC-AC----GAGAAA---T--GAAC-GA--GCTACA-AA---AG-AAA......--A---AG  1073
HIVSF2     ---G-A-AGC-CA----AATAAC---T--GAAC-GA--GTTAAA-AA---AG-GAA......--G---GG  1067
HIVNY5CG   ---G-A-AGC-AA----AATGAC---T--AAAC-GA--GTTACA-AA---AA-GAA......--A---AG  1043
HIVSF162   ---G-G-AGA-AA----AATAAC---T--AAAC-GA--GTTACA-AA---CA-GCA......--A---GG  1052
HIVJFL     ---G-A-AGT-CA----AATAAC---T--AAAC-GA--GTTATA-AG---AG-GAA......--G---GG  1049
HIVCDC4    ---G-A-AGC-CA----AATAAC---T--CAAC-GA--GCTACA-CC---AG-GAA......--A---GG  1091
HIVSF33    ---G-A-AGC-AA----AATAAA---T--GAAC-GG--GCTACA-AA---AG-GAA......--A---GG  1061
HIVHAN     ---G-A-AGC-AG----AATAAA---C--AACC-AA--TTTAGA-AA---AG-GAAATTAGA--A---GA  1070
HIVADA     ---G-A-AAC-AA----AATAAC---T--AATC-AA--GCTACA-AA---AA-GAA......--A---GG  1046
HIVWMJ2    ---G-A-AGC-AA----AATAAC---T--AAAC-GA--GTTGAG-AA---AG-GAA......--A---..  1044
HIVRF      ---G-A-AGC-CA----AATAAC---T--AAAC-GG--GTTACA-AA---AG-GAA......--A---GA  1097
CONSENSUS2 TtAgtagAacaGaaTGGaATaaaaCttTAcaacAGGTAGCT??AaAAttAggaaac......cttcTT??  1018
HIVELI     -T-GTAG-GCACAA---AG-AAAA-TT--CAAC-A------AG-A--T--GGAACC......CTTC--..  1056
HIVZ2Z6    -T-GTAA-AAT-AA---A--AAGA-AT--CAAC--------AT-A--T--GGAAAC......CTTC--..  1056
HIVNDK     -T-GCAG-GCA-AA---A--AAAG-TT--CAAC--------AC-A--C--GGAAAC......CTTC--..  1050
HIVJY1     -T-GTGC-GCA-CA---A--AAAA-TT--CAAC--------AA-A--T--GGAGAC......CTTC--..  1092
HIVMAL     -T-ATGA-ACA-AA---G--AAAA-TT--CAAC--------GT-A--C--GGAAGC......CTTC--..  1068
HIVU455    -C-GTAG-AGG-AC---A--AGAA-TA--CAAC--------GA-C--T--AAGAAA......AAGT--..  1056
HIVZ321    -C-GTAG-ACA-AA---A--GACA-TT--TCCA--------GC-C--T--AGAAAA......CACT--GT  1061
SIVCPZ     -A-ATGGGACCACA---A--CGAA-GG-GGAAG-A---AAAAAGGCTT--GCAACC............TC  1046

CONSENSUS1 g???AATAAaACA.........aTAgtc...TTTaatcaaTCCTCAGGaGGgGAcCCAGAAATTtGTaatg  1096
HIVLAI     AAAT-----A---.........A--ATC...---AAGCAA--------A--G--C---------T--AACG  1134
HIVHXB2R   AAAT-----A---.........A--ATC...---AAGCAA--------A--G--C---------T--AACG  1119
HIVNL43    AAAT-----A---.........A--ATC...---AAGCAA--------A--G--C---------T--AACG  1113
HIVMFA     AAAT-----A---.........G--ATC...---AAGCAA--------A--G--C---------T--AACG  1113
HIVMN      .AAG-----A---.........A--GTC...---AATCAA--------A--G--C---------T--AATG  1128
HIVBRVA    .AAG-----A---.........A--GTC...---AATCGA--------A--G--C---------T--AATG  1113
HIVSC      G...-----A---.........A--ATC...---AATCGA--------A--A--C---------T--AATG  1113
HIVJH3     G...-----A---.........A--GTC...---AATCGA--------A--G--C---------C--AATG  1149
HIVALA1    .AAG-----A---.........A--GTC...---AATCAA--------G--G--C---------T--AACA  1107
HIVBAL1    G...-----A---.........A--GTC...---AAGCAC--------A--G--C---------T--GACG  1125
HIVJRCSF   .AAT-----A---.........A--GTC...---ACTCAC--------A--G--T---------T--AATG  1107
HIVOYI     G...-----A---.........A--GCC...---GATCGA--------A--G--T---------T--AATG  1128
HIVSF2     GAAT-----A---.........A--GTC...---AATCAA--------A--G--C---------T--AATG  1125
HIVNY5CG   G...-----A---.........A--GTC...---AATCAA--------A--G--C---------T--AATG  1098
HIVSF162   G...-----A---.........A--GTC...---AAGCAA--------A--G--C---------T--AATG  1107
HIVJFL     G...-----A---.........A--GTC...---AAGCAA--------A--G--C---------T--AATG  1104
HIVCDC4    G...-----A---.........A--GCC...---AATCAA--------A--G--C---------T--AATG  1146
HIVSF33    G...-----A---.........A--GTA...---AAACAA--------A--A--C---------T--AATG  1116
HIVHAN     A...-----A---.........A--GTC...---AATCGA--------A--A--C---------T--AATG  1125
HIVADA     GAAT-----A---.........A--GTC...---AATCAA--------A--G--C---------T--AATG  1104
HIVWMJ2    .AAG-----A---.........A--GTC...---AATCAT--------A--G--C---------T--AACG  1101
HIVRF      C...-----A---.........A--GTC...---ACGTCA--------A--G--C---------T--ACTT  1152
CONSENSUS2 ????AaCaaaACAaca...???ATAAtt...TTTaatccatCCtCAGGAGGgGAcccAGAAATTACAaCA  1075
HIVELI     ....-A-AAA---.........----TAAAG---AAACCAT--T-------G--CCC----------A--  1113
HIVZ2Z6    ....-A-AAA---ACA......----TT...---AAACCAT--T-------G--CCC----------A--  1113
HIVNDK     ....-A-AAA---ACA......----CT...  -AAGCCAT--T-------G--CCC----------T--  1107
HIVJY1     ....-A-CAG---ACA......----TT...---AAACCAC--G-------A--CCC----------A--  1149
HIVMAL     ....-A-AAA---AAA......----TT...---AATTCAT--T-------G--CCC----------A--  1125
HIVU455    .AAT-A-AAA---.........----TC...---GCTAGCT--T-------G--TAT----------A--  1113
HIVZ321    AAAT-C-AGT---.......GAT----TC...---GCTAACT--T-------G--TGT----------A--  1122
SIVCPZ     ATCA-ATCGA---GCAGCCAAC----CA...C-CAACAGAG--T-T-----A--TCC---GG-C---CAT  1113
```

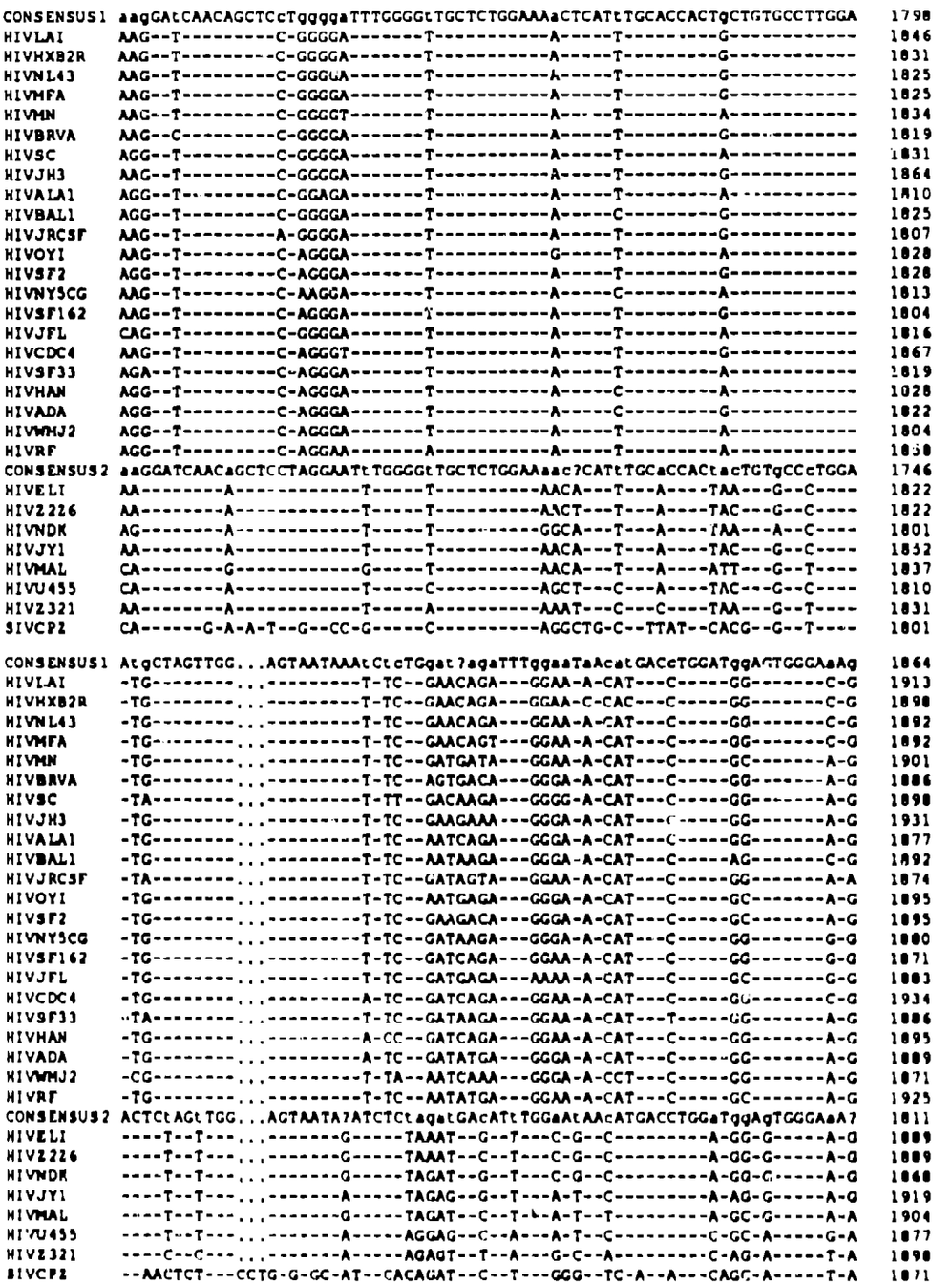


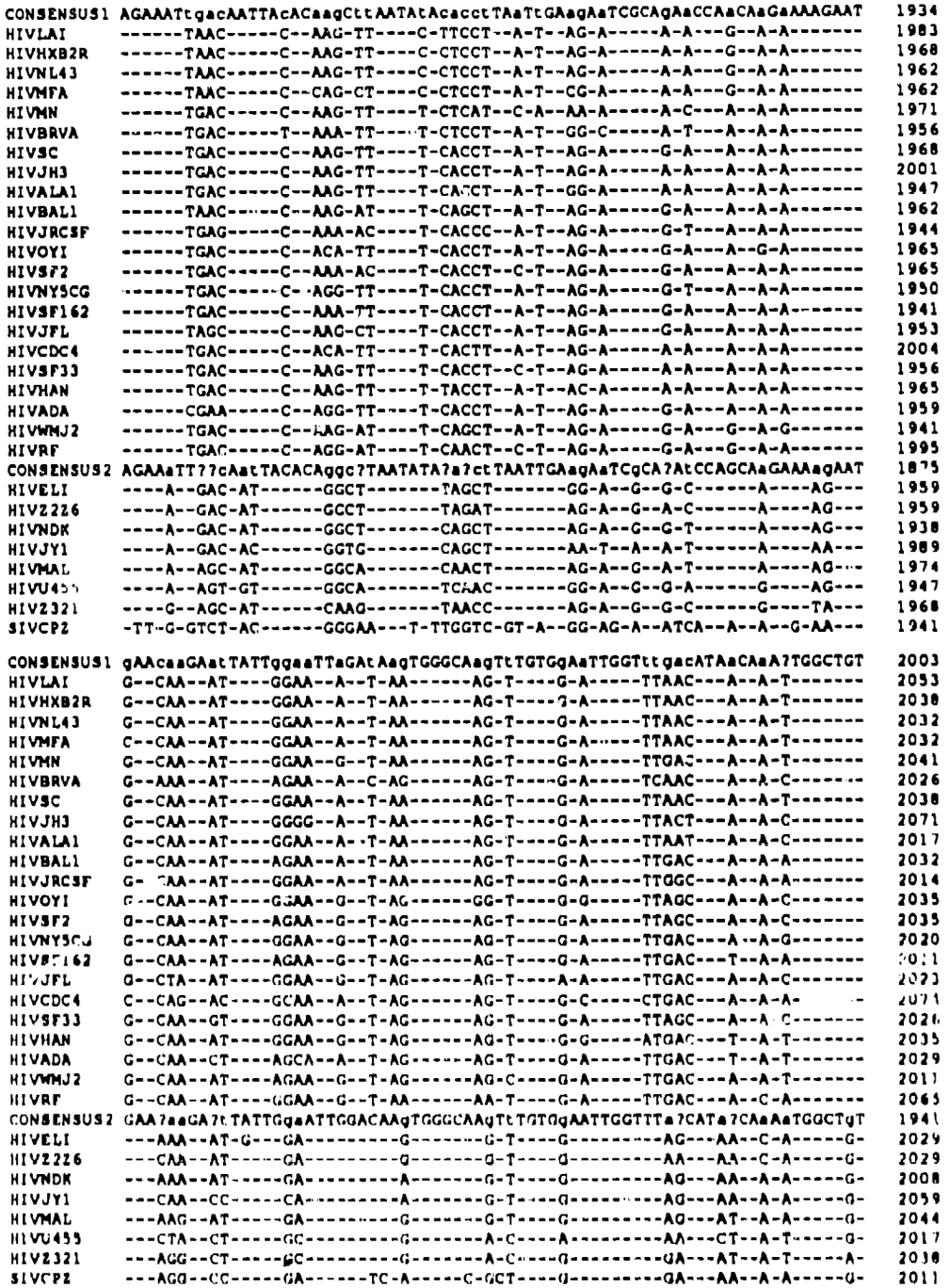


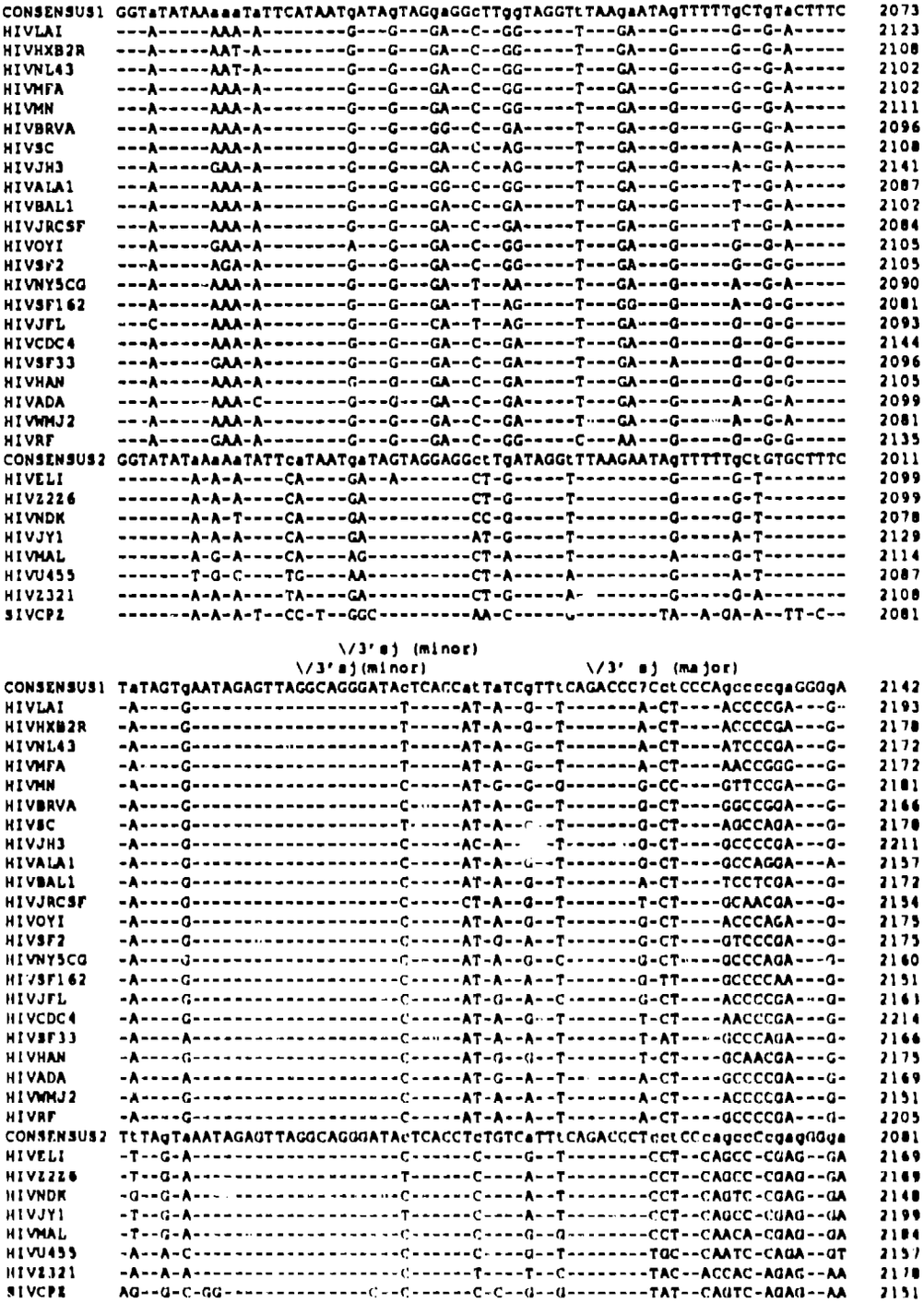


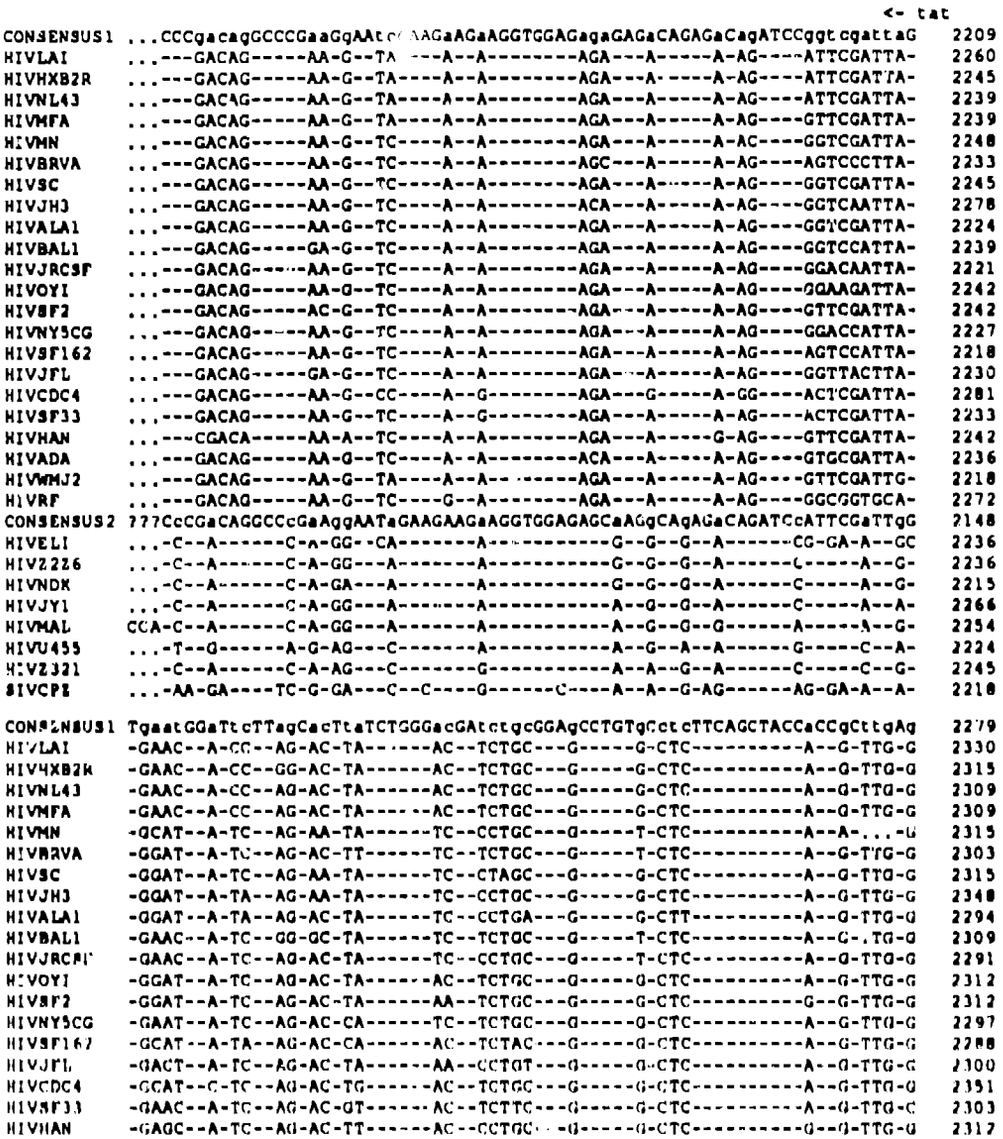

gp120 \/ gp41
<- RRE or CAR


RRE or CAR ->
CONSENSUS1
HIVLAI
HIVHXB2R
HIVNL43
HIVMFA
HIVMN
HIVBRVA
HIVSC
HIVJH3
HIVALA1
HIVBAL1
HIVJRCSF
HIVOYI
HIVSF2
HIVNY5CG
HIVSF162
HIVJFL
HIVCDC4
HIVSF33
HIVHAN
HIVADA
HIVMMJ2
HIVRF
CONSENSUS2
HIVELI
HIVZ226
HIVNDK
HIVJY1
HIVMAL
HIVU455
HIVZ321
SIVCPZ


CONSENSUS1 aagGAtCAACAGCTCcTgqggaTTTGGGGtTGCTCTGGAAAaCTCATtTGCACCACTgCTGTGCCTTGGA 1798
HIVLAI AAG--T---------C-GGGGA-------T----------A-----T---------G---------- 1846
HIVHXB2R AAG--T---------C-GGGGA-------T----------A-----T---------G---------- 1831
HIVNL43 AAG--T---------C-GGGGA-------T----------A-----T---------G---------- 1825
HIVMFA AAG--T---------C-GGGGA-------T----------A-----T---------G---------- 1825
HIVMN AAG--T---------C-GGGGT-------T----------A-----T---------A---------- 1834
HIVBRVA AAG--C---------C-GGGGA-------T----------A-----T---------G---------- 1819
HIVSC AGG--T---------C-GGGGA-------T----------A-----T---------A---------- 1831
HIVJH3 AAG--T---------C-GGGGA-------T----------A-----T---------G---------- 1864
HIVALA1 AGG--T---------C-GGAGA-------T----------A-----T---------A---------- 1810
HIVBAL1 AGG--T---------C-GGGGA-------T----------A-----C---------G---------- 1825
HIVJRCSF AAG--T---------A-GGGGA-------T----------A-----T---------G---------- 1807
HIVOYI AAG--T---------C-AGGGA-------T----------G-----T---------A---------- 1828
HIVSF2 AGG--T---------C-AGGGA-------T----------A-----T---------G---------- 1828
HIVNY5CG AAG--T---------C-AAGGA-------T----------A-----C---------A---------- 1813
HIVSF162 AAG--T---------C-AGGGA-------T----------A-----T---------G---------- 1804
HIVJFL CAG--T---------C-GGGGA-------T----------A-----T---------A---------- 1816
HIVCDC4 AAG--T---------C-AGGGT-------T----------A-----T---------G---------- 1867
HIVSF33 AGA--T---------C-AGGGA-------T----------A-----T---------A---------- 1819
HIVHAN AGG--T---------C-AGGGA-------T----------A-----C---------A---------- 1828
HIVADA AGG--T---------C-AGGGA-------T----------A-----C---------G---------- 1822
HIVWMJ2 AGG--T---------C-AGGGA-------T----------A-----T---------A---------- 1804
HIVRF AGG--T---------C-AGGAA-------A----------A-----T---------A---------- 1858
CONSENSUS2 aaGGATCAACaGCTCCTAGGAATtTGGGGtTGCTCTGGAAaac?CATtTGCaCCACtacTGTgCCcTGGA 1746
HIVELI AA--------A----------T-----T----------AACA---T---A----TAA---G--C---- 1822
HIVZ226 AA--------A----------T-----T----------AACT---T---A----TAC---G--C---- 1822
HIVNDK AG--------A----------T-----T----------GGCA---T---A----TAA---A--C---- 1801
HIVJY1 AA--------A----------T-----T----------AACA---T---A----TAC---G--C---- 1852
HIVMAL CA--------G----------G-----T----------AACA---T---A----ATT---G--T---- 1837
HIVU455 CA--------A----------T-----C----------AGCT---C---A----TAC---G--C---- 1810
HIVZ321 AA--------A----------T-----A----------AAAT---C---C----TAA---G--T---- 1831
SIVCPZ CA------G-A-A-T--G--CC-G-----C----------AGGCTG-C--TTAT--CACG--G--T---- 1801
CONSENSUS1 AtgCTAGTTGG...AGTAATAAAtCtcTGgat?agaTTTggaaTaAcatGACcTGGATggAGTGGGAaAg 1864
HIVLAI -TG--------...---------T-TC--GAACAGA---GGAA-A-CAT---C-----GG-------C-G 1913
HIVHXB2R -TG--------...---------T-TC--GAACAGA---GGAA-C-CAC---C-----GG-------C-G 1898
HIVNL43 -TG--------...---------T-TC--GAACAGA---GGAA-A-CAT---C-----GG-------C-G 1892
HIVMFA -TG--------...---------T-TC--GAACAGT---GGAA-A-CAT---C-----GG-------C-G 1892
HIVMN -TG--------...---------T-TC--GATGATA---GGAA-A-CAT---C-----GC-------A-G 1901
HIVBRVA -TG--------...---------T-TC--AGTGACA---GGGA-A-CAT---C-----GG-------A-G 1886
HIVSC -TA--------...---------T-TT--GACAAGA---GGGG-A-CAT---C-----GG-------A-G 1898
HIVJH3 -TG--------...---------T-TC--GAAGAAA---GGGA-A-CAT---C-----GG-------A-G 1931
HIVALA1 -TG--------...---------T-TC--AATCAGA---GGGA-A-CAT---C-----GG-------A-G 1877
HIVBAL1 -TG--------...---------T-TC--AATAAGA---GGGA-A-CAT---C-----AG-------C-G 1892
HIVJRCSF -TA--------...---------T-TC--GATAGTA---GGAA-A-CAT---C-----GG-------A-A 1874
HIVOYI -TG--------...---------T-TC--AATGAGA---GGGA-A-CAT---C-----GC-------A-G 1895
HIVSF2 -TG--------...---------T-TC--GAAGACA---GGGA-A-CAT---C-----GC-------A-G 1895
HIVNY5CG -TG--------...---------T-TC--GATAAGA---GGGA-A-CAT---C-----GG-------G-G 1880
HIVSF162 -TG--------...---------T-TC--GATCAGA---GGAA-A-CAT---C-----GG-------G-G 1871
HIVJFL -TG--------...---------T-TC--GATGAGA---AAAA-A-CAT---C-----GC-------G-G 1883
HIVCDC4 -TG--------...---------A-TC--GATCAGA---GGAA-A-CAT---C-----GG-------C-G 1934
HIVSF33 -TA--------...---------T-TC--GATAAGA---GGAA-A-CAT---T-----GG-------A-G 1886
HIVHAN -TG--------...---------A-CC--GATCAGA---GGAA-A-CAT---C-----GG-------A-G 1895
HIVADA -TG--------...---------A-TC--GATATGA---GGGA-A-CAT---C-----GG-------A-G 1889
HIVWMJ2 -CG--------...---------T-TA--AATCAAA---GGGA-A-CCT---C-----GG-------A-G 1871
HIVRF -TG--------...---------T-TC--AATATGA---GGAA-A-CAT---C-----GC-------A-G 1925
CONSENSUS2 ACTCtAGtTGG...AGTAATA?ATCTCtagatGAcATtTGGaAtAAcATGACCTGGaTggAgTGGGAaA? 1811
HIVELI ----T--T---...-------G-----TAAAT--G--T---C-G--C----------A-GG-G-----A-G 1889
HIVZ226 ----T--T---...-------G-----TAAAT--C--T---C-G--C----------A-GG-G-----A-G 1889
HIVNDK ----T--T---...-------G-----TAGAT--G--T---C-G--C----------A-GG-G-----A-G 1860
HIVJY1 ----T--T---...-------A-----TAGAG--G--T---A-T--C----------A-AG-G-----A-G 1919
HIVMAL ----T--T---...-------G-----TAGAT--C--T---A-T--T----------A-GC-G-----A-A 1904
HIVU455 ----T--T---...-------A-----AGGAG--C--A---A-T--C----------C-GC-A-----G-A 1877
HIVZ321 ----C--C---...-------A-----AGAGT--T--A---G-C--A----------C-AG-A-----T-A 1898
SIVCPZ --AACTCT---CCTG-G-GC-AT--CACAGAT--C--T---GGG--TC-A--A---CAGC-A-----T-A 1871




\/3' sj (minor)
\/3' sj(minor)
\/3' sj (major)

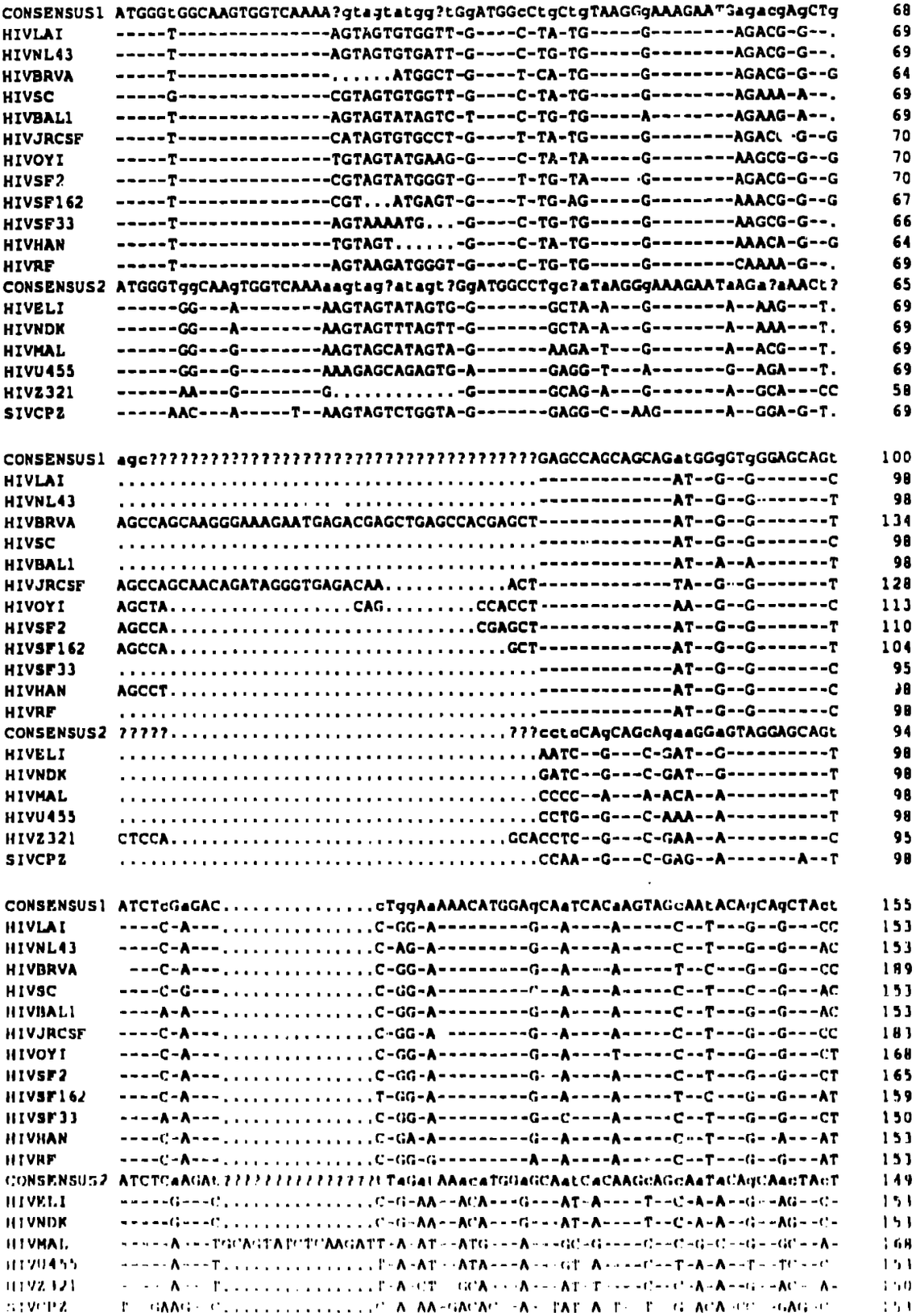

<- tat
CONSENSUS1 ...CCCgacagGCCCGaaGgAAtc AAGaAGaAGGTGGAGagaGAGaCAGAGaCagATCCggtcgattaG 2209
HIVLAI ...---GACAG-----AA-G--TA ---A--A---------AGA---A-----A-AG----ATTCGATTA- 2260
HIVHXB2R ...---GACAG-----AA-G--TA----A--A---------AGA---A-----A-AG----ATTCGATTA- 2245
HIVNL43 ...---GACAG-----AA-G--TA----A--A---------AGA---A-----A-AG----ATTCGATTA- 2239
HIVMFA ...---GACAG-----AA-G--TA----A--A---------AGA---A-----A-AG----GTTCGATTA- 2239
HIVMN ...---GACAG-----AA-G--TC----A--A---------AGA---A-----A-AC----GGTCGATTA- 2248
HIVBRVA ...---GACAG-----AA-G--TC----A--A---------AGC---A-----A-AG----AGTCCCTTA- 2233
HIVSC ...---GACAG-----AA-G--TC----A--A---------AGA---A-----A-AG----GGTCGATTA- 2245
HIVJH3 ...---GACAG-----AA-G--TC----A--A---------ACA---A-----A-AG----GGTCAATTA- 2278
HIVALA1 ...---GACAG-----AA-G--TC----A--A---------AGA---A-----A-AG----GGTCGATTA- 2224
HIVBAL1 ...---GACAG-----GA-G--TC----A--A---------AGA---A-----A-AG----GGTCCATTA- 2239
HIVJRCSF ...---GACAG-----AA-G--TC----A--A---------AGA---A-----A-AG----GGACAATTA- 2221
HIVOYI ...---GACAG-----AA-G--TC----A--A---------AGA---A-----A-AG----GGAAGATTA- 2242
HIVBF2 ...---GACAG-----AC-G--TC----A--A---------AGA---A-----A-AG----GTTCGATTA- 2242
HIVNY5CG ...---GACAG-----AA-G--TC----A--A---------AGA---A-----A-AG----GGACCATTA- 2227
HIVBF162 ...---GACAG-----AA-G--TC----A--A---------AGA---A-----A-AG----AGTCCATTA- 2218
HIVJFL ...---GACAG-----GA-G--TC----A--A---------AGA---A-----A-AG----GGTTACTTA- 2230
HIVCDC4 ...---GACAG-----AA-G--CC----A--G---------AGA---G-----A-GG----ACTCGATTA- 2281
HIVBF33 ...---GACAG-----AA-G--TC----A--G---------AGA---A-----A-AG----ACTCGATTA- 2233
HIVHAN ...---CGACA-----AA-A--TC----A--A---------AGA---A-----G-AG----GTTCGATTA- 2242
HIVADA ...---GACAG-----AA-G--TC----A--A---------ACA---A-----A-AG----GTGCGATTA- 2236
HIVWMJ2 ...---GACAG-----AA-G--TA----A--A---------AGA---A-----A-AG----GTTCGATTG- 2218
HIVRF ...---GACAG-----AA-G--TC----G--A---------AGA---A-----A-AG----GGCGGTGCA- 2272
CONSENSUS2 ???CcCGaCAGGCCcGaAggAATaGAAGAAGaAGGTGGAGAGCaAGgCAgAGaCAGATCcATTCGaTTgG 2148
HIVELI ...-C--A------C-A-GG--CA-------A----------G--G--G--A------CG-GA-A--GC 2236
HIVZ2Z6 ...-C--A------C-A-GG---A-------A----------G--G--G--A------C-----A--G- 2236
HIVNDK ...-C--A------C-A-GA---A-------A----------G--G--G--A------C-----A--G- 2215
HIVJY1 ...-C--A------C-A-GG---A-------A----------A--G--G--A------C-----A--A- 2266
HIVMAL CCA-C--A------C-A-GG---A-------A----------A--G--G--G------A-----A--G- 2254
HIVU455 ...-T--G------A-G-AG---C-------A----------A--G--A--A------G-----C--A- 2224
HIVZ321 ...-C--A------C-A-AG---C-------G----------A--A--G--A------C-----C--G- 2245
SIVCPZ ...-AA-GA----TC-G-GA---C--C----G------C----A--A--G-AG------AG-GA-A--A- 2218

CONSENSUS1 TgaatGGaTtcTTagCacTtaTCTGGGacGAtctgcGGAgCCTGTgCctcTTCAGCTACCaCCgCttgAg 2279
HIVLAI -GAAC--A-CC--AG-AC-TA------AC--TCTGC---G-----G-CTC----------A--G-TTG-G 2330
HIVHXB2R -GAAC--A-CC--GG-AC-TA------AC--TCTGC---G-----G-CTC----------A--G-TTG-G 2315
HIVNL43 -GAAC--A-CC--AG-AC-TA------AC--TCTGC---G-----G-CTC----------A--G-TTG-G 2309
HIVMFA -GAAC--A-CC--AG-AC-TA------AC--TCTGC---G-----G-CTC----------A--G-TTG-G 2309
HIVMN -GCAT--A-TC--AG-AA-TA------TC--CCTGC---G-----T-CTC----------A--A-...-G 2315
HIVBRVA -GGAT--A-TC--AG-AC-TT------TC--TCTGC---G-----T-CTC----------A--G-TTG-G 2303
HIVSC -GGAT--A-TC--AG-AA-TA------TC--CTAGC---G-----G-CTC----------A--G-TTG-G 2315
HIVJH3 -GGAT--A-TA--AG-AA-TA------TC--CCTGC---G-----G-CTC----------A--G-TTG-G 2348
HIVALA1 -GGAT--A-TA--AG-AC-TA------TC--CCTGA---G-----G-CTT----------A--G-TTG-G 2294
HIVBAL1 -GAAC--A-TC--GG-GC-TA------TC--TCTGC---G-----T-CTC----------A--G-.TG-G 2309
HIVJRCSF -GAAC--A-TC--AG-AC-TA------TC--CCTGC---G-----T-CTC----------A--G-TTG-G 2291
HIVOYI -GGAT--A-TC--AG-AC-TA------AC--TCTGC---G-----G-CTC----------A--G-TTG-G 2312
HIVBF2 -GGAT--A-TC--AG-AC-TA------AA--TCTGC---G-----G-CTC----------G--G-TTG-G 2312
HIVNY5CG -GAAT--A-TC--AG-AC-CA------TC--TCTGC---G-----G-CTC----------A--G-TTG-G 2297
HIVBF162 -GCAT--A-TA--AG-AC-CA------AC--TCTAC---G-----G-CTC----------A--G-TTG-G 2288
HIVJFL -GACT--A-TC--AG-AC-TA------AA--CCTGT---G-----G-CTC----------A--G-TTG-G 2300
HIVCDC4 -GCAT--C-TC--AG-AC-TG------AC--TCTGC---G-----G-CTC----------A--G-TTG-G 2351
HIVBF33 -GAAC--A-TC--AG-AC-GT------AC--TCTTC---G-----G-CTC----------A--G-TTG-C 2303
HIVHAN -GAGC--A-TC--AG-AC-TT------AC--CCTGC---G-----G-CTC----------G--G-TTG-G 2317
HIVADA -GGAT--A-TC--AG-AC-TT------AC--CCTGC---G-----G-CTC----------A--G-TTG-G 2306
HIVWMJ2 -GCAT--A-TC--AG-AC-TA------AC--TCTGC---G-----G-CTC----------A--G-TTG-G 2288
HIVRF -GAAT--A-TC--GA-AC-TA------AC--TCTGT---C-----G-AGC----------A--G-TTG-G 2347
CONSENSUS2 TGAaCGGATTcT?agCacTt?tCTGGGACGAcCTg/GGAaCCTGTgCCTcTTCaGcTACCACCGcTTGAG 2215
HIVELI ---A-------C-CGG-AC-TAT----------C--GC---G---------C---A-C--------C----- 2306
HIVZ2Z6 ---A-------C-CAG-AC-TAT----------T--GA---A---------C---A-C--------C----- 2306
HIVNDK ---A-------A-TTG-AC-TTT----------C--GA---A---------C---A-C--------C----- 2285
HIVJY1 ---A-------C-CAG-AC-CTT----------T--CC---A---------C---A-C--------C----- 2336
HIVMAL ---A-------C-CAG-AC-TAT----------C--GA---A---------C---A-T--------C----- 2324
HIVU455 ---G-------C-TAG-GA-TGC----------C--GC---A---------C---A-C--------A----- 2294
HIVZ321 ---G-------C-TAC-AC-TGC----------C--AC---G---------T---T-C--------A----- 2315
SIVCPZ --GAA--G-GC-TGC-AC-TAT----------C--GC-A-A----G-GA-T-GGA-C-----GA-C----C 2288

```
CONSENSUS1 AGACTTAcTcTTGAttGtaacgAgGAtTGTGGAActtCTGGGa.....???cgCAGG  2328
HIVLAI     -------C-C----TT-TAACG-G--T-------CTT-----A.....CG----  2379
HIVHXB2R   -------C-C----TT-TAACG-G--T-------CTT-----A.....CG----  2364
HIVNL43    -------C-C----TT-TAACG-G--T-------CTT-----A.....CG----  2358
HIVMFA     -------C-G----TT-TAACG-G--T-------CTT-----A.....CG----  2358
HIVMN      -------C-C----TT-CAGCG-G--T-------CTT-----A.....CG----  2364
HIVBRVA    -------C-C----TT-TAACG-G--T-------CTT-----A.....CG----  2352
HIVSC      -------C-C----TT-TAACG-G--T-------CTT-----A.....CG----  2364
HIVJH3     -------C-C----TT-TAACG-G--T-------CTT-----A.....CG----  2397
HIVALA1    -------C-C----TT-TAACG-G--T-------CTT-----A.....CG----  2343
HIVBAL1    -------C-C----TT-TGATG-G--T-------CTT-----A.....CTA.G----  2360
HIVJRCSF   -------C-C----CT-TAACG-G--T-------CTT-----A.....CG----  2340
HIVOYI     -------A-C----TT-TAGCG-G--T-------CTT-----A.....CG----  2361
HIVSF2     -------C-C----TT-CAGCG-G--C-------ATT-----G.....CA----  2361
HIVNY5CG   -------C-C----TT-TAGCG-G--T-------CTT-----A.....CG----  2346
HIVSF162   -------A-C----TT-CAGCG-G--T-------CTT-----A.....CG----  2337
HIVJFL     -------C-C----TT-TAACG-G--T-------CTC-----A.....CG----  2349
HIVCDC4    -------C-C----TT-TAGCG-G--T-------CTT-----A.....CG----  2400
HIVSF33    -------C-C----TT-TAGCG-G--T-------CTT-----A.....CG----  2352
HIVHAN     -------C-C----TT-TAGCG-A--T-------ACT-----A.....CG----  2361
HIVADA     -------C-C----TT-TAGCG-G--T-------CTT-----A.....CG----  2355
HIVWMJ2    -------C-C----TT-TGAAG-G--T-------CTT-----A.....CG----  2337
HIVRF      -------C-C----TA-TAGTG-G--T-------CTT-----A.....CG----  2391
CONSENSUS2 AGACTtaatcTTaATTGcAgCGAGgattGTGGAactTCTGgga??????????????????????cGGagG  2264
HIVELI     -----TAATC--A----C-GT---GATT--A--ACT----GGA.....C-CAG-  2355
HIVZ2Z6    -----TAATC--A----C-G----GATT-----GCT----GGA.....C-CAG-  2355
HIVNDK     -----CAATC--A----C-G----GATT-----ACT----GGA.....C-CAG-  2334
HIVJY1     -----TAATC--A----C-A----GATT-----ACT----GGA.....C-CAG-  2385
HIVMAL     -----TACTC--A----C-A----GATT-----ACT----GGA.....C-CAG-  2373
HIVU455    -----TCGCC--G----T-G----AGCT-----ACT----GGACGCAGCAGCCTCAAGGGACTGA-ACT-  2364
HIVZ321    -----GCGCA--G----C-G----GATC-----GAC----.....ATCC-CAG-  2364
SIVCPZ     -AG--TAGCC-GC-A--TGTG---ACAAC-CA-GACGT--CGACATCTAATACTTCACAGCCTCA-ACT-  2358

                                                                  <- rev
CONSENSUS1 G?GGTGGGAAgcCCTcAAaTAtTGgtGGAaTCTCCTqCagTATTGGagtcAGGAaCTAAaGAaTAGTGct  2397
HIVLAI     -.--------GC---C--A--T--GT---A------A-AG------AGTC----A----A--A-----CT  2448
HIVHXB2R   -.--------GC---C--A--T--GT---A------A-AG------AGTC----A----A--A-----CT  2433
HIVNL43    -.--------GC---C--A--T--GT---A------A-AG------AGTC----A----A--A-----CT  2427
HIVMFA     -.--------GC---C--A--T--GT---A------A-AA------AGTC---G----A--A-----CT  2427
HIVMN      -.--------GT---C--A--T--GT---A------A-AG------AGTC----A----A--G-----CT  2433
HIVBRVA    -.--------GT---C--A--T--GT---A------A-AG------AGTC----A----A--A-----CT  2421
HIVSC      -.--------GC---C--A--T--GT---A------A-AG------AGTC----A----G--A-----CT  2433
HIVJH3     -.--------GC---C--A--T--GT---A------A-AA------AGTC----A----A--A-----CT  2466
HIVALA1    -.--------GT---C--A--C--GT---A------G-AG------AGTC----A----A--A-----CT  2412
HIVBAL1    -G--------GT---C--A--T--GT---A------A-AG------AGTC----A----A--A-----CT  2430
HIVJRCSF   -.--------AT---G--A--C--GT---A------A-AG------AGTC----A----A--A-----CT  2409
HIVOYI     -.--------GT---G--A--T--GT---A------G-AG------AGCC----A----A--A-----TC  2430
HIVSF2     -.--------GC---C--A--T--GT---G------G-AG------ATTC----A----A--A-----CT  2430
HIVNY5CG   -.--------GC---C--A--T--TT---A------G-AG------GGTC----A----A--A-----CT  2415
HIVSF162   -.--------GC---C--G--T--GG---A------G-AG------ATTC----A----A--A-----CT  2406
HIVJFL     -.--------GC---C--A--T--GT---A------A-AG------AGTC----A----A--A-----CT  2418
HIVCDC4    -.--------GT---C--A--T--GT---A------G-AG------AGTC----A----A--A-----CT  2469
HIVSF33    -.--------GT---C--A--T--GT---A------G-TG------AGTC----A----A--A-----CT  2421
HIVHAN     -.--------GT---C--G--T--GT---A------G-AG------AGCC----A----A--A-----CT  2430
HIVADA     -.--------GT---C--G--T--GT---A------G-AG------AGTC----A----G--A-----CT  2424
HIVWMJ2    -.--------GC---C--G--T--GT---A------G-AG------AGCA----A----A--A-----CT  2406
HIVRF      -.--------GC---C--G--T--GT---A------G-AG------AGTC----G----A--A-----CT  2460
CONSENSUS2 G.GGTGGGAagcccTCAAATAtCTGtGGAatCTccaTgcagTAtTGG?gTCaGGAACTgAagAatAGTgcT  2332
HIVELI     -.-------CATCC-------T---T-G-AT--CC-ACAG--T---AG--A------G-GG-AC---GC-  2424
HIVZ2Z6    -.-------AGCCC-------T---T-G-AT--CC-ACAG--T---AG--G------G-AG-AG---GC-  2424
HIVNDK     -.-------AGCCC-------C---T-G-AT--CC-GCAG--T---AG--A------G-GG-AT---GC-  2403
HIVJY1     -.-------AGCCA-------T---T-G-GT--CC-GCAG--T---AC--A------G-AG-AT---TT-  2454
HIVMAL     -.-------AGCCC-------T---T-G-AT--CC-GCAA--T---GG--A------G-AG-AT---GC-  2442
HIVU455    -.-------GGGCC-------T---T-G-AT--TC-GCTG--T---GG--G------A-AA-TT---GC-  2433
HIVZ321    -.-------GACTC-------C---G-A-AT--TG-GATC--C---GG--A------A-AG-AT---GC-  2411
SIVCPZ     C.TAC----GAGAC-GTGCCTG---G-GGGAA-TA-TCAA--T---GGAAA---G--A-AA-TT---GC-  2427
```

```
CONSENSUS1 gTtagcTtgcTcaAtgccAcaGCtaTAGCAGTAgcTGAggGgACAGATAGGgtTaTAgAAgTAgtaCAaa  2467
HIVLAI     G-TAGC-TGC-CA-TGCC-CA--CA--------GC---GG-G---------GT-A--G--G--GTA--AG  2518
HIVHXB2R   G-TAGC-TGC-CA-TGCC-CA--CA--------GC---GG-G---------GT-A--G--G--GTA--AG  2503
HIVNL43    G-TAAC-TGC-CA-TGCC-CA--CA--------GC---GG-G---------GT-A--G--G--TTA--AG  2497
HIVMFA     G-TAGC-TGC-CA-TGCC-CA--TA--------GC---GG-G---------GT-A--G--G--GTA--AG  2497
HIVMN      G-TAGC-TGC-TA-TGCC-CA--TA--------GC---GG-G---------GT-A--G--G--CTG--AA  2503
HIVBRVA    G-TAGC-TGC-CA-CGCC-CA--CA--------GG---GA-G---------GC-A--G--G--GTA--AA  2491
HIVSC      G-TAGC-TCG-CA-TGCC-CA--CA--------GC---GG-G---------GT-A--G--T--TTA--AA  2503
HIVJH3     G-TAGC-TGT-CA-TGCC-TA--CA--------GC---GG-G---------GT-C--A--A--TTA--AA  2536
HIVALA1    G-TAGC-TGC-CA-CGCC-CA--CA--------GC---GG-G---------GT-A--G--G--TTA--GA  2482
HIVBAL1    G-TAGC-TGC-CA-TGCC-CA--TG--------GC---AG-G---------GT-A--G--G--TTA--GA  2500
HIVJRCSF   G-TAGC-TGC-TA-TGCC-CA--TA--------GC---GG-G---------AT-A--G--G--GTA--AA  2479
HIVOYI     A-TAGC-TGC-CA-CGCC-CA--TA--------GC---GG-G---------GT-A--G--A--GTA--AA  2500
HIVBF2     G-TAGC-GGC-CA-CGCC-CA--TA--------AC---GG-G---------GT-A--G--G--GCA--AA  2500
HIVNY5CG   A-TAGT-TGC-TA-TGCC-CA--TG--------GC---GG-G---------GT-A--G--G--GCA--AA  2485
HIVBF167   G-TAGT-TGT-TG-TGCC-TA--TA--------GC---GG-G---------AT-A--G--G--GCA--AA  2476
HIVJFL     G-TAGT-TGC-TA-CGCT-CA--TA--------GC---GG-G---------AT-A--G--G--GTA--AA  2488
HIVCDC4    G-TAGC-TGG-TA-TGTC-CA--TA--------GC---GG-A---------GT-A--G--G--GTA--AA  2539
HIVBF33    G-TAGC-TGC-CA-CGCC-CA--TA--------GC---AG-G---------GT-A--G--G--GTA--AA  2491
HIVHAN     G-CAGC-TGT-CA-TACT-TA--TA--------GC---GG-A---------GT-A--G--G--GTA--AA  2500
HIVADA     G-TAGC-TGC-TA-TGCC-CA--TA--------GC---GG-G---------GT-A--G--G--GTA--AA  2494
HIVWMJ2    G-TGGG-TGC-TA-TGCC-TT--TA--------GC---GG-G---------GT-A--G--G--GTA--AA  2476
HIVRF      G-TAGC-TGC-TA-TACC-CA--AA--------GC---AG-G---------AT-A--G--G--GCA--AA  2530
CONSENSUS2 AtTAgcTTGCTTgAtaCcatAGCAaTAGCAGTAGCTGagtGgAcAGATAGGgTTATAGAAaTAgtACaAA  2402
HIVELI     -G--GC---T--G-TG-CAT----A----------AGG-G-C-------AG--------A--AT--A--  2494
HIVZ2Z6    -G--GC------G-TA-CAT----A----------AGG-G-C-------G--------A--GT--G--  2494
HIVNDK     -G--GC------G-TA-CAT----A----------AGA-G-C-------G--------G--GT--A--  2473
HIVJY1     -T--GC------A-TG-CAC----A----------AGG-G-C-------A--------C--AT--G--  2524
HIVMAL     -T--GC------A-TA-CAC----A----------AAT-C-C-------G--------A--GG--A--  2512
HIVU455    -T--CT------G-CG-TGT----G----------GGT-G-T-------G--------A--GG--A--  2503
HIVZ321    -T--AT------G-TA-AGT----A----------ACT-G-C-------G--------G--GT--A--  2503
SIVCPZ     -TA-GCC-----G-TG-AAC---TA-T--------A-AAG-T-C------AA-A------GCTTTT-A-G  2497

CONSENSUS1 gAgctt?TAGaGctaTtctccAcATaCcTagAAGAaTaAGACAgGGCTTgGAAAGGgctTTGCta***  2534
HIVLAI     G-GCTTG---A-CTA-TCGCC-C--A-C-AG----A-A-----G-----G------ATT----TA---  2586
HIVHXB2R   G-GCTTG---A-CTA-TCGCC-C--A-C-AG----A-A-----G-----G------ATT----TA---  2571
HIVNL43    C-GCTTA---A-CTA-TCGCC-C--A-C-AG----A-A-----G-----G------ATT----TA---  2565
HIVMFA     G-GCTTA---A-CTA-TCGCC-C--A-C-AG----A-A-----G-----G------ATT----TA---  2565
HIVMN      G-GCTGG---A-CTA-TCTCC-C--A-C-AC----A-A-----G-----G------GCT----TA---  2571
HIVBRVA    G-GCTTT---A-CTA-CCTGC-C--A-C-AG----A-A-----G-----G------GCT----AA---  2559
HIVSC      G-GCTTT---A-CTA-TCTCC-C--A-C-AC----A-A-----G-----G------GCT----AA---  2571
HIVJH3     G-GCTTT---A-CTA-TCTCC-C--A-C-AC----A-A-----G-----G------GCT----TA---  2604
HIVALA1    G-GCTTT---A-CTA-CCTTC-C--A-C-AG----A-A-----G-----C------GCT----TA---  2550
HIVBAL1    G-GCTGT---A-CTA-TCTCC-C--A-C-AG----A-A-----G-----G------GCT----TA---  2568
HIVJRCSF   G-GTTTA---G-CTA-TCTCC-C--A-C-AC----A-A-----G-----G------GCT----TA---  2547
HIVOYI     G-GCTTA---A-CTT-TCTCA-T--A-C-AG----A-A-----G-----G------GCT----TA---  2568
HIVBF2     G-GCTTA---A-CTA-TCTCC-C--A-A-AG----A-T-----G-----G------CTT----TA---  2568
HIVNY5CG   G-ATTTG---G-GTA-CCTCC-T--C-C-AG----A-A-----A-----A------CTT----TA---  2553
HIVBF167   G-ATTGG---A-CTT-TCTCC-C--A-C-AG----A-A-----G-----T------GCT----TA---  2544
HIVJFL     G-ATTGG---A-CTA-CCTCC-C--A-C-GT----A-A-----G-----A------GCT----TA---  2556
HIVCDC4    G-ATTTA---A-CTT-TCTCC-C--A-C-AG----A-A-----G-----C------GCT----TA---  2607
HIVBF33    G-GTGGG---A-CTA-TCTCC-C--A-C-AC----A-A-----G-----T------GCT----TA---  2559
HIVHAN     G-GCTTG---A-CAA-TCTCC-C--A-C-AG----G-G-----G-----G------GCC----TA---  2560
HIVADA     G-ATTTA---G-CTA-TCTCC-C--A-C-AC----A-A-----G-----G------CTT----TA---  2562
HIVWMJ2    G-ATCTG---A-CTA-TATCC-C--A-C-AG----A-A-----G-----G------GCT----TA---  2544
HIVRF      G-ATTCT---A-CTT-TCTTC-C--A-C-AG----A-A-----G-----A------GCT----TG---  2598
CONSENSUS2 gAgcT?gtAGAGCTaTTCT??ACaTACCcAqAaGAaTaAGACAgGGcTTaGAAAGggcTTTgCTa***  2467
HIVELI     G-GC-TGC------G----TA--A----C-G-A--A-A-----G--C--A-----GTC---A--T---  2562
HIVZ2Z6    G-GC-TGC------G----TC--A---C-C-A--A-A-----G--C--A-----GCT---G--T---  2562
HIVNDK     G-GC-TGC------A----TA--G----C-G-A--A-A-----G--C--G-----GCT---G--A---  2541
HIVJY1     G-GC-TTT------G----TC--A----C-G-C--G-A-----G--C--G-----GGC---A--A---  2592
HIVMAL     G-TT-GGT------A----GC--A----T-G-A--A-T-----G--C--C-----GGC---G--A---  2580
HIVU455    C-AT-GGT------A---- CA--A----T-G-A--A-C-----A--C--A-----GGC---G--A---  2571
HIVZ321    G-GC-GGT------T----CA--A----A-G-A--A-C-----G--A--A-----AGC---G--A---  2571
SIVCPZ     TTACACTT---AT-A--AGAA--A-C--T-GGC-CA-A-----A--GC-G-----AGC---A--T---  2565
```

HIV-1 NEF



```
CONSENSUS1 AAtGCTGaTTGtgCCTGGCTAGAAGCAcAa??????GAGGA?GAgGAgGtgGGtTTTCCAGTCAgACCtC  218
HIVLAI     --T----C---TG--------------C-A......-----G--G--G-TG--T----------C---T-  217
HIVNL43    --T----C---TG--------------C-A......-----G--A--G-TG--T----------C---T-  217
HIVBRVA    --T----A---TG--------------C-A......-----T--G--A-TA--T----------A---T-  253
HIVSC      --T----A---TG--------------C-A......-----G--G--G-TG--T----------G---T-  217
HIVBAL1    --T----A---TG--------------C-A......-----T--A--A-TG--T----------G---T-  217
HIVJRCSF   --T----A---TG--------------T-T......-----T--G--A-TG--T----------G---T-  247
HIVOYI     --T----A---TG--------------C-A......-----T--A--G-TG--T----------G---T-  232
HIVSF2     --T----A---TG--------------C-A......-----G--A--G-TG--T----------G---T-  229
HIVSF162   --T----A---TG--------------C-A......-----C--G--T-TG--C----------G---T-  223
HIVSF33    --T----A---TG--------------C-A......-----T--G--G-TG--T----------A---T-  214
HIVHAN     --C----C---TG--------------C-A......-----A--G--A-AG--T----------G---A-  217
HIVRF      --T----C---CA--------------C-AGAGGAT-----T--G--G-TG--T----------G---T-  223
CONSENSUS2 AATgcta?tTGTgcctggctgGAAgCaCaAga?...GAg???GagGAgGTaGGCTTtCCAGTcaGaCC?C  210
HIVELI     ---GCTGAC---GCCTGGCTA---G-A-A-GAA...--GAGC-AC--G--G-----T-----CA-A--C-  220
HIVNDK     ---GATACC---GCATGGCTG---G-A-A-GAG...--GAGC-AG--G--G-----T-----TA-A--T-  220
HIVMAL     ---GCTAGT---.........---C-A-C-GAA...--AGAG-AG--G--A-----T-----CC-T--T-  226
HIVU455    ---GCCAGT---GCCTGGCTG---G-G-A-......--GGAA-GA--C--A-----C-----TA-G--A-  217
HIVZ321    ---CCTGAT---GCCTGGTTG---G-A-A-GAG...--GTCA-AG--G--A-----T-----CA-A--A-  217
SIVCPZ     ---CAAACTCTAGCTTGGCTA--GGAAATG......--CAAT-AA--A--A--A--C-----AA-A--A-  217

CONSENSUS1 AGGTACCtTTAAGaCCaATGACttaCAagGcAGCT?TAGATcTTAGCCAcTTTTTaAAagAAaAGGGGGG  287
HIVLAI     -------T-----A--A-----TTA--AG-C----G-----C-------C-----A--AG--A-------  287
HIVNL43    -------T-----A--A-----TTA--AG-C----G-----C-------C-----A--AG--A-------  287
HIVBRVA    -------T-----A--A-----CTA--AG-C----G-----C-------C-----A--AG--A-------  323
HIVSC      -------T-----A--A-----TTA--AG-C----G-----C-------T-----A--AG--C-------  287
HIVBAL1    -------T-----A--A-----TCG--GT-C----A-----C-------C-----T--GA--A-------  287
HIVJRCSF   -------T-----A--A-----TTA--AG-C----A-----C-------C-----A--AG--A-------  317
HIVOYI     -------T-----A--A-----TTA--AG-G----T-----C-------C-----A--AG--A-------  302
HIVSF2     -------T-----A--A-----TTA--AG-C----T-----A-------C-----A--AG--A-------  299
HIVSF162   -------T-----A--A-----TTA--AG-C----T-----C-------C-----A--AG--A-------  293
HIVSF33    -------T-----A--A-----TTA--AA-C----T-----C-------C-----A--AG--A-------  284
HIVHAN     -------T-----A--A-----TTA--AG-G----T-----C-------C-----A--AG--A-------  287
HIVRF      -------T-----G--T-----TTT--AG-C----G-----C-------C-----A--AG--A-------  293
CONSENSUS2 AGGTACCTtTaAGACCAATGACttatAAaG?aGCTtTtGATCTcAGCcaCTTTTTAAAAGAAAAGGGGGG  279
HIVELI     --------T-A-----------T-AC--A-AA---C-A-----C---CA--------------------  290
HIVNDK     --------T-A-----------T-AC--A-AA---G-T-----T---CA--------------------  290
HIVMAL     --------T-A-----------T-AT--A-GA---T-T-----C---CA--------------------  296
HIVU455    --------T-A-----------T-AT--G-CA---T-T-----C---TT--------------------  287
HIVZ321    --------C-G-----------C-TT--G-GT---T-T-----T---TT--------------------  287
SIVCPZ     ----T--AACG-----------T-AT--A-CA---T-T-----TTCACA--------------------  287

      3' LTR -> (for more information, see the LTR alignment)
CONSENSUS1 ACTGGAaGGGcTAaTttacTCcCaaaaAAGACAaGAtATcCTTGATcTGTGGaTctAcCACACACAAGGC  357
HIVLAI     ------A---C--A-TCAC--C-AACG------A--T--C------C-----A-CT-C-----------  357
HIVNL43    ------A---C--A-TCAC--C-AAAG------A--T--C------C-----A-CT-C-----------  357
HIVBRVA    ------A---C--A-TCAC--C-AACA------A--T--C------T-----G-CT-T-----------  393
HIVSC      ------A---C--A-.CAC--C-AGAG------A--T--C------C-----A-CT-C-----------  356
HIVBAL1    ------A---C--A-TCAC--C-AAAA------A--T--C------T-----G-CT-C-----------  357
HIVJRCSF   ------A---C--A-TTAC--A-AGAA------A--T--C------C-----A-CT-C-----------  387
HIVOYI     ------A---C--A-TTAC--C-AGAA------A--T--T------C-----G-CT-T-----------  372
HIVSF2     ------A---C--A-TTGG--C-AAAG------A--G--C------C-----A-CT-C-----------  369
HIVSF162   ------A---T--A-TTAC--C-AAAA------A--T--C------C-----A-CC-C-----------  361
HIVSF33    ------A---C--G-TTAC--C-AAAA------A--T--C------C-----A-CT-C-----------  354
HIVHAN     ------A---T--A-TTAC--C-CAAA------A--G--C------C-----G-CT-C-----------  357
HIVRF      ------T---C--G-GTTC--C-AGAA------A--T--C------C-----G-TT-C-----------  361
CONSENSUS2 ACTGGAtGGGcTAaTTtgtTCcaagAAaAGACAAGAaATCCTTGAtCTgTGGGTCTAccAcACACAAGGc  349
HIVELI     -----A---C--A--TGG--CAAA--G-------G-------T--T-------CA-C-------C     360
HIVNDK     -----A---C--A--TGG--CAAG--A-------G-------T--T-------CA-C-------C     360
HIVMAL     -----T---T--G--TGG--CCCA--A-------A-------T--G-------CC-C-------C     366
HIVU455    ---T---T--A--CAC--ACAG--A-------A-------T--G-------TC-C-------A     357
HIVZ321    ---T---C--A--TAC--CAAG--A-------A-------C--G-------TC-T-------C     357
SIVCPZ     ---A---T--G--TAC--CAGG-GA-------G-------C--C-------TC-C-------C     357
```

HIV-1 NEF

```
CONSENSUS1 TAcTTCCCTGAtTGGCAGAAcTACACAcCAGGgCCAGGGatCAGATaT?CCACTGaccTTTGGATGGTGC   426
HIVLAI     --C---------T---------C-------C----G-------GT-----A-.------ACC-----------   426
HIVNL43    --C---------T---------C-------C----G-------GT-----A-.------ACC-----------   426
HIVBRVA    --C---------T---------C-------C----G-------GT-----A-.------ACC-----------   462
HIVSC      --C---------T---------C-------C----G-------AT-----A-C------TGT-----------   426
HIVBAL1    --C---------T---------C-------C----G-------AC-----T-.------ACC-----------   426
HIVJRCSF   --C---------T---------C-------G----A-------GT-----T-.------ACC-----------   456
HIVOYI     --T---------T---------C-------C----G-------AT-----A-.------TGC-----------   441
HIVSF2     --C---------T---------T-------C----G-------AT-----A-.------ACC-----------   438
HIVSF162   --C---------T---------C-------C----G-------AT-----A-.------ACC-----------   432
HIVSF33    --C---------T---------C-------C----G-------GT-----T-.------ACC-----------   423
HIVHAN     --C---------T---------C-------C----G-------GT-----A-.------ACC-----------   426
HIVRF      --C---------C---------C-------C----G-------AC-----A-.------ACC-----------   432
CONSENSUS2 ttCTTCCCtGATTGGCAgAAcTACACACCAGGGCCAGGgAtcAGATac.CCACTgaccTTtGGATGGTGC   418
HIVELI     AT-------T---------A--C-----------------G-TC----AT.-----AACC--T---------   429
HIVNDK     AT-------C---------G--C-----------------G-TC----AC.-----GACC--T---------   429
HIVMAL     TA-------T---------G--T-----------------G-TT----TC.-----GACC--C---------   435
HIVU455    TT-------T---------G--T-----------------A-TC----AC.-----AACA--T---------   426
HIVZ321    TT-------T---------C--C-----------------G-CC----AC.-----GTGT--T---------   426
SIVCPZ     TT-------T--C-----G--C-------A----A-----A-CA----TC.-----CTGT--T---------   426

                                                                          <- NFAT-1
CONSENSUS1 TtCAAGcTAGTACCAgTtgAgCCAGAgaaggtaGAAGAggccaatgaaggaGAGAACaaCaGctTGTtAC   496
HIVLAI     -A----C---------G-TG-G-----TAAGGTA-----GGCCAATAAAGGA------AC-A-CT---T--   496
HIVNL43    -A----C---------G-TG-G-----TAAGGTA-----GGCCAATAAAGGA------AC-A-CT---T--   496
HIVBRVA    -T----C---------G-TG-G-----GAAGATA-----GGCCAATGAAGGA------AA-A-TT---T--   532
HIVSC      -T----C---------G-TA-A-----GAAGATA-----AGCCAATGAAGGA------AA-A-CT---T--   496
HIVBAL1    -T----C---------G-TG-G-----GAAGGTA-----GGCCAATGAAGGA------AA-T-CT---C--   496
HIVJRCSF   -T----C---------G-TG-T-----GAAGGTA-----GGCCAATGAAGGA------AA-T-CT---T--   526
HIVOYI     -T----T---------A-GG-T-----CCAGGTA-----GGCCAACGAAGGA------AA-A-CT---T--   511
HIVSF2     -T----C---------G-TG-G-----GAAGGTA-----GGCCAATGAAGGA------AA-A-CT---T--   508
HIVSF162   -T----C---------G-TG-T-----TTATGTA-----GGCCAATGCAGGA------AA-A-CT---T--   502
HIVSF33    -T----T---------G-AG-G-----GAAAGTA-----GGCCAATGAAGGA------AA-A-CT---T--   493
HIVHAN     -T----T---------G-TG-A-----T.....-----A.............------AG-A-CC---T--   478
HIVRF      -T----C---------G-TG-G-----TAAGGTA-----GGCCACTGAAGGA------AA-A-CT---T--   502
CONSENSUS2 TtcaAGcTAGTACCAgTtgaTCCA?agGAaGTAGAgGAaGccAcTgaAgGAGAGaaCAACtGccTgtTAC   487
HIVELI     -ACG--C---------G-TGA----CAG--G-----A--A-AC-C-GA-G-----AC----A-CT-GT---   499
HIVNDK     -TCC--C---------G-TGA----CAG--G-----A--G-CC-C-GA-A-----GA----T-CT-GT---   499
HIVMAL     -TTA--T---------A-GAG----GAG--A-----G--G-CC-A-GA-G-----AA----T-TC-GT---   505
HIVU455    -ACA--C---------G-TGA----GCT--A-----G--A-CT-C-GG-G-----AA----A-TC-AC---   496
HIVZ321    -TCA--C---------G-TGA----AGA--A-----G--A-CC-A-AC-G-----AA----T-CC-GT---   496
SIVCPZ     -TCA-AC----G--CC-GACAGAGGAGC-G-----AC-A-CC-A-GA-G----TAA----T-CC-GT-G-   496
```

```
                                NFAT-1 ->
CONSENSUS1 AcCCtatgaGCCtGCATGGgAtgGAtGACcCgGAGAaaGAAGTgtTagtgTGGAaGTTTGACAgCcgcCT   566
HIVLAI     -C--TGTGA---T------A-TG--T---C-T----GA-----GT-AGAG----G--------G-CGC--   566
HIVNL43    -C--TGTGA---T------A-TG--T---C-T----GA-----GT-AGAG----G--------G-CGC--   566
HIVBRVA    -C--TATGA---A------A-TG--T---C-G----GA-----GC-GCAG----G--------G-CGC--   602
HIVSC      -C--GATGA---T------G-TG--G---C-G----GA-----GT-AGAG----G--------A-CGC--   566
HIVBAL1    -C--TATGA---T------G-TG--T---C-G----AA-----GT-AGCA----A--------G-AGC--   566
HIVJRCSF   -C--TATGA---A------A-TG--C---C-A----AG-----GT-AGTG----A--------G-AAG--   596
HIVOYI     -T--TATAA---T------G-TG--T---C-A----AA-----GT-AGTG----A--------G-CGC--   581
HIVSF2     -C--TATGA---T------G-TG--G---G-G----AA-----GT-AGTG----G--------G-AAA--   578
HIVSF162   -C--TATGA---A------G-TG--T---C-G----AA-----AT-AGTG----G--------G-CGC--   572
HIVSF33    -C--TATGA---T------G-TG--G---C-G----AA-----GT-AGTG----A--------G-CAC--   563
HIVHAN     -C--TGCGA---T------G-CA--G---A-G----GA-----GT-AAAG----A--------G-CAT--   548
HIVRF      -C--TATAT---T------G-TG--T---C-A----AA-----GT-AGTG----A--------G-CGC--   572
CONSENSUS2 ACCCtAT?tGCCAgCAtGGAaTgGAgGAcgc?GAgAgAgAAGTGtTAAtgTGGAagTTTgACAGcagcCT   555
HIVELI     ----T--AT----G--T---A-G--G--CCCG--G-G-C-----T---AA----GA---A----CAGA--   569
HIVNDK     ----T--GT----G--A---A-G--G--CCCG--G-G-C-----T---TG----GA---A----CAGA--   569
HIVMAL     ----T--TA----A--T---A-G--G--CGCA--A-G-G-----C---AA----AG---G----CAGC--   575
HIVU455    ----T--AT----A--T---G-A--T--TGAA--G-A-G-----T---TG----AG---G----TACC--   566
HIVZ321    ----C--TA----G--T---A-G--T--TGAT--A-G-G-----T---TG----AG---G----CTCC--   566
SIVCPZ     -T--C-  T-T--G--T--GA-G--A--TGAG--C-A-G-G--CT-GGTC---CGC---G----CAGG--   566

              <- IL2/NRE ->
            <-     NRF    ->
CONSENSUS1 aGCATTtCaTCACatgGCCCGAGAgctGCATCCGGAGTaCTaCAAagACTGc???    621
HIVLAI     A-----T-A----GTG--------GCT----------A--T---GA----C***    621
HIVNL43    A-----T-A----GTG--------GCT----------A--T---GA----C***    621
HIVBRVA    A-----T-A----ATG--------GCT----------A--A---GA----C***    657
HIVSC      A-----T-A----ATG--------TCT----------A--A---AG----T***    621
HIVBAL1    A-----C-A----GTG--------A***                               594
HIVJRCSF   A-----G-A----GTG--------GCT----------A--A---GG----C***    651
HIVOYI     A-----T-G----ATG--------GGT----------A--A---AG----C***    636
HIVSF2     A-----T-A----ATG--------GCT----------A--A---AG----C***    633
HIVSF162   G-----T-A----ATG--------GCT----------A--A---AG----C***    627
HIVSF33    A-----T-G----ATG--------GCT----------A--A---AG----C***    618
HIVHAN     A-----T-A----AAG--------GCT----------A--A---AG----C***    603
HIVRF      C-----T-A----GTC--------GAA----------A--A---AG----C***    627
CONSENSUS2 aGCActaaaaCACa?aGCccgaGAgctgCATCCGGAGTtCTAcAAAgAC?G        607
HIVELI     A---TTTGAG---AAG--CCGA--GATG----------T---C---A--***      621
HIVNDK     A---CTTGAG---AAG--CCGA--GCTG----------T---C---G--T-C***   624
HIVMAL     A---CTAAGA---AGA--CAGA--ACAA----------A---C---G--T-C***   630
HIVU455    G---TTAAAA---AGA--TTAT--GCTG----------T---T---G--***      618
HIVZ321    A---CGAAAA---CTA--CCGA--GATG----------T---C---G--***      618
SIVCPZ     T--CTTAAGA--TATT--CAGA--ACAA----------A---C---G--***      618
```

```
LOCUS       HIVLAI       9229 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate LAI, complete
            genome (original LAV-1; formerly HIVBRU).
ACCESSION   K02013 M64186-190 M64194-198 M64203-223
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate LAI proviral
            DNA clone lambda-J19 (Infectious clone available; formerly HIVBRU).
REFERENCE   1  (bases 1 to 9229)
  AUTHORS   Wain-Hobson,S., Sonigo,P., Danos,O., Cole,S. and Alizon,M.
  TITLE     Nucleotide sequence of the AIDS Virus, LAV
  JOURNAL   Cell 40, 9-17 (1985)
REFERENCE   2  (bases 1712 to 1749; correction of [1] in gag-pol cds)
  AUTHORS   Alizon,M., Wain-Hobson,S., Montagnier,L. and Sonigo,P.
  TITLE     Genetic variability of the AIDS virus: Nucleotide sequence analysis
            of two isolates from African patients
  JOURNAL   Cell 46, 63-74 (1986)
REFERENCE   3  (gag, tat-rev, env V1 and env V3 regions of LAI samples)
  AUTHORS   Wain-Hobson,S., Vartanian,J.-P., Henry,M., Chenciner,N., Cheynier,R.,
            Delassus,S., Pedroza Martins,L., Sala,M., Nugeyre,M.-T., Guetard,D.,
            Barre-Sinoussi,F. and Montagnier,L.
  TITLE     LAV revisited: origins of the early viral HIV-1 isolates from
            Institut Pasteur
  JOURNAL   Science 252, 961-964 (1991)
COMMENT     The original LAV, sometimes called LAV-1 to distinguish it from
            HIV2(LAV-2) and formerly designated HIVBRU, is now designated HIVLAI
            to reflect the finding that it originated with the French AIDS
            patient LAI (and not the patient BRU) as explained in [3]. An infec-
            tious clone of this virus has been constructed by Keith Peden, NIH-
            NIAID, Bethesda, 301-496-6730. HIVNL43 is also an infectious clone
            having for its 3' half a clone of the LAI isolate. The tat sequence
            for the L20 clone is available (see 1989: I-A-54 and following pages
            in this release drawing upon PCR-generated sequence sets [3]).
            All LAV, BRU and IIIb viral samples are now believed to have been
            derived from this source, LAI [3]; viruses from the person BRU are
            now designated HIVJB.
BASE COUNT     3289 a   1656 c   2232 g   2052 t
ORIGIN      Cap site of genomic RNA.
        1 GGtctctctg gttagaccag atttgagcct gggagctctc tggctaacta gggaacccac
mRNA start->
       61 tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt
                                              <- R rpt
      121 gtgactctgg taactagaga tccctcagac cctttagtc agtgtggaaa atctctagca
                                                                    <-LTR
      181 gtggcgcccg aacagggact tgaaagcgaa agggaaacca gaggagctct ctcgacgcag
           <- lys-tRNA pbs
      241 gactcggctt gctgaagcgc gcacggcaag aggcgagggg aggcgactgG Tgagtacgcc
                                                  major 5' sj /\
      301 aaaaattttg actagcggag gctagaagga gagagATGgg tgcgagagcg tcagtattaa
                                       gag start ->
      361 gcgggggaga attagatcga tgggaaaaaa ttcggttaag gccagggggA aagaaaaaat
      421 ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg
      481 gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa ccatcccttc
      541 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc
      601 atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa
      661 acaaaagtaa gaaaaaagca cagcaagcag cagctgacac aggacacagc agccaggtca
      721 gccaaaatta ccctatagtg cagaacatcc aggggcaaat ggtacatcag gccatatcac
      781 ctagaacttt aaatgcatgg gtaaaagtag tagaagagaa ggctttcagc ccagaagtga
      841 tacccatgtt ttcagcatta tcagaaggag ccaccccaca agatttaaac accatgctaa
      901 acacagtggg ggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgaggaag
      961 ctgcagaatg ggatagagtg catccagtgc atgcagggcc tattgcacca ggccagatga
     1021 gagaaccaag gggaagtgac atagcaggaa ctactagtac ccttcaggaa caaataggat
     1081 ggatgacaaa taatccacct atcccagtag gagaaatcta taaaagatgg ataatcctgg
     1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac
     1201 caaaagaacc ctttagagac tatgtagacc ggttctataa aactctaaga gccgagcaag
     1261 cttcacagga ggtaaaaaat tggatgacag aaaccttgtt ggtccaaaat gcgaacccag
```

```
1321 attgtaagac tattttaaaa gcattgggac cagcagctac actagaagaa atgatgacag
1381 catgtcaggg agtgggagga cccggccata aggcaagagt tttggctgaa gcaatgagcc
1441 aagtaacaaa ttcagctacc ataatgatgc aaagaggcaa ttttaggaac caaagaaaga
1501 ttgttaagtg tttcaattgt ggcaaagaag ggcacatagc cagaaattgc agggccccta
1561 ggaaaaaggg ctgttggaaa tgtggaaagg aaggacacca aatgaaagat tgtactgaga
1621 gacaggcTAA ttttttaggg aagatctggc cttcctacaa gggaaggcca gggaatttte
       pol orf -> (NH2-terminus uncertain)
1681 ttcagagcag accagagcca acagccccac catttcttca gagcagacca gagccaacag
                                        -> bases added in revision [2]
1741 ccccaccaga agagagcttc aggtctgggg tagagacaac aactccctct cagaagcagg
          ->
1801 agccgataga caaggaactg tatcctttaa cttccctcag atcactcttt ggcaacgacc
1861 cctcgtcaca aTAAagatag gggggcaact aaaggaagct ctattagata caggagcaga
                  <- gag end
1921 tgatacagta ttagaagaaa tgagtttgcc aggaagatgg aaaccaaaaa tgatagggag
1981 aattggaggt tttatcaaag taagacagta tgatcagata ctcatagaaa tctgtggaca
2041 taaagctata ggtacagtat tagtaggacc tacacctgtc aacataattg gaagaaatct
2101 gttgactcag attggttg a ctttaaattt tcccattagt cctattgaaa ctgtaccagt
2161 aaaattaaag ccaggaatgg atggcccaaa agttaaacaa tggccattga cagtagaaaa
2221 aataaaagca ttagtagaaa tttgtacaga aatggaaaag gaagggaaaa tttcaaaaat
2281 tgggcctgaa aatccataca atactccagt atttgccata aagaaaaaag acagtactaa
2341 atggagaaaa ttagtagatt tcagagaact taataagaga actcaagact tctgggaagt
2401 tcaattagga ataccacatc ccgcagggtt aaaaaagaaa aaatcagtaa cagtactgga
2461 tgtgggtgat gcatattttt cagttcccct agatgaagac ttcaggaagt atactgcatt
2521 taccatacct agtataaaca atgagacacc agggattaga tatcagtaca atgtgcttcc
2581 acagggatgg aaaggatcac cagcaatatt ccaaagtagc atgacaaaaa tcttagagcc
2641 ttttagaaaa caaaatccag acatagttat ctatcaatac atggatgatt tgtatgtagg
2701 atctgactta gaaatagggc agcatagaac aaaaatagag gagctgagac aacatctgtt
2761 gaggtgggga cttaccacac cagacaaaaa acatcagaaa gaacctccat tcctttggat
2821 gggttatgaa ctccatcctg atAaaatggac agtacagcct atagtgctgc cagaaaaaga
2881 cagctggact gtcaatgaca tacagaagtt agtgggaaaA ttgaattggg caagtcagat
2941 ttacccaggg attaaagtaa ggcaattatg taaactcctt agaggaacca aagcactaac
3001 agaagtaata ccactaacag aagaagcaga gctagaactg gcagaaaaca gagagattct
3061 aaaagaacca gtacatggag tgtattatga cccatcaaaa gacttaatag cagaaataca
3121 gaagcagggg caaggccaat ggacatatca aatttatcaa gagccattta aaaatctgaa
3181 aacaggaaaa tatgcaagaa cgaggggtgc ccacactaat gatgtaaaac aattaacaga
3241 ggcagtgcaa aaaataacca cagaaagcat agtaatatgg ggaaagactc ctaaatttaa
3301 actacccata caaaaggaaa catgggaaac atggtggaca gagtattggc aagccacctg
3361 gattcctgag tgggagtttg tcaatacccc tcctttagtg aaattatggt accagttaga
3421 gaaagaaccc atagtaggag cagaaacgtt ctatgtagat ggggcagcta gcagggagac
3481 taaattagga aaagcaggat atgttactaa tagaggaaga caaaaagttg tcaccctaac
3541 tgacacaaca aatcagaaga ctgagttaca agcaattcat ctagctttgc agggttcggg
3601 atttgaagta aatatagtaa cagactcaca atatgcatta ggaatcattc aagcacaacc
3661 agataaaagt gaatcagagt tagtcaatca aataatagag cagttaataa aaaaggaaaa
3721 ggtctatctg gcatgggtac cagcacacaa aggaattgga ggaaatgaac aagtagataa
3781 attagtcagt gctggaatca ggaaagtact atttttagat ggaatagata aggcccaaga
3841 tgaacatgag aaatatcaca gtaattggag agcaatggct agtgatttta acctgccacc
3901 tgtagtagca aaagaaatag tagccagctg tgataaatgt cagctaaaac gagaagccat
3961 gcatggacaa gtagactgta gtccaggaat atggcaacta gattgtacac atttagaagg
4021 aaaagttatc ctggtagcag ttcatgtagc cagtggatat atagaagcag aagttattcc
4081 agcagaaaca gggcaggaaa cagcatactt tcttttaaaa ttagcaggaa gatggccagt
4141 aaaaacaata catacagaca atggcagcaa tttcaccagt actacggtta agggccgcctg
4201 ttggtgggcg ggaatcaagc aggaatttgg aattccctac aatccccaaa gtcaaggagt
4261 agtagaatct atgaataaag aattaaagaa aattataggc caggtaagag atcaggctga
4321 acatcttaag acagcagtac aaatggcagt attcatccac aattttaaaa gaaaaggggg
4381 gattggggga tacagtgcag gggaaagaat agtagacata atagcaacag acatacaaac
4441 taaagaatta caaaaacaaa ttacaaaaat tcaaaatttt cgggtttatt acAGggacag
                                                          /\ 3'sj
4501 cagagatcca ctttggaaag gaccagcaaa gctcctctgg aaagGTgaag gggcagtagt
                                        5'sj /\
4561 aatacaagat aatagtgaca taaaagtagt gccaagaaga aaagcaaaga tcattagtga
4621 ttATGgaaaa cagatgtcag gtgattgatg tgtggcaagt agacaggata tagaTAGaa
   vif ->                                                      <-pol
4681 catggaaaag tttagtaaaa caccatatgt atgttttcag gaaagctagg gtatgttttt
```

```
   4741 atagacatca ctatgaaagc cctcatccaa gaataagttc agaagtacac atcccactag
   4801 gggatgctag attggtaata acaacatatt ggggtctgca tacaggagaa agagactggc
   4861 atctgggtca gggagtctcc atagaatgga ggaaaaagag atatagcaca caagtagacc
   4921 ctgaactagc agaccaacta attcatctgt attactttga ctgtttttcA Gactctgcta
                                                              /\3'sj
   4981 taagaaagc cttattagga catatagtta gccctaggtg tgaatatcaa gcaggacata
   5041 acaagGTagg atctctacaa tacttggcac tagcagcatt aataacacca aaaaagataa
      5'sj /\
   5101 agccaccttt gcctagtgtt acgaaactga cagaggatag ATGgaacaag ccccagaaga
                                        vpR start ->
   5161 ccaagggcca cagagggagc cacacaatga atggacacTA Gagcttttag aggagcttaa
                                                  <- vif end
   5221 gaatgaagct gttagacatt ttcctaggat ttggctccat ggcttagggc aacatatcta
   5281 tgaaacttat ggggatactt gggcaggagt ggaagccata ataagaattc tgcaacaact
   5341 gctgtttatc catttcAGaa ttgggtgtcg acatagcaga ataggcgtta ctcaacagag
                        /\ 3'sj
   5401 gagagcaaga aATGgagcca gtagatccTA Gactagagcc ctggaagcat ccaggaagtc
          tat start ->                  <- vpR end
   5461 agcctaaaac tgcttgtacc acttgctatt gtaaaaagtg ttgctttcat tgccaagttt
   5521 gtttcacaac aaaagcctta ggcatctcct ATGgcAGgaa gaagcggaga cagcgacgaa
                                rev start ->    /\ 3'sj
   5581 gacctcctca aggcagtcag actcatcaag tttctctatc aaagcaGTaa gtagtacatg
                                                  5' sj /\
   5641 taATGcaacc tatacaaata gcaatagcag cattagtagt agcaataata atagcaatag
vpU start ->
   5701 ttgtgtggtc catagtaatc atagaatata ggaaaatatt aagacaaaga aaaatagaca
   5761 ggttaattga tagactaata gaaagagcag aagacagtgg caATGagagt gaaggagaaa
                                       env start ->
   5821 tatcagcact tgtggagatg ggggtggaaa tggggcacca tgctccttgg gatattgatg
   5881 atctgTAGtg ctacagaaaa attgtgggtc acagtctatt atggggtacc tgtgtggaag
            <- vpU
   5941 gaagcaacca ccactctatt ttgtgcatca gatgctaaag catatgatac agaggtacat
   6001 aatgtttggg ccacacatgc ctgtgtaccc acagacccca acccacaaga agtagtattg
   6061 gtaaatgtga cagaaaattt taacatgtgg aaaaatgaca tggtagaaca gatgcatgag
   6121 gatataatca gtttatggga tcaaagccta aagccatgtg taaaattaac cccactctgt
   6181 gttagtttaa agtgcactga tttggggaat gctactaata ccaatagtag taataccaat
   6241 agtagtagcg gggaaatgat gatggagaaa ggagagataa aaaactgctc tttcaatatc
   6301 agcacaagca taagaggtaa ggtgcagaaa gaatatgcat ttttttataa acttgatata
   6361 ataccaatag ataatgatac taccagctat acgttgacaa gttgtaacac ctcagtcatt
   6421 acacaggcct gtccaaaggt atcctttgag ccaattccca tacattattg tgccccggct
   6481 ggttttgcga ttctaaaatg taataataag acgttcaatg gaacaggacc atgtacaaat
   6541 gtcagcacag tacaatgtac acatggaatt aggccagtag tatcaactca actgctgttg
   6601 aatggcagtc tagcagaaga agaggtagta attagatctg ccaatttcac agacaatgct
   6661 aaaaccataa tagtacagct gaaccaatct gtagaaatta attgtacaag acccaacaac
   6721 aatacaagaa aaagtatccg tatccagagg ggaccagggg gagcatttgt tacaatagga
   6781 aaaataggaa atatgagaca agcacattgt aacattagta gagcaaaatg gaatgccact
   6841 ttaaaacaga tagctagcaa attaagagaa caatttggaa ataataaaac aataatcttt
   6901 aagcaatcct caggaggggg cccagaaatt gtaacgcaca gttttaattg tggaggggaa
   6961 ttttctact gtaattcaac acaactgttt aatagtactt ggtttaatag tacttggagt
   7021 actgaagggt caaataacac tgaaggaagt gacacaatca cactcccatg tagaataaaa
   7081 caatttataa acatgtggca ggaagtagga aaagcaatgt atgccctcc catcagcgga
   7141 caaattagat gttcatcaaa tattacaggg ctgctattaa caagagatgg tggtaataac
   7201 aacaatgggt ccgagatctt cagacctgga ggaggagata tgagggacaa ttggagaagt
   7261 gaattatata aatataaagt agtaaaaatt gaaccattag gagtagcacc caccaaggca
   7321 aagagaagag tggtgcagag agaaaaaaga gcagtgggaa taggagcttt gttccttggg
   7381 ttcttgggag cagcaggaag cactatgggc gcacggtcaa tgacgctgac ggtacaggcc
   7441 agacaattat tgtctggtat agtgcagcag cagaacaatt tgctgagggc tattgaggcg
   7501 caacagcatc tgttgcaact cacagtctgg ggcatcaagc agctccaggc aagaatcctg
   7561 gctgtggaaa gatacctaaa ggatcaacag ctcctggggg tttggggttg ctctggaaaa
   7621 ctcatttgca ccactgctgt gccttggaat gctagttgga gtaataaatc tctggaacag
   7681 atttggaata cattgaccta gatggagtgg gacagagaaa ttaacaatta cacaagctta
   7741 atacactcct taattgaaga atcgcaaaac cagcaagaaa agaatgaaca agaattattg
   7801 gaattagata aatgggcaag tttgtggaat tggtttaaca taacaaattg gctgtggtat
   7861 ataaaaatat tcataatgat agtaggaggc ttggtaggtt taagaatagt ttttgctgta
```

```
7921 ctttctatag tgaatagagt taggcaggga tattcaccat tatcgtttcA Gacccacctc
                                                           /\ 3'aj
7981 ccaacccccga ggggacccga caggcccgaa ggaaTAGaag aagaaggtgg agagagagac
                                          <- tat end
8041 agagacagat ccattcgatt agtgaacgga tccttagcac ttatctggga cgatctgcgg
8101 agcctgtgcc tcttcagcta ccaccgcttg agagacttac tcttgattgt aacgaggatt
8161 gtggaacttc tgggacgcag ggggtgggaa gccctcaaat attggtggaa tctcctacag
8221 tattggagtc aggaactaaa gaaTAGtgct gttagcttgc tcaatgccac agccatagca
                                  <- rev
8281 gtagctgagg ggacagatag ggttatagaa gtagtacaag gagcttgtag agctattcgc
8341 cacataccta gaagaataag acagggcttg gaaaggattt tgctaTAAgA TGggtggcaa
                                                        <- env
                                                 nef start ->
8401 gtggtcaaaa agtagtgtgg ttggatggcc tactgtaagg gaaagaatga gacgagctga
8461 gccagcagca gatggggtgg gagcagcatc tcgagacctg gaaaaacatg gagcaatcac
8521 aagtagcaat acagcagcta ccaatgctgc ttgtgcctgg ctagaagcac aagaggagga
8581 ggaggtgggt tttccagtca cacctcaggt acctttaaga ccaatgactt acaaggcagc
8641 tgtagatctt agccactttt taaaagaaaa ggggggactg gaagggctaa ttcactccca
                                       3' LTR ->
8701 acgaagacaa gatatccttg atctgtggat ctaccacaca caaggctact tccctgattg
8761 gcagaactac acaccagggc caggggtcag atatccactg acctttggat ggtgctacaa
8821 gctagtacca gttgagccag ataaggtaga agaggccaat aaaggagaga acaccagctt
8881 gttacaccct gtgagcctgc atggaatgga tgacccgag agagaagtgt tagagtggag
8941 gtttgacagc cgcctagcat ttcatcacgt ggcccgagag ctgcatccgg agtacttcaa
9001 gaactgcTGA catcgagctt gctacaaggg actttccgct ggggactttc cagggaggcg
          <- nef
9061 tggcctgggc gggactgggg agtggcgagc cctcagatgc tgcaTATAAG cagctgcttt
                                                signal ->
9121 ttgcctgtac tgGGtctctc tggttagacc agatttgagc ctgggagctc tctggctaac
         R rpt ->
9181 tagggaaccc actgcttaag cctcAATAAA gcttgccttg agtgcttca
                       signal ->                        <- mRNA end
```

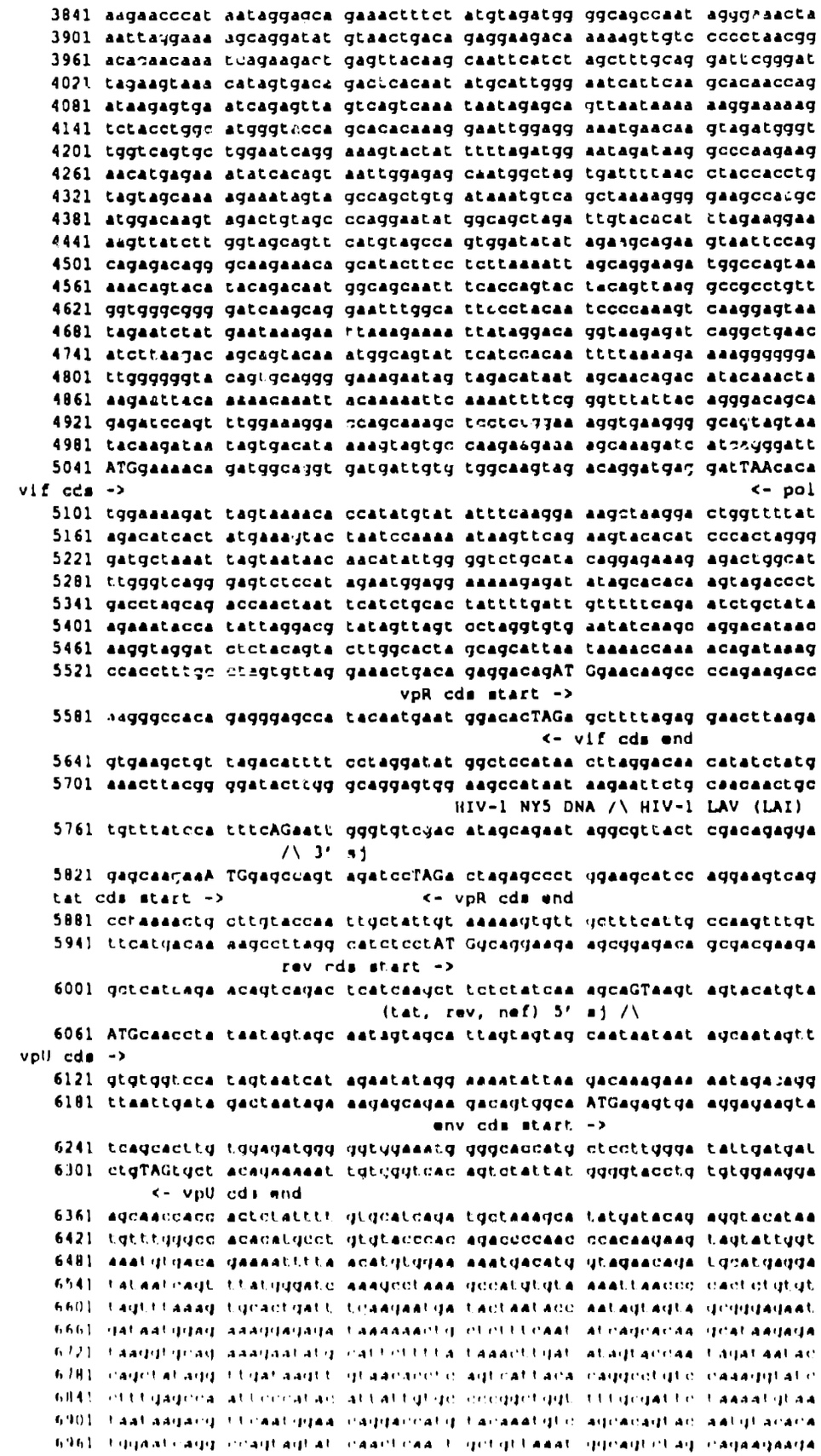
HIVLAI. The following sequences represent the V3 loop coding sequences from LAI, as reported in Wain-Hobson et al., Science 252, 961, 1991 (fig.1).

```
CONSENSUS TGTACAAGACCCAACAAcaatACAaGAAAAaGTATccgTATccagagaGgaCCAgggAGagCATTTgtTa   70
L7        ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
L3        ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
L2        ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
L6        ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AA-----GT-A   70
L9        ----------------CAAT---A-----A----CCG---CCATAGA-GA---GGG--AG-----TA-A   70
L4        ----------------CAAT---A-----A----CCC---A.......-GA---GGG--AG-----TA-A   64
L20       ----------------CAAT---A-----A----CCA---A.......-GA---GGG--AG-----TA-A   64
L16       ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----TA-A   70

E70       ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
E79       ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
E69       ----------------CAAT---A-----A----CCG---CCAGAGA-GG---GGG--AG-----GT-A   70
E86       ----------------CAAT---G-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
E74       ----------------TAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70

M401      ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
M415      ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGA--AG-----GT-A   70
M416      ----------------CAAT---A-----A----CCG---CCAGAGG-GA---GGG--AG-----GT-A   70

K140      ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
K139      ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
K138      ----------------CAAT---A-----A----CCG---CCAAGAA-AA---AAG--AG-----GT-A   70
K145      ----------------CAAT---A-----A----CTG---CCAGAGA-GA---GGG--AG-----GT-A   70
K147      ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
K155      ----------------CAAT---A-----G----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
K144      ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGA--AG-----GT-A   70
K152      ----------------CAAT---A-----A----ACA---A.......-GA---GGG--AG-----TA-G   64

C8        ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
C5        ----------------CAAT---A-----A----CCG---CCGGAGA-GA---GGG--AG-----GT-A   70

G93       ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
G104      ----------------CAAT---G-----A----CCG---CCAGAGG-GA---GGG--AG-----GT-A   70

H117      ----------------  AAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
H116      ----------------CAAT---A-----A----CCG---CCAGAGA-GA---GGG--GG-----GT-A   70
H127      ----------------CAAT---A-----A----CCG---CCAGAGG-GA---GGG--AG-----GT-A   70
H130      ----------------CGAT---A-----A----CCG---CCAGAGA-GA---GGG--AG-----GT-A   70
H135      ----------------T...---A-----A----CCG---CCAGAGG-GA---GGG--AG-----GT-A   67

CONSENSUS CAatAqGAaAAATA???qqAaATaTqaGACAAGCACATTqTl    108
L7        --AC-G--A-----...GG-G--A-GA----------G-1      108
L3        --AT-G--A-----...GG-A--A-GA----------G-1      108
L2        --AC-G--A-----...GG-A--A-GA----------G-1      108
L6        --AT-G--A-----...GG-A--A-GA----------G-1      108
L9        --AT-G--A-----...GG-A--A-GA----------G-1      108
L4        --AC-G--G-----ATAGG-G--A-AA----------G-1      105
L20       --AC-G--G-----ATAGG-A--A-AA----------G-1      105
L16       --AC-G--G-----ATAGG-G--A-AA----------G-1      111

E70       --AT-G--A-----...GG-A--A-GA----------G-1      108
E79       --AT-G--A-----...GG-G--A-GA----------G-1      108
E69       --AT-G--A-----...AG-A--A-GA----------G-1      108
E86       --AT-G--A-----...GG-A--A-GA----------G-1      108
E74       --AT-G--A-----...GG-A--A-GA----------G-1      108
```

```
M401      --AT-G--A-----...GG-A--A-GA----------G-1    108
M415      --AT-G--A-----...GG-G--A-GA----------G-1    108
M416      --GT-G--A-----...GG-A--A-GA----------G-1    108

K140      --AT-G--A-----...GG-A--A-GA----------G-1    108
K139      --AT-G--A-----...GG-A--G-GA----------G-1    108
K138      --AT-A--A-----...GA-A--A-AA----------G-1    108
K145      --AT-G--A-----...GG-A--A-GA----------G-1    108
K147      --AT-G--A-----...GG-A--A-AA----------A-1    108
K155      --AT-G--A-----...GG-A--A-GA----------G-1    108
K144      --AT-G--A-----...GG-G--A-GA----------G-1    108
K152      --AC-G--G-----ATAGG-G--G-AG----------G-1    105

C8        --AT-G--A-----...GG-A--A-GA----------G-1    108
C5        --AT-G--A-----...GG-A--A-GA----------G-1    108

G93       --AT-G--A-----...GG-A--A-GA----------G-1    108
G104      --AT-G--A-----...GG-A--A-GA----------G-1    108

H117      --AT-G--A-----...GG-A--A-GA----------G-1    108
H116      --CT-G--A-----...GG-A--A-GA----------G-1    108
H127      --AT-G--A-----...GA-A--A-GA----------G-1    108
H133      --AT-G--A-----...GG-A--A-GA----------G-1    105
K152      --AC-G--G-----ATAGG-G--G-AG----------G-1    105
C5        --AT-G--A-----...GG-A--A-GA----------G-1    108
C8        --AT-G--A-----...GG-A--A-GA----------G-1    108
H130      --AT-G--A-----...GG-A--A-GA----------G-1    108
```

```
LOCUS       HIVNL43      9709 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, NY5/LAI hybrid (LAV-1);
            recombinant clone pNL4-3.
ACCESSION   M19921
SOURCE      Human immunodeficiency virus type 1 (HIV-1), NY5/LAI (LAV-1)
            recombinant clone pNL4-3. Infectious clone.
REFERENCE   1  (bases 1 to 9709)
  AUTHORS   Buckler,C.E., Buckler-White,A.J., Willey,R.L. and McCoy,J.
  TITLE     ;
  JOURNAL   Unpublished (1988).
REFERENCE   2  (sites)
  AUTHORS   Adachi,A., Gendelman,H.E., Koenig,S., Folks,T., Willey,R.,
            Rabson,A. and Martin,M.A.
  TITLE     Production of acquired immunodeficiency syndrome-associated
            retrovirus in human and nonhuman cells transfected with an
            infectious molecular clone
  JOURNAL   J  Virol. 59, 284-291 (1986)
REFERENCE   3  (sites; revisions of [1])
  AUTHORS   Buckler,C.E.
  TITLE     ;
  JOURNAL   Personal communication (1988)
COMMENT     Clean copy of sequence [1] kindly provided by Chuck Buckler, NIAID
            Bethesda, MD, 24-JUN-1988.  The construction of pNL4-3 has been
            described in [2].  pNL4-3 is a recombinant (infectious) proviral
            clone that contains DNA from HIV isolates NY5 (5' half) and LAI
            (3' half of what was formerly called BRU).  The site of recombination
            is the EcoRI site at positions 5743-5748.
            The length and sequence of the vpR coding region corresponds to
            that of the LAI, SC, SF2, MAL and ELI isolates.  The vpR coding
            region of these isolates is about 18 amino acid residues longer
            than the vpR coding region of the IIIb clones.  In HIVNL43,
            this shift is due to a single base deletion (with respect to the
            IIIb at position 5770.  The sequence at this position is "atttc" in
            HIVNL43 and "attttc" in HIVHXB2.
            The original LAI clone, sequenced by Wain-Hobson, et al. (Cell
            40, 9-17 (1985)), and the LAI portion of the pNL4-3 recombinant
            clone are different clones from the same LAI isolate. LAI was
            previously designated BRU but was re-authenticated by Wain-Hobson
            at al., Science 252, 961, 1991.
            [3] made 5 revisions to the sequence reported by [1]:
               1. "at" rather than "tg" at positions 182-183
               2. "g" rather than "c" at position 191
               3. "g" rather than "a" at position 2421
               4. "tcacac" rather than "ctcaca" at positions 8995-9000
               5. "c" rather than "a" at position 9415
            The revised sequence is reported here.  Two of these revisions
            produced changes in amino acid sequences.  The revision at
            2421 changes one amino acid residue from 'R' to 'G' in the pol
            coding region.  The revision at positions 8995-9000 changes
            three amino acid residues from 'AHT' to 'VTP' in the nef coding
            region.
BASE COUNT      3421 a   1756 c   2366 g   2166 t
ORIGIN      5' terminus of NY5 LTR
        1 tggaagggct aatttggtcc caaaaaagac aagagatcct tgatctgtgg atctaccaca
   5' LTR ->
       61 cacaaggcta cttcccttga tggcaagaac acagaccagg gccaggggat cagatatcca
      121 tgacctttgg atggtgcttc aagttagtac cagttgaacc agagcaagta gaagaggcca
      181 aataaggaga gaagaacagc ttgttacacc ctatgagcca gcatgggatg gagggacccgg
      241 agggagaagt attagtgtgg aagtttgaca gcctcctagc atttgtcacc atgcctgtag
      301 agctgcatcc ggagtgtact aaagaagtag acgtaccaaa tatacagcta acactcaaag
      361 attgcttcga agtctcagga gagatgatgg tgcctgcagg agagttgaga aattgcagtc
                          Sp-1          Sp-1          Sp-1
      421 agctacatat agcagcactg tttgtcctgt actgggtctc tctggttaga ccagatctga
                                  R rpt start ->
```

```
                                    mRNA start ->
 481 gcctgggagc tctctggcta actagggaac ccactgctta agcctcaata aagcttgcct
 541 tgagtgctca aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc
           <- R rpt end
 601 agacccttttt agtcagtgtg gaaaatctct agcagtggcg cccgaacagg gacttgaaag
                                          <- 5' LTR
                                             ->  Lys-tRNA pbs  <-
 661 cgaaagtaaa gccagaggag atctctcgac gcaggactcg gcttgctgaa gcgcgcacgg
 721 caagaggcga ggggcggcga ctgGTgagta cgccaaaaat tttgactagc ggaggctaga
    major 5' splice junction /\
 781 aggagagagA TGggtgcgag agcgtcggta ttaagcgggg gagaattaga taaatgggaa
gag cds start ->
 841 aaaattcggt taaggccagg gggaaagaaa caatataaac taaaacatat agtatgggca
 901 agcagggagc tagaacgatt cgcagttaat cctggccttt tagagacatc agaaggctgt
 961 agacaaatac tgggacagct acaaccatcc cttcagacag gatcagaaga acttagatca
1021 ttatataata caatagcagt cctctattgt gtgcatcaaa ggatagatgt aaaagacacc
1081 aaggaagcct tagataagat agaggaagag caaaacaaaa gtaagaaaaa ggcacagcaa
1141 gcagcagctg acacaggaaa caacagccag gtcagccaaa attaccctat agtgcagaac
1201 ctccaggggc aaatggtaca tcaggccata tcacctagaa ctttaaatgc atgggtaaaa
1261 gtagtagaag agaaggcttt cagcccagaa gtaataccca tgttttcagc attatcagaa
1321 ggagccaccc cacaagattt aaataccatg ctaaacacag tgggggggca tcaagcagcc
1381 atgcaaatgt taaaagagac catcaatgag gaagctgcag aatgggatag attgcatcca
1441 gtgcatgcag ggcctattgc accaggccag atgagagaac caaggggaag tgacatagca
1501 ggaactacta gtacccttca ggaacaaata ggatggatga cacataatcc acctatccca
1561 gtaggagaaa tctctaaaag atggataatc ctgggattaa ataaaatagt aagaatgtat
1621 agccctacca gcattctgga cataagacaa ggaccaaagg aacccttcag agactatgta
1681 gaccgattct ataaaactct aagagccgag caagcttcac aagaggtaaa aaattggatg
1741 acagaaacct tgttggtcca aaatgcgaac ccagattgta agactatttt aaaagcattg
1801 ggaccaggag cgcactaga agaaatgatg acagcatgtc agggagtggg ggggacccggc
1861 cataaagcaa gagttttggc tgaagcaatg agccaagtaa caaatccagc taccataatg
1921 atacagaaag gcaatttttag gaaccaaaga aagactgtta agtgtttcaa ttgtggcaaa
1981 gaagggcaca tagccaaaaa ttgcagggcc cctaggaaaa agggctgttg gaaatgtgga
2041 aaggaaggac accaaatgaa agattgtact gagagacagg cTAAtttttt agggaagatc
              pol cds start (NH2-terminus uncertain) ->
2101 tggccttccc acaagggaag gccagggaat tttcttcaga gcagaccaga gccaacagcc
2161 ccaccagaag agagcttcag gtttgggggaa gagacaacaa ctccctctca gaagcaggag
2221 ccgatagaca aggaactgta tcctttagct tccctcagat cactctttgg cagcgacccc
2281 tcgtcacaaT AAagataggg gggcaattaa aggaagctct attagataca ggagcagatg
                 <- gag cds end
2341 atacagtatt agaagaaatg aatttgccag gaagatggaa accaaaaatg atagggggaa
2401 ttggaggttt tatcaaagta ggacagtatg atcagatact catagaaatc tgcggacata
2461 aagctatagg tacagtatta gtaggaccta cacctgtcaa cataattgga agaaatctgt
2521 tgactcagat tggctgcact ttaaattttc ccattagtcc tattgaagct gtaccagtaa
2581 aattaaagcc aggaatggat ggcccaaaag ttaaacaatg gccattgaca gaagaaaaaa
2641 taaaagcatt agtagaaatt ttgtacagaa ttggaaaagg aggaaaaatt tcaaaaattg
2701 ggcctgaaaa tccatacaat actccagtat ttgccataaa gaaaaaagac agtactaaat
2761 ggagaaaatt agtagatttc agagaacttt ataaggatac tcaagatttc tgggaagttc
2821 aattaggaat accacatcct gcaggattaa aacagaaaaa atcagtaaca gtactggatg
2881 tgggcgatgc atattttca gttcccttag ataaagactt caggaagtat actgcattta
2941 ccatacctag tataaacaat gagacaccag ggattagata tcagtacaat gtgcttccac
3001 agggatggaa aggatcacca gcaatattcc agtgtagcat gacaaaaatc ttagagcctt
3061 ttagaaaaca aaatccagac atagtcatct atcaatacat ggatgatttg tatgtaggat
3121 ctgacttaga aatagggcag catagaacaa aaatagagga actgagacaa catctgttga
3181 ggtggggatt taccacacca gacaaaaaac atcagaaaga acctccattc ctttggatgg
3241 gttatgaact ccatcctgat aaatggacag tacagcctat agtgctgcca gaaaaggaca
3301 gctggactgt caatgacata cagaaattag tgggaaaatt gaattgggca agtcagattt
3361 atgcagggat taaagtaagg caattatgta aacttcttag gggaaccaaa gcactaacag
3421 aagtagtacc actaacagaa gaagcagagc tagaactggc agaaaacagg gagattctaa
3481 aagaaccggt acatgtgtat tatgatccat cagacctaga aatagggcag aaaatacaga
3541 aggaggggca aggccaatgg acatatcaaa ttgctccaga aaaagaatgg gaattcgcta
3601 acggtaaaac acctttaaca ggaacttcag ctgtgaaaca tgtagacaga ttaataggaa
3661 aaaatcctgg tatcaaacag ggctatatag ggccaaaaca tgaacaagtt ggaataccac
3721 ataccaaaac aaaaacagga aaatgggtat accaagtagt agatctgtca gacaaatgtg
3781 tcctgaagtt gggaaaacac ctgaacggta tagatttcac agatctaaac aaaggagaca
```


vif cds ->
<- pol
vpR cds start ->
<- vif cds end
HIV-1 NY5 DNA /\ HIV-1 LAV (LAI)
/\ 3' sj
tat cds start ->
<- vpR cds end
rev cds start ->
(tat, rev, nef) 5' sj /\
vpU cds ->
env cds start ->
<- vpU cds end

```
 7021 tgtagtaatt agatctgcca atttcacaga caatgctaaa accataatag tacagctgaa
 7081 cacatctgta gaaattaatt gtacaagacc caacaacaat acaagaaaaa gtatccgtat
 7141 ccagagggga ccagggagag catttgttac aatagggaaaa atagggaaata tgagacaagc
 7201 acattgtaac attagtagag caaaatggaa tgccacttta aaacagatag ctagcaaatt
 7261 aagagaacaa tttggaaata ataaaacaat aatctttaag caatcctcag gaggggaccc
 7321 agaaattgta acgcacagtt ttaattgtgg aggggaattt ttctactgta attcaacaca
 7381 actgtttaat agtacttggt ttaatagtac ttggagtact gaagggtcaa ataacactga
 7441 aggaagtgac acaatcacac tcccatgcag aataaaacaa tttataaaca tgtggcagga
 7501 agtaggaaaa gcaatgtatg cccctcccat cagtggacaa attagatgtt catcaaatat
 7561 tactgggctg ctattaacaa gagatggtgg taataacaac aatgggtccg agatcttcag
 7621 acctggagga ggcgatatga gggacaattg gagaagtgaa ttatataaat ataaagtagt
 7681 aaaaattgaa ccattaggag tagcacccac caaggcaaag agaagagtgg tgcagagaga
 7741 aaaaagagca gtgggaatag gagctttgtt ccttgggttc ctgggagcag caggaagcac
 7801 tatgggctgc acgtcaatga cgctgacggt acaggccaga caattattgt ctgatatagt
 7861 gcagcagcag aacaatttgc tgagggctat tgaggcgcaa cagcatctgt tgcaactcac
 7921 agtctggggc atcaaacagc tccaggcaag aatcctggct gtggaaagat acctaaagga
 7981 tcaacagctc ctggggattt ggggttgctc tggaaaactc atttgcacca ctgctgtgcc
 8041 ttggaatgct agttggagta ataaatctct ggaacagatt tggaataaca tgacctggat
 8101 ggagtgggac agagaaatta acaattacac aagcttaata cactccttaa ttgaagaatc
 8161 gcaaaaccag caagaaaaga atgaacaaga attattggaa ttagataaat gggcaagttt
 8221 gtggaattgg tttaacataa caaattggct gtggtatata aaattattca taatgatagt
 8281 aggaggcttg gtaggtttaa gaatagtttt tgctgtactt tctatagtga atagagttag
 8341 gcagggatat tcaccattat cgtttcAGac ccacctccca atcccgaggg gacccgacag
                                    /\ 3' sj (tat, rev, nef)
 8401 gcccgaagga aTAGaagaag aaggtggaga gagagacaga gacagatcca ttcgattagt
                   <- tat cds end
 8461 gaacggatcc ttagcactta tctgggacga tctgcggagc ctgtgcctct tcagctacca
 8521 cgcttgaga gacttactct tgattgtaac gaggattgtg gaactttctgg gacgcagggg
 8581 gtgggaagcc ctcaaatatt ggtggaatct cctacagtat tggagtcagg aactaaagaa
 8641 TAGtgctgtt aacttgctca atgccacagc catagcagta gctgagggga cagataggat
       <- rev cds end
 8701 tatagaagta ttacaagcag cttatagagc tattcgccac atacctagaa gaataagaca
 8761 gggcttggaa aggattttgc taTAAgATGg gtggcaagtg gtcaaaaagt agtgtgattg
                                    <- env cds end
                      nef cds start ->
 8821 gatggcctgc tgtaagggaa agaatgagac gagctgagcc agcagcagat ggggtgggag
 8881 cagtatctcg agacctagaa aaacatggag caatcacaag tagcaataca gcagctaaca
 8941 atgctgcttg tgcctggcta gaagcacaag aggaggaaga ggtgggtttt ccagtcacac
 9001 ctcaggtacc tttaagacca atyacttaca aggcagctgt agatcttagc cacttttaa
 9061 aagaaaaggg gggactggaa gggctaattc actcccaaag aagacaagat atccttgatc
              3' LTR ->
 9121 tgtggatcta ccacacacaa ggctacttcc ctgattggca gaactacaca ccagggccag
 9181 gggtcagata tccactgacc tttggatggt gctacaagct agtaccagtt gagccagata
 9241 aggtagaaga ggccaataaa ggagagaaca ccagcttgtt acaccctgtg agcctgcatg
 9301 gaatggatga ccctgagaga gaagtgttag agtggaggtt tgacagccgc ctagcatttc
 9361 atcacgtggc ccgagagctg catccggagt acttctagaa ctgcTGAcat cgagcttgct
                                                    <- nef cds end
 9421 acaagggact ttccgctggg gactttccag gggaggcgtg gcctgggcgg actggggagt
 9481 ggcgagccct cagatgctgc aTATAAGcag ctgctttttg cctgtactGG gtctctctgg
                      signal ->                R rpt start ->
 9541 ttagaccaga tctgagcctg ggagctctct ggctaactag ggaacccact gcttaagcct
 9601 cAATAAAgct tgccttgagt gcttcaagta gtgtgtgccc gtctgttgtg tgactctggt
signal ->                        <- mRNA end
                                 <- R rpt end
 9661 aactagagat ccctcagacc ctttttagtca gtgtggaaaa tctctagca
                                                   <- 3' LTR
```

```
LOCUS       HIVJB01       659 bp ss-RNA
DEFINITION  Human immunodeficiency type 1, isolate BRU/LAV, clone JB6B,
            partial env cds.
ACCESSION   X57447
SEGMENT     1 of 3
SOURCE      Human immunodeficiency virus type 1 (HIV-1), BRU/LAV, clone JB6B.
REFERENCE   1  (bases 1 to  659)
  AUTHORS   Guo H,-G., Chermann J,-C., Waters,D., Hall,L., Louie,A., Gallo, R.C.,
            Streicher,H., Reitz,M.S., Popovic, M., and Blattner, W.
  TITLE     Sequence analysis of original HIV-1
  JOURNAL   Nature 349, 745-746 (1991)
REFERENCE   2 (bases 1 to 659; corrections to Fig.2 in [1])
  AUTHORS   Reitz, M., LTCB-NCI-NIH, Bethesda, 301-496-4850
COMMENT     This sequence was kindly provided in electronic form by M. Reitz,
            NIH-NCI-LTCB; it corresponds to the sequence in Fig. 2 of [1], with
            exception of the following corrections to Fig.2 [2]:

                 coordinate          revision of Fig.2 in [1]
                 ----------          ------------------------
                    349               g in figure should be c

            The reference sequence for LAV-1 (formerly BRU, now LAI) has several
            typographical errors in the figure.

            This viral isolate truly represents the patient BRU. Previous entries
            designated HIVBRU are now changed to HIVLAI to reflect their origin
            from the patient LAI as explained in Wain-Hobson et al., Science 252,
            961, 1991. The latter reference also contains JBB sequences over the
            V1 region.
BASE COUNT      237 a    122 c    132 g    168 t
ORIGIN      Approximately 120nt downstream from the env start codon.
        1 gggtacctgt gtggaaagaa gcaaccacca ctctattttg tgcatcagat gctaaagcat
 env ->
       61 atgatacaga ggtacataat gtttgggcca cacatgcctg tgtacccaca gaccccaacc
      121 cacaagaaat agtattgaaa aatgtgacag aaaattttaa catgtggaaa aataacatgg
      181 tagaacagat gcatgaggat ataatcagtt tatgggatca aagcctaaaa ccatgtgtaa
      241 aattaacccc actctgtgtt actttaaatt gcactgattt gacgaatgct actaatacca
      301 ctagtagtag cggggaacg atggagggag gagaaataaa aaactgctct ttcaatatca
      361 ccacaagcat aagagataag gtgcagaaag aatatgcact gttttataaa cttgatatag
      421 taccaataga taataatact agctataggt tgataagttg taacacctca gtcatcacac
      481 aggcctgtcc aaaggtatcc tttgagccaa ttcccataca ttattgtgcc ccggctggtt
      541 ttgcgattct aaagtgtaaa gataagaagt tcaatggaaa aggaccatgt acaaatgtca
      601 gcacagtaca atgtacacat ggaattaggc cagtagtatc aactcaactg ctgttaaat
                                                                    env ->
```

HIVJB01. Guo et al., Nature 349, 745, 1991. The original LAV sequence, now designated LAI, is shown for comparison. These sequences start at about 120nt downstream from the env start codon. See also <HIVJB02> and <HIVJB03>.

```
CONSENSUS GGGTACCTGTGTGGAAaGAAGCAACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGA   70
LAI       ----------------G-------------------------------------------------------   70
HIVJB4B   ----------------A-------------------------------------------------------   70
HIVJB6B   ----------------A-------------------------------------------------------   70
HIVJB66   ----------------A-------------------------------------------------------   70

CONSENSUS GGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAaTAGTATTGaaA  140
LAI       ------------------------------------------------------------G--------GT-  140
HIVJB4B   ------------------------------------------------------------A--------AA-  140
HIVJB6B   ------------------------------------------------------------A--------AA-  140
HIVJB66   ------------------------------------------------------------A--------AA-  140

CONSENSUS AATGTGACAGAAAATTTTAACATGTGGAAAAATaACATGGTAGAACAGATGCATGAGGATATAATCAGTT  210
LAI       ---------------------------------G-------------------------------------  210
HIVJB4B   ---------------------------------A-------------------------------------  210
HIVJB6B   ---------------------------------A-------------------------------------  210
HIVJB66   ---------------------------------A-------------------------------------  210

CONSENSUS TATGGGATCAAAGCCTAAA?CCATGtGTAAAATTAACCCCACTCTGTGTTAcTTTAAAtTGCACTGATTT  279
LAI       -------------------G-----T-------------------------G------G-----------  280
HIVJB4B   -------------------G-----C-------------------------C------T-----------  280
HIVJB6B   -------------------A-----T-------------------------C------T-----------  280
HIVJB66   -------------------A-----T-------------------------C------T-----------  280

CONSENSUS GacGAATGCTACTAATACCAcTAGTAGT????????????????AGCGGGGgAAcGATG???GAGggAGgA  331
LAI       -GG-----------------A-------AATACCAATAGTAGT-------A--T----ATG---AA--G-  350
HIVJB4B   -AC-----------------C-------.....------G--C----...---GG--A-  332
HIVJB6B   -AC-----------------C-------.....------G--C----...---GG--G-  332
HIVJB66   -AC-----------------C-------.....------G--C----...---GG--G-  332

CONSENSUS GAaATAAAAAACTGCTCTTTCAAtATCAcCACAAGCATAAGAGaTAAGGTGCAGAAAGAATATGCAcTqT  401
LAI       --G--------------------T----G--------------G----------------------T-T-  420
HIVJB4B   --A--------------------C----C--------------A----------------------C-G-  402
HIVJB6B   --A--------------------T----C--------------A----------------------C-G-  402
HIVJB66   --A--------------------T----C--------------A----------------------C-G-  402

CONSENSUS TTTATAAACTTGATATAqTACCAATAGAT???AAT??????ACtAGCTATAqGTTGAtAAGTTGTAACAC  462
LAI       -----------------A-----------...---GATACT--C----  --C-----C------------  481
HIVJB4B   -----------------G-----------...---...AAT--T-------G-----T------------  466
HIVJB6B   -----------------G-----------AAT---.....--T-------G-----T------------  466
HIVJB66   -----------------G-----------AAT---.....--T-------G-----T------------  466

CONSENSUS CTCAGTCATcACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCT  512
LAI       ---------T-------------------------------------------------------------  557
HIVJB4B   ---------C-------------------------------------------------------------  536
HIVJB6B   ---------C-------------------------------------------------------------  536
HIVJB66   ---------C-------------------------------------------------------------  536

CONSENSUS GGTTTTGCGATTCTAAAqTGTAAaqATAAGAaGTTCAATGGAAaAGGACCATGTACAAATGTCAGCACAG  602
LAI       -----------------A-----TA------C-----------C-------------------------  627
HIVJB4B   -----------------G-----AG------A-----------A-------------------------  606
HIVJB6B   -----------------G-----AG------A-----------A-------------------------  606
HIVJB66   -----------------G-----AG------A-----------A-------------------------  606
```

HIVJB01

```
CONSENSUS TACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTaAAT    655
LAI       ----------------------------------------------G---    680
HIVJB4B   ----------------------------                       636
HIVJB6B   ----------------------------------------------A---    659
HIVJB66   ----------------------------------------------A---    659
```

```
LOCUS       HIVJB02       875 bp ss-RNA
DEFINITION  Human immunodeficiency type 1, isolate BRU/LAV, clone J7,
            partial env cds.
ACCESSION   X57450
SEGMENT     2 of 3
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate BRU/LAV,
            clone J7.
REFERENCE   1  (bases 1 to 875)
  AUTHORS   Guo H,-G., Chermann J,-C., Waters,D., Hall,L., Louie,A., Gallo,R.C.,
            Streicher, H., Reitz,M.S., Popovic,M. and Blattner,W.
  TITLE     Sequence analysis of original HIV-1
  JOURNAL   Nature 349, 745-746 (1991)
REFERENCE   2  (bases 1 to 875; correction of Figure 1 sequence in [1])
  AUTHORS   Reitz,M., LTCB/NCI/NIH, Bethesda, 301-496-4850.
COMMENT     This sequence kindly provided in electronic form by M. Reitz.  It
            was originally designated combJ7 and it corresponds to the sequence
            in Figure 1 of [1] with exception of the following corrections:

                 coordinate      revision of Figure 1 in [1]
                 ----------      ---------------------------
                    194          s in figure should be a
                    278          a in figure should be c
                    493          a in figure should be g
                    610          ggggacaat in figure should be ggqacaatt
                    749          g in figure should be c

              The reference sequence for LAV-1 (formerly BRU, now LAI) has
              several typographical errors in the figure.

              This viral isolate truly represents the patient BRU. Previous
              entries designated HIVBRU are now changed to HIVLAI to reflect
              their origin from the patient LAI as explained in Wain-Hobson et al.,
              Science 252, 961, 1991. The latter reference also contains JBB
              sequences over the V3 region.
BASE COUNT      347 a    139 c    196 g    193 t
ORIGIN      Approximately 20nt downstream from <HIVJB01> covering env V3.
        1 tagtaattag atctgacaat ttctcggaca atgctaaaac cataatagta cagctgaacg
  env ->
       61 aaactgtaga aattaattgt acaagaccca acaacaatac aagaaaaagt atacatatag
      121 gaccagggag agcattttat gcaacaggag aaataatagg agatataaga caagcacatt
      181 gtaacattag tagagcaaaa tggaataaca ctttaaaaca gatagttaga aaattaaaag
      241 aacaatataa gaataaaaca atagtctttа agcaatcctc aggaggggac ccagaaattg
      301 taatgcacag ttttacttgt ggaggggaat ttttctictg taattcaaca caactgttta
      361 atagtacttg gaatgatact gaagtgtcaa atgacactaa ggaaaatgac acaatcatac
      421 tcccatgcag aataaaacaa attataaaca tgtggcagga agtaggaaaa gcaatgtatg
      481 cccctcccat cagaggacaa attagatgtt catcaaatat tacagggctg ctattaacaa
      541 gagatggtgg taataacaag aacaacacca acaacaccga gatcttcaga cctggaggag
      601 gaaaatatgag ggacaattgg agaagtgaat tatataaata taaagtagta aaaattgaac
      661 cattaggagt agcacccacc agggcaaaga gaagagtggt gcagagagaa aaaagagcag
      721 tgggaatagg agctgtgttc cttgggttct tgggggcagc aggaagcact atgggcgcac
      781 ggtcaatgac gctgacggta caggccagac tattattgtc tggtatagtg caacagcaga
      841 acaatttgct gagggctatt gaggcgcaac agcat
                                          env ->
```

HIVJB02. These sequences start approximately 20 nt downstream from <HIVJB01> in the env cds; they cover the V3 region. LAI, the original LAV, is shown for comparison.

```
CONSENSUS TAGTAATTAGATCTGacAATTTCtCggACAATGCTAAAACCATAATAGTACAGCTGAAcgAAaCTGtAGA   70
LAI       ---------------CC------A-AG-------------------------------CC--T---T---   70
HIVJB61   ---------------AC------T-GG-------------------------------CG--A---T---   70
HIVJB7    ---------------AC------T-GG-------------------------------CG--A---T---   70
HIVJBH41A ---------------AC------T-GG-------------------------------CG--A---T---   70
HIVJBH4A  ---------------AC------T-GG-------------------------------CG--A---T---   70
HIVJB51   ---------------AC------T-GG-------------------------------CG--A---T---   70
HIVJBD4LL                                                                          0
HIVJB31   ---------------AC------T-AG-------------------------------CG--A---T---   70
HIVJB21                                                                            0
HIVJB11                                                                            0
HIVJBH14A ---------------AC------T-GG-------------------------------CG--A---T---   70
HIVJBD221 ---------------AA------T-GA-------------------------------AG--C---N---   70
HIVJB71                                                                            0
HIVJBC306                                                                          0

CONSENSUS ATTAATTGTACAAGACCCAACAACAATACAAGAAaAaGTATacaTATa??????GGACCAggGAGAGCA   134
LAI       ------------------------------------A-A----CCG---C-AGAGG------GG-------   140
HIVJB61   ------------------------------------A-A----ACA---A......------GG-------   134
HIVJB7    ------------------------------------A-A----ACA---A......------GG-------   134
HIVJBH41A ------------------------------------A-A----ACA---A......------GG-------   134
HIVJBH4A  ------------------------------------A-A----ACA---A......------GG-------   134
HIVJB51   ------------------------------------A-A----ACA---A......------GG-------   134
HIVJBD4LL                                                                          0
HIVJB31   ------------------------------------A-A----ACA---A......------AA-------   134
HIVJB21                                                                            0
HIVJB11                                                                            0
HIVJBH14A ------------------------------------A-A----ACA---A......------GG-------   134
HIVJBD221 ------------------------------------G-C----AAG---A......------GG-------   134
HIVJB71                                                                            0
HIVJBC306                                                                          0

CONSENSUS TTTtaTgCAacAGgAgAAAATAatagGAgATATaAGACaAGCACATTGTAACATTAGTaGAGCAAAATqGA   204
LAI       ---GT-A--AT--G-A------...G--A----G----A-------------------A---------G--   207
HIVJB61   ---TA-G--AC--G-G-----ATAG--G----A----A-------------------A---------G--   204
HIVJB7    ---TA-G--AC--G-G-----ATAG--G----A----A-------------------A---------G--   204
HIVJBH41A ---TA-G--AC--G-G-----ATAG--G----A----A-------------------A---------G--   204
HIVJBH4A  ---TA-G--AC--G-G-----ATAG--G----A----A-------------------A---------G--   204
HIVJB51   ---TA-G--AC--G-G-----ATAG--G----A----A-------------------A---------G--   204
HIVJBD4LL                                                                          0
HIVJB31   ---TA-G--AC--A-G-----ATAA--G----A----A-------------------A---------A--   204
HIVJB21                                                                            0
HIVJB11                                                                            0
HIVJBH14A ---TA-G--AC--G-G-----ATAG--G----A----A-------------------A---------G--   204
HIVJBD221 ---CG-G--GC--G-C-----ATAG--G----A----G-------------------G---------G--   204
HIVJB71                                                                            0
HIVJBC306                                                                          0
```

```
CONSENSUS ATaaCACTTTAAAACAGATAGtTaGaAAATTAAaAGAACAATaTaagaatAAaAcAata??agTCTTTAA   272
LAI       --GC----------------C-A-C-------G--------T-GGAAAT--T-A-ACAATAA-------   277
HIVJB61   --AA----------------T-A-A-------G--------A-GAG...--T-A-ACAATAG-------   271
HIVJB7    --AA----------------T-A-A-------A--------A-AAG...--T-A-ACAATAG-------   271
HIVJBH41A --AA----------------T-A-A-------A--------A-AAG...--T-A-ACAATAG-------   271
HIVJBH4A  --AA----------------T-A-A-------A--------A-AAGAAT--A-C-ATA...G-------   271
HIVJB51   --AA----------------T-A-A-------A--------A-AAGAAT--A-C-ATA...G-------   271
HIVJBD4LL                                                                        0
HIVJB31   --AA----------------T-A-A-------A--------A-AAGAAT--A-C-ATA...G-------   271
HIVJB21                                                                          0
HIVJB11                                                                          0
HIVJBH14A --AA----------------T-A-A-------A--------A-AAGAAT--A-C-AT...AG-------   271
HIVJBD221 --AA----------------T-G-A-------G--------A-AGGAAT--A-C-...ATAG-------   271
HIVJB71                                                                          0
HIVJBCJ06  -AA----------------T-A-A-------A--------A-AAGAAT--A-C-ATA...G-------    66

CONSENSUS gCAATCCTCAGGAGgGGACCcAGAAATTGTAAtGCACAGTTTTAcTTGTgGAgggGAATTTTTCTACTGT   342
LAI       G-------------G-----C----------C----------A----G--GGG---------------   347
HIVJB61   G-------------G-----C----------T----------T----G--GGG---------------   341
HIVJB7    G-------------G-----C----------T----------C----G--GGG---------------   341
HIVJBH41A G-------------G-----C----------T----------C----G--GGG---------------   341
HIVJBH4A  G-------------G-----C----------T----------C----G--GGG---------------   341
HIVJB51   G-------------G-----C----------T----------C----G--GGG---------------   341
HIVJBD4LL                ----------T----------C----G--GGG---------------          48
HIVJB31   G-------------A-----C----------T----------C----A--AAA---------------   341
HIVJB21                                                                          0
HIVJB11                                                                          0
HIVJBH14A G-------------G-----C----------T----------C----G--GGG---------------   341
HIVJBD221 T-------------G-----T----------T----------C----G--GGG---------------   341
HIVJB71                                                                          0
HIVJBCJ06 G-------------G-----C----------T----------C----G--GGG---------------   136

CONSENSUS AATtCAAcACAACTGTTTAATAGTACTTgG???.???AaTgaTACT????????????GAAGtG?CAAAT   392
LAI       ---T---C-------------------G-....TTT-A-AG----TGGAGT...ACT----G-T-----   410
HIVJB61   ---T---C-------------------G-.......-A-GA----...........----T-C-----   392
HIVJB7    ---T---C-------------------G-.......-A-GA----...........----T-T-----   392
HIVJBH41A ---T---C-------------------G-.......-A-GA----...........----T-C-----   392
HIVJBH4A  ---T---C-------------------G-.......-A-GA----...........----T-C-----   392
HIVJB51   ---T---C-------------------G-.......-G-GA----...........----T-T-----   392
HIVJBD4LL ---A---T-------------------G-CTG.GTT-A-AG----TGGAATGATACT----G-T-----   117
HIVJB31   ---T---C-------------------A-.......-A-GA----...........----T-C-----   392
HIVJB21                                                                          0
HIVJB11                                                                          0
HIVJBH14A ---T---C-------------------G-.......-A-GA----...........----T-C-----   392
HIVJBD221 ---A---T-------------------G-CTG.GTT-A-AG----TGGAATGATACT----G-T-----   410
HIVJB71                                                          -T-T-----       9
HIVJBCJ06 ---T---C-------------------G-.......-A-GA----...........----T-C-----   187
```

HIVJB02

```
CONSENSUS aACActaatgaaaatgACACAATCAtACTCCCATGCAGAATAAAACAAaTTaTAAACATGTGGCAGGAAG  462
LAI       A---CTGAAGGAAGTG---------C----------------------T--A-----------------  480
HIVJB61   A---CTAATGAAAATG---------T----------------------A--A-----------------  462
HIVJB7    G---CTAAGGAAAATG---------T----------------------A--A-----------------  462
HIVJBH41A A---CTAATGAAAATG---------T----------------------A--A-----------------  462
HIVJBH4A  A---CTAATGAAAATG---------T----------------------A--A-----------------  462
HIVJB51   G---CTAAGGAAAATG---------T----------------------A--A-----------------  462
HIVJBD4LL G---.............---------C----------------------A--G-----------------  175
HIVJB31   A---CTAATGAAAATG---------T----------------------A--A-----------------  462
HIVJB21                                                                            0
HIVJB11                                                                            0
HIVJBH14A A---CTAATGAAAATG---------T----------------------A--A-----------------  462
HIVJBD221 G---.............---------C----------------------A--G-----------------  468
HIVJB71   G---CTAAGGAAAATG---------T----------------------A--A-----------------   79
HIVJBC306 A---CTAATGAAAATG---------T----------------------A--A-----------------  257

CONSENSUS TAgGAAAAGCAATGTATGCCCCTCCCATcagaGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCT  532
LAI       --G-------------------------CAGC-------------------------------------  550
HIVJB61   --G-------------------------CAGA-------------------------------------  532
HIVJB7    --G-------------------------CAGA-------------------------------------  532
HIVJBH41A --G-------------------------CAGA-------------------------------------  532
HIVJBH4A  --G-------------------------CAGA-------------------------------------  532
HIVJB51   --G-------------------------CAGA-------------------------------------  532
HIVJBD4LL --G-------------------------TGAA-------------------------------------  245
HIVJB31   --A-------------------------CAGA-------------------------------------  532
HIVJB21                                                                            0
HIVJB11                                                                            0
HIVJBH14A --G-------------------------CAGA-------------------------------------  532
HIVJBD221 --G-------------------------TGAA-------------------------------------  538
HIVJB71   --G-------------------------CAGA-------------------------------------  149
HIVJBC306 --G-------------------------CAGA-------------------------------------  327

CONSENSUS ATTAACAAgAGATgG.gGT???AAtAacaAqAAcaacaccaaca???acaCCGAGATcTTCAGACCTgGA  596
LAI       --------G----G--G--...--T-ACA-C--T.........G...GGT-------C---------G--  605
HIVJB61   --------G----G--G--...--T-ACA-G--CAACCCCAACA...ACA-------C---------G--  596
HIVJB7    --------G----G--G--...--T-ACA-G--CAACACCAACA...ACA-------C---------G--  596
HIVJBH41A --------G----G--G--...--T-ACA-G--CAACACCAACA...ACA-------C---------G--  596
HIVJBH4A  --------G----G--G--...--T-ACA-G--CAACACCAACA...ACA-------C---------G--  596
HIVJB51   --------G----G--G--...--T-GCA-G--CAACACCAACA...ACA-------C---------G--  596
HIVJBD4LL --------G----G--G--...--T-ACC-G--CAACACCAATAATGATA-------T---------G--  312
HIVJB31   --------A----A--A--...--T-ACA-G--CAACACCAACA...ACA-------C---------A--  596
HIVJB21                                                                            0
HIVJB11                                                                            0
HIVJBH14A --------G----G--G--...--T-ACA-G--CAACACCAACA...ACA-------C---------G--  596
HIVJBD221 --------G----G--G--...--T-ACC-G--CAACACCAATAATGATA-------T---------G--  605
HIVJB71   --------G----G--G--...--T-ACA-G--CAACACCAACA...ACA-------C---------G--  213
HIVJBC306 --------G----G--G--AAT--C-AGA-C--CCCCAACAACA.....-------C---------G--  391
```

```
CONSENSUS gGAtGAaATATGAggGACAATTgGAGAAGTGAATTAtATAAATATAAAGTAGTaAAAATTGAACCATTAG    666
LAI       G--  --G------GG-------G-----------------------------A----------------    675
HIVJB61   G--G--A------GG-------G-----------------------------A----------------    666
HIVJB7    G--G--A------GG-------G-----------------------------A----------------    666
HIVJBH41A G--G--A------GG-------G-----------------------------T----------------    666
HIVJBH4A  G--G--A------GG-------G-----------------------------T----------------    666
HIVJB51   G--G--A------GG-------G-----------------------------A----------------    666
HIVJBD4LL G--G--A------AG-------G-----------------------------A----------------    382
HIVJB31   A--A--A------AG-------A-----------------------------A----------------    666
HIVJB21                                                                             0
HIVJB11                                                                             0
HIVJBH14A G--G--A------GG-------G-----------------------------A----------------    666
HIVJBD221 G--G--A------AG-------G--     630
HIVJB71   G--G--A------GG-------G-----------------------------A----------------    283
HIVJBC306 G--G--A------AA-------G-----------------------------A----------------    461

CONSENSUS GACTAGCAcCcACCAgGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTgT    736
LAI       --------C-C----A-------------------------------------------------T-    745
HIVJB61   --------C-C----G-------------------------------------------------G-    736
HIVJB7    --------C-C----G-------------------------------------------------G-    736
HIVJBH41A --------C-.----G-------------------------------------------------G-    735
HIVJBH4A  --------.-C----G-------------------------------------------------G-    735
HIVJB51   --------C-C----G-------------------------------------------------G-    736
HIVJBD4LL --------C-C----A-------------------------------------------------G-    452
HIVJB31   --------C-C----G-------------------------------------------------G-    736
HIVJB21                                                                             0
HIVJB11                                                                             0
HIVJBH14A --------C-C----G-------------------------------------------------G-    736
HIVJBD221
HIVJB71   --------C-C----G-------------------------------------------------G-    353
HIVJBC306 --------C-C----G-------------------------------------------------G-    531

CONSENSUS GTTCCTTGGGTTCTTGGGgGCAGCAGGAAGCACTATGCGCgcacggtCaATgACgCTgaCGGTACAGgCC    806
LAI       ------------------A-----------------GCACGGT-A--G--G--GA--------G--    815
HIVJB61   ------------------G-----------------GCACGGT-A--G--G--GA--------G--    806
HIVJB7    ------------------G-----------------GCACGGT-A--G--G--GA--------G--    806
HIVJBH41A ------------------G-----------------GCACG.T-A--G--G--GA--------G--    804
HIVJBH4A  ------------------G-----------------GCAGGCT-A--G--G--GA--------G--    805
HIVJB51   ------------------G-----------------GCACGGT-G--G--G--GA--------G--    806
HIVJBD4LL ------------------A-----------------GCAGCCT-A--G--G--GA--------G--    522
HIVJB31   ------------------G-----------------GCTG. . -A--G--G--GA--------G--    803
HIVJB21                                       --GCAGCGT-A--G--.--GA--------G--     31
HIVJB11                                         TTTTCGT-A--C--.--..--------T--     27
HIVJBH14A ------------------G-----------------GCAGGCT-A--G--G--GA--------G--    806
HIVJBD221
HIVJB71   ------------------G-----------------GCACGGT-A--G--G--GA--------G--    423
HIVJBC306 ------------------G-----------------GCACGGT-G--G--G--GA--------G--    601
```

HIVJB02

```
CONSENSUS AGACtATTATTGTCTGGTATAGTgCAaCAGCAgAACAATTTgCTGAGGGCTATTGAGGCGCAACAgCAT   875
LAI       ----A----------------G--G-----G--------G----------------------G---   884
HIVJB61   ----T----------------G--A-----G--------G----------------------G--    874
HIVJB7    ----T----------------G--A-----G--------G----------------------G---   875
HIVJBH41A ----T----------------A--A-----A--------A----------------------A--    872
HIVJBH4A  ----T----------------A--A-----A--------A----------------------A---   874
HIVJB51   ----T----------------G--A-----A--------G----------------------G---   875
HIVJBD4LL ----T----------------G--A-----G--------G----------------------G---   591
HIVJB31   ----T----------------G--A-----G--------G----------------------G---   872
HIVJB21   ----T----------------G--A-----G--------G----------------------G---   100
HIVJB11   ----T----------------G--A-----G--------G----------------------G---    96
HIVJBH14A ----T----------------G--A-----G--------G----------------------G---   875
HIVJBD221
HIVJB71   ----T----------------G--A-----G--------G----------------------G---   492
HIVJBC306 ----T----------------G--A-----G--------G----------------------G---   670
```

```
LOCUS       HIVJB03      1021 bp ss-RNA
DEFINITION  Human immunodeficiency type 1, isolate BRU/LAV, clone C306nef4,
            partial env and complete nef cds.
ACCESSION   X57465
SEGMENT     3 of 3
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate BRU/LAV, clone
            C306nef4.
REFERENCE   1  (bases 1 to 1021)
  AUTHORS   Guo H,-G., Chermann J,-C., Waters, D., Hall, L., Louie, A., Gallo, R.C.,
            Streicher, H., Reitz, M.S., Popovic, M. and Blattner, W.
  TITLE     Sequence analysis of original HIV-1
  JOURNAL   Nature 349, 745-746 (1991)
REFERENCE   2  (bases 1 to 1021; correction of Figure 3 in [1])
COMMENT     This sequence kindly provided in electronic form by M. Reitz, LTCB-
            NCI-NIH, Bethesda; it corresponds to Fig.3 in [1] with exception of
            the following correction by [2]:

                 coordinate        revision of Fig.3 in [1]
                ------------      --------------------------
                    393             c in figure should be g

            This viral isolate truly represents the patient BRU. Previous entries
            designated HIVBRU are now changed to HIVLAI to reflect their origin
            from the patient LAI as explained in Wain-Hobson et al., Science 252,
            961, 1991.
BASE COUNT      289 a    209 c    301 g    222 t
ORIGIN      Approximately 210nt upstream (in env) of the nef start codon.
        1 cggggtgggg agccctcaaa tattggtgga atcttctaca gtattggatt caggaactaa
   env ->
       61 agaatagtgc tgttagcttg ctcaatgcca cagccatagc agtagctgag gggacagata
      121 gggtcataga agtattacaa agagcttata gagctcttct ccacatacct agaagaatca
      181 gacagggcgc ggaaaggggt ttgataTAAg ATGggtggca aatggtcaaa atgtagtgtg
                                    <- env end
                              nef start ->
      241 ggtgggtggc ctactgtaag ggaaagaatg agacgagctg aaccagcagc ggatggggtg
      301 ggggcagtat ctcgagacct ggaaagacat ggagcaatca caagtagcaa tacagcagct
      361 aacaatgctg attgtgcctg gctagaagca caggaggagg agaaggtggg ttttccagtc
      421 agacctcagg taccttaaag accaatgact tataagggag ctttagatct tagccacttt
      481 ttaagaaaaa aggggggact ggaagggcta atttactccc aaaaaagaca agatatcctt
      541 gatctgtggg tctaccacac acaaggctac ttccctgatt gggataacta cacaccaggg
      601 ccagggatca gatatccact gacctttgga tggtgcttca agctagtacc agttgaacca
      661 gagaaggtag aagaggccaa tgaaggagag aacaacagct tgttacaccc tatgagccag
      721 catgggatag aggacccgga gagagaagtg ttagagtgga ggtttgacag ccgcctagca
      781 tttcgtcacg tggcccgaga tctgcatccg gaatactaca aggactgcTG Acatcgagct
                                                              <- nef
      841 ttctacaagg gactttccgc tgggactttc cagggaggc gtggcctggg cgggactggg
      901 gagtggcgag ccctcagatg ctgcaTATAA Gcagctgctt ttgcctgtac tGGGtctctc
                              signal ->                    R rpt ->
      961 tggttagacc agatctgagc ctgggagctc tctggctaac tagggaaccc actgcttaag
     1021 c
```

HIVJB03. The following sequence start in the env cds approximately 210nt upstream from the nef start codon. The LAI sequence is shown for comparison.

```
CONSENSUS GGGGGTGGGAAGCCCTCAAATATTGGTGGAATCTcCTACAGTATTGGAtTCAGGAACTAAAGAATAGTGC   70
LAI       ----------------------------------C------------G--------------------   70
HIVJBNC31 ----------------------------------C------------T--------------------   70
HIVJBNP11 ----------------------------------C------------T--------------------   70
HIVJBNP41 ----------------------------------C------------T--------------------   70
HIVJBC4   ----------------------------------T------------T--------------------   70
HIVJBC5   ----------------------------------C------------T--------------------   70

CONSENSUS TGTTAGCTTGCTCAA??CCACAGCCaTAGCAGTAGCTGAGgGGACAGATAGGGTcATAGAAGTAtTACAA  138
LAI       ---------------TG--------A-------------G-------------T---------G-----  140
HIVJBNC31 ---------------CA--------G-------------G-------------C---------T-----  140
HIVJBNP11 ---------------CA--------A-------------G-------------C---------T-----  140
HIVJBNP41 ---------------CA--------A-------------G-------------C---------T-----  140
HIVJBC4   ---------------TG--------A-------------G-------------C---------T-----  140
HIVJBC5   ---------------TG--------A-------------A-------------C---------T-----  140

CONSENSUS aGAGCTTgTAGAGCTaTTCtCCACATACCTAGAAGAATcAGACAGGGCgcGGAAAGGgcTTTGaTATAAG  208
LAI       G------G-------A---G-----------------A---------TT-------AT----C------  210
HIVJBNC31 A------G-------A---T-----------------C---------GC-------GC----A------  210
HIVJBNP11 A------G-------A---T-----------------C---------GC-------GC----A------  210
HIVJBNP41 A------G-------A---T-----------------C---------GC-------GC----A------  210
HIVJBC4   A------A-------C---T-----------------C---------GC-------GC----A------  210
HIVJBC5   A------A-------C---T-----------------C---------GC-------GC----A------  210

CONSENSUS ATGGGTGGCAAaTGGTCAAAAcGT?GT?TGGgTGGaTGGCCTACTGTAAGGGAAAGAATGAGACGAGCTG  276
LAI       ----------G---------A--A--G---T---A-----------------------------------  280
HIVJBNC31 ----------A---------C--T--T---G---A-----------------------------------  280
HIVJBNP11 ----------A---------C--T--T---G---A-----------------------------------  280
HIVJBNP41 ----------A---------C--T--T---G---A-----------------------------------  280
HIVJBC4   ----------A---------T--A--G---G---G-----------------------------------  280
HIVJBC5   ----------A---------C--A--G---G---G-----------------------------------  280

CONSENSUS AaCCAGCAGCgGATGGGGTGgGaGCAGtAtCtcgagacctggaaagacatggagcaatcacaagtagcAA  346
LAI       -G--------A---------G-A----C-T-TCGAGACCTGGAAAAACATGGAGCAATCACAAGTAGC--  350
HIVJBNC31 -A--------G---------G-A----T-T-TCGAGACCTGGAAAGACATGGAGCAATCACAAGTAGC--  350
HIVJBNP11 -A--------G---------G-A----T-G-.....................................--  313
HIVJBNP41 -A--------G---------G-A----T-T-TCGAGACCTGGAAAGACATGGAGCAATCACAAGTAGC--  350
HIVJBC4   -A----    ----G---------G-G----T-T-TCGAGACCTGGAAAGACATGGAGCAATCACAAGTAGC--  350
HIVJBC5   -A--------G---------A-A----T-T-TCGAGACCTGGAAAGACATGGAGCAATCACAAGTAGC--  350

CONSENSUS TACAGCAGCTAaCAATGCTGaTTGTGCCTGGCTAGAAGCACAqGAGGAGGAGgAGGTGGG1TTTCCAGTC  416
LAI       ----------C---------C--------------------A---------G-----------------  420
HIVJBNC31 ----------A---------A--------------------G---------G-----------------  420
HIVJBNP11 ----------A---------A--------------------G---------G-----------------  383
HIVJBNP41 ----------A---------A--------------------G---------G-----------------  420
HIVJBC4   ----------A---------A--------------------G---------A-----------------  420
HIVJBC5   ----------A---------A--------------------G---------A-----------------  420

CONSENSUS AqACTTCAGGTACCTTTAAGACCAATqACTTAtAAGGqAGCTtTAGATCTTaqCCACTTTTTAAqAqAAA  486
LAI       -C----------------------G-----C----C----G---------AG----------A-G---  490
HIVJBNC31 -G----------------------C-----T----G----T---------AG----------G-G---  490
HIVJBNP11 -G----------------------G-----T----G----T---------GA----------G-G---  454
HIVJBNP41 -G----------------------G-----T----G----T---------GA----------G-G---  490
HIVJBC4   -G----------------------G-----T----G----T---------AG----------G-A---  490
HIVJBC5   -G----------------------G-----T----G----T---------AG----------G-G---  490
```

```
CONSENSUS AGGGGGGACTGGAAGGGCTAATTtACTCCCA?aaAAG?CAAGATATCCTTGAT?TGTGGgTCTACCACAC   553
LAI       ----------------------C-------ACG---A---------------C-----A----------   560
HIVJBNC31 ----------------------T-------GAA---G---------------T-----G----------   560
HIVJBNP11 ----------------------T-------GAA---G---------------T-----G----------   523
HIVJBNP41 ----------------------T-------GAA---G---------------T-----G----------   560
HIVJBC4   ----------------------T-------AAA---A---------------C-----G----------   560
HIVJBC5   ----------------------T-------AAA---A---------------C-----G----------   560

CONSENSUS ACA?GGCTACTTCCCTGATTGGgAtAACTACACACCAGGGCCAGGGaTCAGATATCCACTGACCTTTGGa   622
LAI       ---A------------------C-C-------------------G---------------------A   630
HIVJBNC31 ---G------------------G-T-------------------A---------------------A   630
HIVJBNP11 ---G------------------G-T-------------------A---------------------A   593
HIVJBNP41 ---G------------------G-T-------------------A---------------------A   630
HIVJBC4   ---A------------------G-T-------------------A---------------------A   630
HIVJBC5   ---A------------------G-T-------------------A---------------------G   630

CONSENSUS TGGTGCTtCAAGCTAGT?CCAGT?GAaCCAGAgAAGGTAGAAGAGGCCAATgAAGGAGAGAACAaC?GCT   689
LAI       -------A---------A-----T--G-----T-----------------A-----------C-A---   700
HIVJBNC31 -------T---------G-----G--A-----G-----------------G-----------A-T---   700
HIVJBNP11 -------C---------G-----G--A-----G-----------------G-----------A-T---   663
HIVJBNP41 -------T---------G-----G--A-----G-----------------G-----------A-T---   700
HIVJBC4   -------T---------A-----T--A-----G-----------------G-----------A-A---   700
HIVJBC5   -------T-------  -A-----T--A-----G-----------------G-----------A-A---   700

CONSENSUS TGTTACACCCTaTGAGCCaGCATGGgATgGAtGACCCgGAGAGAGAAGTGTTAG?GTGGAGGTTTGACAG   758
LAI       -----------G------T------A--G--T-----T---------------A---------------   770
HIVJBNC31 -----------A------A------G--G--T-----G---------------T---------------   770
HIVJBNP11 -----------A------A------G--G--T-----G---------------T---------------   733
HIVJBNP41 -----------A------A------G--G--T-----G---------------T---------------   770
HIVJBC4   -----------A------A------G--A--G-----G---------------A---------------   770
HIVJBC5   -----------A------A------G--A--G-----G---------------A---------------   770

CONSENSUS CCGCCTAGCATTTCgTCACGTGGCCCGAGAg???CATCCGGAgTACTaCAAGgACTGCTGACATCGAGCT   825
LAI       -------------A---------------GCTG--------G----T----A---------------   840
HIVJBNC31 -------------G---------------GAAA--------G----A----G---------------   840
HIVJBNP11 -------------G---------------GAAA--------G----A----G---------------   803
HIVJBNP41 -------------G---------------GAAA--------G----A----G---------------   840
HIVJBC4   -------------G---------------TCTG--------A----A----G---------------   840
HIVJBC5   -------------G---------------TCTG--------A----A----G---------------   840

CONSENSUS TtCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGCGTGG?CTGGGCGGGACTGGGGAGTGGCGAg   894
LAI       -G------------------------------------------C-----------------------G   910
HIVJBNC31 -T------------------------------------------A-----------------------G   910
HIVJBNP11 -T------------------------------------------A-----------------------G   873
HIVJBNP41 -T------------------------------------------A-----------------------G   910
HIVJBC4   -T------------------------------------------C-----------------------G   910
HIVJBC5   -T------------------------------------------C-----------------------A   910

CONSENSUS CCCTCAGATGCTGCATATAAGCAGCTGCTTTTtGCCTGTACTGGGTCTCTCTGGTTAGACCAGATtTGAG   964
LAI       ------------------------------T-------------------------------T----   980
HIVJBNC31 ------------------------------T-------------------------------T----   980
HIVJBNP11 ------------------------------T-------------------------------T----   943
HIVJBNP41 ------------------------------T-------------------------------T----   980
HIVJBC4   ------------------------------.-------------------------------C----   979
HIVJBC5   ------------------------------.-------------------------------C----   979

CONSENSUS CCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGC   1006
LAI       ---------- ----------------------------     1022
HIVJBNC31 ---------- ----------------------------     1021
HIVJBNP11 ---- ---------------------------- ----      985
HIVJBNP41 ---------- ----------------------------     1022
HIVJBC4   ---------- ----------------------------     1021
HIVJBC5   - ---------- ----------------------------   1021
```

LOCUS HIVBAL 3808 bp ss-RNA
DEFINITION Human immunodeficiency virus type 1, isolate BAL, clone 1; tat, rev, vpu, env and nef coding sequences.
ACCESSION
SOURCE Human immunodeficiency virus type 1, BAL isolate, lambda clones 1 and 2 (partial).
REFERENCE 1 (bases 1 to 3808)
AUTHORS Reitz,M., Popovic,M., Gartner,S., Gallo,R., Read-Connole,E. and Beaver,B.
TITLE ;
JOURNAL Unpublished (1990)
REFERENCE 2 (sites)
AUTHORS Gartner,S., Markovits,P., Markowitz,D.M., Kaplan,M.H., Gallo,R.C. and Popovic,M.
TITLE The Role of Mononuclear Phagocytes in HTLV-III/LAV Infection
JOURNAL Science 233, 215-219 (1986)
COMMENT Described in [2] to be a macrophage-tropic virus isolated from lung alveolar tissue from an upstate New York patient. This sequence kindly provided prior to publication oy M. Reitz, Laboratory of Tumor Cell Biology, NCI, NIH, Bethesda MD.

A sequence representing a second clone, BAL2, is annotated below. The BAL virus is not available as a full-length molecular clone at this time.

coordinate	variation
167	start of clone 2 sequence
172	ctg in 1; cg in 2
181	ct in 1; tc in 2
192	c in 1; t in 2
201	g in 1; a in 2
212	a in 1; g in 2
219	ca in 1; c in 2
234	ttca in 1; ctc in 2
423	a in 1; g in 2
509	t in 1; a in 2
817	cg in 1; gc in 2
898	g in 1; a in 2
904	a in 1; g in 2
942	c in 1; a in 2
954	a in 1; g in 2
1014	c in 1; t in 2
1043-1093	attt...tagca in 1; taagcg in 2
1111	g in 1; a in 2
1189	a in 1; g in 2
1202	c in 1; t in 2
1218	tagtaataaga in 1; aaatgatagc in 2
1378	a in 1; t in 2
1410	a in 1; c in 2
1484	ccaatttcgcgg in 1; aaaatttcacga in 2
1507	gt in 1; ac in 2
1582	c in 1; a in 2
1778	gt in 1; cg in 2
1927	g in 1; a in 2
2241	gg in 1; at in 2
2361	c in 1; t in 2
2424	g in 1; a in 2
2451	c in 1; t in 2
2529	a in 1; g in 2
2938	at in 1; ca in 2
2964	g in 1; a in 2
2978	tg in 1; ca in 2
3092	tagcag in 1; gca in 2
3147	g in 1; a in 2

```
                3250            cgt in 1; tgg in 2
                3261            ctgtaaggaa in 1; aggtaagag in 2
                3281            a in 1; c in 2
                3304            a in 1; g in 2
                3315            tatcta in 1; catctc in 2
                3375            c in 1; t in 2
                3400            agaggat in 1; gtaggag in 2
                3412            end of clone 2 sequence
BASE COUNT     1308 a    662 c    937 g    901 t
ORIGIN      About 20nt downstream from the vpR start.
        1 gatcaagggc cacagaggga gccacacaat gaatggacac TAGagctttt agaggagctt
   vpR ->                                              <- vif
       61 aagagtgaag ctgttagaca ctttcctagg atatggcttc atggcttagg gcaacatatc
      121 tatgaaactt atggggatac ttgggcagga gtggaagcca taataagaat tctgcaacaa
      181 ctgctgttta tccatttcAG gattgggtgc caacatagca agaataggta ttattcaaca
                              /\ 3' sj
      241 gaggagagca agaaATGgag ccagtagatc cTAAactaga gccctggaag catccaggaa
              tat start ->                  <- vpR
      301 gtcagcctaa gactgcttgt accacttgct attgtaaaaa gtgttgcttt cattgccaag
      361 tttgcttcat aacaaaaggc ttaggcatct cctATGgcag gaagaagcgg agacagcgac
                                   rev start ->
      421 gaagagctcc tcaagacagt gagactcatc aagtttctct atcaaagcaG Taagtagtac
                                                        5' sj /\
      481 atgtaATGca agctttacaa atatcagcta tagtaggatt agtagtagca gcaataatag
     vpU start ->
      541 caatagttgt gtggaccata gtattcatag aatataggaa aatattaagg caaagaaaaa
      601 tagacaggtt aattgataga ataacagaaa gagcagaaga cagtggcaAT Gagagtgacg
                                                  env start ->
      661 gagatcagga agagttatca gcactggtgg agatggggca tcatgctcct tgggatatta
      721 atgatctgTA Atgctgaaga aaaattgtgg gtcacagtct attatggggt acctgtgtgg
                  <- vpU
      781 aaagaagcaa ccaccactct attttgtgca tcagatcgta aagcatatga tacagaggta
      841 cataatgttt gggccacaca tgcctgtgta cccacagacc ccaacccaca agaagtagaa
      901 ttgaaaaatg tgacagaaaa ttttaacatg tggaaaaata acatggtaga acaaatgcat
      961 gaggatataa tcagtttatg ggatcaaagc ctaaagccat gtgtaaaatt aaccccactc
     1021 tgtgttactt taaattgcac tgatttgagg aatgctacta atgggaatga cactaatacc
     1081 actagtagta gcaggggaat ggtgggggga ggagaaatga aaaattgctc tttcaatatc
     1141 accacaaaca taagaggtaa ggtgcagaaa gaatatgcac ttttttataa acttgatata
     1201 gcaccaatag ataataatag taataataga tataggttga taagttgtaa cacctcagtc
     1261 attacacagg cctgtccaaa ggtatccttt gagccaattc ccatacatta ttgtgccccg
     1321 gctggttttg cgattctaaa gtgtaaagat aagaagttca atggaaaagg accatgtaca
     1381 aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac tcaactgctg
     1441 ttaaatggca gtctagcaga agaagaggta gtaattagat ccgccaattt cgcggacaat
     1501 gctaaagtca taatagtaca gctgaatgaa tctgtagaaa ttaattgtac aagacccaac
     1561 aacaatacaa gaaaaagtat acatatagga ccaggcagag cattttatac aacaggagaa
     1621 ataataggag atataagaca agcacattgt aacottagta gagcaaaatg gaatgacact
     1681 ttaaataaga tagttataaa attaagagaa caatttggga ataaaacaat agtctttaag
     1741 cactcctcag gaggggaccc agaaattgtg acgcacagtt ttaattgtgg aggggaattt
     1801 ttctactgta attcaacaca actgtttaat agtacttgga atgttactga agagtcaaat
     1861 aacactgtag aaaataacac aatcacactc ccatgcagaa taaaacaaat tataaacatg
     1921 tggcaggaag taggaagagc aatgtatgcc cctcccatca gaggacaaat tagatgttca
     1981 tcaaatatta cagggctgct attaacaaga gatggtggtc ctgaggacaa caagaccgag
     2041 gtcttcagac ctggaggagg agatatgagg gataattgga gaagtgaatt atataaatat
     2101 aaagtagtaa aaattgaacc attaggagta gcacccacca aggcaaagag aagagtggtg
     2161 cagagagaaa aaagagcagt gggaatagga gctgtgttcc ttgggttctt gggagcagca
     2221 ggaagcacta tgggcgcagc ggcaatgacg ctgacggtac aggccagact attattgtct
     2281 ggtatagtgc aacagcagaa caatctgctg agggctattg aggcgcaaca gcatctgttg
     2341 caactcacag tctggggcat caagcagctc caggcaagag tcctggctgt ggaaagatac
     2401 ctaagggatc aacagctcct ggggatttgg ggttgctctg gaaaactcat ctgcaccact
     2461 gctgtgcctt ggaatgctag ttggagtaat aaatctctga ataagatttg ggataacatg
     2521 acctggatag agtgggacag agaaattaac aattacacaa gcataatata cagcttaatt
     2581 gaagaatcgc agaaccaaca agcaaagaat gaacaagaat tattagaatt agataaatgg
     2641 gcaagtttgt ggaattggtt tgacataaca aaatggctgt ggtatataaa aatattcata
     2701 atgatagtag gaggcttgat aggtttaaga atagtttttt ctgtactttc tatagtgaat
```

HIVBAL

```
2761 agagttaggc agggatactc accattatcg tttcAGaccc acctcccatc ctcgagggga
                                          /\ 3'sj
2821 cccgacaggc ccggaggaat cgaagaagaa ggtggagaga gagacagaga cagatccggt
2881 ccatTAGtga acggattctt ggcgcttatc tgggtcgatc tgcggagcct gttcctcttc
       <- tat
2941 agctaccacc gcttgagaga cttactcttg attgtgatga ggattgtgga acttctggga
3001 ctagcagggg ggtgggaagt cctcaaatat tggtggaatc tcctacagta ttggagtcag
3061 gaactaaaga aTAGtgctgt tagcttgctc aatgccacag ctgtagcagt agctgaaggg
               <- rev
3121 acagataggg ttatagaagt attacagaga gctgttagag ctattctcca catacctaga
3181 agaataagac agggcttgga aagggctttg ctaTAAgATG ggtggcaagt ggtcaaaaag
                                          <- env
                                  nef start ->
3241 tagtatagtc gtatggcctg ctgtaaggaa aagaatgaga agaactgagc cagcagcaga
3301 tggagtagga gcagtatcta gagacctgga aaaacatgga gcaatcacaa gtagcaatac
3361 agcagctaac aatgctgatt gtgcctggct agaagcacaa gaggatgaag aagtgggttt
3421 tccagtcaga cctcaggtac ctttaagacc aatgactcgc agtgcagcta tagatcttag
3481 ccacttttt aagaaaaagg ggggactgga agggctaatt cactcccaaa aaagacaaga
                     3'LTR start ->
3541 tatccttgat ttgtgggtct accacacaca aggctacttc cctgattggc agaactacac
3601 accagggcca gggaccagat ttccactgac ctttggatgg tgcttcaagc tagtaccagt
3661 tgagccagag aaggtagaag aggccaatga aggagagaac aactgcttgt cacaccctat
3721 gagcctgcat gggatggatg acccggagaa agaagtgtta gcatggaagt ttgacagcag
3781 cctagcattc catcacgtgg cccgagaa
                                nef ->
```

```
LOCUS       HIVADA       2562 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1 monocyte-replicative isolate
            ADA, complete env cds.
ACCESSION   M60472
SOURCE      Human immunodeficiency virus type 1, SacI/SacI molecular clone
            from the ADA isolate.
REFERENCE   1  (bases 1 to 2562)
  AUTHORS   Westervelt,P., Gendelman,H.E. and Ratner,L.
  TITLE     Identification of a determinant within the HIV-1 surface envelope
            glycoprotein critical for productive infection of primary monocytes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 3097-3101 (1991)
COMMENT     Hardcopy of sequence in [1] kindly provided by Ratner, 10-FEB-1989.
            Studies reported in [1], using chimeric molecular clones with the
            ADA viral fragment, suggest that the HIV-1 env V3 region contains
            determinants for productive infection of primary monocytes.
BASE COUNT      893 a    425 c    604 g    640 t
ORIGIN      Env cds start.
        1 ATGaaagtga aggggatcag gaagaattat cagcacttgt ggaaatgggg catcatgctc
env start ->
       61 cttgggatgt tgatgatctg TAGtgctgta gaaaatttgt gggtcacagt ttattatggg
                               <- vpU
      121 gtacctgtgt ggaaagaagc aaccaccact ctattttgtg catcagatgc taaagcatat
      181 gatacagagg tacataatgt ttgggccaca catgcctgtg tacccacaga ccccaaccca
      241 caagaagtag tattggaaaa tgtgacagaa aattttaaca tgtggaaaaa taacatggta
      301 gaacagatgc atgaggatat aatcagttta tgggatcaaa gcctaaagcc atgtgtaaaa
      361 ttaaccccac tctgtgttac tttaaattgc actgatttga ggaatgttac taatatcaat
      421 aatagtagtg agggaatgag aggagaaata aaaaactgct ctttcaatat caccacaagc
      481 ataagagata aggtgaagaa agactatgca ctttttata gacttgatgt agtaccaata
      541 gataatgata atactagcta taggttgata aattgtaata cctcaaccat tacacaggcc
      601 tgtccaaagg tatcctttga gccaattccc atacattatt gtaccccggc tggttttgcg
      661 attctaaagt gtaaagacaa gaagttcaat ggaacagggc catgtaaaaa tgtcagcaca
      721 gtacaatgta cacatggaat taggccagta gtgtcaactc aactgctgtt aaatggcagt
      781 ctagcagaag aagaggtagt aattagatct agtaatttca cagacaatgc aaaaaacata
      841 atagtacagt tgaaagaatc tgtagaaatt aattgtacaa gacccaacaa caatacaagg
      901 aaaagtatac atataggacc aggaagagca ttttatacaa caggagaaat aataggagat
      961 ataagacaag cacattgcaa cattagtaga acaaaatgga ataacacttt aaatcaaata
     1021 gctacaaaat taaaagaaca atttgggaat aataaaacaa tagtctttaa tcaatcctca
     1081 ggaggggacc cagaaattgt aatgcacagt tttaattgtg gaggggaatt tttctactgt
     1141 aatttcaacac aactgtttaa tagtacttgg aattttaatg gtacttggaa tttaacacaa
     1201 tcgaatggta ctgaaggaaa tgacactatc acactcccat gtagaataaa acaaattata
     1261 aatatgtggc aggaagtagg aaaagcaatg tatgcccctc ccatcagagg acaaattaga
     1321 tgctcatcaa atattacagg gctaatatta acaagagatg gtggaactaa cagtagtggg
     1381 tccgagatct tcagacctgg gggaggagat atgagggaca attggagaag tgaattatat
     1441 aaatataaag tagtaaaaat tgaaccatta ggagtagcac ccaccaaggc aaaaagaaga
     1501 gtggtgcaga gagaaaaaag agcagtggga acgataggag ctatgttcct tgggttcttg
     1561 ggagcagcag gaagcactat gggcgcagcg tcaataacgc tgacggtaca ggccagacta
     1621 ttattgtctg gtatagtgca acagcagaac aatttgctga gggctattga ggcgcaacag
     1681 catctgttgc aactcacagt ctggggcatc aagcagctcc aggcaagagt cctggctctg
     1741 gaaagatacc taagggatca acagctccta gggatttggg gttgctctgg aaaactcatc
     1801 tgcaccactg ctgtgccttg gaatgctagt tggagtaata aaactctgga tatgatttgg
     1861 gataacatga cctggatgga gtgggaaaga gaaatcgaaa attacacagg cttaatatac
     1921 accttaattg aagaatcgca gaaccaacaa gaaaagaatg aacaagactt attagcatta
     1981 gataagtggg caagtttgtg gaattggttt gacatatcaa attggctgtg gtatataaaa
     2041 atcttcataa tgatagtagg aggcttgata ggtttaagaa tagttttac tgtacttttt
     2101 atagtaaata gagttaggca gggatactca ccattgtcat ttcAGaccca cctcccagcc
                                                        /\ 3'sj
     2161 ccgaggggac ccgacaggcc cgaaggaatc gaagaagaag gtggagacag agacagagac
     2221 agatccgtgc gatTAGtgga tggattctta gcacttttct gggacgacct gcggagcctg
                     <- tat
     2281 tgcctcttca gctaccaccg cttgagagac ttactcttga ttgtagcgag gattgtggaa
     2341 cttctgggac gcagggggtg ggaagtcctc aagtattggt ggaatctcct gcagtattgg
     2401 agtcaggaac taaggaaTAG tgctgttagc ttgcttaatg ccacagctat agcagtagct
                          <- rev
     2461 gagggagacag atagggttat agaagtagta caaagaattt atagggctat tctccacata
     2521 cctacaagaa taagacaggg cttggaaagg cttttgctaT AA
                                                 <- env
```

```
LOCUS       HIVSWB401      59 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate SWB40, gag region.
ACCESSION   M27951
SEGMENT     1 of 2
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate SWB40, clone
            01.
REFERENCE   1  (bases 1 to 59)
  AUTHORS   Goodenow,M., Huet,T., Saurin,W., Kwok,S., Sninsky,J. and
            Wain-Hobson,S.
  TITLE     HIV-1 isolates are rapidly evolving quasispecies: evidence for
            viral mixtures and preferred nucleotide substitutions
  JOURNAL   J. Acq. Immun. Defic. Syndrome 2, 344-352 (1989)
COMMENT     [1] kindly submitted in computer readable form by S. Wain-Hobson
            Sibling sequences are available (see 1989: I-A-177). This sequence
            derives from a "rapid-high" viral isolate taken from a Swedish
            patient.
BASE COUNT      26 a      7 c     10 g     16 t
ORIGIN
        1 ctataaaaga tggataatcc tgggattaaa taaaatagta agaatgtata gccctatca


LOCUS       HIVSWB402     265 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate SWB40, envelope
            region (hypervariable V1 region).
ACCESSION   M27946 M27951
SEGMENT     2 of 2
SOURCE      Human immunodeficiency virus type 1 (HIV-1), proviral DNA, isolate
            SWB40.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Goodenow,M., Huet,T., Saurin,W., Kwok,S., Sninsky,J. and
            Wain-Hobson,S.
  TITLE     HIV-1 isolates are rapidly evolving quasispecies: evidence for
            viral mixtures and preferred nucleotide substitutions
  JOURNAL   J. Acq. Immun. Defic. Syndrome 2, 344-352 (1989)
REFERENCE   2  (bases 1 to 265)
  AUTHORS   Vartanian,J., Meyerhans,A., Asjo,B. and Wain-Hobson,S.
  TITLE     Selection, recombination, and G->A hypermutation of human
            immunodeficiency virus type 1 genomes
  JOURNAL   J. Virol. 65, 1779-1788 (1991)
COMMENT     [1], [2] kindly submitted in computer readable form by S. Wain-
            Hobson, Pasteur Institute, Paris. These sequences derive from
            a "rapid-high" viral isolate taken from a Swedish patient.
            Sibling sequences over this V1 region are available.
BASE COUNT     100 a     48 c     49 g     68 t
ORIGIN
        1 aaattaaccc cactctgtgt tactctaaat tgcactgact tgaggaatga tactaatacc
       61 aatagtagta gctgggggaaa gatggagaga ggagaaataa aaaactgctc tttcaatgtc
      121 accacaagca taagaaataa gatgcagaaa gaatatgcac ttctttataa acttgatgta
      181 gtaccaatag atratgatac taccagctat aggttgataa gttgtaacac ctcagtcatt
      241 acacaggcct gtccaaaggt aacct
```

```
LOCUS       HIVSWB881      59 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate SWB88, gag region.
ACCESSION   M27952
SEGMENT     1 of 4
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate SWB88, clone
            G1.
REFERENCE   1  (bases 1 to 59)
  AUTHORS   Goodenow,M., Huet,T., Saurin,W., Kwok,S., Sninsky,J. and
            Wain-Hobson,S.
  TITLE     HIV-1 isolates are rapidly evolving quasispecies: evidence for
            viral mixtures and preferred nucleotide substitutions
  JOURNAL   J. Acq. Immun. Defic. Syndrome 2, 344-352 (1989)
COMMENT     [1] kindly submitted in computer readable form by S. Wain-Hobson.
            B88 is a so-called "slow virus" isolated in Sweden. Sibling
            sequences available. (see 1989: I-A-174).
BASE COUNT      26 a      8 c     10 g     15 t
ORIGIN
        1 ctataaaaga tggataatcc tgggattaaa taaaatagta agaatgtata gccctacca


LOCUS       HIVSWB882     314 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate B88, tat cds for
            first expressed exon of isolate L1.01.
ACCESSION   M31462
SEGMENT     2 of 4
REFERENCE   1  (bases 1 to 314)
  AUTHORS   Meyerhans,A., Cheynier,R., Albert,J., Seth,M., Kwok,S., Sninsky,J.,
            Morfeldt-Manson,L., Asjo,B. and Wain-Hobson,S.
  TITLE     Temporal fluctuations in HIV quasispecies in vivo are not reflected
            by sequential HIV isolations
  JOURNAL   Cell 58, 901-910 (1989)
COMMENT     Kindly submitted in computer-readable form by S. Wain-Hobson,
            Pasteur Institute, Paris.  The B88 isolate is considered to be a
            "slow-low" viral form.  Sibling sequences from pbl's and viruses
            propogated on pbl's are available. This sequence corresponds to the
            earliest sample major form, L1.01.
BASE COUNT      111 a     57 c     70 g     76 t
ORIGIN
        1 cattattcaa cagaggagag caacaagaaA TGgagccagt agatcctaga ctagagccct
                                      tat ->
       61 ggaagcatcc aggaagtcag cctaagactg cttgtaacaa ttgctattgt aaaaagtgtt
      121 gctttcattg ccaagtttgt ttcacaaaaa aaggcttagg catctcctat ggcaggaaga
      181 agcggagaca gcgacgaaga gctcctcagg acagtgagac tcatcaagtt tctctatcaa
      241 agcaGTaagt agtacatgta atgcaacctt tagaaatatt agcaatagta gcattagtag
       5' sj /\
      301 tagcaataat acta
```

```
LOCUS       HIVSWB883     250 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate SWB88, envelope
            region.
ACCESSION   M27947
SEGMENT     3 of 4
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate SWB88, clone
            01.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Goodenow,M., Huet,T., Saurin,W., Kwok,S., Sninsky,J. and
            Wain-Hobson,S.
  TITLE     HIV1 isolates are rapidly evolving quasispecies: evidence for viral
            mixtures and preferred nucleotide substitutions
  JOURNAL   J. Acq. Immun. Defic. Syndrome 2, 344-348 (1989)
COMMENT     [1] kindly submitted in computer readable form by S. Wain-Hobson.
            This stretch comes from the first hypervariable region of env.
            Sibling sequences are available (see 1989: I-A-175).
BASE COUNT       96 a     40 c     42 g     72 t
ORIGIN
        1 aaattaacac cactctgtgt tactttaaat tgcactaagt tgaagaatgt tacttatgcc
       61 aataatagta gtagggaa.a agaagaaatg aaaaactgct ctttcaccac aagaataggg
      121 aataaggtgc agaaagaata tgcacttttt tataaacttg atgtagtacc aatagataag
      181 gataatacaa gctttatatt gatacattgt aatacctcga ccattacaca ggcttgtcca
      241 aaggtatcct
```

```
LOCUS       HIVSWB884     851 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate SWB88, nef,ltr
            regions. ( nef cds for isolate L1.14).
ACCESSION   M58206
SEGMENT     4 of 4
REFERENCE   1  (bases 1 to 851)
  AUTHORS   Delassus,S., Cheynier,R. and Wain-Hobson,S.
  TITLE     Evolution of human immunodeficiency virus type 1 nef and long
            terminal repeat sequences over 4 years in vivo and in vitro
  JOURNAL   J. Virol. 65, 225-231 (1990)
COMMENT     [1] kindly submitted in computer readable form by S. Wain-Hobson,
            Pasteur Institute, Paris. Sibling sequences of this isolate
            (L1.14) are available.
BASE COUNT      237 a    186 c    251 g    177 t
ORIGIN
        1 ATGggtggca agtggtcaaa aagtagtatg agtggatggc ctgctataag ggaaagaatg
nef start ->
       61 aaacaagctg agccagcagc agatggggtg ggagcagtat ctcgagacct ggaaaaacat
      121 ggagcaatca caagtagcaa tacagcagct actaatgctg attgtgcctg gctagaagca
      181 caagaggagg aagaggtggg ttttccagtc agacctcagg tacctttaag accaatgact
      241 cgcagggcag ctatagatct tagccacttt ttaagagaaa agggggggact ggaagggcta
                                                        3'LTR ->
      301 gttcactccc aaaaaagaca agagatcctt gatctgtggg tctaccacac acaaggctac
      361 ttccctgatt ggcagaacta cacaccaggg ccagggacca gatggccact gaccctgga
      421 tggtgcttca agctagtacc agtcgagcca gagaaggtag aagaggccaa tgaaggagag
      481 aacaacagct tgttacaccc tatgagcctg catggaatgg aggacccgga gaaagaagtg
      541 ttagtgtgga agtttgacag cagcctagca tttcaccaca aggcccgaga gctgcatccg
      601 gagtattaca aagactgcTG Acatcgagtt ttctacgagg gactttccgc tggggacttt
                            <- nef
      661 ccagggaagc gtggcctggg cgggactggg gagtggcgag ccctcagatg ctgcaTATAA
                                                              signal ->
      721 gcagctgctt tttgcctgta ctGGGtctct ctggttagac cagatctgag cctgggagct
                          R start ->
      781 ctctggctag ctagggaacc cactgcttaa gcctcAATAA Agcttgcctt gagtgcttca
                                        signal ->
      841 agtagtgtgt g
```

```
LOCUS       HIVM77        201 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, gag (p24) region.
SOURCE      Human immunodeficiency virus type 1 (HIV-1), direct sequencing of
            PCR amplified proviral DNA from the peripheral blood of an
            HIV-infected hemophiliac (patient 77).
SEGMENT     1 of 3.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     ;
  JOURNAL   Unpublished (1990).
REFERENCE   2  (bases 1 to 201)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Concurrent evolution of human immunodeficiency virus type 1 in
            patients infected from the same source: rate of sequence change and
            low frequency of inactivating mutations
  JOURNAL   J. Virol. 64, 6221-6233 (1990)
REFERENCE   3  (bases 1 to 201)
  AUTHORS   Simmonds,P., Balfe,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Analysis of sequence diversity in hypervariable regions of the
            external glycoprotein of human immunodeficiency virus type 1
  JOURNAL   J. Virol. 64, 5840-5850 (1990)
COMMENT     Kindly submitted in computer readable form by P. Balfe.  These
            sequences were obtained by direct sequencing of PCR amplified DNA
            from the peripheral blood of a member of a cohort of hemophiliacs,
            some of whom were probably infected by the same batch of factor
            VIII.  The gag sequences in reference [1] were used in calculating
            the gag distance matrices in reference [2].
            The sequence in this entry is one of fifteen sequences over the gag
            region obtained from this patient.  An alignment presenting all the
            gag sequences obtained from each patient in this study is presented
            on the following pages. Sequences from the same individual are
            indicated by the same first two digits of the sequence name.
BASE COUNT       75 a     38 c     48 g     40 t
ORIGIN      201 total bases
        1 gaatgcatcc agtgcatgca gggcctattg caccaggcca gatgagagaa ccaaggggaa
       61 gtgacatagc aggaactact agtacccttc aggaacaaat aggatggatg acaaataatc
      121 cacctatccc agtaggagaa atctataaaa gatggataat cctgggatta aataaagatag
      181 taaggatgta tagtcctacc a
```

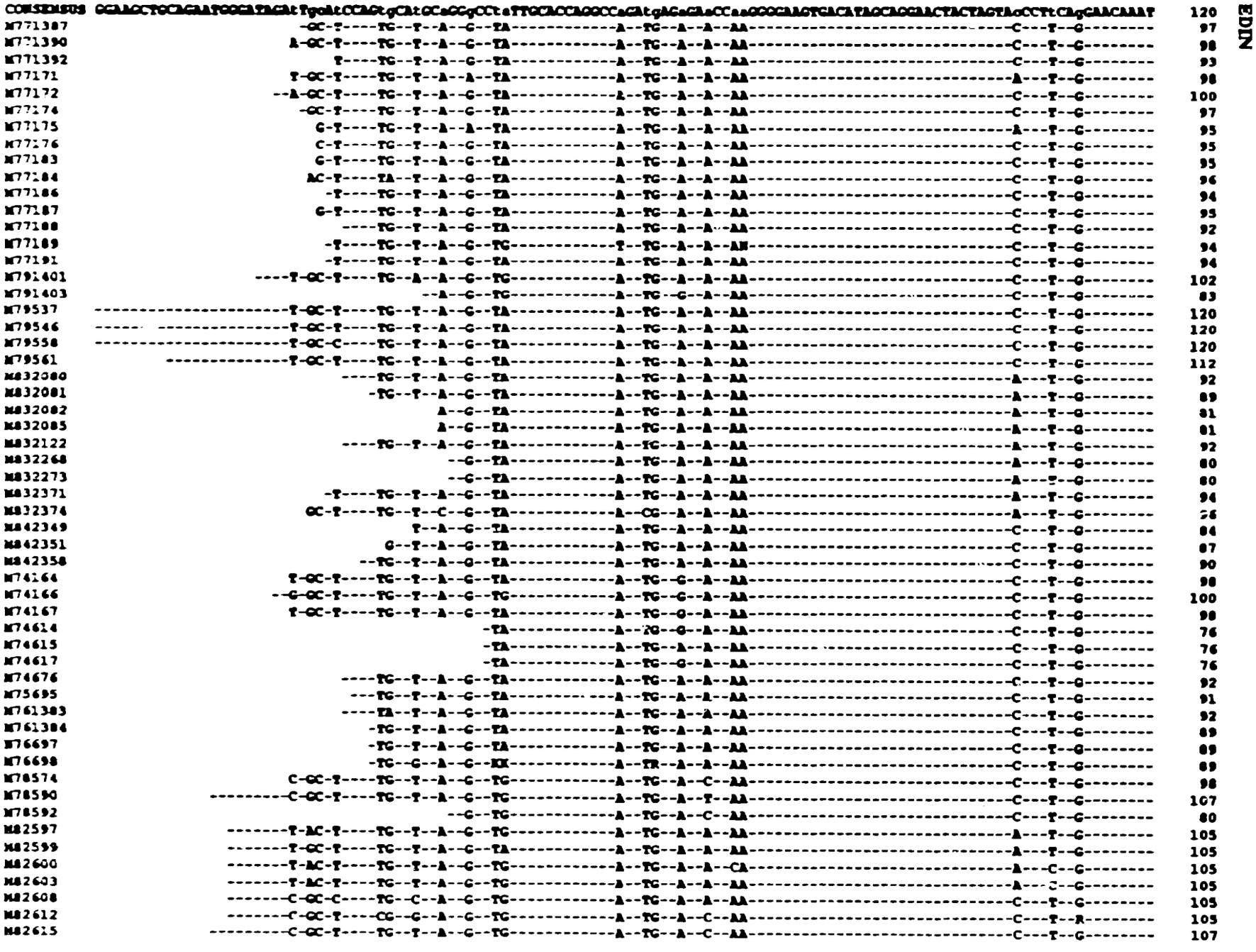
```
LOCUS       HIVV77        282 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, v3 region.
SOURCE      Human immunodeficiency virus type 1, direct sequencing of PCR
            amplified proviral DNA from peripheral blood of an HIV-infected
            hemophiliac (patient 77).
SEGMENT     2 of 3.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Concurrent evolution of human immunodeficiency virus type 1 in
            patients infected from the same source: rate of sequence change and
            low frequency of inactivating mutations
  JOURNAL   J. Virol. 64, 6221-6233 (1990)
REFERENCE   2  (bases 1 to 282)
  AUTHORS   Simmonds,P., Balfe,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Analysis of sequence diversity in hypervariable regions of the
            external glycoprotein of human immunodeficiency virus type 1
  JOURNAL   J. Virol. 64, 5840-5850 (1990)
COMMENT     Kindly submitted in computer readable form by P. Balfe.  These
            sequences were obtained by direct sequencing of PCR amplified DNA
            from the peripheral blood of a member of a cohort of hemophiliacs,
            some of whom were probably infected by the same batch of factor
            VIII.
            The sequence in this entry is one of six sequences over the v3
            region obtained from this patient.  An alignment presenting all the
            v3 sequences obtained from each of the patients in this study is
            presented on the following pages. Sequences from the individual are
            indicated by the same first two digits of the sequence name.
BASE COUNT      126 a     46 c     51 g     59 t
ORIGIN
        1 acggacaatg ctaaaaccat aatagtacag ctgaaggacc ctgtaaacat tacttgtaca
       61 agacccagca acaatacaag aaaaagtata cacatagcac cggggagagt atttcatgca
      121 acaggagaaa taataggaga tataagacaa gcacattgta acattagtag aaaagattgg
      181 aataacactt taaaacagat agttacaaaa ttaaaagaac agtttggaaa taaaacaata
      241 atctttaatc actcctcagg aggggaccca gaaattgtaa tg
```

```
LOCUS       HIVT77        251 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, v4c3v5 region.
SOURCE      Human immunodeficiency virus type 1, direct sequencing of PCR
            amplified proviral DNA from peripheral blood of an HIV-infected
            hemophiliac (patient 77).
SEGMENT     3 of 3.
REFERENCE   1  (bases 1 to 251)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Concurrent evolution of human immunodeficiency virus type 1 in
            patients infected from the same source: rate of sequence change and
            low frequency of inactivating mutations
  JOURNAL   J. Virol. 64, 6221-6233 (1990)
REFERENCE   2  (bases 1 to 251)
  AUTHORS   Simmonds,P., Balfe,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Analysis of sequence diversity in hypervariable regions of the
            external glycoprotein of human immunodeficiency virus type 1
  JOURNAL   J. Virol. 64, 5840-5850 (1990)
COMMENT     Kindly submitted in computer readable form by P. Balfe.  These
            sequences were obtained by direct sequencing of PCR amplified DNA
            from the peripheral blood of a member of a cohort of hemophiliacs,
            some of whom were probably infected by the same batch of factor
            VIII.
            The sequence in this entry is one of thirteen sequences over the
            v4c3v5 region obtained from this patient.  An alignment presenting
            all the v4c3v5 sequences obtained from each of the patients in this
            study is presented on the following pages. Sequences from the same
            individual are indicated by the same first two digits of the
            sequence name.
BASE COUNT       93 a     49 c     53 g     55 t      1 others
ORIGIN
        1 tgtctaatag tacttggctg tttaatagta cttggaatga tactacagag tcaaataaca
       61 ctgaacctat cacactccca tgcagaataa aacaaattat aaacagatgg caggaagtag
      121 gaaaagcaat gtatgccctc cccatcagag gacaaattag atgttcatca aatattacag
      181 ggctgctatt aacaagagat ggtggtagga acaacgagac caacggnacc gagaccttca
      241 gacctggagg a
```

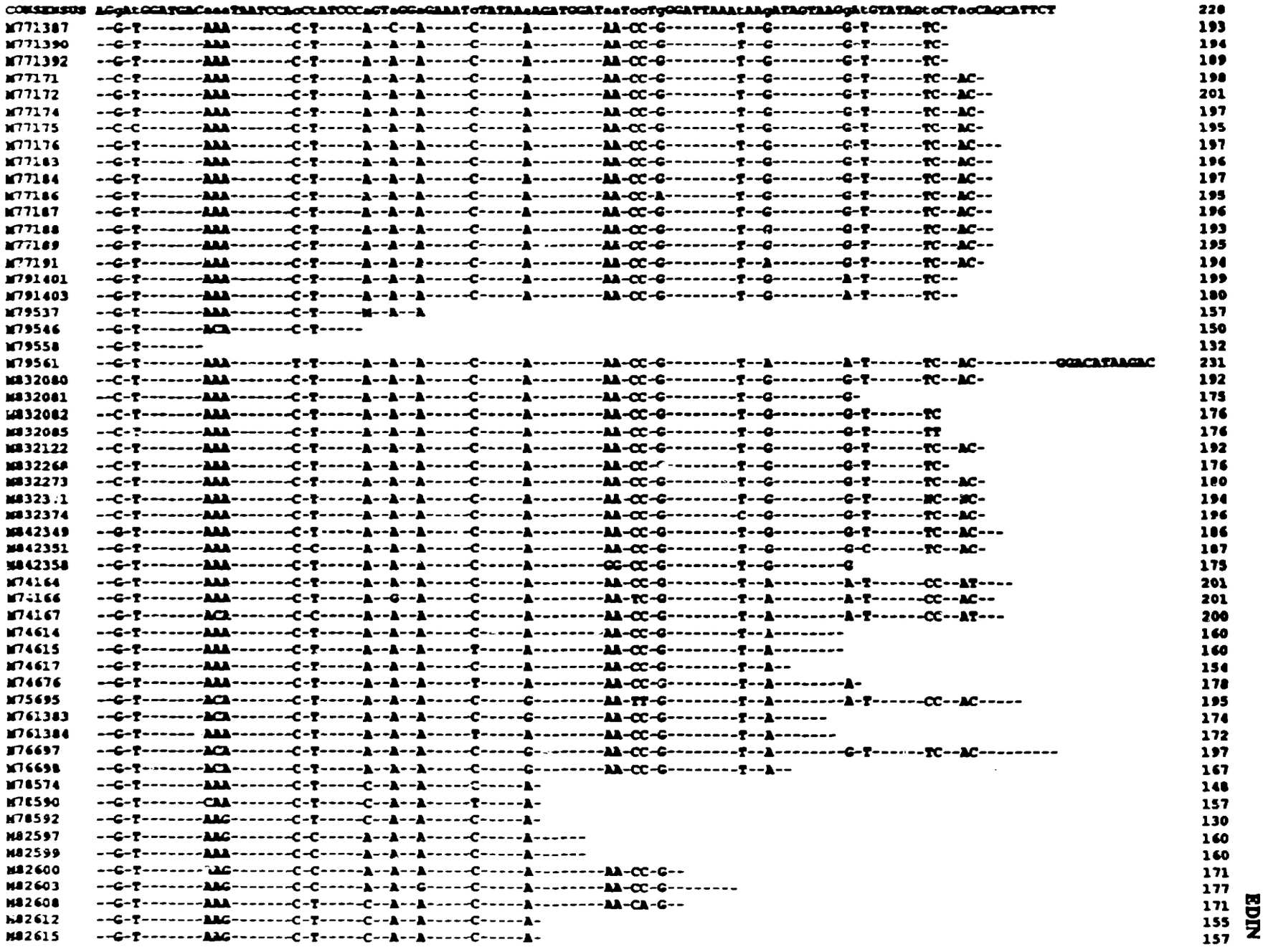
```
LOCUS       HIVM82        177 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, gag (p24) region.
SOURCE      Human immunodeficiency virus type 1 (HIV-1), direct sequencing of
            PCR amplified proviral DNA from the peripheral blood of an
            HIV-infected hemophiliac (patient 82).
SEGMENT     1 of 3.
REFERENCE   1  (bases 1 to 177)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     ;
  JOURNAL   Unpublished (1990).
REFERENCE   2  (bases 1 to 177)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Concurrent evolution of human immunodeficiency virus type 1 in
            patients infected from the same source: rate of sequence change and
            low frequency of inactivating mutations
  JOURNAL   J. Virol. 64, 6221-6233 (1990)
REFERENCE   3  (bases 1 to 177)
  AUTHORS   Simmonds,P., Balfe,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Analysis of sequence diversity in hypervariable regions of the
            external glycoprotein of human immunodeficiency virus type 1
  JOURNAL   J. Virol. 64, 5840-5850 (1990)
COMMENT     Kindly submitted in computer readable form by P. Balfe.  These
            sequences were obtained by direct sequencing of PCR amplified DNA
            from the peripheral blood of a hemophiliac infected by a
            contaminated batch of factor VIII.  The gag sequences in reference
            [1] were used in calculating the gag distance matrices in reference
            [2].
            The sequence in this entry is one of seven sequences over the gag
            region obtained from this patient.  An alignment presenting all the
            gag sequences obtained from each of the patients in this study is
            presented on the following pages. Sequences from the same individual
            are indicated by the same first two digits of the sequence name.
BASE COUNT      64 a     35 c     46 g     32 t
ORIGIN      177 total bases
        1 ggatagatta catccagtgc atgcagggcc tgttgcacca ggccagatga gagaaccaag
       61 gggaagtgac atagcaggaa ctactagtaa cctccaggaa caaataggat ggatgacaag
      121 taatccaccc atcccagtag gggaaatcta taaaagatgg ataatcctgg gattaaa
```

```
LOCUS       HIVV82        306 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, v3 region.
SOURCE      Human immunodeficiency virus type 1, direct sequencing of PCR
            amplified proviral DNA from peripheral blood of an HIV-infected
            hemophiliac (patient 82).
SEGMENT     2 of 3.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Concurrent evolution of human immunodeficiency virus type 1 in
            patients infected from the same source: rate of sequence change and
            low frequency of inactivating mutations
  JOURNAL   J. Virol. 64, 6221-6233 (1990)
REFERENCE   2  (bases 1 to 306)
  AUTHORS   Simmonds,P., Balfe,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Analysis of sequence diversity in hypervariable regions of the
            external glycoprotein of human immunodeficiency virus type 1
  JOURNAL   J. Virol. 64, 5840-5850 (1990)
COMMENT     Kindly submitted in computer readable form by P. Balfe.  These
            sequences were obtained by direct sequencing of PCR amplified DNA
            from the peripheral blood of a hemophiliac infected by an
            HIV-contaminated batch of factor VIII.
            The sequence in this entry is one of six sequences over the v3
            region obtained from this patient.  An alignment presenting all the
            v3 sequences obtained from each of the patients in this study is
            presented on the following pages. Sequences from the same individual
            are indicated by the same first two digits of the sequence name.
BASE COUNT      135 a     43 c     54 g     74 t
ORIGIN
        1 tctagcaatt ttacagacaa tgctaaaacc ataatagtac agctgaatga atctgtagta
       61 attaattgta caagacccaa caacaataca agaaaaggta tacatatagg accagggagt
      121 gcattttatg caacaggagg aataatagga gatataagac aagcacattg taacattagt
      181 agagcaaaat ggaatgaaac tttaaaacag atagttacaa aattaagaga acaatttagg
      241 aataaaacaa tagtctttaa tcaatcctca ggaggggacc cagaaattgt aatgcacagt
      301 tttaat
```

```
LOCUS       HIVT82        260 bp ss-RNA             VRL       23-MAY-1991
DEFINITION  Human immunodeficiency virus type 1, v4c3v5 region.
SOURCE      Human immunodeficiency virus type 1, direct sequencing of PCR
            amplified proviral DNA from peripheral blood of an HIV-infected
            hemophiliac (patient 82).
SEGMENT     3 of 3.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Balfe,P., Simmonds,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Concurrent evolution of human immunodeficiency virus type 1 in
            patients infected from the same source: rate of sequence change and
            low frequency of inactivating mutations
  JOURNAL   J. Virol. 64, 6221-6233 (1990)
REFERENCE   2  (bases 1 to 260)
  AUTHORS   Simmonds,P., Balfe,P., Ludlam,C.A., Bishop,J.O. and Brown,A.J.L.
  TITLE     Analysis of sequence diversity in hypervariable regions of the
            external glycoprotein of human immunodeficiency virus type 1
  JOURNAL   J. Virol. 64, 5840-5850 (1990)
COMMENT     Kindly submitted in computer readable form by P. Balfe.  These
            sequences were obtained by direct sequencing of PCR amplified DNA
            from the peripheral blood of a hemophiliac infected with an
            HIV-contaminated batch of factor VIII.
            The sequence in this entry is one of eleven sequences over the
            v4c3v5 region obtained from this patient.  An alignment presenting
            all the v4c3v5 sequences obtained from each of the patients in this
            study is presented on the following pages. Sequences from the same
            individual are indicated by the same first two digits of the
            sequence name.
BASE COUNT      106 a     45 c     51 g     58 t
ORIGIN
        1 ttaatagtac ttggaattca acatgggatt taacacaact taatagtact cagaataaag
       61 aagaaaatat cacactccca tgtagaataa aacaaattat aaacatgtgg caggaagtag
      121 gaaaagcaat gtatgccctt cccatcaaag gacaaattag atgttcatca aatattacag
      181 ggctattatt aacaagagat ggtggtaata gtggtaacaa aagcaacgac accaccgagt
      241 ccttcagacc tgggggagga
```

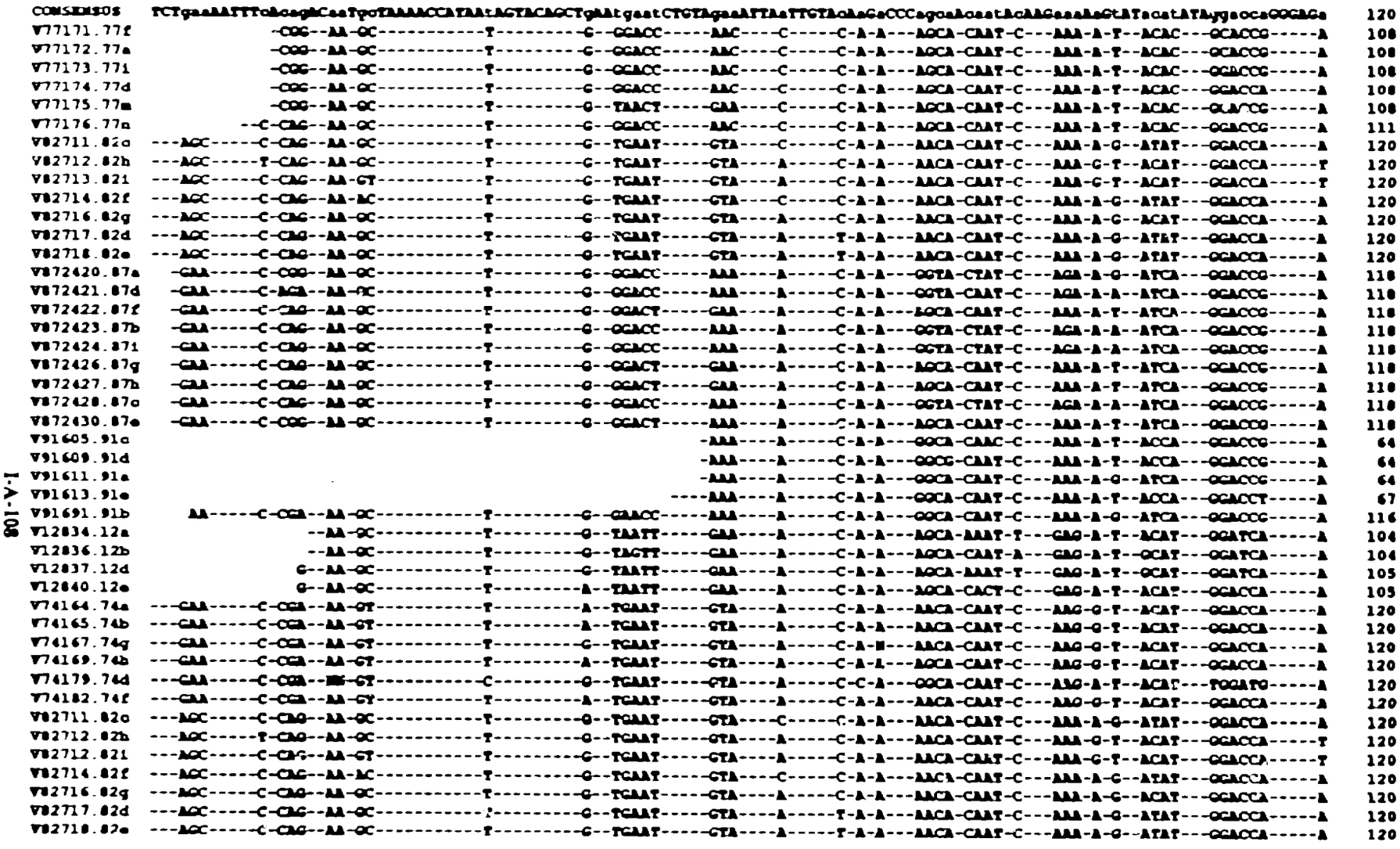


GAG

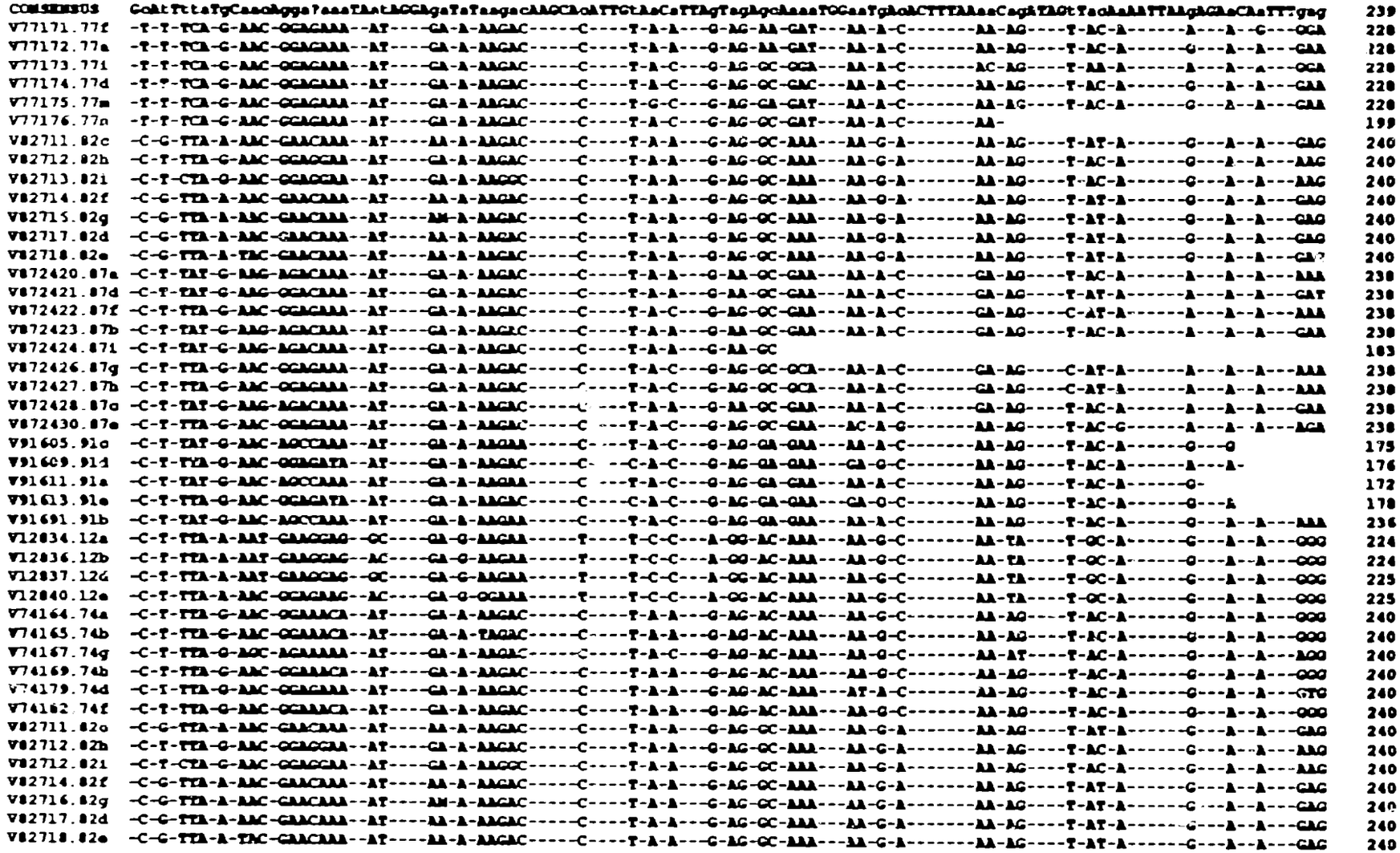

CONSENSUS
M771387
M771390
M771392
M77171
M77172
M77174
M77175
M77176
M77183
M77184
M77186
M77187
M77188
M77189
M77191
M791401
M791403
M79537
M79546
M79558
M79561
M832080
M832081
M832082
M832085
M832122
M832268
M832273
M832314
M832374
M842349
M842351
M842358
M74164
M74166
M74167
M74614
M74615
M74617
M74676
M75695
M761383
M761384
M76697
M76698
M78574
M78590
M78592
M82597
M82599
M82600
M82603
M82608
M82612
M82615
220
193
194
189
198
201
197
195
197
196
197
195
196
193
195
194
199
180
157
150
132
231
192
175
176
176
192
176
180
194
196
186
187
175
201
201
200
160
160
154
178
195
174
172
197
167
148
157
130
160
160
171
177
171
155
157

EDIN

V3



MAY 91
1-A-109



EDIN

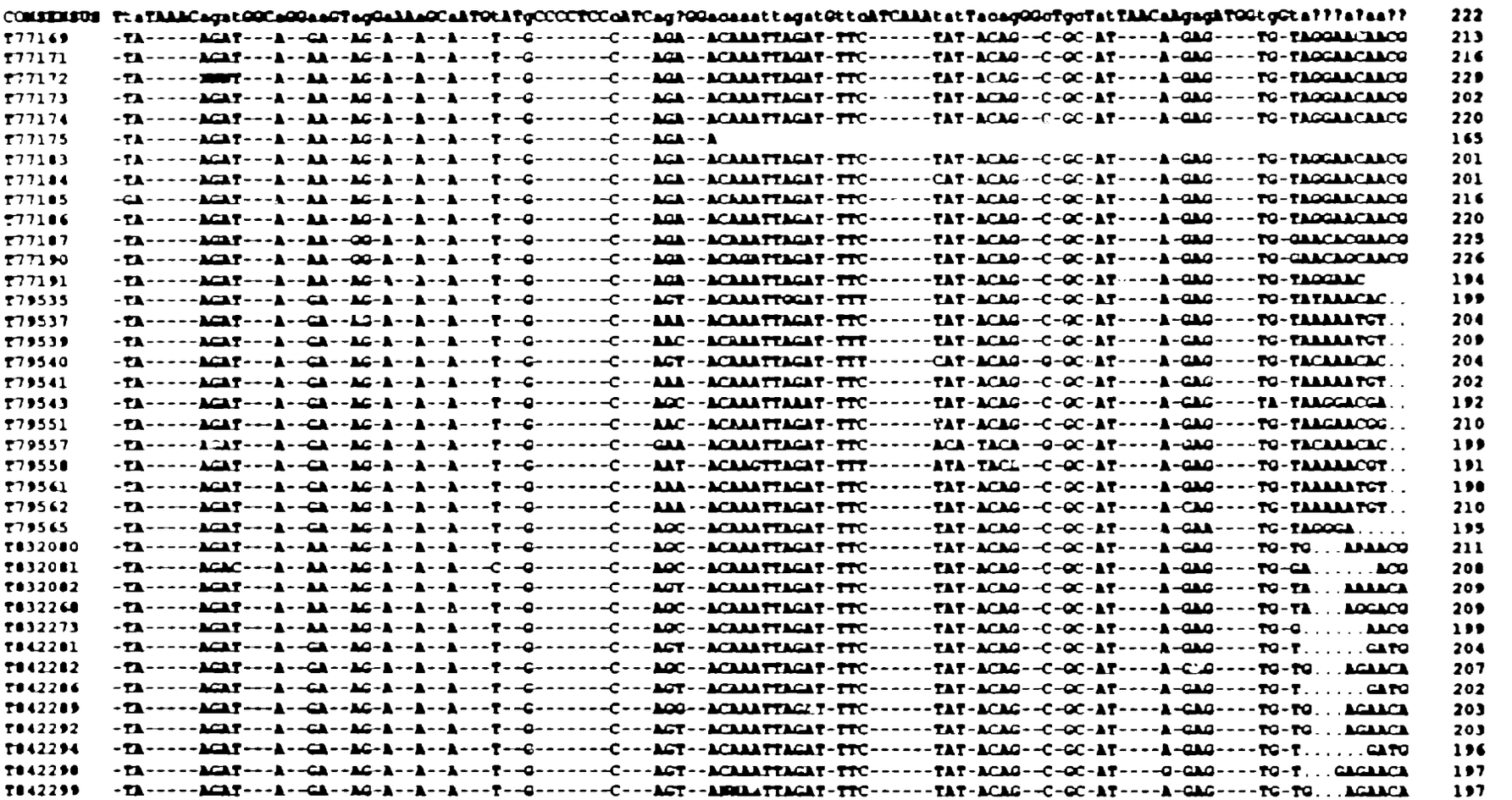
V3

```
CONSENSUS   AATAaaaCaAtAgtcTTTaatCaaTCctCAGCAGGGCACCCAGAAATTgTAATGCACAGTTTTAAT  305
V77171.77f  ----AAA-A-T-ATC---AAT-AC--CT--------------------G-----  282
V77172.77a  ----AAA-A-T-ATC---AAT-AA--TT--------------------G-----  282
V77173.771  ----AAA-A-T-ATC---AAT-AA--CT--------------------G-----  282
V77174.77d  ----AAA-A-T-ATC---AAT-AA--TT--------------------G-----  282
V77175.77m  ----AAA-A-T-ATC---AAT-AA--TT--------------------G-----  282
V82711.82a  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  302
V82712.82b  ----AAA-A-T-GTC---AAT-AA--CT--------------------G----------  306
V82713.821  ----AAA-A-T-GTC---AAT-AA--CT--------------------G----------  306
V82714.82f  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  306
V82716.82g  ----AAG-A-T-GTC---AAG-AA--CT--------------------A----------  299
V82717.82d  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  302
V82718.82e  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  302
V872420.87a ----AAA-A-T-ATC---AAT-AA--CT--------------------G-  288
V872421.87d ----AAA-A-T-ATC---AAT-AA--CT--------------------G-  288
V872422.87f ----CAA-A-T-ATC---AAT-AA--CG--------------------G-  288
V872423.87b ----AAA-A-T-ATC---AAT-AA--CT--------------------G-  288
V872426.87g ----CAA-A-A-ATC---AAT-AA -CT--------------------G-  288
V872427.87h ----CAA-A-A-ATC---AAT-AA--CT--------------------G-  289
V872428.87c ----AAA-A-T-ATC---AAT-AA--CT--------------------G---  290
V872430.87e ----AAA-A-A-ATC---AAT-AA--CT--------------------G-  288
V12834.12a  ----AAA-A-T-CAA---AAG-CA--CT--------------------  272
V12836.12b  ----AAA-A-T-ACA---AAG-CA--CT--------------------  272
V12837.12d  ----AAA-A-T-CAA---AAG-CA--CT--------------------  273
V12840.12e  ----AAA-T-T-AAA---CAG-CA--CT--------------------  273
V74164.74a  ----GCA-A-T-GTC---AAT-AA--CT--------------------G----------  305
V74165.74b  ----GCA-A-T-GTC---AAT-AA--CT--------------------G----------  306
V74167.74g  ----CAA-A-T-GTC---GGT-CA--CT--------------------G----------  306
V74169.74h  ----GCA-A-T-GTC---ACT-AA--CT--------------------G----------  306
V74179.74d  ----AAA-A-T-GTC---AAT-AA--CT--------------------G----------  306
V74182.74f  ----AAA-A-T-GTC---GAT-AA--CT--------------------G----------  306
V82711.82a  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  302
V82712.82b  ----AAA-A-T-GTC---AAT-AA--CT--------------------G----------  306
V82712.821  ----AAA-A-T-GTC---AAT-AA--CT--------------------G----------  306
V82714.82f  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  306
V82716.82g  ----AAG-A-T-GTC---AAG-AA--CT--------------------A----------  299
V82717.82d  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  302
V82718.82e  ----AAA-A-T-GTC---AAG-AA--CT--------------------G----------  302
```

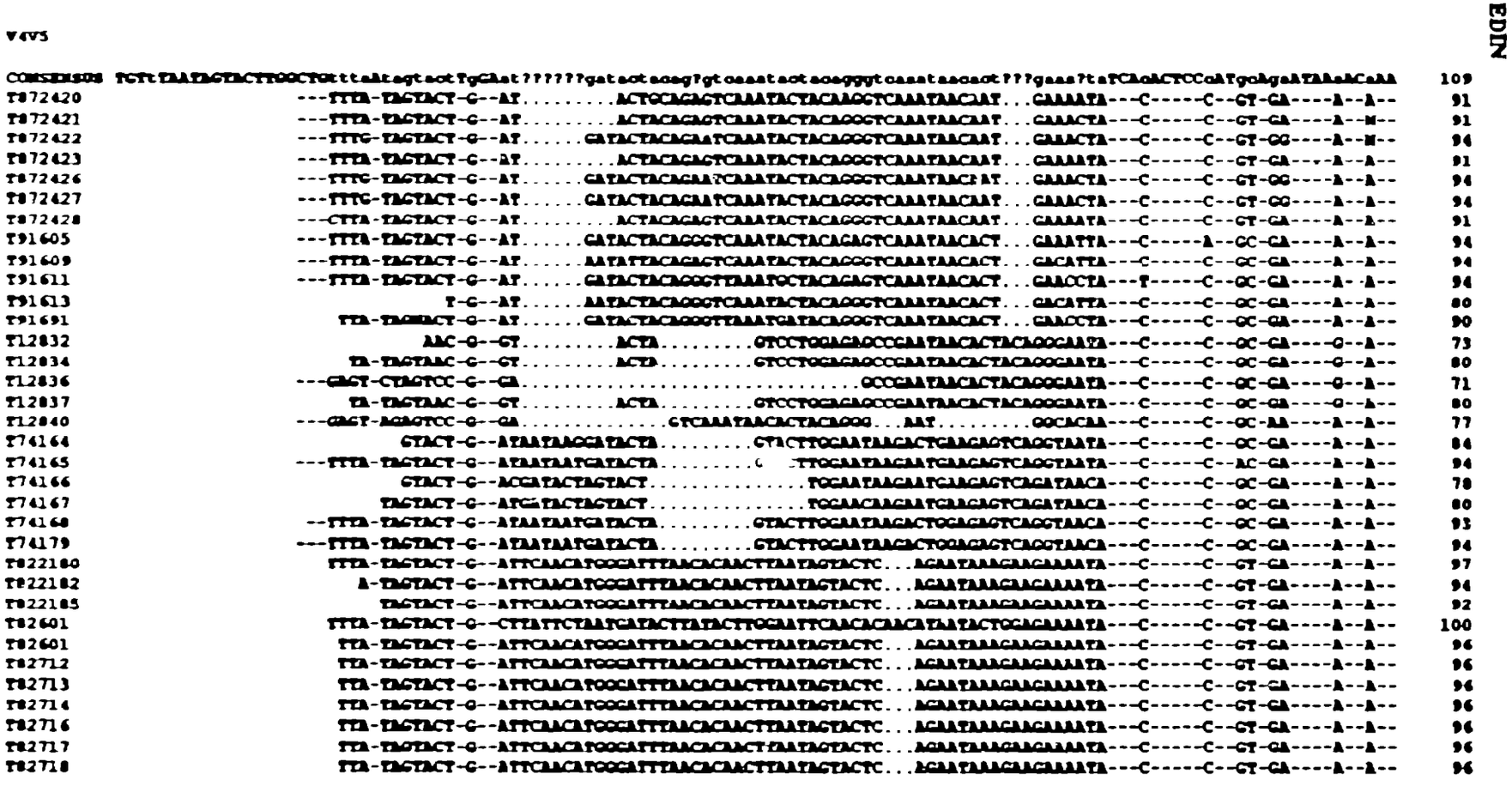
V4V5

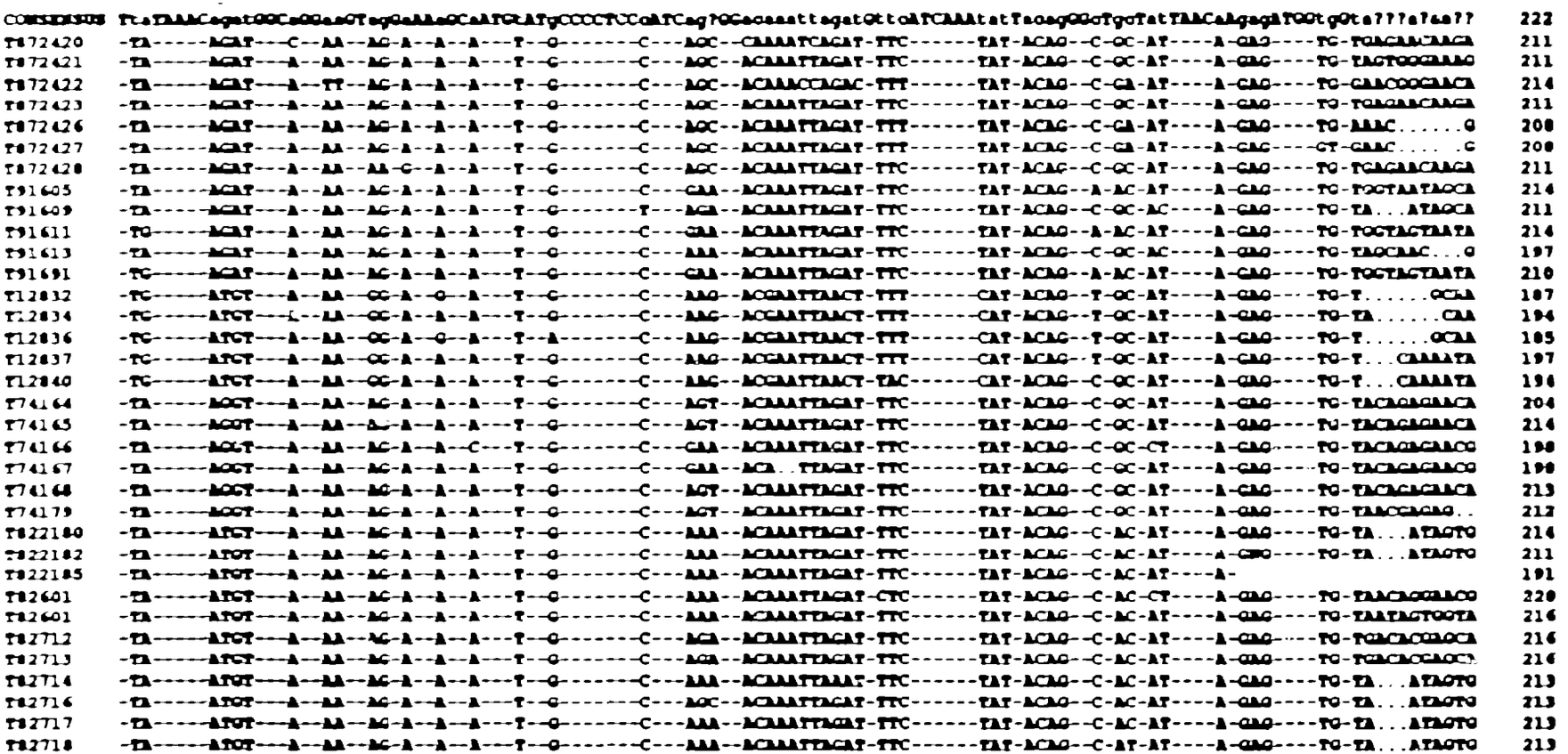
```
CONSENSUS TCTtTAATAGTACTTT20CTGttaAtagtactTgGAat??????gatactacag?gCcaaatactacaggggtcaaaataacact???gaaa?taTCAcAACTCCcAtgcAgaATAAaACaAA  109
T77169                        --TTTA-TAGTACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACAAT...GAAACTA---C-----C--GC-GA----A--A--   93
T77171     ---C----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAAAT..............AACACT...GAACCTA---C-----C--GC-GA----A--A--   96
T77172       -T----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAAATACTACGGAGTCAAATAACACT...GAACCTA---C-----C--GC-GA----A--A--  109
T77173                         TTTA-TAGTACT-G --AT.....GGTACTACAGAGCCAAATATTACAGGGTCAAGTAACA.............---C-----C--GC-GA----A--A--   82
T77174       -T----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAAATACTACAGAGTCAAGTAACA.............---C-----C--GC-GA----A--A--  100
T77175       -T----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAAATACTACAGAGACAAATAACACA...GAACCTA---C-----C--GC-GA----A--A--  109
T77183                        --TTTA-TAGTATT-G--AT.....ACTACAGAGCCAAATATTACAGGATCA.............GAAAATA---C-----C--GC-GA----A--A--   91
T77184                        --TTTA-TAGTATT-G--AT.....ACTACAGAGCCAAATATTACAGGGTCA.............GAAAATA---C-----C--GC-GA----A--A--    
T77185     ---T----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAAAT..............AACACT...GAACCTA---C-----C--GC-GA----A--A--   94
T77186       -T----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAAATACTACAGAGTCAAGTAACA.............---C-----C--GC-GA----A--A--  100
T77187        T----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAACT...ACAGAGTCAAATAACACT...GAACCTA---T-----C--GC-GA----A--A--  105
T77190       -T----------------TTTA-TAGTACT-G--AT.....GATACTACAGAGTCAACT...ACAGAGTCAAATAACACT...GAACCTA---T-----C--GC-GA----A--A--  106
T77191                          TTA-TAGTATT-G--AT.....ACTACAGAGCCAAATATTACAGGGTCA.............GAAAATA---C-----C--GC-GA----A--A--   78
T79535                                CT-G--AT.....GATACTACAGGGTCAAATGCTACAGGGTCAAATGACACT...GAAAC A---C-----C--GC-GA----A--A--   81
T79537                             TAGTACC-G--AT.....GATACTACTGGGTCAAACACTACAGGGTCAAATAACACT...GAACCTA---C-----C--GC-GA----A--A--   86
T79539                          TTTA-TAGTACT-G--AT.....GATACTACAGGGTCAAACACTACAGGGTCAAATAACACT...GAACCTA---C-----C--GC-GA----A--A--   91
T79540                             TAGTACT-G--AT.....GATACTACAGGGTCAATTGCTACAGGATCAAATGACACT...GAAACTA---C-----C--GC-GA----A--A--   86
T79541                              GTACT-G--AT.....GATACTACAGGGTCAAACGCTACAGGGTCAAATAACACT...GAACCTA---C-----C--GC-GA----A--A--   84
T79543                            TA-TAGTACT-A--AT.....GCTACTACAGGGTCAAATA..............ACACT...GAAACTA---C-----C--GC-GA----A--A--   74
T79551                         -TTTA-TAGTACT-G--AT.....GATACTACAGGGTCAAACACTACAGCGTCAAATAACACT...GGACCTA---C-----C--GC-GA----A--A--   92
T79557                                CT-G--AT.....AATGCTACAGGGTCAAATGCTACAGGGTCAAATGACACT...GAAACTA---C-----C--GC-GA----A--A--   81
T79558                              ACT-G--AT.....GATACTACAGGTTCAAACACTACAGGGTCAAATAACA.............---C-----C--GC-GA----A--A--   73
T79561                                 T-G--AT.....GATACTACAGGGTCAAACACTACAGGGTCAAATAACACT...GAACCTA---C-----C--GC-GA----A--A--   80
T79562                         -TTTA-TAGTACT-G--AT.....GATACTACAGGGTCAAACACTACAGGGTCAAATAACACT...GAACCTA---C-----C--GC-GA----A--A--   92
T79565                                 T-G--AT.....GATACTATAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   80
T832080                     ---TTTA-TAGTACT-G--AT.....GATACTACAGGGTTAAATATTACAGGGTCAAATAACAAT...GACACTA---C-----C--GC-GA----A--A--   94
T832081                     ---TTTA-TAGTACT-G--AT.....GATACTACAGGGTTAAATATTACAGGGTCAAATAACATT...GGAAATA---C-----C--GC-GA----A--A--   94
T832082                       -TTTA-TAGTACT-G--AT.....GATACTACAGGGTYAAATATTACAGGGTCAAATAACAAT...GAAACTA---C-----C--GC-GA----A--A--   97
T832268                       -TTTA-TAGTACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GGAAATA---C-----C--GC-GA----A--A--   92
T832273                     ---TTTA-TAGTACT-G--AT.....GATACTACAGGGTCAAATACTACTGGGTC.........T...GGAAATA---C-----C--GC-GA----A--A--   85
T842281                         TTA-TAGCACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   90
T842282                         TTA-TAGCACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   90
T842286                           A-TAGCACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   88
T842289                             TAGCACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   86
T842292                             TAGCACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   86
T842294                                ACT-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAGTAACACT...GAAACTA---C-----C--GC-GA----A--A--   82
T842298                                T-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   80
T842299                                T-G--AT.....GATACTACAGGGTCAAATACTACAGGGTCAAATAACACT...GAAACTA---C-----C--GC-GA----A--A--   80
```

V4V5



CONSENSUS	?ga??????oea?ga????acoqAqaoot?oageCCTgGeOCAOCAGATATG	262
T77169	AGAC.....C.....ACCG-GATCT-	229
T77171	AGAC.....CAACOGN...ACCG-GACCT-CAGA---G-A---	251
T77172	AGAC.....CACC...ACCACCG-GACCT-CAGA---G-A	261
T77173	AGAC.....CAACGAGACCACCG-GACCT-CAGA---G-A---	240
T77174	AGAC.....CAACGAG...ACCG-GACCT-CAGA---	249
T77183	AGAC.....CAACGAG...ACCA	216
T77184	AGAC.....CAACGAG...ACCT	216
T77185	AGAC.....CAACGAGACCACCG-AC	237
T77186	AGAC.....CAACGAGACCACCG-GAC	242
T77187	GGAC.....CAACGAGAC	238
T77190	GGAC.....CAACGAGACC	240
T79535	.GAC.....C.....ACCG-GA	210
T79537	.GAG.....CAACGCG...ACCG-GACCT-	225
T79539	.GAG.....CAAGGAG...ACCG-GAC	227
T79540	.GAG.....CAAGACGTCCACCG-GAC	225
T79541	.GAG.....CAACGGG...ACCG-GA	219
T79543	.GAA.....CAAG.....ACAG-GACCT-CAGA---G-A----------	232
T79551	.GAG.....CAACGAG...ACCG-GA	227
T79557	.GAG.....CATGACCTCCACCG-GA	219
T79558	.GAG.....CAATGAG...ACCG-GA	208
T79561	.GAG.....CAACGGG...ACCG-GA	215
T79562	.GAG.....CAACGGG...ACCG-GA	227
T79565	.CAT.....GAACGAG...ACCG-GACCT-CAGA---G-A----------	238
T832080	GGA.....ATAAC...ACCG-GA	226
T832081	AGA.....ATGGG...ACCG-G	222
T832082	AGA.....ATGAG...ACCG-GA	226
T832268	GGA.....ATGAG...ACCG-GA	224
T832273	AGA.....ATGGG...ACCG-GA	214
T842281	AGA.....ACAGG...AACG-GACC	221
T842282	GGA.....ACGAG...A CG-GA	222
T842286	AGAA.....GAGGAACGGGACCG-GA	223
T842289	GGAC.....C...AAG...ACCG-GA	218
T842292	GGAC.....C...GAG...ACCG-GA	218
T842294	AGAA.....GAGGAACGAGACCG-GA	217
T842298	GGAC.....C...GAG...ACCG-GA	212
T842299	GGAC.....C...GAG...ACCG-GA	212





V4V5

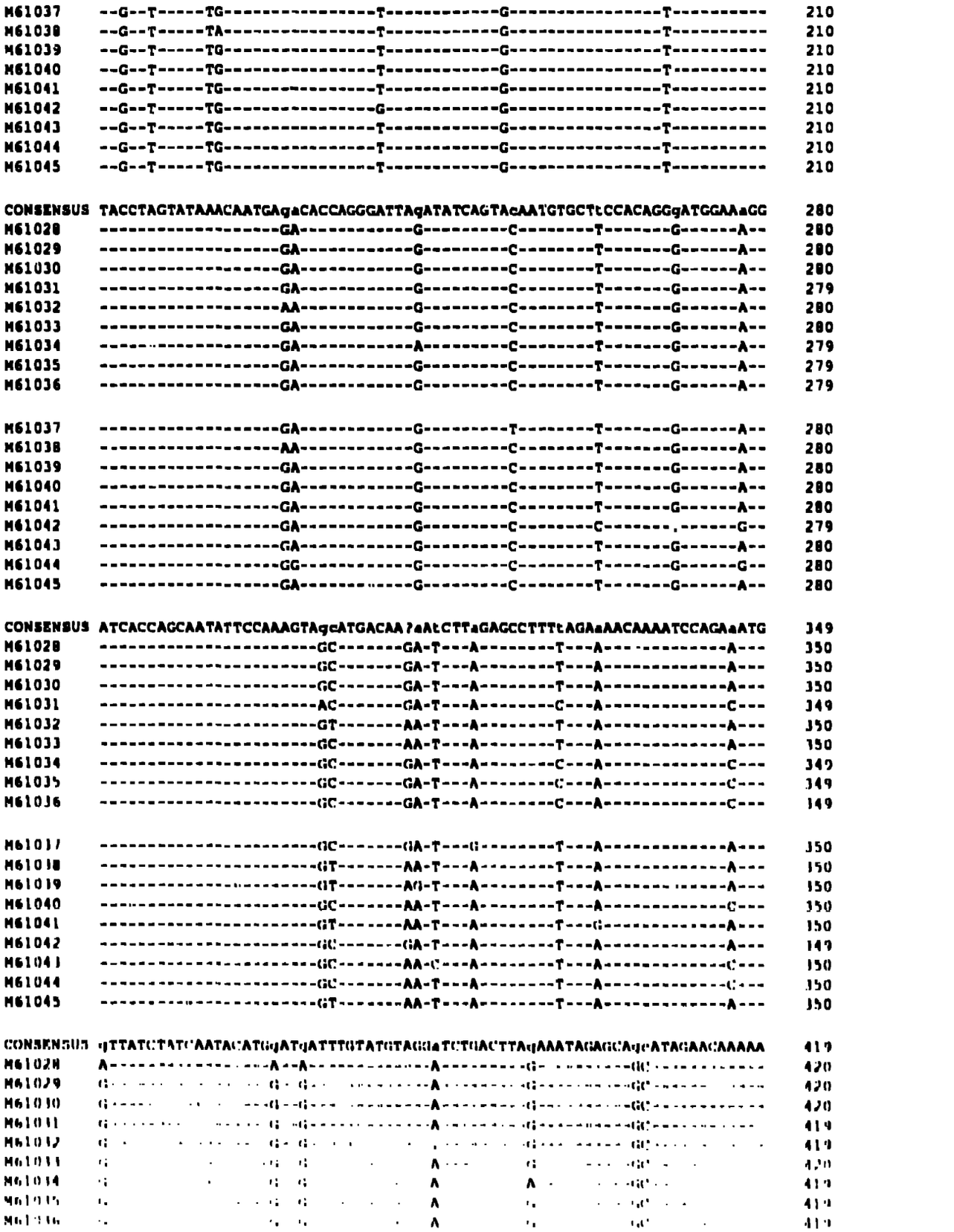
CONSENSUS	hga??????foaahga????aacgWgacctToagaCCTgCaCCACCACATATG	262
T872420	GCAA.....G...ACCCACCCCG-CACC	231
T872421	GCAA.....ACAT.....ACCG-CATCT-	230
T872422	ACAC.....CCACCCACATC	228
T872423	GCAA.....CAGC...CAGCCCG-CACCT-TA	235
T872426	GCAA.....CAAG.....ACCG-GCAGA-CTTC	231
T872427	GCAA.....CGTAGT....ACCG-GCAGA-CTTC	233
T872428	GC.....AACACCCAGCCCG-CACC	231
T91605	GC.....AACG-CACCT-CAGA---G-A----------	248
T91609	GC.....AACG-CACCT-CAGA---G-A----------	245
T91611	GC.....AGCAACG-CACCT-CAGA	236
T91613	GC.....AACG-CACCT-CAGA---G-A---	223
T91691	GC.....AGCAACG-CACCT-CAGA---G-A----------	247
T12832	ATGG.....CACAACCAGCACTG-GCTCT-CAGA---G-A----------	233
T12834	ATGA.....AACACATCAGCACTG-CATCT-TAGA---G-A----------	240
T12836	ATGG.....CACAACCAGCACTG-GCTCT-CAGA---G-A----------	231
T12837	ATGA.....CACCAACCATACTG-AACTT-CAGA---G-A----------	243
T12840	ATGA.....CACAACCATACTG-AATTT-CAGA---G-A----------	240
T74164	GCAC.....ACACAACAGCTCCG-CATCT-CAGA---	236
T74165	GCAC.....ACACAACAGCTCCG-CATCT-CAGA---	246
T74166	GCACCACAGAGAACAGCCACCACCG-CATCT-CAGA---G-A	239
T74167	GCACCACAGAGAACAGCCACCACCG-CATCT-CAGA---T-	238
T74168	GCACACAGA.....ACAGCACCG-CATCT-CAGA---G	246
T74179	.....CACCACCG-CATCT-CA	229
T822180	GTAACAAAGCAACCACACCACCG-CA	241
T822182	GTAACAAAGCAACCACACCACCG-CA	238
T82601	ACAGCA.....ACACCACCG-CATCT-CAGA	246
T82601	ACAAAA...GCAACCACACCACCG-CACCT-CAGA---G-G---	257
T82712	.....ACACCACCG-CACCT-CAGA---G-G---	245
T82713	.....ACACCACCG-CACCT-CAGA---G-	241
T82714	GTAACAAAGCAACCACACCACCG-CACCT-CAGA--	250
T82716	GTAACAAAGCAACCACACCACCG-CACCT-CAGA---G-	253
T82717	GTAACAAAGCAACCACACCACCG-CACCT-CAGA---G-G------	260
T82718	GTAACAAAGCAACCACACCACCG-CACCT-CAGA---G-G------	260

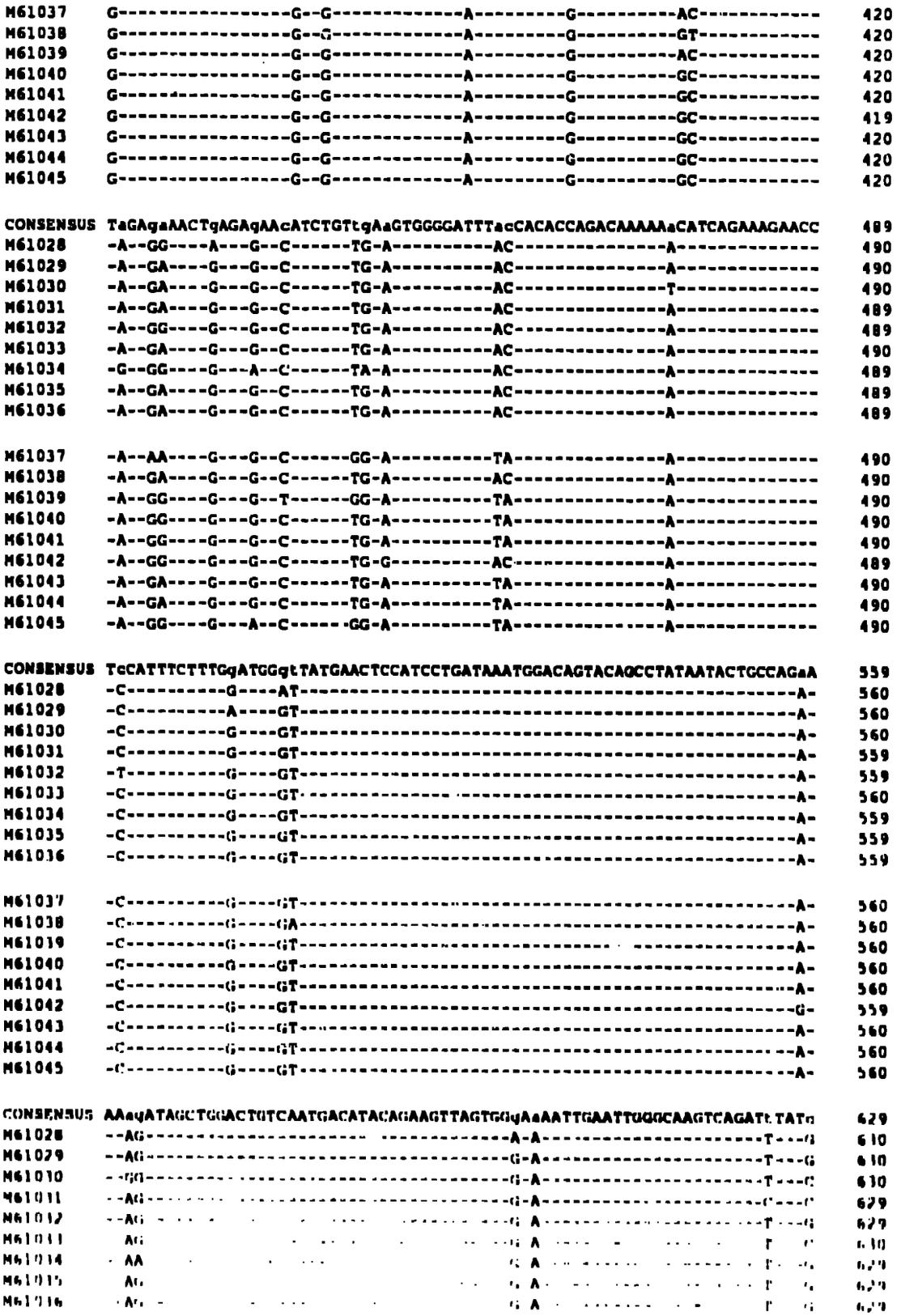
LOCUS HIVNJA 654 bp ss-RNA
DEFINITION Human immunodeficiency virus type 1, partial pol cds; samples from New Jersey patient before and after AZT treatment.
ACCESSION M61028 M61029 M61030 M61031 M61032 M61033 M61034 M61035 M61036 M61037 M61038 M61039 M61040 M61041 M61042 M61043 M61044 M61045
SOURCE Human immunodeficiency virus type 1 (HIV-1) direct amplification and sequencing of 25-year old male patient. Clone A14.
REFERENCE 1 (bases 1 to 654)
AUTHORS Fitzgibbon,J.E., Howell,R.M., Schwartzer,T.A., Gocke,D.J. and Dubin,D.T
TITLE In vivo prevalence of azidothymidine (AZT) resistance mutations in an AIDS patient before and after AZT therapy
JOURNAL AIDS Res. Hum. Retroviruses 7, 265-269 (1991)
COMMENT Blood samples were collected from a 25-year old male patient in 8/88, at the commencement of AZT therapy, and at 4/90. Pol sequences were PCR amplified directly from PBMCs and analyzed. The A14 clone (pre-AZT, 8/88) is shown below. Substitution of 'tac' for 'acc' at coordinate 459 below produces a Thr to Tyr change at residue 215, which is thought to be one of the four steps towards AZT resistance. (The other three crucial residues are intact after 19 months of therapy and the clinical status of the patient has not changed.)
BASE COUNT 272 a 105 c 121 g 156 t
ORIGIN At amino acid residue 63 of the RT cds.

```
        1 ccataaaaaa aaaagacagt accaaatgga aaaaattagt aaatttcaga gaacttaata
       61 aaagaactca agacttctgg gaagttcaat tagaaatacc acatcccgca gggttaaaaa
      121 agaaaaaatc agtaacagta ctggatgtgg gtgatgcata cttttcagtt cccttagata
      181 aggaattcag aaaatatact gcatttacca tacctagtat aaacaatgag acaccaggga
      241 ttagatatca gtacaatgtg cttccacagg gatggaaagg atcaccagca atattccaaa
      301 gtagcatgac aagaatctta gagcctttta gaaaacaaaa tccagaaatg attatctatc
      361 aatacatgaa taatttgtat gtaggatctg acttagaaat agagcagcat agaacaaaaa
      421 tagaggaact aagagaacat ctgttgaagt ggggatttac cacaccagac aaaaaacatc
      481 agaaagaacc tccatttctt tggatggatt atgaactcca tcctgataaa tggacagtac
      541 agcctataat actgccagaa aaagatagct ggactgtcaa tgacatacag aagttagtgg
      601 aaaaattgaa ttggcaagt cagatttatg caggaattca aataaaacag ttat
```

Fitzgibbon et al., AIDS Res. Hum. Retroviruses 7, 265, 1991. The first nine sequences are pre-AZT (M61028 is clone A14); the second nine are post-AZT.

```
CONSENSUS CcATAAAgAAAAAAGACAGTACtAAATGGAgAAAATTAGTAgATTTCAgAGAACTTAATAAAaGAACtCA  70
M61028    -C-----A--------------C--------A----------A------G-------------A----T--  70
M61029    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61030    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61031    -C-----A--------------T--------G----------G------G-------------A----T--  70
M61032    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61033    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61034    -C-----G--------------T--------G----------G------A-------------A----T--  70
M61035    -C-----G--------------T--------G----------G------G-------------.----A--  69
M61036    -C-----G--------------T--------G----------G------G-------------.----A--  69

M61037    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61038    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61039    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61040    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61041    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61042    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61043    -C-----G--------------T--------G----------G------G-------------A----T--  70
M61044    -T-----G--------------T--------G----------G------G-------------A----T--  70
M61045    -C-----G--------------T--------G----------G------G-------------A----T--  70

CONSENSUS aGACTTCTGGGAAGTTCAATTAGgAATACCACATCCCGCAGGGTTAAAAAAgaAAAAaTCAGTAACAGTA  140
M61028    A----------------------A-------------------------GA----A--------------  140
M61029    A----------------------G-------------------------GA----A--------------  140
M61030    A----------------------G-------------------------GA----A--------------  140
M61031    G----------------------G-------------------------AA----.--------------  139
M61032    A----------------------G-------------------------GA----A--------------  140
M61033    A----------------------G-------------------------GA----A--------------  140
M61034    A----------------------G-------------------------AA----.--------------  139
M61035    A----------------------G-------------------------GA----A--------------  139
M61036    A----------------------G-------------------------GA----A--------------  139

M61037    A----------------------G-------------------------GA----A--------------  140
M61038    A----------------------G-------------------------GA----A--------------  140
M61039    A----------------------G-------------------------GA----A--------------  140
M61040    A----------------------G-------------------------GA----A--------------  140
M61041    A----------------------G-------------------------GA----A--------------  140
M61042    A----------------------G-------------------------GG----A--------------  140
M61043    A----------------------G-------------------------GA----A--------------  140
M61044    A---------------------- G-------------------------GA----A--------------  140
M61045    A----------------------G-------------------------GA----A--------------  140

CONSENSUS CTgGAtGTGGGtgATGCATACTTTTCAGTtCCCTTAGATAAGgAATTCAGAAAATATACtGCATTTACCA  210
M61028    --G--T-----TG----------------T------------G----------------T----------  210
M61029    --G--T-----TG----------------T------------G----------------T----------  210
M61030    --G--T-----TG----------------T------------G----------------T----------  210
M61031    --G--T-----TG----------------T------------G----------------T----------  209
M61032    --G--T-----TG----------------T------------G----------------T----------  210
M61033    --G--C-----TG----------------T------------G----------------T----------  210
M61034    --G--T-----TG----------------T------------A----------------C----------  209
M61035    --A--T-----TG----------------T------------G----------------T----------  209
M61036    --A--T-----TG----------------T------------G----------------T----------  209
```





```
M61037    --AG-----------------------------------G-A--------------------T---C    630
M61038    --AG-----------------------------------G-A--------------------T---C    630
M61039    --AG-----------------------------------G-G--------------------T---C    630
M61040    --AG-----------------------------------G-A--------------------T---C    630
M61041    --AG-----------------------------------G-A--------------------T---C    630
M61042    --AG-----------------------------------A-A--------------------T---C    629
M61043    --AG-----------------------------------G-A--------------------T---C    630
M61044    --AG-----------------------------------G-A--------------------T---C    630
M61045    --AG-----------------------------------G-A--------------------T---G    630

CONSENSUS CAGGAATTaAAaTAAaACAGTTAT1    653
M61028    --------C--A---A--------    654
M61029    --------A--A---G--------    654
M61030    --------A--A---A--------    654
M61031    --------A--A---A--------    653
M61032    --------A--A---G--------    653
M61033    --------A--A---A--------    654
M61034    --------C--A---A--------    653
M61035    --------A--A---A--------    653
M61036    --------A--A---A--------    653

M61037    --------A--A---G--------    654
M61038    --------A--A---G--------    654
M61039    --------A--A---A--------    654
M61040    --------A--A---G--------    654
M61041    --------A--A---A--------    654
M61042    --------A--A---A--------    653
M61043    --------A--A---G--------    654
M61044    --------A--A---A--------    654
M61045    --------A--G---G--------    654
```

```
LOCUS       HIVU455      9178 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate U455 from a Ugandan
            person (1985); complete genome.
ACCESSION   M62320
SOURCE      Human immunodeficiency virus type 1 proviral DNA taken from PBLs.
REFERENCE   1  (bases 1 to 9178)
  AUTHORS   Oram,J.D., Downing,R.G., Roff,M., Clegg,J.C.S., Serwadda,D. and
            Carswell,J.W.
  TITLE     Nucleotide sequence of a Ugandan HIV-1 provirus reveals genetic
            diversity from other HIV-1 isolates
  JOURNAL   AIDS Res. Hum. Retroviruses 6, 1073-1078 (1990)
COMMENT     Hardcopy of sequence [1] kindly provided by J.D. Oram, 2/5/91.
            The U455 virus was isolated in 1985 from a 35-year old Ugandan
            male. The sequenced clone has in-phase stops in the vpR and env
            coding sequences and a frameshift, generating a stop codon, in the
            vpU cds. While this virus is related to other African HIV-1 forms
            ( most closely to Z321; see Part III), the serology of the gag
            protein, for example, is unusual according to [1].
BASE COUNT     3265 a   1636 c   2208 g   2069 t
ORIGIN
        1 aagcttgcct tgagtgcttg aagtagtgtg tgcccgtctg ttatttgact ctggtaacta
 R rpt ->                     <- R end
       61 gagatccctc agaccactat agactgtgta aaaatctcta gcagtggcgc ccgaacaggg
      121 actcgaaagc gaaagttcca gagaagttct ctcgacgcag gactcggctt gctgaggtgc
      181 acacagcaag aggcgagagc ggcgactggt gagtacgcct aaaatttttt gactagcgga
      241 ggctagaagg agagaaATGg gtgcgagagc gtcagtatta agcgggaaaa aattagattc
                 gag start ->
      301 atgggagaaa attcggttaa ggccaggggg aaacaaaaaa tatagactga aacatttagt
      361 atgggcaagc agggagctgg aaaaattcac acttaaccct ggccttttag aaacagcaga
      421 aggatgtcag caaatactgg gacaattaca accagctctc cagacaggaa cagaagaact
      481 tagatcatta tataatacag tagcagtcct ctattgtgta catcaaagga tagatgtaaa
      541 agacaccaag gaagctttaa ataaaataga ggaaatgcaa aataagaaca agcaaaggac
      601 acagcaggca gcagctaaca caggaagcag tcaaaattac cccatagtgc aaaatgcaca
      661 agggcaacca gtacaccagg ccttatcacc taggaccttg aatgcatgcg tgaaagtagt
      721 agaagacaag gctttcagcc cagaagtaat acccatgttt tcagcattat cagagggagc
      781 caccccacaa gatttaaata tgatgctgaa tgtagtgggg ggacaccagg cagctatgca
      841 aatgttaaaa gataccatca atgaggaagc tgcagagtgg gacaggttac atccagtgca
      901 tgcagggcct attccaccag gccagatgag agaaccaagg ggaagtgaca tagcaggaac
      961 tactagcacc gttcaagaac aaataggatg gatgacaggc aatccaccta tcccagtggg
     1021 agacatctat agaagatgga taatcctggg attaaataaa atagtaagaa tgtatagccc
     1081 tgttagcatt ttggacataa gacaagggcc aaaagaaccc ttcagggatt atgtagatag
     1141 attctttaaa actctcagag ctgagcaagc tacacaggat gtaaaaaact ggatgacaga
     1201 aaccttgctg gtccaaaatg cgaatccaga ctgtaagtcc attttaagag cattagggcc
     1261 aggggctaca ttagaagaaa tgatgacagc atgccaggga gtgggaggac ccggccataa
     1321 agcaagggtt ttggctgagg caatgagtca agtacaacag acaagcataa tgatgcagag
     1381 aggcaatttt aggggcccga gaagaattaa gtgtttcaac tgtggcaaag aaggacacct
     1441 agccaaaaat tgtagggccc ctaggaaaaa gggctgttgg aaatgcggga aagaaggaca
     1501 ccaaatgaaa gactgcactg agagacaggc TAAtttttta gggaaaattt ggccttccaa
                                pol orf start ->
     1561 caaggggagg ccagggaatt ttcctcagag cagaccagag ccaacagccc caccagcaga
     1621 aatctttggg atgggggaaa agatgacctc ccctgcgaaa caggagctga aagacaggga
     1681 acagactcct ttagtttccc tcaaatcact ctttggcaac gacccctgt cacagTAAaa
                                                                    <- gag
     1741 ataggaggac agctgataga agctctatta gatacaggag cagatgatac agtcttagaa
     1801 gacataaatt tgccaggaaa atggaaacca aaaataatag ggggaattgg aggttttatc
     1861 aaagtaagac agtatgatca gatacttata gaaatttgtg gaaaaaagac tataggtaca
     1921 gtattggtag gacctacacc tgtcaacata attggaagga atatgttgac tcagattggt
     1981 tgtactttaa attttccaat tagtcctatt gaaactgtac cagtaaaatt aaaaccagaa
     2041 atggatggcc caaaggttaa acaatggcca ttgacagaag agaaaataaa agcattaaca
     2101 gaaatttgta atgagatgga aaaggaagga aaaatttcaa aaattgggcc tgaaaatcca
     2161 tacaatactc cagtatttgc tataaagaaa aaggacagca ctaaatggag gaaattagta
     2221 gatttcagag aactcaataa aagaactcag gacttctggg aagttcaatt aggaatcccg
     2281 catacagcgg gtctaaaaaa gaaaaaatca gtaacagtac tagatgtggg ggacgcatat
     2341 ttttcagttc ctttagatga aagctttaga aagtatactg cgttcaccat acctagtata
```

```
2401 aacaatgaga caccaggagt caggtatcag tacaatgtgc ttccgcaggg atggaaagga
2461 tcaccatcaa tattccagag tagcatgaca aaaatcttag agccctttag atcacaacat
2521 ccagacatag ttatctatca atacatggat gacttgtatg tgggatctga tttagaaata
2581 gggcaacata gagcaaaaat agaggaatta agagctcatc tattgagctg gggattcatt
2641 accccagaca aaaagcatca gaaagaacct ccatttctrt ggatggggta tgaacttcat
2701 cctgacaaat ggacagttca gcctatacag ctgccagaaa aagacagctg gactgtcaat
2761 gatatacaga aattagtagg aaaactaaat tgggcaagtc agatttatgc agggattaaa
2821 gtaaagcaac tgtgtaaact tctcagggga gccaaagcac taacagatat agtaaccctg
2881 actgaagaag cagaattaga attggcagag aacagggaga ttctaaaaga ccctgtgcat
2941 ggagtatatt atgacccatc aaaagaccta gtagcagaaa tacagaaaca agggcaagac
3001 caatggacat atcaaattta tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca
3061 agaaaaaggt ctgctcacac taatgatgta aaacaattaa cagaagtggt gcaaaaagtg
3121 tccacagaaa gcatagtaat atgggaaaag atccctaaat ttagactacc catacaaaag
3181 gaaacatggg aagcatggtg gatggagtat tggcaggcta cctggattcc tgaatgggag
3241 tttgtcaata cccctcccct agtaaaatta tggtaccagt tagagaaaga ccccatagca
3301 ggagcagaga cattctatgt agatggggca gctaataggg aaactaagct aggaaaagca
3361 gggtatgtca ctgacagagg aagacaaaag gttgtttccc taactgagac aacaaatcaa
3421 aagactgaat tacatgcaat ccatctagcc ttgcaggatt caggatcaga agtaaatata
3481 gtaacagact cacagtatgc attagggatc attcaggcac aaccagacag gagtgaatca
3541 gaaatagtca atcaaataat agagaaagcta atagaaaagg aaaaagtcta cctgtcatgg
3601 gtaccagcgc acaaagggat tggaggaaat gagcaagtag ataaattagt cagttctgga
3661 atcaggaagg tgctattttt agatgggata gataaggctc aagaggacca tgaaaaatat
3721 cactgcaact ggagagcaat ggctagtgat tttaatctgc cacctgtggt agcgaaggaa
3781 atagtagcca gctgtaataa atgtcaacta aaaggggaag ccatgcatgg acaagtagac
3841 tgtagtccag ggatatggca attagattgc acacatctag aaggaaaagt aatcctagta
3901 gcagtccatg tagccagtgg ctacatagaa gcagaagtta tcccagcaga aacaggacag
3961 gagacagcat actttatact aaaattagca ggaagatggc cagtaaaagt aatacacaca
4021 gacaatggca gcaatttcac cagcgctgca gttaaagcag tctgttggtg ggcaaatatc
4081 caacaggaat ttgggatccc ctacaatccc caaagtcaag gagtagtgga atccatgaac
4141 aaggaattaa agaaaatcat agggcaggta agagagcaag ctgaacacct taagacagca
4201 gtacaaatgg cagtattcat tcacaatttt aaaagaaaag gggggattgg gggatacagt
4261 gcaggggaaa gaataataga cataatagca acagacatac aaactaaaga attacaaaaa
4321 caaatttcaa aaattcaaaa ttttcgggtt tattacaggg acagcagaga ccccatttgg
4381 aaaggaccag caaaactact ctggaaaggt gaaggggcag tagtaatata agacaatagt
4441 gatataaagg tagtaccaag aagaaaagca aagatcatta gggattATGg aaaacagatg
                                                vif start ->
4501 gcaggtgatg attgtatggc aggtagacag gatgaagatT AGaacatgga acagtttagt
                                                <- pol
4561 aaaacatcac atgtatgtct ccaagaaagc tcaaygttgg ttttatagac atcactatga
4621 aagtagacat tcaagagtaa gttcagaagt acacatccca ctaggggaag ctagattagt
4681 agtaagaaca tatggggtc tgcatacagg agaaaaagac tggcacttgg gtcatggggt
4741 ctccatagaa tggaggctga aaagatatag cacacaagta gatcctgacc tagcagacca
4801 tctaattcac ctgcattatt ttgactgttt ttcagaatct gccataagga gagccatatt
4861 aggacaaata gttagacctg ggtgtgaata tcaagcagga cataacaagg taggatctct
4921 acaatatttg gctctgaaag cattagtaac cccaacaagg gcaaagccac ctttgcctag
4981 tgttaagaaa ttaacagagg atagATGgaa caagccccag aagaccaggg gccacagagg
                          vpR ->
5041 gagccgtacg ctgaaTAGgc atTAGagttg ttagaaagag ttaagcatga agatgttaga
                     <- vpR premature stop
                           <- vif
5101 cattttccta ggcagtggct ccatggatta ggacaacata tctataacac ctatggggat
5161 acttgggagg gagttgaagc tataataaga attttgcaac aattactgtt tatccatttc
5221 agaattgggt gccaacatag cagaataggc attattccag ggagaagagg caggaATGga
                                                      tat start ->
5281 gccagtagat ccTAAcctag agccctgcaa acacccggga agtcagccta caactgcttg
                  <- vpR normal stop
5341 tagcaactgt tattgtaaag tgtgttgctg gcattgccaa ctttgctttc tgaaaaaagg
5401 cttaggcatt tcctATGgca agaagaagcg gaaaccccga cgaggaccta ctcaaggcag
         rev start ->
5461 taaggatcat caaactctta taccaaagca GTaagtaatt agcttatgta ATGacacctt
                              5' sj /\      vpU(pseudo-rdy) ->
5521 tggaaatctg gccaataaca gggctgataa tagcactaat cttagcaata gtagtgtgga
5581 ctatagtagg tatagaaTAA aaaaaaaatt gctaaagcaa aaaaatatag acaggttatt
                        <- vpU normal stop
```

```
5641 aaacagaata agagaaagag cagaagacag tggcaATGag agtgatgggg atacagagga
                              env start ->
5701 attatccttg cttgtggaga tggggaacta tgatcttggg gttgataata atttgtaatg
5761 cacaacaatt gtgggtcacg gtctzctatg gggtacctgt gtggaaagat gcagttacca
5821 ccctattttg tgcatcagat gctaaagcat atgatgcaga agtgcataat gtctgggcta
5881 cacatgcctg tgtacccaca gaccccaacc cacaagaaat agatttggtc aatgtgacag
5941 aagaatttaa catgtggaaa aataacatgg tagatcagat gcatgaagat ataatcagtc
6001 tatgggacca aagcctaaag ccatgtgtaa agttaacccc tctctgcgtr acgttagatt
6061 gccataacat caccattaac aacaccaata acaacaccaa tatcacagat ggagtgaggg
6121 aagaaatgaa aaattgctct ttcaatatga ccacagaact aagagataag aaacagaaag
6181 tatattcact tttttataga cttgatatag tacaaatcaa taaaactgat aacaatagtt
6241 atagactaat aaattgtaat acctcaacca ttacacaggc ttgtcccaag gtgtcctttg
6301 agccaattcc catacattat tgtgccccag ctggttttgc gattctaaag tgtaaggatc
6361 ctgagttcaa tggaaaaggg ccatgcagga atgtcagcac agtacaatgt acacatggaa
6421 tcaagccagt ggtatcaact caactgctgt taaatggcag tctagcagaa agagaaataa
6481 ggattagatc tgaaaatttc acaaacaatg caaaaaccat aatagtacag cttgtcaatc
6541 ctgtaaaaat caattgttcc agaccttaca atacaagaaa aaatatacgt aggtatagta
6601 taggatcagg acaagcattc tatgtaacag gtaaaataat aggggatata agacaagcac
6661 attgtaatgt cagtagaagg gactggaata gaactataca acaggtagct gaacaattaa
6721 agaaaaagtt taataacaaa acaataatct ttgctagctc ctcaggaggg gatatagaaa
6781 ttacaacaca tagctttaat tgtggaggag aatttttcta ttgcaacaca tcaggcctgt
6841 ttaatagcat ttggaatggc agcatgtcaa atgacatggg gccaaatggc actataactc
6901 tccaatgtag aataaagcaa attataaata tgtggcagag agtaggacaa gcaatgtatg
6961 cccctcccat ccaaggagta ataaggtgtg aatcaaacat tacaggacta ctattaacaa
7021 gagatgyggg tactaacaat acaaaaaatg agaccttcag acctggagga ggagatatga
7081 gggacaattg gaaaagtgaa ttatataagt ataaagtagt aaaaattgaa ccactaggag
7141 tagcacccac cagggcaaag agaagagtgg tggagagaga aaaaagagca gttggactag
7201 gagctatctt ccttgggttc ttaggagcag ctggaagcac aatgggcgcg gcgtcaataa
7261 cgctgacggt acaggccaga caattattgt ctggtatagt gcaacagcag agcaatctgc
7321 tgagggctat agaggctcaa cagcatctgt tgaaactcac tgtctggggc attaaacagc
7381 tccaggcaag agtcctggct gtggaaagat acctaccgga tcaacagctc ctaggaattt
7441 ggggctgctc tggaaagctc atctgcacca ctactgtgcc ctggaactct agttggagta
7501 ataaatctca ggaggacata tggaataaca tgacctggct gcaatgggag aaagaaatta
7561 gtagttacac aggcataata tatcaactaa ttgaggaatc gcagaaccag caggaaaaga
7621 atgaactaga cttattggca ttggacaagt gggcaaatct gTGAaattgg tttaacatct
                                                  <- premature env stop
7681 caaattggct gtggtatatt agactatttg taataatagt aggaggctta ataggattaa
7741 gaatagtttt tactgtgctt tctataatca atagagttag gcagggatac tcacctctgt
7801 cgtttcAGac ccttgcccca atcccagagg gtctcggcag gccaggaaga atcgaagaag
         /\ 3' sj
7861 aaggtggaga gcaaggcaaa gacagatcga ttcgctTAGt gagcggattc ttagcgattg
                                        <- tat
7921 cctgggacga cctgcggaac ctgtgcctct tcagctacca ccgattgaga gacttcgcct
7981 tgattgtagc gagagctgtg gaacttctgg gacgcagcag cctcaaggga ctgagactgg
8041 ggtgggaggg cctcaaatat ctgtggaatc ttctgctgta ttggggtcgg gaactaaaaa
8101 ttagtgctat tactttgctt gacgctgtag cagtagcagt agctgggtgg atagataggg
8161 ttatagaaat aggacaaaca attggtagag ctattctcaa catacctaga agaatcagac
8221 aaggcttaga aagggctttg ctaTAAcATG ggtggcaagt ggtcaaaaaa gagcngagtg
                              <- normal env end
                             nef ->
8281 gaatggcctg aggttaggaa aagaatgaga gaaactcctg cagcagcaaa aggagtagga
8341 gcagtatctc aagatttaga taaatatgga gcagtcacaa gcagtaatac atcatctact
8401 aatgccagtt gtgcctggct ggaagcgcaa gaggaaggag acgtaggctt cccagttagg
8461 ccacaggtac ctttaagacc aatgacttat aaggcagctt ttgatctcag cttcttttta
8521 aaagaaaagg ggggactgga tgggttaatt cactcacaga taagacaaga aatccttgat
                    3'LTR ->
8581 ctgtgggtct atcacacaca aggattcttc cctgattggc agaattacac accagggcca
8641 ggaatcagat acccactaac attgggatgg tgctacaagc tagtaccagt tgatccagct
8701 gaagtagagg aagctactgg aggagagaac aaccagtcta ctacccctat atgccaacat
8761 ggagtagata tgtaaagaga agaaagtgtta atgtggaagt ttgacagtac cctggcatta
8821 aaacacagag cttatgagct gcatccggag ttctataaag acTGAaagac tgctgacaca
                                                  <- nef end
8881 gaagttgctg acagggactt tccactgggg actttccgag ggaggtgtgg ttggggcgga
8941 gttggggagt ggctagccct cagatgctgc aTATAAGcag ctgctttttg cctgtactGG
```

```
                                signal ->                    R start ->
9001 Gtctctcttg ttagaccaga tcgagcctgg gagctctctg gctagcgagg gaacccactg
9061 cttaagcctc AATAAAgctt gccttgaggt gcttgaagta gtgtgtgccc gtctgttatt
        signal ->                          <- R end
9121 tgactctg   aactagagat ccctcagacc actatagact gtgtaaaaat ctctagca
```

HIVSF1701

```
LOCUS       HIVSF1701     537 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 1, isolate 170 from Rwanda,
            clone 50 LTR
ACCESSION   M66534
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate 170, clone
            50 (biologically active)
SEGMENT     1 of 2
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Evans,L., Higgins,D., Cheng-Mayer,C. and Levy,J.
  JOURNAL   Unpublished (1990)
REFERENCE   2  (sites)
  AUTHORS   Cheng-Mayer,C., Homsy,J., Evans,L.A. and Levy,J.A.
  TITLE     Identification of human immunodeficiency virus subtypes with
            distinct patterns of sensitivity to serum neutralization
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 2815-2819 (1988)
COMMENT     Hardcopy and electronic copy of sequence kindly provided by
            D.Higgins, Chiron Corporation, 415-601-2976. SF170 was isolated
            in Dr. Jay Levy's lab (U.C.S.F.) from a Rwandan sample. The isolate
            exhibits "macrophage tropism" and does not grow in established T-
            cell lines. The env sequence is presented in <HIVSF1702>.
BASE COUNT      153 a    118 c    145 g    121 t
ORIGIN      5' terminus of LTR.
        1 ctggatgggt taatttactc caggagaaga caagaaatcc ttgatctgtg ggtctacaac
      LTR ->
       61 acacaaggct acttccctga ttggcagaat tacacaccag ggccagggac cagattccca
      121 ctaacatttg gatggtgctt caagctagta ccagttgatc cagagggagt agagaaggac
      181 acggtgggag agaccaacag cctattacac cctatatgcc aacatggaat ggatgatgag
      241 gaaagagaaa cattaatgtg gaagtttgac agccgcctgg cattaaaaca cagagcccaa
      301 gagatgcatc cggagttcta caaagactgc tgacacaaaa gttgctgaca gagactttcc
      361 gctggggact ttccagggga ggtgtggttt gggcggagtt ggggagtggc taaccctcag
      421 atgctgcaTA TAAGcagctg cttttcgcct gtactgggtc tctcttgtta gaccagatag
           signal ->
      481 agtctgagaa ctctctggct agcgagggaa cccactgctt aagcctcaat aaagctt
```

```
LOCUS       HIVSF1702     2607 bp ds-DNA
DEFINITION  Human immunodeficiency virus type 1, isolate sf170 from Rwanda;
            env cds.
ACCESSION   M66533
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate sf170,
            proviral DNA, lambda clone 50.
SEGMENT     2 of 2
REFERENCE   1  (bases 1 to 2607)
  AUTHORS   Evans,L., Higgins,D., Cheng-Mayer,C. and Levy,J.
  JOURNAL   Unpublished (1990)
REFERENCE   2  (sites)
  AUTHORS   Cheng-Mayer,C., Homsy,J., Evans,L.A. and Levy,J.A.
  TITLE     Identification of human immunodeficiency virus subtypes with
            distinct patterns of sensitivity to serum neutralization
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 2815-2819 (1988)
COMMENT     Hardcopy and electronic copy of sequence kindly provided by
            D.Higgins., Chiron Corporation, 415-601-2976. SF170 was isolated
            in Dr. Jay Levy's lab (U.C.S.F.) from a Rwandan sample.  The
            isolate exhibits "macrophage tropism" and does not grow in estab-
            lished T-cell lines. Lambda clone #50 is biologically active.
            SF170 is divergent when compared to Zairean HIV-1s but not so
            divergent as other African viral isolates. The entry <HIVSF1701>
            provides the LTR sequence from this viral isolate.
BASE COUNT      915 a    463 c    603 g    626 t
ORIGIN      Start of env cds.
        1 ATGagagtga tggggataca gatgaactgt caaaacttat tgagatgggg aactatgatc
env start ->
       61 ttgggaatgt taataatctg tagtgctaca agcaaattgt gggttactgt ctattatggg
      121 gtacctgtgt ggaaagacgc agagaccacc ttattttgtg catcagatgc taaagcatat
      181 gagagagaag tgcataatgt ctgggcgaca catgcctgtg tacccacaga ccccgaccca
      241 caagaaatat atttggaaaa tgtgacagaa gggtttaaca tgtggaaaaa taacatggta
      301 gagcagatgc atacagatat aatcagtcta tgggaccaga gtctaaagcc atgtgtaaag
      361 ttaacccctc tctgcgttac tttaaattgt agccataaca tcaccaccca taacatcacc
      421 accaacaacc ccaatatcac caaatacatg gaaggagaaa taaagaactg ctcttacaat
      481 atgaccacag aactaaggga taagaaacag aaagtgtatt cactttttta taaacttgat
      541 gtagtaccaa ttgataaaaa taatggtaat aacaatataa ggactcagta tagattaata
      601 aattgtaata cctcagccat tacacaggct tgtccaaaag tatcctttga gccaattccc
      661 atacattatt gtgctccagc tggttttgcg attctaaagt gtaatgatgc agagttcaat
      721 ggaacagggc catgcaagaa tgtcagcaca gtacaatgca cacatggaat caggccagta
      781 atatcaaccc aactgctgtt aaatggcagt ctagcagaag gaagggtaaa gattagatct
      841 gaaaatatca caaacaatgc caaaaccata atagtacaac ttaacaagac tgtagaaatt
      901 aattgtacca gacctaacaa caatacaaga aaaagtgtac gtataggacc aggacaagca
      961 ttctatgcaa caggtgacat aataggggat ataagacaag catattgtaa tgtcagtaga
     1021 gcagactgga ataaaacttt acaaggggta gccaaccaat taaaaagtta ctttagtaac
     1081 aaaacaataa tctttgctag ctcctcagga ggggatttag aaattacaac acatagtttt
     1141 aattgtggag gagaattttt ttattgtgat acatcaggcc tgtttaatag cacatggggg
     1201 ggaaatagca ctgacagcat acaggagtca aatagcacag agtcaaatga cactataatt
     1261 ctcccatgca gaataaagca aattataaat atgtggcaga gagtaggaca agcaatgtat
     1321 gcccctccca tccaaggagt aataaagtgt atatcaaata ttacaggact tctattaaca
     1381 agagatggtg gggataataa cagtgcaaat gaaaccttca gacctacagg aggaaatatg
     1441 agggacaatt ggagaagtga attatataag tataaagtag taaaaattga accactagga
     1501 gtagcaccca ccccggcaaa gagaagagtg gtgcagagag aaaaaagagc agttggaata
     1561 ggagctgtat tcattgggtt cttaggagca gccggaagca ctatgggcgc ggcgtcaata
     1621 acgctgacgg tacaggctag acaattattg tctggcatag tgcaacagca aagcaatttg
     1681 ctgagggcta tagaggctca acaacatctg ttgaaactca cggtctgggg cattaaacag
     1741 ctccaggcaa gagtcctggc tgtggaaaga tacctaaagg atcaacagct cctaggaatt
     1801 tggggctgct ctggaaaact catctgcacc actaatgtgc cctggaactc tagttggagt
     1861 aatagaactc agagtgagat atggaataac atgacctggc tgcaatggga taaagaaatt
     1921 agcaattaca cagacataat atataatcta attgaagaat cgcagatcca gcaggaaaag
     1981 aatgaacaag aattattggc attggacaag tgggcaaatc tgtggaattg gtttgacata
     2041 tcaaagtggc tgtggtatat aaggatattt ataatgatag taggaggctt aataggatta
     2101 agaatagttt ttgctgtgct ttctataata aatagagtta ggcagggata ctcacctctg
     2161 tcatttcaga tccataccccc aaacccaagg ggtcccgaca ggcccgaaag aatcgaagaa
     2221 gaaggtggag agcaaggcag agacagatcg attcgattag tgagcggatt cttagcactt
```

```
2281 gcctgggacg acctgcggag cctgtgcctc ttcagctacc accgcttgag agacttcatc
2341 ttgattgcag cgaggaccgt ggaacttctg aaacgcagca gtctcagggg gctgagactg
2401 gggtgggaag gcctcaagta cctgtggaat cttctgctat attggggtcg ggaactaaaa
2461 aatagtgcta ttaatttgct tgataccata gcaatagcag tagctggctg gacagatagg
2521 gttatagaaa taggacaggc aatttgtaga gctattctca acatacctag aagaatcaga
2581 cagggcttcg aaagggcttt gttaTAA
                              <- env
```

```
LOCUS       HIVCBL41      180 bp ss-RNA
DEFINITION  Human immunodeficiency type 1, isolate CBL-4 (RUT), partial ltr.
ACCESSION   M63474
SEGMENT     1 of 2
REFERENCE   1  (bases 1 to 180)
  AUTHORS   Goudsmit,J., Geelen,J., Keulen,W., Notermans,D., Kuiken,C.,
            Ramautarsing,C., Smit,L., Koole,L., van Genderen,M., Buck,H.,
            Sninsky,J. and Krone,W.
  TITLE     Characterization of the African HIV-1 isolate CBL-4 (RUT) by
            partial sequence analysis and virus neutralization with peptide
            antibody and antisense phosphate-methylated DNA
  JOURNAL   AIDS 4, 559-564 (1990)
COMMENT     This isolate is from a Tanzanian person.
BASE COUNT      36 a     42 c     57 g     45 t
ORIGIN      Approximately 123nt upstream from the mRNA start.
        1 gacaccgagt tttctacaag ggactttccg ctggggactt tccagggagg agtaacgggg
       61 gcgggactgg gagtggctaa ccctcagatg ctgcaTATAA Gcagctgctt tttgcctgta
                                                 signal ->
      121 ctgGGtctct ctggttagac cagatttgag cctgagagct ctctggctag ctagggaacc
  mRNA start ->


LOCUS       HIVCBL42      120 bp ss-RNA
DEFINITION  Human immunodeficiency type 1, isolate CBL-4 (RUT), env v3 region.
ACCESSION   M63475
SEGMENT     2 of 2
REFERENCE   1  (bases 1 to 120)
  AUTHORS   Goudsmit,J., Geelen,J., Keulen,W., Notermans,D., Kuiken,C.,
            Ramautarsing,C., Smit,L., Koole,L., van Genderen,M., Buck,H.,
            Sninsky,J. and Krone,W.
  TITLE     Characterization of the African HIV-1 isolate CBL-4 (RUT) by
            partial sequence analysis and virus neutralization with peptide
            antibody and antisense phosphate-methylated DNA
  JOURNAL   AIDS 4, 559-564 (1990)
COMMENT     This isolate is from a Tanzanian person.
BASE COUNT      55 a     21 c     24 g     20 t
ORIGIN      In the env V3 cds.
        1 tgcacaaggc cctactacaa aaatataaga aaaggtacac atgtaggaca tgtaggacca
       61 ggacgagcaa tctatacaac aaacataaaa ggaagtggaa atataagaca agcacattgt
```

Splice sites have been determined for SIV only.

```
 nef cds ->
CONSENSUS  TGgAaGGGaT?TaTTAcAgTgagagaAGACatAgAaTctTagAcataTactTaGAaAAgGAaGAaGG?AT    68
HIV2ROD    --G-A---A-G-T---C-G-GAAAGA----AT-A-A-CT-AA-TATA-ACT-A--A--G--A--A--G--    70
HIV2NIHZ   --G-A---C-G-T---T-G-CAGAGA----AT-G-A-CT-AG-CTTA-ACT-A--C--G..........    59
HIV2ISY    --G-A---A-G-A---C-G-GAGAGA----AT-G-A-CC-AG-CATA-ACT-A--A--G--A--A--G--    70
HIV2ST     --G-A---C-G-A---C-G-GATAGG----GT-G-G-CC-AG-CATA-ACT-A--A--G--A--G--A--    70
HIV2BEN    --C-A---A-G-T---C-G-AGGAGG----AT-G-A-CC-AG-CATA-ACC-A--A--A--G--A--G--    70
HIV2D194   --G-A---A-A-T---C-G-AGGGAG----AT-G-A-CC-AG-CTTG-TCC-A--A--G--G--A--G--    70
HIV2GH1    --G-T---A-G-T---C-G-AGGGAT----AT-G-A-CC-AG-CTTG-ACC-A--A--G--G--A--G--    70
HIV2D205   --G-A---A-G-A---T-G-GAGAGA----AC-G-A-AT-AG-CACA-ATT-T--G--T--A--A--C--    70
SIVMM251   --G-A---A-T-A---C-G-GCAAGA----AT-G-A-CT-AG-CATG-ACT-A--A--G--A--A--C--    70
SIVMM142   --G-A---A-T-A---C-G-GCAAGA----AT-G-A-CT-AG-CATA-ACT-A--A--G--A--A--C--    70
SIVMM239   --G-A---A-T-A---C-G-GCAAGA----AT-G-A-CT-AG-CATA-ACT-A--A--G--A--A--C--    70
SIVMNE     --G-A---A-T-A---C-G-GAAAGA----AT-A-A-CT-AG-CATG-ACT-A--A--G--A--A--C--    70
SIVSMMH4   --G-A---A-T-A---C-A-GAGAGA----AT-G-A-AT-AG-TATG-ACT-G--A--G--A--A--A--    70
SIVSMMPBJ  --G-A---A-T-A---C-G-GATAGA----AT-A-A-AT-AG-TCTG-ACT-A--A--G--A--A--A--    70

CONSENSUS  aaTaccaGatTGGCAgaatTAcA?ctCa?GGgCcAGGaataAGatAcCcaA?gt?cTttGGgTGGcTaTG   134
HIV2ROD    AA-TGCA-AT-----GAAC--C-.CT-AT--G-C---AGTA--AT-C-CA-TGTTC-TT--G---C-A--   139
HIV2NIHZ   ..........................................................................    59
HIV2ISY    AA-TCCA-AT-----GAAC--T-.CT-AT--G-T---AGTA--GT-C-CA-TGTTC-TT--G---C-A--   139
HIV2ST     AA-TGGA-AC-----GAAC--T-.CT-AT--A-C---AGTA--GT-T-CA-AGTTC-TT--G---T-A--   139
HIV2BEN    AA-ACCA-AT-----GAAT--T-.CT-AT--G-C---AGTA--GT-C-CA-TGTAC-TC--G---C-G--   139
HIV2D194   AA-ACCA-AT-----GAAT--T-.CT-AT--G-C---AACA--GT-C-CA-TGTAC-TC--G---C-G--   139
HIV2GH1    AA-ACCA-AT-----GAAT--T-.CT-AT--G-C---AGTA--GT-C-CA-TGTGT-TC--G---C-A--   139
HIV2D205   TG-GTCT-GC-----AAAC--T-.CT-AT--G-C---GATA--GC-T-CC-AATAC-TT--T---C-G--   139
SIVMM251   CA-ACCA-AT-----GGAT--C-CCT-A.--A-C---AATT--AT-C-CA-AGACA-TT--C---C-A--   139
SIVMM142   CA-ACCA-AT-----GA..--C-.CT-C.--A-C---AATT--AT-C-TA-AGATG-TT--C---C-A--   136
SIVMM239   CA-ACCA-AT-----GGAT--C-CCT-A.--A-C---AATT--AT-C-CA-AGACA-TT--C---C-A--   139
SIVMNE     CA-GCCA-AT-----GAAT--C-CCT-G.--A-C---ACCT--AT-C-CA-AGACA-TT--C---C-A--   139
SIVSMMH4   AA-ACCA-AT-----AAAT--C-CAT-A.--G-C---AATT--AT-C-CT-TGCAC-AT--G---C-C--   139
SIVSMMPBJ  AA-ACCA-AT-----AAAT--C-CAG-A.--G-C---AATC--AT-C-CT-TGTTC-TT--A---C-C--   139

CONSENSUS  GAAa-TaqTacCaGTAgatqTctCAqa?Gaqqcac???AgGA?qatGA???qaq?catT?ctTaqtqCAt   194
HIV2ROD    ---GC-AG-AC-A---GATG-CC--CAA-AAGGGG...-G--CACT--...GACTCAC-GCT-AGTA--T   203
HIV2NIHZ   ..................................................... .................    59
HIV2ISY    ---GC-AG-AC-A---ACTG-CC--CAA-AAGGGG...-G--CACT--...GACTCTC-GCT-AATG--C   203
HIV2ST     ---GC-AG-AC-A---GATG-CC--CAA-AGGGAG...-T--CAGT--...GACTCAC-GCT-AGTG--T   201
HIV2BEN    --.GC-AG-AT-A---GAAC-CT--CAA-AGGCAG...-G--AGAT--...GGCCAAC-GCT-AGTA--C   203
HIV2D194   ---AC-AG-AC-A---GACA-CT--CAA-AGGCAG...-G--AGTA--...GACCAAC-GCT-AGTA--C   203
HIV2GH1    ---GC-AG-AC-A---GATG-CT--CAA-AGGCAG...-G--CGAT--...GACCAAC-ACT-AACA--C   203
HIV2D205   ---GC-GG-AC-A---GAGG-GC--GCA-CGACCCGAG-G--GGAG--...AACCCAT-GCC-AATG--C   206
SIVMM251   ---AT-AG-CC-T---AATG-AT--GAT-AGGCAC...-G--GGAT--...GAGGCAT-ATT-AATG--G   203
SIVMM142   ---AT-AA-CC-T---AATG-AT--GAT-AGGCAC...-G--GGAT--...GGAGCAT-ATT-AGTG--C   200
SIVMM239   ---AT-AG-CC-T---AATG-AT--GAT-AGGCAC...-G--GGAT--...GGAGCAT-ATT-AATG--T   203
SIVMNE     ---AT-AG-CC-T---AATG-AT--GAT-AGGCAC...-G--GGGT--...GGAGAAT-ATT-ACTG--T   203
SIVSMMH4   ---AT-AG-CC-A---GATG-CT--GAT-AAGCTC...-G--AGAC--...GACACAT-GCC-GGTG--T   203
SIVSMMPBJ  ---AT-AG-CC-G---AATG-CT--GAT-AAGCTC...-G--AGAC--...GACACAT-ACT-GATG--T   203
```

HIV2/SIV LTR

```
CONSENSUS  cCaGCaCAaacat?caagtggGATGAccc???TGGgGAgA???acTagt?TGGaagTTTGAtcC?a?gcT  254
HIV2ROD    C-A--A--AACAAGCAAGTTT-----CCCGCA---G--G-..CAC-AGTC---GAG-----TC-CTTGC-  271
HIV2NIHZ   ............................................................................  59
HIV2ISY    T-A--A--AGTAAGCAGATTT-----CCCGCA---G--G-..CAC-AGTC---AAG-----CC-CATGC-  271
HIV2ST     C-A--A--AACAAGCAGGTTT-----CCCGCA---A--A-..CAT-AGTT---AGG-----CC-CACGC-  271
HIV2BEN    C-A--A--AACAAGCAGACAT-----TGAGCA---G--G-..CAT-AGTG---CAG-----CT-CATGC-  271
HIV2D194   C-A--A--AACAAGCAGATAT-----CGAGCA---G--G-..CAC-AGTT---CGG-----CC-CATGC-  271
HIV2GH1    C-A--A--AACAAGCAGACAT-----CGAGCA---A--A-..CAC-ACTT---CGG-----CC-CACGC-  271
HIV2D205   C-G--A--GATCTCCTCATGG-----CATCCA---G--G-..CTC-TATC---CAG-----TT-CCTCC-  274
SIVMM251   C-A--T--AACTTCCAAGTGG-----CCCT..---G--G-GGTTC-AGCG---AAG-----TC-AACTC-  271
SIVMM142   C-A--T--AACTTCCCAGTGG-----CCCT..---G--G-GGTTC-AGCA---AAG-----TC-AACTC-  268
SIVMM239   C-A--T--AACTTCCCAGTGG-----CCCT..---G--G-GGTTC-AGCA---AAG-----TC-AACTC-  271
SIVMNE     C-A--T--AACTTCCCAGTGG-----CCCT..---G--G-GGTTC-AGTA---AAG-----TC-AACTC-  271
SIVSMMH4   C-G--A--GACTTATCAGTGG-----CCCC..---G--G-GGTAC-GGCA---AAG-----TC-AGAAT-  271
SIVSMMPBJ  C-A--A--AACTTCTCAGTGG-----CCCC..---G--G-GGTAC-GGCA---AAG-----TC-AAAGT-  271

CONSENSUS  ?GCctata?tTAtgaGgC?TttatT?ggtacCCAGAaGAgTTtGGg?acaAgTCAGGccTgccAgAggAa  319
HIV2ROD    G--TTATAGT--CGA-G-T-TTAT-CGGTAC-----G--A--T--GCACA-G-----CC-GCC-G-GG-A  341
HIV2NIHZ   ..............G-T-TTAC-CTGTAC-----G--A--T--GCATA-T-----AC-GCC-G-GA-A  113
HIV2ISY    G--TCATGAG--CAC-A-C-TTAT-CTATAC-----G--A--T--GCACA-G-----AA-GGA-G-AG-T  341
HIV2ST     A--TTTTAGC--CGA-G-C-TTAT-CGATAC-----G--G--T--GTACA-G-----CC-GCC-G-GG-T  341
HIV2BEN    G--CTATAAC--CAA-G-C-TCAC-CTGTAC-----A--G--T--GCACA-G-----AT-GCC-G-GA-A  341
HIV2D194   G--CTATAGT--CAA-G-C-TCAT-CTGCAC-----A--A--T--GCACA-G-----AT-GCC-G-GA-A  341
HIV2GH1    G--CTATGAT--CAA-G-T-TCAT-CTGCAT-----A--G--T--GCACA-G-----AT-GCC-G-GA-A  341
HIV2D205   G--ATATGAT--TGT-G-T-TCAA-AGGTTT-----A--G--C--GTACC-G-----GC-ACC-G-GG-G  344
SIVMM251   A--CTACACT--TGA-G-A-ATGC-AGATAC-----A--G--.--AAGCA-G-----CC-GTC-G-GG-A  340
SIVMM142   A--CTACACT--TGA-G-A-ATAT-AGATAC-----A--G--T--AAGCA-G-----CC-GTC-G-GA-A  338
SIVMM239   G--CTACACT--TGA-G-A-ATGT-AGATAC-----A--G--T--AAGCA-G-----CC-GTC-G-GG-A  341
SIVMNE     A--CTACACT--TGA-G-A-ATAT-AGATAC-----A--G--T--AAGCA-G-----CC-GTC-G-GG-A  341
SIVSMMH4   A--TTATAGC--TAA-G-A-TTAT-AAGTAC-----A--G--T--TAGTA-G-----CT-GTC-G-GG-A  341
SIVSMMPBJ  A--TTATAAC--TAA-G-A-TTGT-GAGCAC-----A--G--T--TAGTC-G-----CT-GTC-A-GG-A  341

                          <- nef
CONSENSUS  GAgtggaaggcaAgaCTaaaaGCAAGaGG?aTaCCatt????TAGttAAa?aACAGgAaCAgccAtActt  383
HIV2ROD    --GTGGAAGGCG-GA--GAAA-----A--AA-A--ATT....---TT--A.G----G-A--GCT-T-CTT  406
HIV2NIHZ   --GTGGAAGGCG-GA--GAAA-----G--AA-A--ATT....---TT--A.A----G-A--GCT-T-CTT  178
HIV2ISY    --CTGGAAGGCA-AA--GAAA-----A--GA-A--ATT....---TT--A.A----G-A--ACC-T-CTT  406
HIV2ST     --ATGGAAGGCA-GA--GAAA-----A--GA-A--GTT....---CT--A.A----G-A--GCT-T-CTT  406
HIV2BEN    --ATGGAAGGCA-AA--GAAA-----A--GA-A--ATA....---TG--T.A----G-A--ACC-T-CTT  406
HIV2D194   --GTGGAAGGCA-AA--GAAA-----A--GA-A--ATA....---TG--T.A----G-A--ACC-T-CTT  406
HIV2GH1    --GTGGAAGGCA-AA--GAAA-----A--GA-A--ATA....---TT--C.A----G-A--GCC-T-CTT  406
HIV2D205   --GTGGAAGGCT-GA--AAAA-----A--GA-A--TACAGAT---GC--GAG----C-G--TAA-C-GGA  414
SIVMM251   --GGTTAGAAGA-GG--AACC-----A--CC-T-.....................................  374
SIVMM142   --GGTTAAAAGA-GG--AGCC-----A--CC-T-.....................................  372
SIVMM239   --GGTTAGAAGA-GG--AACC-----A--CC-T-.....................................  375
SIVMNE     --GGTTAGAAGA-GG--AACC-----A--CC-CT....................................  375
SIVSMMH4   --GGTAAAGAGA-GG--AACC-----A--CC-T-.....................................  374
SIVSMMPBJ  --GGTGCAGAGA-GG--AACC-----A--CC-TT....................................  375
```

```
                                                                  enhancer
CONSENSUS  ggTcAgggca?gaaatagct?acaagaag?aaac?aGCtgagactGCAGGGACTTTCCA?AA????????   440
HIV2ROD    GG-C-GGGCAGGAAGTAACT......AACAGAAACA--TGAGACT-------------G--........   462
HIV2NIHZ   GG-C-GGGCAGGAAGTAGCTACTGA......GAACA--TGAGACT-------------G--........   234
HIV2ISY    GG-C-GGACAGGAAGTAGCTACTGA......AAACA--TGAGACT-------------G--........   462
HIV2ST     GG-C-GGGCAGGAAGTAACTAACAGA.....AAACA--TGAGACT-------------G--........   463
HIV2BEN    GG-C-AGGCAGGAAGTAGCTACTAAG.....AAACA--TGAGGCT-------------G--........   463
HIV2D194   GG-C-GGGCAGGAAATAGCTACTAAG.....AA.CA--TGAGACT-------------G--........   462
HIV2GH1    GG-C-AGGCAGGAAGTAGCTACTGAGAACAGCTGAG--TGCAGCT-------------G--........   468
HIV2D205   AC-.-GCTGACACTGCA.....CAAGAAGGAAACTA--AGACACT-------------A--........   470
SIVMM251   .........TTAACATGGCTGACAAGAGGGAAACTC--TGAGATA-------------C--........   427
SIVMM142   .........TTGAAATGGCTGACAGGAAGGAAACTA--TGACACA-------------C--........   425
SIVMM239   .........TTAACATGGCTGACAAGAAGGAAACTC--TGAAACA-------------C--........   428
SIVMNE     .........TAAAAATGGCTGACAAGAGGGAAACTA--TGAGACA-------------T--........   428
SIVSMMH4   .........TTAAAATGGCTGACAAGAAGCAAACAA--TGAGACA-------------C--........   427
SIVSMMPBJ  .........TAAAAATGGCTGACAAGAAGAAAACAA--TAAGACA-------------C--AGACAGCA   436

                           enhancer     Sp-1         Sp-1          Sp-1
CONSENSUS  ??????????????GgGGcTGTaaca?gGgGagGgacatgggaggagctGGtggggaacgccc?tc?ta   492
HIV2ROD    ..............-G--C---AACCAA-G-AG-GACATGGGAGGAGCT--TGGGGAACGCCC.TCATA   516
HIV2NIHZ   ..............-G--C---AACCAG-A-AG-GACATGGGAGGAGCT--TGGGGAACGCCCTTCATA   289
HIV2ISY    ..............-G--C---AACCAG-G-AG-GACATGGGAGGAGCT--TGGGGAACGCCC.TCATA   516
HIV2ST     ..............-G--C---TACCAG-G-AG-GACATGGGAGGAGCC--TGGGGAACGCCC.TCATA   517
HIV2BEN    ..............-G--C---AACCAA-G-AG-GACATGGGAGGAGCT--TGGGGAACGCCC.TCATA   517
HIV2D194   ..............-G--C---AACCAA-G-AG-GACATGGGAGGAGCT--TG..........TCATA   506
HIV2GH1    ..............-G--C---AACCAA-G-AG-GACATGGGAGGAGCT--TGGGGAACGCCC.TCATA   522
HIV2D205   ..............-A--C---AACAT.-G-AG-AGCAAGGGAGGGGTT--TTGGGAACGCCC.ATTAA   523
SIVMM251   ..............-G--A---T..ATG-G-AG-..........AGCC--TCGGGAACACCC.ACTTT   468
SIVMM142   ..............-G--A---C..ATG-G-AG-TACTGGGGAGGAGCC--TTGGGAACACCC.ACTTT   477
SIVMM239   ..............-G--A---T..ACG-G-AG-TACTGGGGAGGAGCC--TCGGGAACGCCC.ACTTT   480
SIVMNE     ..............-G--A---C..ATG-G-AG-TACTGGGGAGGAGCC--TCGGGAACACCC.ACTTT   480
SIVSMMH4   ..............-G--C---C..AT.-G-AG-TACTGGGGAGGAGCT--CTGG.AACGCCC.ACTTA   477
SIVSMMPBJ  GGGACTTTCCACAAA-G--A---T..ACG-G-AG-TACTGGGGAGGAGCT--CTGG.AACGCCC.ACGTA   502

                                                    /<- IG-varH chain similarity, see III-33 APR89
           signal ->                                   mRNA ->          /<- tar core I
CONSENSUS  cTttcTGTATAAATatacccGCt?tcTcqCatTGTA?TTCaGTCgCTCTGcGqAGAGGCTGGCAqATtGA   560
HIV2ROD    T-CTC---------ATACCC--TAGC-TG-AT----C---G---G-----C-G-----------G--T--   586
HIV2NIHZ   C-TTC---------GTACCC--T.GC-TG-AT----.---A---G-----.-C-----------G--T--   356
HIV2ISY    C-TTC---------GTACCC--T.GC-CG-AT----.---A---G-----C-G-----------G--T--   584
HIV2ST     C-TTC---------GTACCC--T.AC-CG-AT----.---A---G-----C-G-----------G--T--   585
HIV2BEN    C-TAC---------GTACCC--T.TC-TG-AT----.---A---G-----C-G-----------G--C--   585
HIV2D194   T-CTC---------GTACCC--T.TC-TG-AT----.---A---G-----C-G-----------G--T--   574
HIV2GH1    C-TAC---------GTACTC--T.TC-TA-AT----.---A---G-----.-C-----------G--T--   589
HIV2D205   A-CTC---------GTACCC--T.TC-CG-AT----.---A---G-----C-G-----------G--T--   591
SIVMM251   C-TGA---------ATCACT--A.TT-CG-TC----.---A---G-----C-G-----------G--1--   536
SIVMM142   C-TGA---------ATCACT--A.TT-CG-TC----.---A---G-----C-G-----------C--T--   545
SIVMM239   C-TGA---------ATCACT--A.TT-CG-TC----.---A---G-----C-G-----------G--T--   548
SIVMNE     C-TGA---------ATCACT--A.TT-CG-TC----.---A---G-----C-G-----------G--T--   548
SIVSMMH4   T-CTC---------ACAACT--A.TT-CG-TC----.---A---G-----C-G-----------G--T--   545
SIVSMMPBJ  T-CTC---------ACAACT--A.TT-CG-TC----.---A---A-----C-G-----------G--T--   570
```

HIV2/SIV LTR

```
                          IG-varH chain similarity, see III-33 APR89   ->/
        tar core I       ->/          /<-       tar core II          ->/
                                     \/ 5' sj
CONSENSUS  GCCCTGgGAGGTTCTCTCCAGCAcTAGCAGGTAGAGCCTGGGTGTTCCCTGCTaGACTCTCACCAGcaCT   630
HIV2ROD    ------G----------------G----------------------------A-----------CA--   656
HIV2NIHZ   ------G----------------C----------------------------A-----------TG--   426
HIV2ISY    ------G----------------C----------------------------G-----------TG--   654
HIV2ST     ------G----------------C----------------------------A-----------TG--   655
HIV2BEN    ------A----------------C----------------------------G-----------TA--   655
HIV2D194   ------G----------------C--------.....................................   606
HIV2GH1    ------G----------------C----------------------------A-----------TG--   659
HIV2D205   ------G----------------C----------------------------A-----------CA--   661
SIVMM251   ------G----------------C----------------------------A-----------CA--   606
SIVMM142   ------G----------------C----------------------------A-----------CA--   615
SIVMM239   ------G----------------C----------------------------A-----------CA--   618
SIVMNE     ------G----------------C----------------------------A-----------CA--   618
SIVSMMH4   ------G----------------C----------------------------A-----------CA--   615
SIVSMMPBJ  ------G----------------C----------------------------A-----------CA--   640

                                                              signal ->
CONSENSUS  tGGCCgGtgcTGGGcACA??GgCtCCACGCTTGCTTGCTTAaA?gacCTCtTcAATAAAgCTGCCa??TT   695
HIV2ROD    T----G-TGC----C---.C-G-C----------------A-.AAC---C-T------G-----AG.--   723
HIV2NIHZ   T----G-CGC----C---.C-G-T----------------A-.GAC---.-C------G-----AA.--   492
HIV2ISY    T----G-CGC----C---.C-G-T----------------A-AGAC---.-T------G-----AG.--   721
HIV2ST     T----G-CAC----C---.C-G-T----------------A-AGAC---.-T------G-----AG.--   722
HIV2BEN    T----G-TAC----C---.C-G-T----------------A-.GAC---T-C------G-----AG.--   722
HIV2D194   .....................----------------A-.GAC---.-T------G-----AG.--   649
HIV2GH1    T----G-CAC----C---.C-G-T----------------A-.GAC---.-T------G-----AA.--   725
HIV2D205   A----A-TGT----C---.C-G-T----------------A-.ACC---.-T------.-----AA.--   726
SIVMM251   T----A-TGC----C---GT-G-T----------------A-.GAC---T-C------G-----ATT--   675
SIVMM142   T----G-TGC----C---GT-G-T----------------A-.GAC---T-C------G-----AT.--   683
SIVMM239   T----G-TGC----C---GT-A-T----------------A-.GCC---T-C------G-----ATT--   687
SIVMNE     T----G-TGC----C---GT-G-T----------------A-.GAC---T-C------G-----.TT--   686
SIVSMMH4   T----G-TGC----C---GT-G-T----------------A-.GAC---T-C------G-----AA.--   683
SIVSMMPBJ  T----G-TGC----C---GT-G-T----------------A-.GAC---T-C------G-----AT.--   708

        <- R rpt                          \/ 3' sj
CONSENSUS  AGAAGcAAGttaAGTGtGTGtTCCCATCTCTCCTAGtCGCcGCCTGGTCA?ctCGGTgcTCa?ctg?aat   762
HIV2ROD    -----C---TTA----T---C----------------T---C---------.TT----GT--ACCTG.AGT   791
HIV2NIHZ   -----C---TTA----T---C----------------T---C---------.TC----GT--ACCTGTA.T   560
HIV2ISY    -----C---TTA----T---T----------------T---C---------.TT----GT--TCCTG.AGT   789
HIV2ST     -----C---TTA----T---C----------------T---C---------.TT----GT--ATCTAAAGT   791
HIV2BEN    -----C---TTA----T---T----------------T---C---------.TT----GT--ATCTG.AGT   790
HIV2D194   -----C---TTA----T---T----------------T---C---------.TT----GT--ATCTG.AGT   717
HIV2GH1    -----C---TTA----T---T----------------T---C---------.TT----GT--ATCTG.AAT   793
HIV2D205   -----C---T.A----T---T----------------T---.---------.CT----GC--CACTGATA.   792
SIVMM251   -----T---CC.----T---T----------------T---C---------ACT----AC--G...GTAAT   741
SIVMM142   -----T---CT.----T---T----------------T---C---------ACT----AC--G...GTAAT   749
SIVMM239   -----T---CT.----T---T----------------C---C---------ACT----AC--A...ATAAT   751
SIVMNE     -----T---CC.----T---C----------------T---C---------ACT----AC--G...ATAAT   752
SIVSMMH4   -----T---CA.----T---T----------------T---C---------TCT----AC--GAC..ACAT   750
SIVSMMPBJ  -----T---CC.----T---T----------------T---C---------TCT----AC--G...GACAT   774
```

```
CONSENSUS  aa?aagaCCCTGGTCTGTTAGGACC??????????????????CTTtctGCTTTGggaaaCcgAgGCaGg??  813
HIV2ROD    AACAAGA----------------.................---CTT------GGAAA-CG-G--A-G..  843
HIV2NIHZ   AACAAGA----------------.................---CTT------GGAAA-CG-G--A-GCG  614
HIV2ISY    AACAAGA----------------.................---CTT------GGAAA-CG-G--A-G..  841
HIV2ST     AACAAGA----------------.................---TCT------GGAAA-CA-G--A-G..  843
HIV2BEN    AACAAGA----------------.................---CTC------GGAAT-CA-G--A-G..  842
HIV2D194   AACAAGA----------------.................---CCC------AGAAT-CA-G--A-G..  769
HIV2GH1    AACAA.A----------------.................---CC.------GGAAT-CA-G--A-G..  843
HIV2D205   .....GA----------------.................---CTT------GGAAA-CG-A--G-G..  039
SIVMM251   AAGAAGA----------------.................---TCT------AGAAA-CG-A--A-G..  793
SIVMM142   AAAAAGA----------------CTGGTCTGTTAGGACC---TCT------GGAAA-CG-A--A-G..  817
SIVMM239   AAGAAGA----------------.................---TCT------GGAAA-CG-A--A-G..  805
SIVMNE     AAGAAGA----------------.................---TCT------GGAAA-CG-A--A-G..  804
SIVSMMH4   AAGAAGA----------------.................---TCT------GAAAA-CG-G--A-G..  802
SIVSMMPBJ  AAGAAGC----------------.................---TCT------GGAAA-CG-G--A-G..  826

                          <- lys-tRNA pbs
CONSENSUS  ?AAAATCCCTAGCAGgTTGGCGCCcgAACAGGgACtTGaaggaGa?gTGAGAgactcct?GAgcacgG?C  879
HIV2ROD    .--------------G--------TG------G--T--AAGAA-A.C-----.AGTCTTG--ACACG-.-  909
HIV2NIHZ   G--------------G--------CG------.--T--AAGAA-A.C-----.AGCCTTG--ACAAG-G-  681
HIV2ISY    .--------------G--------CG------G--T--AGGAA-A.C-----.AGCCTTG--ACACG-.-  907
HIV2ST     .--------------G--------CG------G--T--AAGAA-A.C-----.AGCCTTG--ACACG-.-  909
HIV2BEN    .--------------G--------CG------G--T--AAGAG-A.C-----.AGCCCTG--ACTCG-.-  908
HIV2D194   .--------------G--------CG------G--T--AAAGA-GAC-----.AGCCCTG--ACACG-.-  836
HIV2GH1    .--------------G--------C.------G--T--AAAGG-A.C-----.AGCCCTG--ACACG-.-  908
HIV2D205   .--------------A--------CG------G--C--.GGAA-A.G-----GTCTTCAA--GCACA-.-  905
SIVMM251   .--------------A--------CG------.--T--AAGGA-A.G-----GACTCCT.--GTACG-.-  858
SIVMM142   .--------------A--------CG------G--T--AAGGA-A.G-----GACTCCT.--GTACG-.-  883
SIVMM239   .--------------A--------TG------G--T--AAGGA-A.G-----GACTCCT.--GTACG-.-  871
SIVMNE     .--------------A--------CG------G--T--AAGGA-A.G-----GACTCCT.--GTACG-.-  870
SIVSMMH4   .--------------G--------CG------G--T--AAGGA-A.G-----G.CTCCT.--GTACG-.-  867
SIVSMMPBJ  .--------------G--------CG------G--T--AAGGA-A.G-----G.CTCCT.--GTACG-.-  891

CONSENSUS  TgaGTGAAGGcAGtAAGGgCGGCAGGAAcaAACCAcgaCGGAGtGCTCCTagAAAgGCGCgGGccgaggt  949
HIV2ROD    -GA--------C--T----G---------CA-----CGA-----T------AG---G----G--CCGAGGT  979
HIV2NIHZ   -GG--------C--T----G---------CA-----GCA-----T------AG---G----G--CCGAGGT  751
HIV2ISY    -GA--------C--T----G---------CA-----CGA-----T------AG---G----A--CCAAGGT  977
HIV2ST     -GA--------C--T----G---------CA-----CGA-----T------AG---A----A--CCGAGGT  979
HIV2BEN    -GA--------C--T----G---------CA-----CGA-----T------AG---G----G--CCGAGGT  978
HIV2D194   -GA--------C--T----G---------CA-----CGA-----T------AG---A----G--CCGAGGT  906
HIV2GH1    -.A--------T--T----.---------CA-----CGA-----T------AG---A----G--CCGAGGT  976
HIV2D205   -GA--------C--T----G---------TC-----CGA-----A------GT---A----A--CCG..GT  973
SIVMM251   -GA--------C--T----G---------CC-----CGA-----T------AT---G----G--TCG..GT  926
SIVMM142   -GA--------C--T----G---------CC-----CGA-----T------AG---G----G--TCG..GT  951
SIVMM239   -GA--------C--T----G---------CC-----CGA-----T------AT---G----G--TCG..GT  939
SIVMNE     -GA--------C--T----G---------CC-----CGG-----T------AG---G----G--TCG..GT  938
SIVSMMH4   -GA--------C--T----G---------CA-----CGA-----A------AG---G----G--CCG..GT  935
SIVSMMPBJ  -GA--------C--C----G---------CA-----CCA-----T------AT---G----G--CCG..GT  959
```

HIV2/SIV LTR

```
                                                major 5' sj \/
CONSENSUS  aCca??ggcgGcGtGaGGagcgggagt?gaaGAGgCcTCCgGgt?gaag?GtAAGTgc?t?ca?caaaaa  1011
HIV2ROD    A-CAAAGGCA-C-T-T--AGCGGGAGGAGAA---G-C---G-GT.GAAG.-T----ACCTACACCAAAAA  1047
HIV2NIHZ   A-CAAAGCG.-C-T-T--A.CGGGAGTCGA.---G-C---G-GT.GAAGG-T----ACCTACACCAAAAA   817
HIV2ISY    A-CAAAGGCG-C-T-T--AGCGGGAGTCAA.---G-C---G-GT.GAAG.-T----ACCTACACCAAAAA  1044
HIV2ST     A-CAAGGGCG-C-T-T--AGCGGGAGTGAAA---G-C---G-GT.GAAG.-T----GCCTACACCAAATA  1047
HIV2BEN    A-CAAAGGCG-C-T-T--AGCGGGAGTGAAA---G-C---G-GT.GAAG.-T----ACCTACACCAAAAA  1046
HIV2D194   A-CGAAG.CG-C-T-T--AGCGGGAGTGAAA---G-C---G-GT.GAAG.-T----ACCTACACCGAAAA   973
HIV2GH1    A-CAAAGGCG-C-T-T--AGCGGGAGTGAAA---G-C---G-GT.GAAG.-T----GCCTACACCGAAAA  1044
HIV2D205   A-CA..GGCA-C-T-A--AGCGGGAGGAGAA---G-.---G-GAGTAAG.-T----ACTCC......AAA  1033
SIVMM251   A-CA..GACG-C-T-A--AGCGGGAGAGGAG---G-C---G-TT.GCAG.-T----G....CAACACAAA   988
SIVMM142   A-CA..GACG-C-T-A--AGCGGGGAGAGAA---G-C---T-GTTGCAG.-T----G....CAACACAAA  1014
SIVMM239   A-CA..GACG-C-T-A--AGCGGGAGAGGAA---G-C---G-TT.GCAG.-T----G....CAACACAAA  1001
SIVMNE     A-CA..GACG-C-T-A--AGCGGGAG.AGAA---G-C---G-TT.GCAG.-T----G....CAACACAAA   999
SIVSMMH4   A-CA..GGCG-C-T-A--AGCGGGAGTCGGA---G-C---G-TT.GCAG.-T----G....CAACAGAAA   997
SIVSMMPBJ  A-CA..GGCG-A-T-A--AGCGGGAGTCGGA---G-C---G-TT.GCAG.-T----G....CAACAAAAA  1021

                                                                    gag start ->
CONSENSUS  ?agtagcc?gaaa?ggc?????ttg?Tatcctaccttt?aga?ag???gtAgaagat???GTGGGAG      1062
HIV2ROD    CTGTAGCC.GAAAGGGC.....TTGC-ATCCTACCTTT.AGACAG...GT-GAAGATT..-------      1102
HIV2NIHZ   CTGTAGCCAAAGAGGGC.....TTGT-ATCCTACCTTT.AGACAG...GT-GAAGATT..-------       873
HIV2ISY    CTGTAGCCGGAAAAGGC.....TTGT-ATCCTACCTTT.AGACAG...GT-GAAGATT..-------      1100
HIV2ST     CAGTAGCCAGAA.GGGC.....TTGT-ATCCTACCTTT.AGACGG...GT-GAAGATT..-------      1102
HIV2BEN    CTGTAGCCAGAAAAGGC.....TTGT-ATCCTACCTTT.AGACAG...GT-GAAGATT..-------      1102
HIV2D194   CTGTAGCCAGAAAAGGC.....TTGT-ATCCTACCTTT.AGACAG...GT-GAAGATT..-------      1029
HIV2GH1    CCGTAGCCAGAAAAGGC.....TAGT-ATCCTATCTTTTAGACAG...GT-GAAGATT..-------      1101
HIV2D205   TTTT.GC.................AT-GTC.TACTCTAAAGAGGGA..GT-GGGCATAA.-------      1078
SIVMM251   AAA......GAAATAGC.TGTCTTGT-ATCC.AGGAAGG.GATA......-TAAGATAGA-------      1040
SIVMM142   AAG......GAAATAGC.TGTCTT.T-ATCC.AGGAAGG.GATA......-TAAGATAGA-------      1065
SIVMM239   AAA......GAAATAGC.TGTCTT.T-ATCC.AGGAAGG.GGTA......-TAAGATAGA-------      1052
SIVMNE     AAA......GAGATAGC.TGTCTT.T-ATCC.AGGAAGG.GATA......-TAAGATAGA-------      1050
SIVSMMH4   .AGT......CATAGGACTGAG...T-CCCT.ACTTTTGAGGAAAGA.GT-GGAGA....-------      1048
SIVSMMPBJ  GAGT......CATAGGACTGAG...T-CCCTACTTTTGAGAAAAGA..GT-GGAGA....-------      1073
```

```
CONSENSUS  ATGGGCGcGAgAaaCTCCGTCtTG??AGGGAA?aAaqCAGAtGAATTAGAAAaaaTtAGgtTACGgCCCg   67
HIV2ROD    -------C--G-AA-------T--AG------AA-AG----T----------GAA-C--GT----G---G   70
HIV2NIHZ   -------C--G-AA-------T--AG------AA-AG----T----------AAA-T--GT----G---G   70
HIV2ISY    -------C--A-AA-------T--AG------AA-GG----T----------AAA-T--GT----G---G   70
HIV2ST     -------C--G-AA-------T--AG------AA-AG----C----------AGA-T--GT----G---G   70
HIV2BEN    -------C--G-AA-------T--AG------AA-AG----C----------AAG-T--GT----G---G   70
HIV2D194   -------C--G-AA-------T--AG------AA-AG----C----------AAG-T--GT----G---A   70
HIV2GH1    -------C--G-AA-------T--AG------AA-AG----T----------AGA-T--GT----G---A   70
HIV2D205   -------C--G-GG-------C--TC------GA-AA----T----------AAG-T--GT----G---G   70
SIVMM251   -------C--G-AA-------T--TC------GA-AG----T----------AAA-T--GC----A---G   70
SIVMM142   -------C--G-AA-------T--TC------GA-AG----T----------AAA-T--AC----A---G   70
SIVMM239   -------T--G-AA-------T--TC------GA-AG----T----------AAA-T--GC----A---A   70
SIVMNE     -------C--G-AA-------T--TC------GA-AG----T----------AAA-T--GC----A---G   70
SIVSMMH4   -------C--G-AA-------T--TC------GG-AG----T----------AAG-T--GT----G---A   70
SIVSMMPBJ  -------C--G-AA-------T--TC------GA-AG----T----------AAA-T--GT----G---G   70

CONSENSUS  gCGGaAAGAAAAagTAca?gtTaAAaCAT?T?GT?TGGGCAGcgAAtgAATTgGAcAgATT?GGATTagC  132
HIV2ROD    G---A-------AG--CAGGC-A--A---A-T--G-------CG--TA----G--C-G---C-----AG-  140
HIV2NIHZ   G---A-------AG--CAAGT-A--A---A-T--G-------CG--TG----G--C-G---C-----AG-  140
HIV2ISY    G---G-------AA--CAGAT-A--A---A-T--G-------CG--TG----G--C-G---C-----AA-  140
HIV2ST     G---A-------AA--TAGGC-A--A---A-T--G-------CG--TG----G--C-G---C-----GG-  140
HIV2BEN    G---A-------AG--CAGGT-A--A---A-T--G-------CG--TG----G--T-A---C-----GG-  140
HIV2D194   A---A-------GA--CAGGT-A--A---G-T--G-------CG--TG----A--C-G---C-----GG-  140
HIV2GH1    G---A-------AG--CAGGC-A--A---A-T--G-------CG--CG----G--C-A---C-----GG-  140
HIV2D205   G---A-------AG--TATGC-A--A---G-A--A-------TG--CG----A--C-G---T-----AG-  140
SIVMM251   G---A-------AG--CATGT-G--G---G-A--A-------CA--TG----A--T-G---T-----AG-  140
SIVMM142   G---A-------AG--CATGT-G--G---G-A--A-------CA--TG----A--T-G---T-----AG-  140
SIVMM239   A---A-------AG--CATGT-G--G---G-A--A-------CA--TG----A--T-G---T-----AG-  140
SIVMNE     G---G-------AG--CATGT-G--G---G-A--A-------CA--TG----A--T-G---T-----AG-  140
SIVSMMH4   A---A-------AG--TATGT-G--G---G-A--A-------CA--TG----G--C-G---T-----AG-  140
SIVSMMPBJ  G---A-------GG--CCAGC-G--G---A-A--A-------CA--TG----G--C-G---T-----AG-  140

CONSENSUS  AGAgAGCCTGTTGGAgtcaAAAGAaGGtTGtCAAAAaATtcTt?caGTtTTAG?tCCAtTaGTaCCaACA  200
HIV2ROD    ---G-----------GTCA-----G--T--T-----A--TC-TACA--T----AT---A-G--A--G---  210
HIV2NIHZ   ---G-----------GTCA-----A--T--C-----A--TC-TACA--T----AT---T-A--A--G---  210
HIV2ISY    ---G-----------GTCA-----A--T--C-----A--TA-TTCA--T----AA---T-A--A--A---  210
HIV2ST     ---G-----------GTCA-----G--T--C-----A--TC-TACA--T----AT---T-A--A--G---  210
HIV2BEN    ---G-----------GTCA-----A--T--C-----G--TC-CAGA--T----AT---T-A--A--A---  210
HIV2D194   ---G-----------ATCA-----A--T--C-----G--TC-TAAA--T----AA---T-A--A--A---  210
HIV2GH1    ---G-----------GTCA-----A--T--C-----G--TC-CACA--T----AT---T-A--A--A---  210
HIV2D205   ---G-----------ATCA-----A--A--T-----A--TC-TAAA--C----CT---T-A--A--A---  210
SIVMM251   ---A-----------GAAC-----A--A--T-----A--AC-TTCG--C----CT---T-A--G--A---  210
SIVMM142   ---A-----------GAAC-----A--A--T-----A--AC-TTCG--C----CT---T-A--G--A---  210
SIVMM239   ---A-----------GAAC-----A--A--T-----A--AC-TTCG--C----CT---T-A--G--A---  210
SIVMNE     ---A-----------GAAC-----A--A--T-----A--AC-TTCG--C----CT---T-A--G--A---  210
SIVSMMH4   ---A-----------TAAC-----A--T--T-----A--TC-TTCA--T----CT---T-A--T--G---  210
SIVSMMPBJ  ---A-----------GAAC-----A--T--T-----A--TC-TTCA--T----CT---T-A--T--G---  210
```

HIV2/SIV GAG

```
CONSENSUS  GG?TCAGAAAAtTTAAAaAGcCTTT?TAAtActgtCTGCGTCaTtTggTGcaT?CACGCAGAAGAGAAAG  267
HIV2ROD    --T--------T-----A--T----T---T-CTGT-------A-T-GG--CA-A----------------  280
HIV2NIHZ   --T--------T-----A--C----T---T-CTGT-------A-C-GG--CA-A----------------  280
HIV2ISY    --G--------T-----A--C----A---T-CTAC-------A-T-GG--CT-G----------------  280
HIV2ST     --G--------T-----A--C----T---T-CTGT-------A-T-GG--TA-A----------------  280
HIV2BEN    --G--------T-----A--C----T---T-CCGT-------A-T-GG--CT-G----------------  280
HIV2D194   --G--------T-----A--C----T---T-CCGT-------A-T-GG--CT-G----------------  280
HIV2GH1    --G--------T-----A--C----T---C-CAGT-------A-T-GG--CT-G----------------  280
HIV2D205   --G--------T-----A--C----T---T-TCGT-------A-T-TT--CC-G----------------  280
SIVMM251   --C--------T-----A--C----A---T-CTGT-------A-C-GG--CA-T----------------  280
SIVMM142   --C--------T-----A--C----A---T-CTGT-------A-C-GG--CA-T----------------  280
SIVMM239   --C--------T-----A--C----A---T-CTGT-------A-C-GG--CA-T----------------  280
SIVMNE     --C--------T-----G--C----A---T-CTGT-------A-C-GG--CA-T----------------  280
SIVSMMH4   --T--------C-----G--C----A---T-CTGT-------A-T-GG--CA-T----------------  280
SIVSMMPBJ  --T--------T-----G--C----A---C-CTGT-------C-T-GG--CA-T----------------  280

CONSENSUS  tGAAAgAtACtGA?GaAGCAAAacaaaataGt?caGAGACAtCTAGtGgc?GAaACaggaactgcaGAaAa  334
HIV2ROD    T----G-T--T--A-G------ACAAATA-TGCG------T----T-GCA--A--AGGAACTGCA--G-A  350
HIV2NIHZ   T----G-T--T--A-G------ACAAATA-TACA------T----T-GCA--A--AGGAACTGCA--G-A  350
HIV2ISY    T----G-T--T--A-A------AAGAATA-TAGG------T----T-GCA--A--AGAAACTGCA--G-A  350
HIV2ST     C----G-T--T--A-A------ACAAAAG-TACA------T----T-GCA--A--AAAAACTACA--A-A  350
HIV2BEN    T----G-T--T--G-A------GAAACTA-CACA------T----T-GCA--A--TGGAACTGCA--G-A  350
HIV2D194   T----G-T--T--A-A------GAAACTA-CACA------T----T-GCA--A--AGGAACTGCA--G-A  350
HIV2GH1    T----G-T--T--A-A------AAAACTA-TACA------T----G-GCA--A--AGGAACTGCA--A-A  350
HIV2D205   T----G-T--A--G-A------AAAGATA-CACA------T----C-GCG--C--A.........--A-A  341
SIVMM251   T----C-C--T--G-A------ACAGATA-TGCA------C----T-GTG--A--AGGAACAGCA--A-C  350
SIVMM142   T----C-C--T--G-A------ACAGATA-TGCA------C----T-ATG--A--AGGAACAGCA--A-C  350
SIVMM239   T----C-C--T--G-A------ACAGATA-TGCA------C----T-GTG--A--AGGAACAACA--A-C  350
SIVMNE     T----C-C--T--G-A------ACAGATA-TGCA------C----T-GTG--A--AGGAACAGCA--A-C  350
SIVSMMH4   T----C-T--T--G-A------ACAAATA-TGCA------T----T-GTG--A--AGGAACAGCA--C-G  350
SIVSMMPBJ  T----C-C--T--A-A------ACAAATA-TGCA------T----T-GTG--A--AGGAACAGCA--C-A  350

CONSENSUS  aATGCCAaatAcAAgtAgACCAACAGCACCAccTAGcGg?aaaggaggaAAtTacCCaGTgCAaCAa???  400
HIV2ROD    A----·--AGC-C--GT-G------------TC---C-AGAAGGGAGGA--T-AC--A--G--A--T...  417
HIV2NIHZ   A------AAT-C--GT-G------------CC---C-GGAAA.....--C-TC--C--G--G--A...  411
HIV2ISY    A------AAT-T--GT-G---------·---CC---C-GGAAAGGGGGA--C-TC--C--G--A--A...  417
HIV2ST     A------AGT-C--GT-G------------CC---C-GGAACGGAGGA--C-TC--C--A--A--A...  417
HIV2BEN    A------AAT-C--GT-G------------CC---T-GGAAAAGAGGA--C-AC--C--G--A--A...  417
HIV2D194   A------AAT-T--GT-G------------CC---T-GGAAAGGAGGA--C-TC--C--G--A--G...  417
HIV2GH1    A------AGT-C--GT-G------------CC---C-GGAGAGGAAGA--C-TC--T--A--A--GACA  420
HIV2D205   A------GCT-C--AT-A------------CC---C-GC......GGA--T-AT--A--G--G--A...  402
SIVMM251   T------AAA-C--GT-G------------TC---C-GCAGAGGAGGA--T-AC--A--A--A--A...  417
SIVMM142   T------AAA-C--GT-G------------TT---C-GCAGAGCAGGA--T-AC--A--A--A--A...  417
SIVMM239   T------AAA-C--GT-G------------TC---C-GCAGAGGAGGA--T-AC--A--A--A--A...  417
SIVMNE     T----·--AAA-C--GT-G------------TC---T-GCAGAGGAGGA--T-AC--A--A--A--A...  417
SIVSMMH·   A------GCA-C--GC-G------------CC---T-GCAGAGGAGGA--T-AC--A--G--G--A...  417
SIVSMMPBJ  A------GCA-C--GT-G------------CC---T-GCAAAGGAGGA--T-AC--A--G--G--A...  417
```

```
CONSENSUS  gtaGgtGGcAAcTAtatCCAccTgCCacTaAGcCCccGAAC??TAAATGCtTGGGTAAAatTagTaGAgG  468
HIV2ROD    GTA-GC--C--C--CAC---TA-A--GC-G--T--CC----CC-------C--------AT-AG-A--G-  487
HIV2NIH2   GTA-CT--C--C--TAC---TA-A--AC-G--T--CG----TC-------T--------AT-AG-G--G-  481
HIV2ISY    ATA-GC--C--C--TGT---TC-G--GC-G--T--CC----CC-------T--------GT-AG-A--G-  487
HIV2ST     GTG-CC--C--C--TAC---TG-G--AC-A--T--CC----CC-------T--------AC-AG-A--G-  487
HIV2BEN    GCG-GT--C--C--TGT---TG-G--AC-G--C--CC----TC-------A--------AT-AG-G--G-  487
HIV2D194   GCA-GC--C--C--TAT---TG-G--GC-G--C--CC----TC-------T--------AT-AG-A--G-  487
HIV2GH1    GGC-GT--C--C--TAT---CG-G--AC-G--C--CC----TC-------T--------AT-AG-A--G-  490
HIV2D205   CTG-CT--C--C--CGT---CC-G--GC-A--C--CC----CT-------T--------GT-AG-A--A-  472
SIVMM251   ATA-GT--T--C--TGT---CC-G--AT-A--C--GA----AT-------C--------AT-GA-A--G-  487
SIVMM142   ATA-GT--T--C--TAC---CC-A--AT-A--C--GA----AT-------C--------AT-AA-A--G-  487
SIVMM239   ATA-GT--T--C--TGT---CC-G--AT-A--C--GA----AT-------C--------AT-GA-A--G-  487
SIVMNE     GTA-GT--T--C--TAC---CC-A--AT-A--C--GA----AT-------C--------AT-GA-A--G-  487
SIVSMMH4   GTA-GT--C--T--TGT---CC-A--CT-A--T--AA----AT-------T--------AT-AG-A--A-  487
SIVSMMPBJ  ATA-GT--C--T--TAC---CC-A--TT-A--T--AA----AT-------T--------AT-GA-A--A-  487

CONSENSUS  AaAAgAA?TTcGGgGCAGAaGTAGTgcCaGGATTTCAgGCaCTcTCAGAAGGcTGCacgCCcTATGAtAT  537
HIV2ROD    -A--A--G--C--G-----A-----GC-A--------G--A--C--------C---ACG--C-----T--  557
HIV2NIH2   -A--G--A--C--G-----A-----GC-A--------G--A--C--------C---ACG--C-----T--  551
HIV2ISY    -G--A--G--C--G-----A-----GC-G--------G--A--C--------C---ACG--C-----T--  557
HIV2ST     -A--G--G--C--G-----A-----GC-A--------G--A--C--------C---ACG--C-----T--  557
HIV2BEN    -A--G--G--C--G-----A-----GC-A--------G--A--C--------C---ACG--C-----T--  557
HIV2D194   -A--G--G--C--G-----A-----GC-A--------G--A--C--------C---ACG--C-----T--  557
HIV2GH1    -C--G--G--C--G-----A-----GC-A--------A--A--C--------C---ACG--C-----T--  560
HIV2D205   -A--G--G--C--G-----A-----AC-A--------G--A--A--------A---ACC--T-----T--  542
SIVMM251   -A--G--A--T--A-----A-----GC-A--------G--A--G--------C---ACC--C-----C--  557
SIVMM142   -G--G--A--T--A-----A-----GT-A--------G--A--G--------C---CTC--C-----C--  557
SIVMM239   -A--G--A--T--A-----A-----GC-A--------G--A--G--------T---ACC--C-----C--  557
SIVMNE     -G--G--A--T--A-----A-----GC-A--------G--A--G--------C---ACC--C-----C--  557
SIVSMMH4   -G--A--A--T--G-----G-----GC-A--------A--G--A--------C---ACT--C-----T--  557
SIVSMMPBJ  -G--A--A--T--G-----G-----GC-A--------A--G--A--------C---ACT--C-----C--  557

CONSENSUS  tAAtCAaATGcT?AATTGTGTgGG?GAcCAtCAaGCaGCtATGCAaATaATcAGaGAaATtAT?AAtGAg  604
HIV2ROD    C--C--A---C-T--------G--C--C--T--A--A--C-----G--A--C--G--G--T--C--T--G  627
HIV2NIH2   T--T--G---C-T--------G--C--C--T--A--A--T-----A--A--C--G--A--T--T--T--G  621
HIV2ISY    T--T--A---C-T--------G--C--C--T--A--A--G-----A--A--C--A--A--T--T--T--A  627
HIV2ST     T--T--A---C-T--------G--C--C--T--A--A--T-----A--A--C--G--A--T--T--T--A  627
HIV2BEN    T--T--A---C-T--------G--C--T--C--A--A--T-----A--A--C--A--G--T--T--T--A  627
HIV2D194   C--T--A---C-T--------G--C--T--C--A--A--T-----A--A--C--A--A--T--T--T--G  627
HIV2GH1    C--C--A---C-T--------G--C--T--C--A--A--T-----A--A--C--A--G--T--C--T--C  630
HIV2D205   A--T--C---C-A--------A--A--A--T--G--A--C-----A--T--T--A--A--A--C--T  C  612
SIVMM251   T--T--G---T-A--------G--A--C--T--A--G--T-----G--T--C--A--T--T--A--T--G  627
SIVMM142   T--T--G---T-A--------G--A--C--T--A--G--T-----G--C--C--A--T--T--A--T--G  627
SIVMM239   T--T--G---T-A--------G--A--C--T--A--G--T-----G--T--C--A--T--T--A--C--G  627
SIVMNE     T--T--G---T-A--------G--A--A--T--A--A--T-----G--T--C--A--A--T--T--A--C--G  627
SIVSMMH4   T--T--A---C-A--------A--A--A--T--G--A--C-----A--T--T--A--G--T--T--A--T--A  627
SIVSMMPBJ  C--T--A---C-A--------A--A--G--T--G--A--C-----A--T--T--A--A--T--T--A--T--A  627
```

HIV2/SIV GAG

```
CONSENSUS  GAaGCaGCaGAtTGGGAtttgca?CAcCCaataCcAGgcCCcttaCcagcaCG?CAgCTtAGaGAgCCaa  672
HIV2ROD    --A--A--A--A-----TGTGCAA--T--AATA-C--GC--CTTA-CAGCG--G--G--T--A--G--AA  697
HIV2NIHZ   --A--A--A--C-----TGTTGCA--T--CATA-C--GC--CTTA-CAGCG--G--G--T--A--G--AA  691
HIV2ISY    --A--A--A--C-----TGTACAA--T--AATA-C--GC--CTTA-CAGCG--G--G--C--A--T--AC  697
HIV2ST     --A--A--A--T-----CGCACAA--C--AATA-C--GC--CTTA-CAGCG--G--G--C--G--G--AA  697
HIV2BEN    --A--A--A--C-----TTCGCAG--C--AATA-C--GC--CTTA-CAGCA--A--G--C--A--C--AA  697
HIV2D194   --A--A--A--T-----TGCGCAG--C--AATA-C--GC--CTTA-CAGCA--G--G--T--A--C--AA  697
HIV2GH1    --A--A--A--T-----TGCACAG--C--AATA-C--GC--CTTA-CAGCA--G--G--T--A--C--AA  700
HIV2D205   --A--A--A--C-----CCAGCAA--C--GTCA-C--GC--AATG-CGGCA--A--A--T--G--C--AA  682
SIVMM251   --G--T--A--T-----CTTGCAG--C--ACAA-C--CT--A...-AACAA--A--G--T--G--G--GT  694
SIVMM142   --G--T--A--T-----CTTGCAG--C--ACAA-A--CT--A...-AACAA--A--G--T--G--G--GT  694
SIVMM239   --G--T--A--T-----CTTGCAG--C--ACAA-C--CT--A...-AACAA--A--A--T--G--G--GT  694
SIVMNE     --G--T--A--T-----CTTGCAG--C--ACAA-A--CT--A...-AACAA--A--G--T--G--G--GT  694
SIVSMMH4   --A--T--C--T-----TTTACAA--C--GCAA-C--GT--ACTA-CAGCA--G--A--T--A--G--AA  697
SIVSMMPBJ  --A--T--T--T-----TTTACAA--C--GCAG-C--GT--GATA-CACCA--A--A--T--A--G--AA  697

CONSENSUS  gaGGaTCtGAcATaGCAGGaACaACaAGcaCAGTAGA?GAaCAgATccAgTGGATGTatAGgcaacaaAA  741
HIV2ROD    GG--A--T--C--A-----G--A--A--CA-------A--A--G--CC-G-------TT--GCCACAA--  767
HIV2NIHZ   GA--G--T--C--A-----G--A--A--CA-------A--G--G--CC-G-------TT--ACCACAA--  761
HIV2ISY    GA--A--T--C--A-----G--A--A--CA-------G--A--G--CG-A-------AT--GCAAGAA--  767
HIV2ST     GG--A--T--C--A-----G--A--A--CA-------A--G--G--CC-G-------TT--GCCACAA--  767
HIV2BEN    GA--G--T--C--A-----A--A--A--CA-------T--A--G--CC-G-------AT--GCCACAA--  767
HIV2D194   GG--G--T--C--A-----A--A--A--CA-------T--A--G--CC-G-------AT--GCAACCA--  767
HIV2GH1    GG--G--T--C--A-----A--A--A--CA-------A--A--G--CC-G-------AT--GCCACAA--  770
HIV2D205   GA--G--A--T--A-----A--C--C--CA-------G--A--G--AC-G-------AC--GGCCCAA--  752
SIVMM251   CA--A--A--T--T-----A--A--T--TT-------T--A--A--CC-G-------AC--ACAACAG--  764
SIVMM142   CA--A--A--T--T-----A--A--T--TA-------A--A--A--CC-G-------AC--ACAACAG--  764
SIVMM239   CA--A--A--T--T-----A--A--T--TT-------T--A--A--CC-G-------AC--ACAACAG--  764
SIVMNE     CA--A--A--C--T-----A--A--T--TA-------T--A--A--CC-G-------AC--ACAACAG--  764
SIVSMMH4   GA--A--A--C--T-----A--T--T--TA-------T--A--A--CC-A-------AC--GCAACAA--  767
SIVSMMPBJ  GA--A--C--C--T-----A--T--T--CA-------C--A--A--TC-A-------AT--GCAGCAA--  767

CONSENSUS  tCCcgTaCCaGTaGG?AaCATcTAcAG?AGaTGGATcCAaaTaGGgcTgCAgAAaTGtGTcaGaAtGTAc  809
HIV2ROD    T--TG-A--A--A--A-A---C--T--A--A-----C--GA-A--AT-G--G--G--T--CA-G-T---C  837
HIV2NIHZ   T--TG-A--A--A--C-A---C--T--G--A-----T--GA-A--AC-G--A--G--C--CA-G-T---C  831
HIV2ISY    T--TG-A--A--A--A-A---C--T--G--A-----C--GA-A--AC-G--G--G--T--CA-A-T---C  837
HIV2ST     T--TG-A--A--A--A-G---C--T--A--A-----C--GA-A--GC-A--G--G--C--CA-G-T---C  837
HIV2BEN    T--CG-A--G--A--G-A---C--C--A--A-----C--AA-A--GC-G--A--G--T--CA-A-A---C  837
HIV2D194   T--CG-G--G--A--G-A---C--C--G--A-----C--GA-A--GC-A--G--A--T--CA-G-T---C  837
HIV2GH1    T--CG-G--G--A--G-A---C--C--A--A-----C--GA-A--GC-A--G--G--T--CA-G-T---C  840
HIV2D205   T--TG-C--A--G--A-A-- T--T--A--A-----T--AT-A--AT-G--G--A--T--CC-A-T---C  822
SIVMM251   C--CA-A--A--A--C-A---T--C--G--A-----C--AC-G--GT-G--A--A--T--CA-A-T---T  834
SIVMM142   C--CA-A--A--A--C-A---T--C--G--A-----C--AC-G--GT-G--A--A--T--CA-A-T---T  834
SIVMM239   C--CA-A--A--A--C-A---T--C--G--A-----C--AC-G--GT-G--A--A--T--CA-A-T---T  834
SIVMNE     C--CA-A--A--A--C-A---T--C--G--A-----C--AC-G--GT-G--A--A--T--CA-A-T---T  834
SIVSMMH4   C--CA-A--A--A--C-A---T--T--A--G-----C--AT-A--GC-G--G--A--T--AA-A-T---T  837
SIVSMMPBJ  C--TA-A--A--A--T-A---C--C--A--G-----C--AC-A--GC-G--G--A--T--AA-A-T---T  837
```

```
CONSENSUS  AAcCCaAC?AaCATtcTaGA?gTaAAaCAaGGaCCAAAaGAgcCaTTcCA?AGcTAtGTaGA?AG?TTCT   874
HIV2ROD    --C--G--C-A---CC-A--CA-A--A--G--A-----G--GC-G--C--A--C--T--A--T--A----   907
HIV2NIHZ   --T--A--C-A---CC-A--CA-A--T--A--A-----G--AC-A--C--A--C--T--A--T--G----   901
HIV2ISY    --T--A--C-A---TC-A--CA-A--A--A--A-----A--GT-G--C--A--C--T--G--T--A----   907
HIV2ST     --C--A--C-A---CC-A--CA-A--A--G--A-----G--GC-A--C--G--T--T--A--T--A----   907
HIV2BEN    --C--A--T-A---CT-A--CA-A--A--G--A-----A--AC-G--C--A--C--T--A--C--G----   907
HIV2D194   --C--A--T-A---CT-A--TG-G--G--G--A-----A--AT-G--C--G--C--T--A--C--A----   907
HIV2GH1    --C--A--T-A---CT-A--CG-A--G--G--A-----G--AC-G--C--G--C--T--G--C--G----   910
HIV2D205   --T--T--C-A---AT-A--CA-A--G--G--A-----G--GC-C--C--A--C--T--A--T--A----   892
SIVMM251   --C--A--A-A---TC-A--TG-A--A--A--G-----A--GC-A--T--G--C--T--A--C--G----   904
SIVMM142   --C--A--A-A---TC-A--TG-A--A--A--G-----A--GC-A--T--G--C--T--A--C--G----   904
SIVMM239   --C--A--A-A---TC-A--TG-A--A--A--G-----A--GC-A--T--G--C--T--A--C--G----   904
SIVMNE     --C--A--A-G---TC-A--TG-A--A--A--G-----A--GC-A--T--G--C--T--A--C--G----   904
SIVSMMH4   --C--A--A-A---TT-A--TG-G--A--A--A-----A--GC-A--T--A--C--C--A--T--A----   907
SIVSMMPBJ  --T--A--A-A---TT-G--TG-G--A--A--A-----A--GC-A--T--A--C--T--A--T--A----   907

CONSENSUS  AcAAaAGctTaaGgGCAGAaCAaACAGAtcCaGCaGTaAAgAAtTGGATGACcCAaACaCTGCT?aTtCA   943
HIV2ROD    -C--A--CT-GA-G-----A--A-----TC-A--A--G--G--T--------C--A--A-----AG-A--   977
HIV2NIHZ   -C--A--CT-GA-G-----A--A-----TC-A--A--A--A--T--------C--A--A-----AA-A--   971
HIV2ISY    -C--A--CT-AA-G-----A--G-----TG-A--A--G--G--T--------C--G--G-----AG-G--   977
HIV2ST     -C--G--CT-GA-G-----A--A-----TC-A--A--A--A--T--------C--A--A-----AG-G--   977
HIV2BEN    -C--A--CT-GA-G-----A--A-----CC-A--A--A--A--T--------C--A--G-----AA-A--   977
HIV2D194   -C--A--CC-AA-G-----A--A-----CC-G--T--A--A--T--------C--A--G-----AA-A--   977
HIV2GH1    -T--A--CT-GA-G-----A--A-----TC-G--A--A--G--C--------C--A--G-----AA-A--   980
HIV2D205   -C--A--CT-AC-G-----A--A-----CC-A--A--G--A--T--------A--A--A-----GA-T--   962
SIVMM251   -C--A--CT-AA-A-----A--A-----TG-A--A--A--G--T--------T--A--A-----GA-T--   974
SIVMM142   -C--A--TT-AA-A-----A--A-----TC-A--A--A--G--T--------T--A--A-----GA-T--   974
SIVMM239   -C--A--TT-AA-A-----A--G-----TG-A--A--A--G--T--------T--A--A-----GA-T--   974
SIVMNE     -C--A--CT-AA-A-----A--A-----TC-A--A--A--G--T--------T--A--A-----GA-T--   974
SIVSMMH4   -C--A--TC-AA-A-----G--A-----TC-C--A--A--G--T--------C--A--A-----GA-T--   977
SIVSMMPBJ  -C--G--TC-AA-A-----A--A-----CC-G--A--G--A--T--------T--A--A-----GA-T--   977

CONSENSUS  aaatgCcAACCCaGA?TGcAAgtTaGTgcTaAAgGGacTgGG?atgAAtCCcACc?TAGAaGAaATGcTg  1010
HIV2ROD    AAATG-C-----A--C--T--AT-A--GC-A--A--AC-A--GATG--C--T--CT----A--G---C-G  1047
HIV2NIHZ   GAATG-C-----A--T--T--AT-A--AC-A--A--AT-A--AATG--T--T--CT----A--A---C-G  1041
HIV2ISY    A...T-G-----A--C--T--GT-A--AC-A--G--AC-A--GATG--T--T--CT----A--G---C-A  1044
HIV2ST     GAATG-C-----A--C--T--GT-A--AC-A--A--AC-A--GATA--T--T--CT----A--A---C-A  1047
HIV2BEN    GAATG-C-----A--C--C--GT-A--AC-A--A--AC-G--GATG--T--C--CC----A--G---C-A  1047
HIV2D194   GAATG-C-----A--C--C--GT-A--AT-A--A--AC-A--GATG--T--C--CC----G--G---C-G  1047
HIV2GH1    GAATG-C-----A--C--C--GT-A--AC-A--A--AC-G--GATG--T--C--CC----A--G---C-G  1050
HIV2D205   GAATG-T-----A--T--C--GT-A--GC-T--G--CT-G--AATG--T--C--CT----G--A---C-A  1032
SIVMM251   AAATG-T-----A--T--C--GC-A--GC-G--G--GC-G--TGTG--T--C--CC----A--A---C-G  1044
SIVMM142   AAATG-T-----A--T--C--GC-A--GC-G--G--GC-G--TACG--T--C--CC----A--A---C-G  1044
SIVMM239   AAATG-T-----A--T--C--GC-A--GC-G--G--GC-G--TGTG--T--C--CC----A--A---C-G  1044
SIVMNE     AAATG-T-----A--T--C--GC-A--GC-G--G--GC-G--TATG--T--C--CC----A--A---C-G  1044
SIVSMMH4   AAATG-T-----A--T--T--AC-A--GC-C--G--TC-G--CATG--T--C--TT----A--A---C-G  1047
SIVSMMPBJ  AAATG-C-----G--C--T--AT-G--GC-C--G--TC-G--TATA--T--C--TT----A--A---T-G  1047
```

HIV2/SIV GAG

```
CONSENSUS  AC?gCcTGtCAgGG?gTAGG?GGaCCagGcCAgAAaGCtAGatTAATGGCaGAaGCccT?AAaGAgGccc  1076
HIV2ROD    --CG-C--T--G--GG----T--G--AG-C--G--A--T--AT-------A--G--CC-G--A--G-TCA  1117
HIV2NIHZ   --CA-C--T--G--GG----G--A--AG-C--G--A--T--AT-------A--A--CC-A--G--G-TCA  1111
HIV2ISY    --CG-C--T--A--GA----T--A--AG-C--G--G--C--AC-------A--A--CT-A--A--G-CCA  1114
HIV2ST     --CG-C--T--G--GG----T--A--AG-C--G--A--C--AT-------A--A--CT-A--G--G-CCA  1117
HIV2BEN    --CG-C--C--G--GG----C--A--AG-C--G--A--C--GC-------T--A--CC-A--A--G-CTA  1117
HIV2D194   --TG-C--C--G--AG----C--A--AA-C--G--A--C--AC-------T--A--CC-A--G--G-CTT  1117
HIV2GH1    --TG-C--T--G--GG----T--A--AG-C--A--A--C--AC-------C--A--CC-G--G--G-CTT  1120
HIV2D205   --GG-C--C--A--GA----A--C--AG-G--G--G--A--GC-------C--A--CT-A--A--G-CCC  1102
SIVMM251   --GG-T--T--A--AG----G--A--AG-A--G--G--T--AT-------A--A--CC-G--A--G-CCC  1114
SIVMM142   --GG-C--T--A--AG----G--G--AG-A--G--G--T--AT-------A--A--CC-G--A--G-CCC  1114
SIVMM239   --GG-T--T--A--AG----G--G--GG-A--G--G--T--AT-------A--A--CC-G--A--G-CCC  1114
SIVMNE     --GG-T--T--A--AG----A--A--AG-A--A--G--A--AT-------A--A--CC-G--A--G-CCC  1114
SIVSMMH4   --AG-C--T--G--AG----A--G--AG-A--A--A--T--GT-------A--A--AT-G--A--G-CAC  1117
SIVSMMPBJ  --AG-C--T--G--AG----A--G--AG-A--A--A--T--AC-------A--G--AT-A--A--T-CAC  1117

CONSENSUS  TggcaCcagccCc?aTcCCaTTtGCaGCaGccCAacAg????????????AgaaaG?cAaTtAagTgtTG  1132
HIV2ROD    -AGGA-CTGCC-CTA-C--A--C--A--A-CC--GC-G.............-GAAA-GC-T-T-AA-GC--  1175
HIV2NIHZ   -GGCA-CAGCC-CCA-C--A--T--A--A-CC--GC-G.............-GAAA-AC-T-C-AG-GC--  1169
HIV2ISY    -GCGA-CAGCC-CTA-C--A--T--A--A-CC--AC-G.............-AAAG-GC-A-T-AG-GT--  1172
HIV2ST     -GGCA-CAGCC-CCA-C--A--T--A--A-CC--AC-G.............-GAAG-AC-A-T-AG-GC--  1175
HIV2BEN    -GGGA-CAAGC-CTA-C--A--T--A--A-CC--AC-A.............-GAAA-GC-A-T-GG-AT--  1175
HIV2D194   -GACG-CAGCC-CTA-C--A--T--A--A-CC--AC-A.............-GAAG-GC-A-T-GG-GT--  1175
HIV2GH1    -GACA-CACCC-CTA-C--A--T--A--A-CC--AC-A.............-GAAA-GT-A-T-GG-GT--  1178
HIV2D205   -AACA-CTGCA-CCA-A--G--T--T--C-TT--AC-AAAAGCAGGGAAG-GAGG-AC-G-G-CA-GC--  1172
SIVMM251   -CGCA-CAGTG-CAA-C--T--T--A--A-CC--GA-GAGGGGACCA...-GAAA-CC-A-T-AG-GT--  1181
SIVMM142   -CGCA-CAGCG-CAA-C--T--T--A--A-CC--AC-GAAGGGACCA...-GAAA-CC-A-T-AG-GT--  1181
SIVMM239   -CGCA-CAGTG-CAA-C--T--T--A--A-CC--AC-GAGGGGACCA...-GAAA-CC-A-T-AG-GT--  1181
SIVMNE     -TGCA-CAGGG-CAC-C--T--T--A--A-CC--AC-GAAGGGACCA...-GAAA-CC-A-T-AG-GT--  1181
SIVSMMH4   -GAGG-CAGAT-AAC-C--A--T--A--A-TC--AC-GAAAGGACAA...-GGAA-AC-A-C-AG-GT--  1184
SIVSMMPBJ  -GACG-AAGGG-CAC-C--A--T--A--A-TC--AC-GAAAGGACAA...-GAAA-AT-A-T-AG-GT--  1184

CONSENSUS  GAAtTGtGGaAAgqAgGGaCACtCgGCaAgaCAaTGcaGagCcCCtAGAAGaCAgGGcTGCTGGAAaTGT  1202
HIV2ROD    ---C--T--A--GG-A--G---T-G--A-GA--A--CC-AG-A--T-----G--G--C---------G---  1245
HIV2NIHZ   ---T--T--A--GG-A--A---T-G--A-GA--A--GA-CG-G--T-----A--G--C---------G---  1239
HIV2ISY    ---T--T--A--GG-A--G---T-G--A-GA--A--CC-AG-G--T-----A--G--C---------A---  1242
HIV2ST     ---T--C--A--GG-A--G---T-G--A-GA--A--CC-AG-A--T-----A--A--C---------A---  1245
HIV2BEN    ---C--T--A--GG-G--A---T-G--A-GA--G--CC-AG-A--T-----A--G--C---------G---  1245
HIV2D194   ---T--T--A--GG-G--A---T-G--G-AA--G--CC-AG-A--C-----A--G--C---------G---  1245
HIV2GH1    ---C--C--A--GG-G--A---T-G--G-GA--G--CC-AG-G--T-----A--G--C---------A---  1248
HIV2D205   ---C--T--C--AC-G--A---A-A--C-GG--A--CA-GG-C--T-----A--G--A---------A---  1242
SIVMM251   ---T--T--G--GG-G--A---T-T--A-GG--A--CA-AG-C--A-----A--G--A---------A---  1251
SIVMM142   ---T--T--G--GG-A--A---T-T--A-GG--A--CA-AG-C--A-----A--G--A---------A---  1251
SIVMM239   ---T--T--G--AG-G--A---T-T--A-GG--A--CA-AG-C--A-----A--G--A---------A---  1251
SIVMNE     ---T--T--G--AG-G--A---T-T--A-GG--A--CA-AA-C--A-----A--G--C---------A---  1251
SIVSMMH4   ---T--T--A--GG-G--A---T-T--A-AA--A--CA-GG-C--T-----A--G--C---------G---  1254
SIVSMMPBJ  ---C--T--A--GG-A--A---T-T--A-GA--A--CA-AG-C--T-----A--G--C---------G---  1254
```

```
                                                pol orf ->
CONSENSUS  GGaaA?acaGgaCA?gTcATGgCaAAaTGCCCagAaAGACAgGCgGGTTTTTTAGGgcTtGGcCcatGGG  1270
HIV2ROD    --TA-GCCA-GA--CA-C---A-A--C-----AG-T-----G--A----------AC-G--C-CTT---  1315
HIV2NIHZ   --TA-GTCA-GA--TG-C---G-A--T-----AG-T-----G--T----------GA-T--C-ACT---  1309
HIV2ISY    --CA-GTCA-GA--CA-C---G-A--C-----AG-T-----G--T----------GC-T--A-CAT---  1312
HIV2ST     --CA-GGCA-GA--CA-C---G-A--A-----AG-A-----G--G----------GT-G--C-CAT---  1315
HIV2BEN    --CA-GCCA-GA--CA-C---G-A--C-----GG-A-----G--A----------GT-G--C-CAC---  1315
HIV2D194   --CA-GTCA-GA--CA-C---G-A--C-----GG-A-----G--A----------GA-G--C-CAC---  1315
HIV2GH1    --CA-GACA-GA--CG-C---G-A--A-----AG-A-----G--A----------GA-G--T-CAT---  1318
HIV2D205   --AA-AACA-GA--CA-C---T-A--A-----AG-A-----G--G----------GT-A--A-CCT---  1312
SIVMM251   --AA-AATG-AC--TG-T---G-C--A-----AG-C-----G--G----------CC-T--T-CAT---  1321
SIVMM142   --AA-AATG-AC--TG-T---G-C--A-----AA-C-----G--G----------CC-T--C-CTT---  1321
SIVMM239   --AA-AATG-AC--TG-T---G-C--A-----AG-C-----G--G----------CC-T--T-CAT---  1321
SIVMNE     --AC-AATG-GC--TG-T---G-C--A-----AG-C-----G--A----------CT-T--C-CAT---  1321
SIVSMMH4   --AA-AACG-GT--TG-T---G-C--A-----TG-A-----G--G----------GC-T--C-CAT---  1324
SIVSMMPBJ  --AA-AGCA-GC--TG-T---G-T--A-----CG-A-----A--G----------GC-T--C-CAT---  1324

CONSENSUS  GAAAGaAGc?cCGCAA?TTCCCc?tGgccCaagtgcctCaGGGgcTg??????????????????????  1315
HIV2ROD    -----A--C-C-----C-----CGT-GCC-AAGTTCCG-A---GC-G.......................  1362
HIV2NIHZ   -----A--G-C-----C-----CGC-GCC-AAGTCCCG-A---GC-A.......................  1356
HIV2ISY    -----A--C-C-----C-----CGT-GTC-CAAGTTCG-A---GC-A.......................  1359
HIV2ST     -----A--C-C-----T-----TGT-GCC-AAATCCCG-A---GC-G.......................  1362
HIV2BEN    -----A--C-T-----C-----CGT-ACC-AAGCCCCT-A---GC-G.......................  1362
HIV2D194   -----C--C-C-----C-----CGC-GCC-AAGCTCCT-A---GC-G.......................  1362
HIV2GH1    -----A--C-C-----C-----CGT-GCC-AAGCTCCT-C---GC-G.......................  1361
HIV2D205   -----A--C-T-----C-----CAT-ACC-AAGTGCCT-A---AG-GACACCATCTGCACCCCCGATGAA  1382
SIVMM251   -----A--C-C-----T-----CAT-GCT-AAGTGCAT-A---GC-G.......................  1368
SIVMM142   -----A--C-C-----T-----CAT-GCT-AAGTGCAT-A---GC-G.......................  1368
SIVMM239   -----A--C-C-----T-----CAT-GCT-AAGTGCAT-A---GC-G.......................  1368
SIVMNE     -----A--C-C-----T-----CAT-GCC-AAATGCAT-A---GC-G.......................  1368
SIVSMMH4   -----A--C-C-----T-----CAT-GCC-AGATGCCT-A---GC-G.......................  1371
SIVSMMPBJ  -----A--C-C-----T-----CAT-GCC-AGATGCCT-A---GC-G.......................  1371

CONSENSUS  ????????????'????????????????AcaCCAaC?GCaCCcCCaq?qGAtCCAGcaGtGGAtcT?  1351
HIV2ROD    ..............................-CA---A-A--A--C--AGTG--T----CA-T---TC-A  1401
HIV2NIHZ   ..............................-CA---A-A--A--C--ATTG--T----CA-T---CC-A  1395
HIV2ISY    ..............................-CA---A-A--A--C--AATG--T----CA-T---CC-A  1398
HIV2ST     ..............................-CA---A-A--A--C--GATA--C----TA-A---CC-A  1401
HIV2BEN    ..............................-TA---A-A--A--T--GGCA--T----CA-C---AC-G  1401
HIV2D194   ..............................-TA---A-A--A--C--AATA--T----CA-T---CC-G  1401
HIV2GH1    ..............................-TA---A-A--A--C--AGCA--T----CA-T---TC-A  1404
HIV2D205   CCCAGCAGAGGGCATGACACCTCGGGGGGCG-CA---T-T--G--C--TGCA--T----CA-T---GA-G  1452
SIVMM251   ..............................-CG---A-T--T--C--AGAG--C----CT-T---TC-G  1407
SIVMM142   ..............................-CG---A-T--T--C--AGAA--A----CT-T---TC-G  1407
SIVMM239   ..............................-TG---A-T--T--C--AGAG--C----CT-T---TC-G  1407
SIVMNE     ..............................-CG---A-T--T--C--AGAG--C----CT-T---TC-G  1407
SIVSMMH4   ..................... .........-TA---A-T--T--C--AGAG--T----CT-T---TC-A  1410
SIVSMMPBJ  ..............................-CA---A-T--T--C--AGAG--T----CT-T---TC-A  1410
```

HIV2/SIV GAG

```
CONSENSUS  cTg?AgAa?TAcATGcAgc?gGGgA?aaagCAgAGaGAg????????????a?cagagagAgaCC?TACA  1403
HIV2ROD    C-GG-G-AA--T---C-GCAA--G-AAAGA--G--A--G...CAG.........AGAGAG-GA--A----  1459
HIV2NIHZ   C-GG-G-AA--T---C-GCAA--A-AAAGG--A--G--G...CAG.........AGAGAG-GA--G----  1453
HIV2ISY    C-GG-G-AG--C---C-GCAA--G-GAAAA--G--A--G...CAG.........AGACAA-GA--A----  1456
HIV2ST     C-AG-G-AG--C---C-GCAA--G-AAAGG--G--A--G...CAG.........AGAGAG-GG--A----  1459
HIV2BEN    T-GG-G-GA--T---C-GCAA--G-GAAAG--G--G--G...CAG.........AGGGAG-GA--A----  1459
HIV2D194   T-GG-G-AA--T---C-GCAA--G-GAAAG--G--A--G............CAGAGGGAG-GA--A----  1459
HIV2GH1    T-GG-A-GA--T---C-GCAG--G-GAGAG--G--G--G...CAG.........AGGGAG-GA--A----  1462
HIV2D205   C-GA-A-GT--C---C-GATG--G-GACAA--G--A--G............AGCCGAGAG-GA--C----  1510
SIVMM251   C-AA-G-AC--C---C-GTTG--C-AGCAG--G--A--A............AGCAGAGAG-AG--T----  1465
SIVMM142   C-AA-G-AC--C---C-CTTG--C-AGCAG--G--A--A............AGCAGAGGG-AG--T----  1465
SIVMM239   C-AA-G-AC--C---C-GTTG--C-AGCAG--G--A--AAAGCAGAGAGAAAGCAGAGAG-AG--T----  1477
SIVMNE     C-AA-A-AC--C---C-GTTG--C-AACAG--G--A--A............AGCAAAAGG-AG--T----  1465
SIVSMMH4   C-GA-G-AT--C---A-GATG--C-GAAAG--G--G--G............AACAGAGAG-GA--T----  1468
SIVSMMPBJ  C-GA-G-AT--C---A-GATG--C-GAAGG--G--G--G............AACAGAGAG-GA--T----  1468

CONSENSUS  AgGAgGTg????????????????????????????????????????????????????????ACaGAGGA  1419
HIV2ROD    -G--A--GACAGAGGACTTACTGCACCTCGAGCAGGGGGAGACACCATACAGGGAGCCACCA--A-----  1529
HIV2NIHZ   -G--A--GACAGAGGACTTACTGCGTTTCGAGCAGGCAGAGACACCATGCAGGGAG...ACG--A-----  1520
HIV2ISY    -A--A--GACAGAGGACTTGCTGCATCTCGAGCAAGGAGAGACACCACACAGAGAG...ACG--A-----  1523
HIV2ST     -A--A--GACAGAGGACTTCCTGCAGCTCGAGAAACAAGAGACACCATGCAGAGAG...ACG--A-----  1526
HIV2BEN    -A--G--GACGGAGGACTTGCTGCACCTCGAGCAGACAGAGACACCTCACAGAGAG...GAG--A-----  1526
HIV2D194   -G--G--GACGGAGGACTTACTGCACCTCGAGCAGGGAGAGACGCCCCACAGAGGG...GCG--A-----  1526
HIV2GH1    -A--G--GACGGAAGACTTACTGCACCTCGAGCAGGGAAAAGCACCTCACAGAGAG...GCG--A-----  1529
HIV2D205   -G--G--G.....................................................--A-----  1526
SIVMM251   -G--G--G.....................................................--A-----  1491
SIVMM142   -G--G--G.....................................................--A-----  1481
SIVMM239   -G--G--G.....................................................--A-----  1493
SIVMNE     -G--G--G.....................................................--A-----  1481
SIVSMMH4   -G--G--G.....................................................--A-----  1484
SIVSMMPBJ  -G--G--A.....................................................--G-----  1484

CONSENSUS  ?TTGCTgCACCTCAATTCTCTCTTTGGA?aaGACCAg***   1455
HIV2ROD    C-----G--------------------AAA-----G---   1569
HIV2NIHZ   C-----G--------------------AAA-----G---   1560
HIV2ISY    C-----G--------------------AAC-----G---   1563
HIV2ST     C-----G--------------------AAA-----G---   1566
HIV2BEN    C-----G--------------------AAA-----G---   1566
HIV2D194   C-----A--------------------AAA-----G---   1566
HIV2GH1    C-----G--------------------AAA-----G---   1569
HIV2D205   T-----G--------------------GAA-----G---   1566
SIVMM251   T-----G--------------------GGA-----G---   1521
SIVMM142   T-----G--------------------GGA-----G---   1521
SIVMM239   T-----G--------------------GGA-----G---   1533
SIVMNE     T-----G--------------------GAA-----G---   1521
SIVSMMH4   T-----G--------------------GAA-----G---   1524
SIVSMMPBJ  T-----G--------------------GAA-----A---   1524
```

Splice sites have been determined for SIVs only.

```
CONSENSUS  gTGtTGGAATTGtGGgAAgGagGGaCACtCtGCaAggCAaTGcagAgccCCaAGAAGACAgGGcTGcTGG    70
HIV2ROD                                                                              0
HIV2NIHZ                                                                             0
HIV2ISY                                                                              0
HIV2ST     G--C--------C--A--G-AA--G---T-G--A-GA--A--CCG-GCA--T--------A--C--C---    70
HIV2BEN                                                          ------G--C--C---    16
HIV2D194   G--T--------T--A--G-AG--A---T-G--G-AA--G--CCG-GCA--C--------G--C--C---    70
HIV2GH1                                                          ------G--C--C---    16
HIV2D205                                                         ------G--A--C---    16
SIVMM251   G--T--------T--G--G-AG--A---T-T--A-GG--A--CAG-GCC--A--------G--A--C---    70
SIVMM142   G--T--------T--G--G-AA--A---T-T--A-GG--A--CAG-GCC--A--------G--A--C---    70
SIVMM239   G--T--------T--G--A-AG--A---T-T--A-GG--A--CAG-GCC--A--------G--A--C---    70
SIVMNE     G--T--------T--G--A-AG--A---T-T--A-GG--A--CAG-ACC--A--------G--C--C---    70
SIVSMMH4                                                                             0
SIVSMMPBJ                                                                            0

CONSENSUS  AAaTGTGGaaAaac?Gg?CA?gtcATGgC?AAaTGCCCagAaAGACAgGcgGGTTTTTTAGGg?TtGGcC   135
HIV2ROD                                                ---G-CA-----------AC-G--C-    26
HIV2NIHZ                                               ---G-CT-----------GA-T--C-    26
HIV2ISY                                                ---G-CT-----------GC-T--A-    26
HIV2ST     --A-----CA-GGCA-GA--CATC---G-A--A-----AG-A-----G-CG-----------GT-G--C-   140
HIV2BEN    --G-----CA-GCCA-GA--CATC---G-A--C-----GG-A-----G-CA-----------GT-G--C-    86
HIV2D194   --G-----CA-GTCA-GA--CATC---G-A--C-----GG-A-----G-CA-----------GA-G--C-   140
HIV2GH1    --A-----CA-GACA-GA--CGTC---G-A--A-----AG-A-----G-CA-----------GA-G--T-    86
HIV2D205   --A-----AA-AACA-GA--CATC---T-A--A-----AG-A-----G-CG-----------GT-A--A-    86
SIVMM251   --A-----AA-AATG-AC--TGTT---G-C--A-----AG-C-----G-CG-----------CC-T--T-   140
SIVMM142   --A-----AA-AATG-AC--TGTT---G-C--A-----AA-C-----G-CG-----------CC-T--C-   140
SIVMM239   --A-----AA-AATG-AC--TGTT---G-C--A-----AG-C-----G-CG-----------CC-T--T-   140
SIVMNE     --A-----AC-AATG-GC--TGTT---G-C--A-----AG-C-----G-CA-----------CT-T--C-   140
SIVSMMH4                                            A-----G-CG-----------GC-T--C-    29
SIVSMMPBJ                                   A-----CG-A-----A-CG-----------GC-T--C-    38

CONSENSUS  catGGGGAAAGaAGcCcCgCAAcTTCCCcatGgccCaagt?cctCaGGGgcT?????????????????   186
HIV2ROD    CTT--------A--C-C-G---C-----CGT-GCC-AAGTTCCG-A---GC-..................    78
HIV2NIHZ   ACT--------A--G-C-G---C-----CGC-GCC-AAGTCCCG-A---GC-..................    78
HIV2ISY    CAT--------A--C-C-G---C-----CGT-GTC-CAAGTTCG-A---GC-..................    78
HIV2ST     CAT--------A--C-C-G---T-----TGT-GCC-AAATCCCG-A---GC-..................   192
HIV2BEN    CAC--------A--C-T-G---C-----CGT-ACC-AAGCCCCT-A---GC-..................   138
HIV2D194   CAC--------C--C-C-G---C-----CGC-GCC-AAGCTCCT-A---GC-..................   192
HIV2GH1    CAT--------A--C-C-G---C-----CGT-GCC-AAGCTCCT-C---GC-..................   138
HIV2D205   CCT--------A--C-T-G---C-----CAT-ACC-AAGTGCCT-A---AG-GACACCATCTGCACCCCC   156
SIVMM251   CAT--------A--C-C-G---T-----CAT-GCT-AAGTGCAT-A---GC-..................   192
SIVMM142   CTT--------A--C-C-G---T-----CAT-GCT-AAGTGCAT-A---GC-..................   192
SIVMM239   CAT--------A--C-C-G---T-----CAT-GCT-AAGTGCAT-A---GC-..................   192
SIVMNE     CAT--------A--C-C-G---T-----CAT-GCC-AAATGCAT-A---GC-..................   192
SIVSMMH4   CAT--------A--C-C-G---T-----CAT-GCC-AGATGCCT-A---GC-..................    81
SIVSMMPBJ  CAT--------A--C-C-G---T-----CAT-GCC-AGATGCCT-A---GC-..................    90
```

HIV2/SIV POL

```
CONSENSUS  ????????????????????????????????????gAcaCCAaCtGC?CCcCCag?gGAtCCAGcaGtG  218
HIV2ROD    ..................................G-CA---A-A--A--C--AGTG--T----CA-T-  112
HIV2NIHZ   ..................................A-CA---A-A--A--C--ATTG--T----CA-T-  112
HIV2ISY    ..................................A-CA---A-A--A--C--AATG--T----CA-T-  112
HIV2ST     ..................................G-CA---A-A--A--C--GATA--C----TA-A-  226
HIV2BEN    ..................................G-TA---A-A--A--T--GGCA--T----CA-C-  172
HIV2D194   ..................................G-TA---A-A--A--C--AATA--T----CA-T-  226
HIV2GH1    ..................................G-TA---A-A--A--C--AGCA--T----CA-T-  172
HIV2D205   GATGAACCCAGCAGAGGGCATGACACCTCGGGGGGCG-CA---T-T--G--C--TGCA--T----CA-T-  226
SIVMM251   ..................................G-CG---A-T--T--C--AGAG--C----CT-T-  226
SIVMM142   ..................................G-CG---A-T--T--C--AGAA--A----CT-T-  226
SIVMM239   ..................................G-TG---A-T--T--C--AGAG--C----CT-T-  226
SIVMNE     ..................................G-CG---A-T--T--C--AGAG--C----CT-T-  226
SIVSMMH4   ..................................G-TA---A-T--T--C--AGAG--T----CT-T-  115
SIVSMMPBJ  ..................................G-CA---A-T--T--C--AGAG--T----CT-T-  124

CONSENSUS  GAtcTgcTgaagAa?TAcATGcAgctgGG?Aaaaa???????????gCAgAGaGAgaacagagagAgaC  274
HIV2ROD    --TC-AC-GGAG-AA--T---C-GCAA--G-AAAG.............A--G--A--GCAGAGAGAG-GA-  170
HIV2NIHZ   --CC-AC-GGAG-AA--T---C-GCAA--A-AAAG.............G--A--G--GCAGAGAGAG-GA-  170
HIV2ISY    --CC-AC-GGAG-AG--C---C-GCAA--G-GAAA.............A--G--A--GCAGAGACAA-GA-  170
HIV2ST     --CC-AC-AGAG-AG--C---C-GCAA--G-AAAG.............G--G--A--GCAGAGAGAG-GG-  284
HIV2BEN    --AC-GT-GGAG-GA--T---C-GCAA--G-GAAA.............G--G--G--GCAGAGGGAG-GA-  230
HIV2D194   --CC-GT-GGAG-AA--T---C-GCAA--G-GAAA.............G--G--A--GCAGAGGGAG-GA-  284
HIV2GH1    --TC-AT-GGAA-GA--T---C-GCAG--G-GAGA.............G--G--G--GCAGAGGGAG-GA-  230
HIV2D205   --GA-GC-GAAA-GT--C---C-GATG--G-GACA. ...........A--G--A--GAGCCGAGAG-GA-  284
SIVMM251   --TC-GC-AAAG-AC--C---C-GTTG--C-AGCA.............G--G--A--AAGCAGAGAG-AG-  284
SIVMM142   --TC-GC-AAAG-AC--C---C-CTTG--C-AGCA.............G--G--A--AAGCAGAGGG-AG-  284
SIVMM239   --TC-GC-AAAG-AC--C---C-GTTG--C-AGCAGCAGAGAGAAAAG--G--A--AAGCAGAGAG-AG-  296
SIVMNE     --TC-GC-AAAA-AC--C---C-GTTG--C-AACA.............G--G--A--AAGCAAAAGG-AG-  284
SIVSMMH4   --TC-AC-GAAG-AT--C---A-GATG--C-GAAA.............G--G--G--GAACAGAGAG-GA-  173
SIVSMMPBJ  --TC-AC-GAAG-AT--C---A-GATG--C-GAAG.............G--G--G--GAACAGAGAG-GA-  182

CONSENSUS  C?TACAAqGAqGT?????????????????????????????????????????????????????gAC  289
HIV2ROD    -A-----G--A--GACAGAGGACTTACTGCACCTCGAGCAGGGGGAGACACCATACAGGGAGCCACCA--  240
HIV2NIHZ   -G-----G--A--GACAGAGGACTTACTGCGTTTCGAGCAGGCAGAGACACCATGCAGGGAGAC...G--  237
HIV2ISY    -A-----A--A--GACAGAGGACTTGCTGCATCTCGAGCAAGGAGAGACACCACACAGAGAGAC...G--  237
HIV2ST     -A-----A--A--GACAGAGGACTTCCTGCAGCTCGAGAAACAAGAGACACCATGCAGAGAGAC...G--  351
HIV2BEN    -A-----A--G--GACGGAGGACTTGCTGCACCTCGAGCAGAGAGAGACACCTCACAGAGAGGA...G--  297
HIV2D194   -A-----G--G--GACGGAGGACTTACTGCACCTCGAGCAGGGAGAGACGCCCCACAGAGGGGC...G--  351
HIV2GH1    -A-----A--G--GACGGAAGACTTACTGCACCTCGAGCAGGGAAAAGCACCTCACAGAGAGGC...G--  297
HIV2D205   -C-----G--G--.....................................................G--  300
SIVMM251   -T-----G--G--.....................................................G--  300
SIVMM142   -T-----G--G--.....................................................G--  300
SIVMM239   -T-----G--G--.....................................................G--  312
SIVMNE     -T-----G--G--.....................................................G--  300
SIVSMMH4   -T-----G--G--.....................................................G--  189
SIVSMMPBJ  -T-----G--G--.....................................................A--  198
```

```
                                                    <- gag
CONSENSUS  aGAGGAtTTGCTgCACCTCAATTCTCTCTTTGGAgaaGACCAgTAGTCAcaGC?tacaTtGAgGg?CAGc  357
HIV2ROD    A-----C-----G--------------------AAA-----G------CA--ATACA-T--G-GT---C  310
HIV2NIH2   A-----C-----G--------------------AAA-----G------CA--ACACA-T--G-GT---C  307
HIV2ISY    A-----C-----G--------------------AAC-----G------CA--ATACA-T--G-AT---C  307
HIV2ST     A-----C-----G--------------------AAA-----G------CA--ACATG-T--G-GC---C  421
HIV2BEN    A-----C-----G--------------------AAA-----G------CA--GTACA-C--G-AT---C  367
HIV2D194   A-----C-----A--------------------AAA-----G------CA--ATTCA-C--G-AT---C  421
HIV2GH1    A-----C-----G--------------------AAA-----G------CA--ATACA-C--G-TA---C  367
HIV2D205   A-----T-----G--------------------GAA-----G------AA--ATGTA-C--G-GT---T  370
SIVMM251   A-----T-----G--------------------GGA-----G------CT--TCATA-T--A-GA---C  370
SIVMM142   A-----T-----G--------------------GGA-----G------CT--TCATA-T--A-GA---C  370
SIVMM239   A-----T-----G--------------------GGA-----G------CT--TCATA-T--A-GA---C  382
SIVMNE     A-----T-----G--------------------GAA-----G------CT--TCATA-T--G-GA---C  370
SIVSMMH4   A-----T-----G--------------------GAA-----G------CT--CTACA-T--A-AA---C  259
SIVSMMPBJ  G-----T-----G--------------------GAA-----A------CG--CTACA-T--A-AA---C  268

CONSENSUS  C?GtAGAAGTatTatTaGAcACAgGgGcTgAtGAcTCaATaGTagCaGGaaTagAgtTaGGg??caAtTA  424
HIV2ROD    -A-T------CT-GT-A--C---G-G-C-G-C--C--A--A--AG-A--AA-AG-GT-A--GAACA-T--  380
HIV2NIH2   -A-T------TT-GT-A--C---A-G-C-A-C--C--A--A--AG-A--AA-AG-GT-A--GAGCA-T--  377
HIV2ISY    -A-T------TT-AC-A--C---G-G-C-G-C--C--A--A--AG-A--AA-AG-GT-A--GAGCA-T--  377
HIV2ST     -A-T------TT-GC-A--C---G-G-C-G-C--C--A--A--AG-A--CG-AG-GT-A--GAGCA-T--  491
HIV2BEN    -G-T------CT-AC-A--C---G-G-C-G-T--C--A--A--AG-A--AA-AG-AT-A--GGACA-T--  437
HIV2D194   -G-T------CT-AC-A--C---G-A-C-G-T--C--A--A--AG-A--AA-AG-GT-A--GGACA-T--  491
HIV2GH1    -G-T------CT-AT-A--C---G-G-C-G-T--C--A--A--GG-A--AA-AC-GT-A--GGACA-T--  437
HIV2D205   -A-T------AT-AC-A--C---G-A-T-G-C--C--A--A--AG-A--GA-AG-AT-A--TAGCA-T--  440
SIVMM251   -T-T------AT-AT-G--T---G-G-C-G-T--T--T--T--AA-A--AA-AG-GT-A--TCCAC-T--  440
SIVMM142   -T-T------AT-AT-A--T---G-G-C-G-T--T--T--T--AA-A--AA-AG-GT-A--TCCAC-T--  440
SIVMM239   -T-T------AT-AC-G--T---G-G-C-G-T--T--T--T--AA-A--AA-AG-GT-A--TCCAC-T--  452
SIVMNE     -T-C------AT-AT-A--T---G-G-C-G-T--T--T--T--AG-A--AA-AG-GT-A--TCCAC-T--  440
SIVSMMH4   -C-T------AT-AT-A--T---G-G-C-G-C--T--A--T--AG-A--GA-AG-AT-G--TCCAA-T--  329
SIVSMMPBJ  -C-T------AT-AT-A--T---G-G-C-G-T--T--A--T--AG-A--GA-TG-AT-G--TCCAA-T--  338

CONSENSUS  taccCCaAAaaTAGTaGGaGGaATAGGaGGaTTcATaAAtACcAAaGAatatAAaaATgTAgAaATAgaa  494
HIV2ROD    TAGC--A--AA----A--G--A-----G--A--C--A--T--C--G--ATAT--AA--G--G-A---GAA  450
HIV2NIH2   TAGT--A--AA----A--A--A-----G--A--C--A--T--C--G--ATAT--AA--G--G-A---GAA  447
HIV2ISY    TAGT--A--AA----A--G--A-----A--A--C--A--T--C--A--ATAT--AG--G--G-A---AGA  447
HIV2ST     TAGT--A--GA----A--G--A-----G--A--C--A--T--C--A--ATAT--AA--G--G-A---AGA  561
HIV2BEN    CACT--A--AA----A--G--A-----G--A--T--A--C--C--A--ATAC--AA--G--G-A---AAA  507
HIV2D194   CACT--A--AA----G--G--A-----G--A--C--A--T--C--A--ATAT--AA--G--G-A---AAG  561
HIV2GH1    CGTT--A--AA----G--G--A-----G--A--C--A--C--C--A--GATT--AA--A--G-A---AAA  507
HIV2D205   CACC--A--AA----A--A--G-----A--G--C--A--T--C--A--ATAC--AG--G--G-A---GAA  510
SIVMM251   TACC--A--AA----A--A--A-----A--T--T--T--T--T--A--ATAC--AA--G--A-A---GAA  510
SIVMM142   TACC--A--AA----A--A--A-----A--T--T--T--T--T--A--ATAC--AA--G--G-A---GAA  510
SIVMM239   TACC--A--AA----A--A--A-----A--T--T--T--T--T--A--ATAC--AA--G--G-A---GAA  522
SIVMNE     TACC--A--AA----A--A--A-----A--T--T--T--T--T--A--ATAC--AA--G--A-A---GAA  510
SIVSMMH4   TACC--T--AA----A--A--A-----A--C--C--T--C--C--A--ATAT--AG--G--A-A---AAA  399
SIVSMMPBJ  TACC--T--AA----A--A--A-----A--C--C--T--C--C--A--ATAT--AG--G--A-A---AAA  408
```

HIV2/SIV POL

```
CONSENSUS  GT?tTagg?AAaagagTaaaaGc?AC?ATaATGACaGG?GAtACcCCaAT?AAcATtTTTGGcAGaAAt?  557
HIV2ROD    --TC-AAAT--AAAGG-ACGG-CC--C--A-----A--C--C--C--A--C--C--T-----C--A--TA  520
HIV2NIHZ   --CC-AGGT--AAGGG-AAAA-CC--C--A-----A--T--T--C--G--C--C--T-----C--A--TG  517
HIV2ISY    --GC-AAAT--AAAGG-AAGA-CC--C--A-----A--T--T--C--A--C--C--T-----C--A--TA  517
HIV2ST     --AT-AAAT--AAGAG-AAGA-CC--C--A-----A--T--T--C--A--C--C--T-----C--A--CA  631
HIV2BEN    --AC-AAAT--AAGAG-AAGA-CC--C--A-----A--A--T--C--A--C--C--C-----C--A--TA  577
HIV2D194   --AC-AAAT--AAGAG-AAGA-CC--C--A-----A--A--T--C--A--C--C--T-----C--A--TA  631
HIV2GH1    --AC-AAAT--AAGAG-AAGA-CC--C--A-----A--A--T--C--A--C--C--C-----C--A--TA  577
HIV2D205   --AG-GGGA--AAGAG-AAGG-CA--T--A-----A--A--T--C--A--A--C--T-----C--A--TA  580
SIVMM251   --TT-AGGC--AAGGA-TAAA-GG--A--C-----A--G--C--T--G--T--C--T-----T--G--TT  580
SIVMM142   --TT-AGGC--AAGGA-TAAA-GG--A--C-----A--G--C--C--G--T--C--T-----T--A--TT  580
SIVMM239   --TT-AGGC--AAGGA-TAAA-GG--A--C-----A--G--C--C--G--T--C--T-----T--A--TT  592
SIVMNE     --TT-AGGC--AAGGA-TAAA-GG--A--C-----A--G--C--C--G--T--C--T-----T--A--TT  580
SIVSMMH4   --CT-AGGC--GGTAA-TAAG-GA--A--T-----G--A--T--C--A--T--T--T-----C--A--TT  469
SIVSMMPBJ  --AT-AGGC--AGTAA-TAAG-GA--A--T-----A--A--T--C--A--T--T--T-----C--A--TC

CONSENSUS  T?cTgacagCttTgGgcATGtCatTAAAtcT?CCagT?GC?AAggTaGAgCCaaTAAAagtaac?tTaAA  622
HIV2ROD    -TC-GACAG-CT-A-GC---T-AT----TC-A--AG-C--C--AG-A--G--AA----AATAATGC-A--  590
HIV2NIHZ   -TC-GACAG-TT-A-GC---T-AT----CC-G--AG-T--C--GA-A--A--AA----AATAATGC-A--  587
HIV2ISY    -CC-GACAG-CT-G-GC---T-AT----TT-A--AG-C--C--AA-A--A--AG----AGTAACAT-A--  587
HIV2ST     -TC-GACAG-CT-A-GC---T-AT----TC-A--AG-C--C--GA-A--A--AA----AATAATGC-G--  701
HIV2BEN    -TC-GACAG-CT-A-GC---T-AT----TT-A--AG-T--C--GA-A--G--AA----AGTAACAT-G--  647
HIV2D194   -TC-GGCAA-CT-A-GC---T-AT----CC-A--AG-C--C--GT-A--C--AA----AGTAACAT-G--  701
HIV2GH1    -CC-GACAG-CT-A-GC---T-AT----TT-A--AA-C--C--GA-A--A--AA----AGTAACGT-G--  647
HIV2D205   -TT-AAATA-CT-G-CC---A-TT----TT-C--AG-G--A--GG-A--A--AG----AGTTGAGT-A--  650
SIVMM251   -GC-AACAG-TC-G-GG---T-TC----TC-T--CA-A--T--GG-A--G--TG----AGTCACCT-A--  650
SIVMM142   -AC-AACAG-TC-G-GG---T-TC----TC-T--CA-A--T--GG-A--G--TG----GTCGCCCT-A--  650
SIVMM239   -GC-AACAG-TC-G-GG---T-TC----TT-T--CA-A--T--AG-A--G--TG----AGTCGCCT-A--  662
SIVMNE     -GC-AACAG-TC-G-AG---T-TC----TT-C--CA-A--T--GG-A--G--TG----AGTCACCT-A--  650
SIVSMMH4   -GC-AACAG-TA-G-GC---T-TT----TC-C--CA-A--T--GG-G--G--TA----AGTAACAC-A--  539
SIVSMMPBJ  -GC-AACAG-TA-G-GT---T-TC----TC-C--CA-A--T--GG-A--G--TA----AGTAACAC-A--  548

CONSENSUS  gCCaGGaAAaGatGGaCCAAaa?tgAgaCAaTGGCC?tTAtCaAaaGAaAAgATAga?GCacTaAgAGAa  689
HIV2ROD    G--A--G--A-AT--A----AACTG-GA--A-----CT--A-A-AA--A--A---GAA--AC-A-A---A  660
HIV2NIHZ   G--A--G--A-AT--A----GACTA-AA--A-----TT--A-A-AA--A--A---GAA--AC-A-A---A  657
HIV2ISY    G--A--A--A-AT--G----AACAA-GA--A-----CT--A-A-GA--A--A---GAA--AC-A-G---A  657
HIV2ST     G--A--A--G-AT--A----AACTG-GA--A-----CT--A-A-AA--A--A---GAG--AC-A-A---G  771
HIV2BEN    G--A--G--A-AT--A----GGCTG-AA--A-----CC--A-A-AA--G--A---GAA--AC-A-A---G  717
HIV2D194   G--A--G--A-AT--A----GGCTG-AA--A-----CC--A-A-AA--A--A---GAA--AC-A-A---A  771
HIV2GH1    G--A--G--A-AT--A----GACTG-GA--A-----CC--A-A-AA--G--A---GAA--AC-A-G---G  717
HIV2D205   A--T--A--A-AT--G----AGATC-GA--A-----TC--T-C-GG--A--G---CTA--CC-C-A---A  720
SIVMM251   G--A--A--G-TT--A----AATTG-AG--G-----AT--T-A-AA--A--G---GTT--AT-A-G---A  720
SIVMM142   G--A--A--G-AT--A----AATTG-AG--G-----AT--T-A-AA--A--G---GTT--AT-A-G---A  720
SIVMM239   G--A--A--G-AT--A----AATTG-AG--G-----AT--T-A-AA--A--G---GTT--AT-A-G---A  732
SIVMNE     G--A--A--A-AT--A----AATTG-GG--G-----AT--T-A-AA--A--G---GTT--AT-A-G---A  720
SIVSMMH4   A--A--A--A-AA--A----AATTA-GA--G-----GC--T-A-AA--A--G---ATT--AT-A-G---A  609
SIVSMMPBJ  A--A--A--A-AT--A----AATTA-GA--G-----GC--T-A-AA--A--G---ATT--AT-G-G---A  618
```

```
CONSENSUS  ATcTGtGAaAAaATGGAaAaaGA?GGcCAgcTaGAgGAaGC?CCtCCaACtAATCC?TAtAAtACcCCcA  756
HIV2ROD    --C--T--A--A-----A-AA--A--C--GC-A--G--A--A--T--A--T-----T--T--T--C--C-  730
HIV2NIHZ   --C--T--A--A-----A-AA--A--C--GC-A--A--A--A--T--A--T-----T--T--T--C--C-  727
HIV2ISY    --C--T--A--A-----A-GA--A--T--GC-A--A--A--G--T--A--T-----C--T--T--C--T-  727
HIV2ST     --C--T--G--A-----A-GA--G--C--GC-A--G--G--A--T--A--T-----T--T--T--C--C-  841
HIV2BEN    --C--T--A--A-----A-AA--G--C--GC-A--A--G--A--T--A--T-----T--T--T--C--C-  787
HIV2D194   --T--T--A--A-----A-GG--G--C--AC-A--A--A--A--T--A--T-----T--T--T--C--C-  841
HIV2GH1    --C--C--A--A-----A-AG--G--C--AC-A--A--G--A--T--A--T-----T--T--T--C--C-  787
HIV2D205   --C--T--A--A-----A-AG--G--A--AT-A--A--A--A--C--T--T-----A--C--C--A--C-  790
SIVMM251   --C--T--A--G-----A-AG--T--T--GT-G--G--A--T--C--G--C-----A--C--C--C--C-  790
SIVMM142   --C--T--A--G-  --A-AA--T--T--GT-G--G--A--T--C--G--C-----A--T--C--C--C-  790
SIVMM239   --C--T--A--G-----A-AG--T--T--GT-G--G--A--T--C--G--C-----A--C--C--C--C-  802
SIVMNE     --C--T--A--G-----A-AG--T--T--GT-G--G--A--T--C--G--C-----A--C--C--C--C-  790
SIVSMMH4   --C--T--A--A-----A-AA--T--C--GT-A--G--A--C--T--A--C-----G--T--C--C--C-  679
SIVSMMPBJ  --C--T--A--A-----G-AA--T--C--GT-A--G--A--C--T--A--C-----A--T--C--C--T-  688

CONSENSUS  CaTTTGC?ATaAAGAAaAAgGAcAAgAAcAAaTGGAGaATgCTaATAGAtTTtAGaGAacTaAAcAagGT  825
HIV2ROD    -A-----A--C-----A--G--C--A--C--A-----G--G--A-----T--C--A--AC-A--C-AG--  800
HIV2NIHZ   -A- ---A--C-----A--G--C--A--C--A-----A--G--A-----C--T--G--GC-A--C-AG--  797
HIV2ISY    -A-----A--T-----A--G--C--A--C--A-----G--G--G-----T--T--A--AC-A--C-AG--  797
HIV2ST     -A-----A--C-----A--G--C--A--C--A-----A--G--A-----T--T--A--AC-A--C-AG--  911
HIV2BEN    -A-----A--T-----A--G--C--G--C--A-----G--G--G-----T--T--A--AC-A--T-AG--  857
HIV2D194   -A-----A--T-----A--G--C--G--C--A-----A--G--A-----T--T--A--AC-A--C-GG--  911
HIV2GH1    -A-----A--T-----A--G--C--G--C--A-----G--G--A-----C--T--A--AC-A--C-GG--  857
HIV2D205   -C-----C--A-----G--A--T--G--C--A-----A--G--A-----C--T--A--GT-A--C-AG--  860
SIVMM251   -A-----C--A-----A--A--T--G--C--A-----A--G--G-----T--T--G--AC-A--T-GG--  860
SIVMM142   -A-----T--A-----A--G--T--A--C--A-----A--G--G-----T--T--G--AC-A--T-GG--  860
SIVMM239   -A-----T--A-----A--G--T--G--C--A-----A--G--G-----T--T--G--AC-A--T-GG--  872
SIVMNE     -A-----C--A-----A--G--C--G--C--A-----A--A--G-----T--T--G--AC-A--T-AG--  860
SIVSMMH4   -T-----T--A-----G--A--C--A--T--A-----G--G--A-----T--T--A--AT-G--T-AG--  749
SIVSMMPBJ  -A-----T--A-----G--A--C--G--T--A-----G--G--A-----T--T--A--AT-A--C-AG--  758

CONSENSUS  ?ACtCAaGAtTTcACaGAagT?CAgtTaGGaA?tCCaCACCC?GCAGGacTaGCaaAaAagAaaaGgATt  892
HIV2ROD    A--T--A--T--C--A--AA-T--GT-A--A--T--A-----A-----GT-G--CA-G-AG-GAA-A--T  870
HIV2NIHZ   A--T--G--T--C--A--AA-T--GT-A--A--T--A-----A-----AT-G--CA-A-AG-GGA-A--T  867
HIV2ISY    A--T--A--T--C--A--GG-T--GT-A--A--T--A-----A-----AT-A--CA-G-AA-GAA-A--T  867
HIV2ST     A--T--A--C--C--A--AA-C--GT-A--A--T--A-----A-----AC-A--C?-G-AG-AAC-A--T  981
HIV2BEN    G--T--A--T--C--A--AA-T--GC-A--A--T--A-----G-----AC-A--CA-A-AG-AAA-G--C  927
HIV2D194   G--T--A--T--C--A--AA-T--GC-A--A--T--A-----G-----AT-A--CA-A-AG-AAA-G--T  981
HIV2GH1    G--T--A--T--C--A--AA-T--GC-A--A--T--A-----G-----AT-A--CA-A-AG-AAA-G--C  927
HIV2D205   A--C--G--C--C--A--AG-.--AC-G--T--T--C-----G-----..-G--AG-A-AA-GGA-A--A  927
SIVMM251   C--T--G--C--T--A--AG-C--AT-A--A--A--A-----T-----AC-A--AA-A-GG-AAA-G--T  930
SIVMM142   C--T--A--C--T--G--AG-C--AT-A--A--A--A-----T-----AC-A--AA-A-GG-AAA-G--T  930
SIVMM239   C--T--G--C--T--G--AG-C--AT-A--A--A--A-----T-----AC-A--AA-A-GG-AAA-A--T  942
SIVMNE     C--T--G--C--T--A--AG-C--AT-G--A--A--A-----T-----AC-A--AA-A-GG-AGA-G--C  930
SIVSMMH4   C--T--A--C--T--A--AG-A--GT-A  ..  -A--A-----T-----AC-A--AA-G-GA-GGA-G--C  819
SIVSMMPBJ  C--T--A--T--T--A--AG-A--GT-A--G--A--A-----T-----AC-A--AA-G-GA-GAA-G--C  828
```

HIV2/SIV POL

```
CONSENSUS  aCagTacTaGATgTaGG?GAtGC?TAtTTttccATaCC?cTAgATgaaga?TTtAGgCagTAtACtGCaT   958
HIV2ROD    A-TG-AC-A---G-A--G--T--T--C--TTCC--A--AC--C--GAGGAC--T--A-CA--T--T--A-   940
HIV2NIHZ   A-TG-AC-A---G-A--G--T--T--C--TTCC--A--AC--C--GAGGAC--T--A-AG--T--T--A-   937
HIV2ISY    A-TG-GT-A---G-A--A--T--C--C--TTCC--A--CC--T--GAGGAT--T--A-AG--T--T--A-   937
HIV2ST     A-TG-CC-A---G-A--G--T--T--C--TTCC--A--AC--C--GAGGAT--T--A-AG--T--T--A-  1051
HIV2BEN    T-TA-AT-A---G-A--G--T--C--T--TTCC--A--AC--C--GAAGAT--T--G-AG--T--T--A-   997
HIV2D194   A-TG-AC-A---G-A--G--T--C--C--TTCC--A--AC--C--GAAGAT--T--G-AA--T--T--A-  1051
HIV2GH1    A-TG-AC-A---G-A--G--T--C--C--TTCC--A--GC--C--GAAGAT--T--G-AG--T--T--A-   997
HIV2D205   A-AG-AA-A---G-G--A--T--C--C--CAGT--C--AC--G--CCAAAC--C--A-AG--T--A--A-   997
SIVMM251   A-AG-AC-G---A-A--T--T--A--T--CTCC--A--TC--G--GAAGAA--T--G-AG--C--T--C-  1000
SIVMM142   A-AG-AC-G---A-A--T--C--A--T--CTCT--A--TC--G--GAAGAA--T--G-AG--C--T--C-  1000
SIVMM239   A-AG-AC-G---A-A--T--T--A--T--CTCC--A--TC--G--GAAGAA--T--G-AG--C--T--C-  1012
SIVMNE     A-AG-AC-G---G-A--T--C--A--T--CTCC--A--TC--G--GAAGAA--T--G-AG--C--T--T-  1000
SIVSMMH4   A-AG-AT-G---G-A--T--T--A--T--CTCC--A--TC--G--GAGGAA--C--G-AG--C--T--C-   889
SIVSMMPBJ  A-AG-GT-G---G-A--T--T--A--T--CTCC--A--TT--G--GAAGAA--C--G-AA--C--T--C-   898

CONSENSUS  TtACttTaCCAtCagTaAAcAATgcaGAaCCAGgAAAaaGATAcaT?TAtAAgGT?ttaCCaCAgCGaTG  1026
HIV2ROD    -T--TC-A---T-AG-G--C---GCA--A----G---AA----CA-A--T--A--CTTG--A--G--A--  1010
HIV2NIHZ   -C--TC-A---T-AG-A--C---GCA--A----G---AA----TA-A--C--A--CCTA--A--G--G--  1007
HIV2ISY    -T--TC-G---T-AG-A--C---GCA--A----G---AA----TA-A--C--A--CTTA--A--G--A--  1007
HIV2ST     -T--TC-A---T-AA-A--C---GCT--A----G---AA----CA-A--T--A--CTCA--A--G--A--  1121
HIV2BEN    -T--CC-A---G-AG-A--C---ATG--A----G---AA----TA-A--T--A--CTTG--A--A--A--  1067
HIV2D194   -T--CC-A---T-AG-A--C---GCA--G----A---AA----TG-A--T--G--CTTA--A--A--A--  1121
HIV2GH1    -T--CC-A---T-AG-A--C---GCA--A----G---AA----TA-A--T--G--CTTA--A--A--A--  1067
HIV2D205   -C--CT-G---T-AG-A--C---GCA--G----G---GA----CA-T--T--A--TCTA--A--A--G--  1067
SIVMM251   -T--TT-A---T-AG-A--T---GCA--G----G---AC----CA-T--T--G--TCTG--T--G--A--  1070
SIVMM142   -T--TT-A---T-AG-A--T---GCA--G----G---AC----CA-T--T--G--TCTG--T--G--A--  1070
SIVMM239   -T--TT-A---T-AG-A--T---GCA--G----G---AC----CA-T--T--G--TCTG--T--G--A--  1082
SIVMNE     -T--TT-A---T-AG-A--T---GCA--A----G---AC----CA-T--T--G--TCTG--T--G--G--  1070
SIVSMMH4   -T--TT-A---T-AG-A--T---GCA--A----G---AA----CA-C--T--G--ATTA--T--A--G--   959
SIVSMMPBJ  -T--CT-A---T-GG-A--C---GCA--A----G---AA----CA-C--T--G--ATTA--T--A--G--   968

CONSENSUS  GAAgGGgTCaCcagCaATtTttCAatacac?ATGAGacA?gT?tTaGAaCC?TTcAGaAaaGCaAAccca  1092
HIV2ROD    ---G--A--A-CAG-A--T-TT--ACACACA-----AC-GG-AT-A--A--A--C--A-AA--A--CAAG  1080
HIV2NIHZ   ---G--G--A-CAG-A--T-TT--ATACACA-----GC-GA-CT-A--A--A--C--A-AA--A--CGAG  1077
HIV2ISY    ---G--G--A-CAG-A--T-TT--ATACACA-----GC-AG-CT-A--A--A--C--A-AA--A--CCCA  1077
HIV2ST     ---G--A--A-CAG-A--T-TT--GTACACA-----GC-GG-CT-A--A--A--C--A-AA--A--CCCG  1191
HIV2BEN    ---G--A--A-CAG-A--T-TT--ATACACA-----GC-AG-CT-A--A--T--C--A-AA--A--CCCA  1137
HIV2D194   ---A--A--A-CAG-A--C-TT--ATTCATG-----GC-AA-CT-A--A--T--C--A-AA--A--CCCA  1191
HIV2GH1    ---G--A--A-CAG-A--T-TT--ACACACA-----GC-AG-CT-A--A--T--C--A-AA--A--CCCA  1137
HIV2D205   ---G--G--C-AGT-A--T-GT--ATACTCC-----AA-GG-AT-A--T--T--T--A-AG--C--CAGC  1137
SIVMM251   ---G--G--A-CAG-C--C-TC--ATACACT-----AC-TG-GC-A--A--C--C--G-AG--A--TCCA  1140
SIVMM142   ---G--G--A-CAG-C--C-TC--ATACACT-----AC-TG-GC-A--A--C--C--G-AG--A--TCCA  1140
SIVMM239   ---G--G--A-CAG-C--C-TC--ATACACT-----AC-TG-GC-A--A--C--C--G-AG--A--TCCA  1152
SIVMNE     ---G--G--A-CAG-C--C-TC--ACACACT-----AA-TG-GC-G--A--C--C--G-AG--A--TCCA  1140
SIVSMMH4   ---G--G--A-CAG-T--T-TT--GTATACT-----AA-TG-AT-A--A--T--C--A-AA--A--TCCA  1029
SIVSMMPBJ  ---A--A--A-CAG-T--T-TT--GCATACT-----AA-TG-CT-A--A--C--T--A-AA--A--TCCA  1038
```

```
CONSENSUS  GAtgTcAtt?T?aTcCAgTAcATGGATGAtATctTaaTAGCtAGtGAcAGgAcaGatttaGAaCAtGAcA  1160
HIV2ROD    --TG-C-TTA-CA-T--G--C---------T--CT-AA----T--T--C--G-CA-ATT-A--A--T--T-  1150
HIV2NIHZ   --TG-C-TTA-CA-T--G--C---------T--CT-AA----C--C--C--G-CA-ACT-A--A--T--C-  1147
HIV2ISY    --TG-C-TTA-CG-T--G--C---------T--CT-AA----T--T--C--G-CA-ATT-G--A--T--C-  1147
HIV2ST     --TA-C-TTC-CA-T--G--C---------T--CT-GA----C--C--C--G-CA-ATT-A--A--T--C-  1261
HIV2BEN    --TG-C-TTC-CA-C--G--C---------T--CT-AA----T--T--C--G-CA-GTT-A--G--T--C-  1207
HIV2D194   --CG-C-TTC-CA-C--A--C---------T--CT-AA----T--T--C--G-CG-GTT-A--G--T--C-  1261
HIV2GH1    --TG-C-TTC-CA-C--A--C---------T--TT-AA----T--T--C--G-CA-GTT-A--G--T--C-  1207
HIV2D205   --TG-C-TTA-AA-T--G--C---------C--CC-TA----A--T--C--A-GT-ATC-G--G--C--C-  1207
SIVMM251   --TG-G-CCT-AG-C--G--T---------C--CT-AA----T--T--C--G-CA-ACC-G--A--T--C-  1210
SIVMM142   --TG-G-CCT-AG-C--G--T---------C--CT-AA----T--T--C--G-CA-ACC-G--A--T--C-  1210
SIVMM239   --TG-G-CCT-AG-C--G--T---------C--CT-AA----T--T--C--G-CA-ACC-G--A--T--C-  1222
SIVMNE     --TG-G-CCT-AG-C--G--T---------C--CT-AG----T--T--C--G-CA-ACC-G--A--T--C-  1210
SIVSMMH4   --TG-G-CCC-GA-C--A--C---------C--CT-AA----T--T--T--A-CA-ATT-A--G--T--C-  1099
SIVSMMPBJ  --TG-G-CCC-GA-C--A--C---------C--CT-AA----T--T--T--A-CA-ATT-A--A--T--C-  1108

CONSENSUS  ggGTaGTtttaCAgctaAAgGAacTtcTaAAtggc?TaGGaTTtTCtaCcCCaGAaGAgAAgTTcCAaAA  1229
HIV2ROD    GG--A--CCTG--GCTC--G--AC-TC-A--TGGCC-A--A--T--TA-C--A--T--G--G--C--A--  1220
HIV2NIHZ   AA--G--CCTG--GCTA--A--AC-TC-A--TGGAC-A--A--T--CA-C--A--T--G--A--C--G--  1217
HIV2ISY    AA--A--CCTA--GCTA--G--AC-TC-A--TGGCC-G--A--T--CA-C--A--C--A--G--C--A--  1217
HIV2ST     GA--G--TCTG--GCTA--G--AC-TC-A--TGGCC-G--A--T--CA-C--A--T--G--G--C--A--  1331
HIV2BEN    AA--G--CCTG--GCTA--A--AC-TC-A--TGGCC-A--G--T--TA-T--A--T--G--G--C--A--  1277
HIV2D194   AA--A--CCTG--ACTA--A--AC-TC-G--TGGCC-A--G--C--TA-C--A--T--G--G--C--A--  1331
HIV2GH1    AA--A--CCTG--GCTA--A--AC-TC-G--TGGCC-A--G--C--TA-C--A--T--G--G--C--A--  1277
HIV2D205   GG--A--GTCC--ACTA--A--GT-AT-A--TGACA-G--A--C--TA-C--A--A--A--G--C--A--  1277
SIVMM251   GG--A--TTTA--GCTA--G--AC-CT-A--TAGCA-A--G--C--TA-C--A--A--G--A--C--A--  1280
SIVMM142   GG--A--TTTA--GTTA--A--AC-CT-A--TAGCA-A--G--T--AT-C--A--A--G--A--C--A--  1280
SIVMM239   GG--A--TTTA--GTCA--G--AC-CT-G--TAGCA-A--G--T--TA-C--A--A--G--A--C--A--  1292
SIVMNE     GG--A--TTTA--GTTA--G--AC-CT-A--TAGCA-A--G--T--TA-C--A--A--G--G--C--A--  1280
SIVSMMH4   GG--A--TTTA--GTTA--G--AC-TC-G--CGGCA-A--A--C--TA-C--A--A--G--G--C--G--  1169
SIVSMMPBJ  GG--A--TTTA--GTTG--G--AC-TC-G--CAGTA-A--A--T--CA-C--A--A--G--G--C--G--  1178

CONSENSUS  aGAcCCtCCatttcaaTGGATGGG?TatGAatTgTGGCCaAc?AAaTGGAAgcTGCAaAaaaTAcAgtTg  1297
HIV2ROD    A--C--T--ATACCAC--------C-AT--AC-A-----A-CT--A-----GT----G-AAA--C-GT-G  1290
HIV2NIHZ   A--C--T--ATATCGC--------C-AT--AT-A-----A-CT--A-----GT----A-AAA--C-GC-G  1287
HIV2ISY    G--C--T--ATACCAA--------C-AT--AC-G-----A-CC--A-----AT----A-AAA--C-AT-G  1287
HIV2ST     A--C--T--ATACCAA--------C-AT--AC-G-----A-CT--A-----GC----A-GAA--C-AT-G  1401
HIV2BEN    A--C--T--ATTTCAA--------C-GT--AC-A-----A-CT--A-----GC----G-AAC--C-AC-G  1347
HIV2D194   G--C--T--GTTTCAA--------C-AT--AT-G-----A-CT--A-----AC----G-AAA--C-AT-A  1401
HIV2GH1    A--C--T--GCTTCAA--------C-AT--AC-A-----A-CT--A-----GC----G-AAT--C-AC-G  1347
HIV2D205   A--C--T--GTTCAAA--------T-AT--GC-C-----A-AA--G-----AC----A-AAA--C-AC-G  1347
SIVMM251   A--T--C--ATTTCAA--------G-AC--AT-G-----G-CA--A-----GT----A-AGA--G-GT-G  1350
SIVMM142   A--T--C--ATTTCAA--------G-AC--AT-G-----G-CA--A-----GT----A-AGA--G-GT-G  1350
SIVMM239   A--T--C--ATTTCAA--------G-AC--AT-G-----A-CA--A-----GT----A-AGA--G-GT-G  1362
SIVMNE     A--T--C--ATTTCAA--------G-AT--AT-G-----A-CA--A-----GT----A-AGA--G-GT-G  1350
SIVSMMH4   A--T--C--ATTCCAG--------A-AT--AT-G-----A-CC--A-----AC----G-AAA--G-GT-G  1219
SIVSMMPBJ  A--T--C--ATTCCAG--------A-AT--AT-G-----C-CC--A-----AC----A-AAA--G-GT-G  1248
```

HIV2/SIV POL

```
CONSENSUS  CC?cAaAaaGAaat?TGGACAGT?AAtGacATcCAaAAgcTaGTaGGaGT?tTaAAtTGGGCaGCaCAaa  1363
HIV2ROD    --CC-A-AA--AATA--------C--T-AC--C--G--GC-A--G--T--CC-A--T-----A--A--AC  1360
HIV2NIHZ   --CC-A-AA--AGTA--------C--T-AC--C--A--GC-A--G--T--CC-A--T-----A--A--AA  1357
HIV2ISY    --CC-A-AG--AGTA--------T--T-AC--C--G--AC-A--G--T--CC-A--C-----G--A--AA  1357
HIV2ST     --CC-A-AG--AGTA--------C--T-AC--C--A--AC-G--G--T--CC-A--T-----A--A--AA  1471
HIV2BEN    --CC-G-AA--CATA--------C--T-AC--C--A--GC-A--G--A--CT-A--T-----G--A--AA  1417
HIV2D194   --TC-G-AA--AATA--------C--T-AC--C--A--AC-A--A--A--TT-G--C-----G--G--GA  1471
HIV2GH1    --CC-G-AA--AATA--------C--C-AC--C--G--GC-G--G--A--CT-A--T-----G--A--AA  1417
HIV2D205   --AG-A-AA--AGTT--------G--T-CA--T--A--AC-G--A--A--AT-A--C-----A--T--AC  1417
SIVMM251   --AC-A-GA--GACC--------G--T-AT--A--G--GT-A--A--A--AT-A--T-----A--T--AA  1420
SIVMM142   --AC-A-GA--GACC--------G--T-AT--A--G--GT-A--A--A--AT-A--T-----A--T--AA  1420
SIVMM239   --AC-A-GA--GACC--------G--T-AT--A--G--GT-A--A--A--AT-A--T-----A--T--AA  1432
SIVMNE     --AC-A-AA--GACC--------G--T-AT--A--G--GT-A--A--A--AT-A--T-----A--T--AA  1420
SIVSMMH4   --AC-A-GA--GACC--------A--T-AC--A--A--AT-A--A--A--GC-A--T-----A--A--AA  1309
SIVSMMPBJ  --AC-A-GA--AACC--------A--T-AT--A--A--AT-A--A--A--GC-A--T-----A--A--AA  1318

CONSENSUS  TcTatcCaGG?ATaAAaACcAaacAtcT?TGtAggtTaATtAgaGGaAAaATGaCact?ACaGAaGAaGT  1430
HIV2ROD    -C-ACC-A--G--A--G--C-AAC-CT-A--T-GGT-A--C-GA--A--A---A-ACTC--A--A--A--  1430
HIV2NIHZ   -C-ACC-A--G--A--G--C-AAC-CC-A--T-GGC-A--T-GA--A--A---A-ACTC--A--A--G--  1427
HIV2ISY    -C-ACC-A--A--A--G--C-AAC-CT-A--T-AGC-A--T-GA--A--G---A-ACCC--G--A--A--  1427
HIV2ST     -C-ACC-A--G--A--G--C-GAA-CT-A--T-CGT-A--C-GA--A--A---A-ACTC--A--A--G--  1541
HIV2BEN    -C-ATT-A--A--A--A--C-AAC-CT-A--T-GAC-A--T-GA--A--A---A-ACTC--A--A--A--  1487
HIV2D194   -C-ATC-A--G--A--A--C-AGC-TT-A--T-AAT-G--T-GA--A--A---A-ACTC--A--G--A--  1541
HIV2GH1    -C-ATC-A--A--A--A--C-AAC-TT-G--T-GGC-A--C-AA--A--A---A-ACTC--A--A--A--  1487
HIV2D205   -C-TTC-T--A--T--G--A-GGC-CA-A--C-AAC-A--T-GG--A--G---A-CCTA--A--A--A--  1487
SIVMM251   -T-ATC-A--T--A--A--C-AAC-TC-C--T-GGT-A--T-GA--A--A---A-TCTA--A--G--A--  1490
SIVMM142   -T-ATC-A--T--A--A--C-AAC-TC-C--T-GGT-A--T-GA--A--A---A-TCTA--A--G--A--  1490
SIVMM239   -T-ATC-A--T--A--A--C-AAC-TC-C--T-GGT-A--T-GA--A--A---A-TCTA--A--G--A--  1502
SIVMNE     -T-ATC-A--T--A--A--C-AAC-TC-C--T-GGT-A--T-GA--A--A---A-TCTA--A--G--A--  1490
SIVSMMH4   -T-ATC-A--A--A--G--T-AAC-TC-T--C-GAC-A--C-GA--A--A---A-TTTA--A--A--G--  1379
SIVSMMPBJ  -T-ATC-A--A--A--G--T-AAC-TC-T--C-GAT-A--C-GA--A--A---A-TTTA--A--A--G--  1388

CONSENSUS  ?CAGTGGACaGAg?TaGCAGAaGCaGAact?gaaGAaAAcAaaAT?ATccT?agtCAggAaCAaGAaGGa  1495
HIV2ROD    A--------A--AT-A-----A--A--GCTAGAA--A--C-GA--T--CC-AAGC--GG-A--A--G--A  1500
HIV2NIHZ   A--------A--AT-A-----A--A--GCTAGAG--A--C-GA--T--CT-AAGC--GA-A--A--A--A  1497
HIV2ISY    A--------A--AT-A-----A--A--GCTGGAG--A--C-AA--T--CT-AAGC--GG-A--G--G--A  1497
HIV2ST     A--------A--AT-A-----A--G--ACTAGAA--A--C-AA--C--CT-AAGC--GG-A--A--A--A  1611
HIV2BEN    G--------A--AC-A-----A--A--GCTAGAA--A--C-AA--T--CT-GAGC--GG-A--A--A--A  1557
HIV2D194   A--------A--GT-A-----G--A--ACTAGAA--A--C-AA--T--CT-AAGT--GG-A--A--G--A  1611
HIV2GH1    G--------A--GT-A-----A--A--GCTAGAA--A--C-AA--T--CT-AAGC--GG-A--A--A--A  1557
HIV2D205   A--------A--AC-A-----A--A--GCTACAG--G--T-AA--C--CT-AGAA--GG-A--A--A--A  1557
SIVMM251   T--------T--GA-G-----A--A--ATATGAG--A--T-AG--A--TC-CAGT--GG-A--A--A--A  1560
SIVMM142   T--------T--GA-G-----A--A--ATATGAG--A--T-AA--A--TC-CAGT--GG-A--A--A--A  1560
SIVMM239   T--------T--GA-G-----A--A--ATATGAG--A--T-AA--A--TC-CAGT--GG-A--A--A--A  1572
SIVMNE     T--------T--GA-G-----G--A--ATATGAG--A--T-AA--A--TC-CAGT--GG-A--A--A--A  1560
SIVSMMH4   T--------T--GA-G-----G--A--ATATGAA--A--C-AG--A--TC-CAGT--AG-A--A--A--A  1449
SIVSMMPBJ  T--------T--GA-G-----G--A--ATATGAA--A--C-AG--A--TC-CAGT--AG-A--A--A--A  1458
```

```
CONSENSUS  t??TAtTAccaaGAagaaaaaccatTAGAaGCaACagT??taAAgaatCaaGacaATCAgTGG?CaTAtA  1560
HIV2ROD    CAC--T--CCAA--AGAAAAAGAGC----A--A--AG-CCAA--GGAT-AA-AGA----G---A-A--T-  1570
HIV2NIHZ   CAC--T--CCAG--AGAAAAAAAGT----A--A--AG-CCAG--AGAT-AA-ACA----G---A-A--T-  1567
HIV2ISY    CAC--T--CCAA--GGAAAAAGAGT----A--A--AG-TCAA--GGAT-AA-ACA----G---A-A--T-  1567
HIV2ST     TGC--T--CCAA--GGAAAAGGAGC----A--A--AG-CCAA--AGAT-AA-ACA----G---A-A--T-  1681
HIV2BEN    TAT--T--CCAA--AGAAAAAGAAT----G--A--AA-CCAA--AAGC-AA-GAC----A---A-A--C-  1627
HIV2D194   TCC--C--TCAG--AGAACAAGAAC----A--A--AG-CATC--AAGC-AA-ACA----G---G-A--C-  1681
HIV2GH1    TAC--T--CCAA--AGAAAAAGAAT----G--A--AA-CCAA--AAAT-AA-ACA----G---A-A--C-  1627
HIV2D205   TCC--C--CAAG--AAGGGTACCGC----A--A--AG-ACAG--AAAC-TA-CAA----G---A-A--C-  1627
SIVMM251   TGT--T--CCAA--AGGCAAGCCAT----A--C--GG-AATA--GAGT-AG-ACA----G---T-T--T-  1630
SIVMM142   TGT--T--CCAA--AAGCAAGCCAT----A--C--GG-GATA--GAGT-AG-ACA----G---T-T--T-  1630
SIVMM239   TGT--T--CCAA--AGGCAAGCCAT----A--C--GG-AATA--GAGT-AG-ACA----G---T-T--T-  1642
SIVMNE     TGT--T--CCAA--AGGCAAGCCAT----G--C--GG-AATA--GAAT-AG-ACA----G---T-T--T-  1630
SIVSMMH4   TGT--C--CCAA--AGGAAAACCAA----G--A--AG-AATA--GAGT-AG-ATA----A---T-A--T-  1519
SIVSMMPBJ  TGT--C--TCAG--GGGAAAGCCAC----G--A--AG-AATA--GAGT-AG-ATA----A---T-A--T-  1528

CONSENSUS  AaaTaCAcCAgGaaga?AaaaT?cTaAAAGtaGGaAAatttgcaAAgataAAaAAtac?CataccaAtgG  1627
HIV2ROD    -AA-A--C--G-AAGAA-AAA-TC-A----TA--A--ATATGCA--GGTG--A--CACC-ATACCA-TG-  1640
HIV2NIHZ   -GG-A--C--G-GGGAG-AAA-TC-C----GT--G--AATATGC--AGAT--A--ATAC-CATACC-AC-  1637
HIV2ISY    -AG-A--C--G-GAGAA-AAA-TC-A----TA--A--ATATGCA--GATA--A--TACC-ATACCA-CG-  1637
HIV2ST     -GA-A--C--G-GAGGA-AAA-TC-A----TA--A--ATATGCA--GGTA--A--TACC-ACACCA-CG-  1751
HIV2BEN    -AA-A--C--G-AAGAG-AAA-CC-A----TA--A--GTATGCA--GATA--A--TACC-ATACCA-TG-  1697
HIV2D194   -AA-A--C--G-GAGAG-GGG-TC-A----TA--A--GTATGCG--GATA--A--TACT-ATACCA-TG-  1751
HIV2GH1    -AA-A--C--G-AAGAG-AAA-CC-A----TA--G--ATATGCA--GATA--A--TACC-ATACCA-TG-  1697
HIV2D205   -AA-T--T--G-GAAAT-AAG-CC-A----TA--A--ATATGCA--GGTT--A--CACG-ACACCA-CG-  1697
SIVMM251   -AA-T--C--A-AAGAC-AAA-AC-G----TA--A--ATTTGCA--GATA--G--TACA-ATACCA-TG-  1700
SIVMM142   -AA-T--C--A-AAGAC-AAA-AC-G----TA--A--ATTTGCA--GATA--G--TACA-ATACCA-TG-  1700
SIVMM239   -AA-T--C--A-AAGAC-AAA-AC-G----TA--A--ATTTGCA--GATA--G--TACA-ATACCA-TG-  1712
SIVMNE     -GA-T--C--A-AAGAC-AAA-AC-A----TA--A--ATTTGCA--GATA--A--TACA-ATACCA-TG-  1700
SIVSMMH4   -AA-T--C--A-AAGAC-AAG-AC-G----TA--T--ATTTGCA--GGTT--A--TACA-ATACAA-TG-  1589
SIVSMMPBJ  -AA-T--C--A-AAGAC-AAA-AC-G----TA--C--ATTTGCA--GATT--A--TACA-ATACAA-TG-  1598

CONSENSUS  agTcAGatTatTaGCACA?GTagTtCAgAAaATAGGaAAaGAaGCacTaGTcATtTGGGGAcagaT?CCa  1695
HIV2ROD    AA-C--AT-GT-A-----G--AG-T--G--A-----A--A--A--AC-A--C--T------CGAA-A--A  1710
HIV2NIHZ   GG-C--AT-GT-A-----G--AG-T--A--A-----A--A--A--AC-A--C--T------CGGA-A--A  1707
HIV2ISY    GG-C--GT-GT-G-----G--AG-T--G--A-----A--A--A--AC-A--C--T------CGAA-A--A  1707
HIV2ST     AG-C--AC-CC-A-----A--AG-T--A--A-----A--A--A--AC-A--C--T------CGAA-A--A  1821
HIV2BEN    GG-C--AT-AC-A-----G--AG-T--G--A-----A--A--G--AC-A--C--T------CGGA-A--A  1767
HIV2D194   GG-C--AC-AC-A-----A--AG-C--A--A-----A--G--A--AC-G--C--T------CGAG-G--A  1821
HIV2GH1    GG-C--AT-AC-A-----G--AG-T--G--A-----A--A--A--AC-A--C--C------CGGA-A--A  1767
HIV2D205   GG-A--AC-AC-G-----T--AG-T--G--A-----C--A--A--CC-A--C--C------GAGA-A--A  1767
SIVMM251   AG-T--AC-AT-A-----T--AA-A--G--A-----A--G--A--AA-A--G--C------CAGG-C--A  1770
SIVMM142   AG-T--AC-AT-A-----T--AA-A--G--A-----A--G--A--AA-A--G--C------CAGG-C--A  1770
SIVMM239   AG-G--AC-AT-A-----T--AA-A--G--A-----A--G--A--AA-A--G--C------CAGG-C--A  1782
SIVMNE     AG-T--AC-AT-A-----T--AA-A--G--A-----A--G--A--AA-A--G--C------CAGG-C--A  1770
SIVSMMH4   AG-C--AT-AC-A-----C--AG-G--G--A-----A--A--A--AC-A--A--T------GAGG-G--A  1659
SIVSMMPBJ  AG-C--AT-AT-A-----T--AG-G--G--A-----A--A--A--AA-A--A--T------CAGG-G--G  1668
```

HIV2/SIV POL

```
CONSENSUS  aaaTTcCActTaCCaGTaGAgAgaGA?atcTGGGAaCAgTGGTGGgcagAcTAtTGGCAaGTaACaTGGa  1764
HIV2ROD    AAA--T--CC-A--A--A--G-GA--AATC-----G--G------GATA-C--C-----A--G--A---A  1780
HIV2NIHZ   AAA--T--CC-A--A--A--G-GA--TACC-----G--G------GATA-C--C-----A--A--A---A  1777
HIV2ISY    AAA--T--CC-A--A--A--A-GA--GACC-----A--G------GATA-C--T-----A--G--A---A  1777
HIV2ST     AAA--T--CC-A--A--A--A-GA--TACC-----A--G------GATA-C--C-----A--G--A---A  1891
HIV2BEN    AAA--T--CC-G--A--G--G-GA--GACC-----G--G------GATA-C--C-----A--G--A---A  1837
HIV2D194   AAA--T--CC-A--G--A--G-GA--CACC-----G--A------GATA-C--C-----A--A--A---G  1891
HIV2GH1    AAA--C--CT-A--G--G--G-GA--GACC-----G--G------GATA-C--C-----A--G--A---A  1837
HIV2D205   GTG--C--TC-G--A--A--A-GA--GACA-----C--G------ACAG-T--C-----A--A--C---A  1837
SIVMM251   AAA--C--CT-A--A--T--G-GG--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1840
SIVMM142   AAA--C--CT-A--A--T--G-AG--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1840
SIVMM239   AAA--C--CT-A--A--T--G-AG--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1852
SIVMNE     AAA--C--CT-A--A--T--G-AA--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1840
SIVSMMH4   AAA--C--TT-G--A--A--A-GA--AATT-----A--A------ACAG-T--T-----A--A--C---A  1729
SIVSMMPBJ  AGA--T--TT-G--A--A--G-GA--AATT-----A--A------ACAG-T--T-----A--A--T---A  1738

CONSENSUS  TcCCagA?TGGGAcTTtgT?TC?ACcCCaCCacTagTcAGatTagt?TttAAccTaGTaaaaGA?CCtaT  1829
HIV2ROD    -C--AG-C-----C--CG-G--T--C--A--AC-GG-C--GT-AGCG-TT--CC-G--AGGG--T--TA-  1850
HIV2NIHZ   -C--AG-C-----C--TG-A--T--C--A--AC-GG-C--GC-AGCA-TT--CC-G--AGGA--G--TG-  1847
HIV2ISY    -C--AG-C-----C--CG-A--C--C--A--GT-GG-C--GT-AGCA-TT--CC-G--AAAA--T--TA-  1847
HIV2ST     -C--AG-C-----C--CA-A--T--C--G--AC-GG-C--AT-AGTA-TT--CC-G--GAAA--T--CA-  1961
HIV2BEN    -C--AG-G-----C--TG-A--T--C--A--AC-GG-C--GT-AACA-TT--CC-A--AGGA--T--TA-  1907
HIV2D194   -C--AG-G-----C--CG-A--T--C--A--AC-GG-C--GT-GACA-TT--CT-G--AGGA--T--TA-  1961
HIV2GH1    -C--AG-G-----C--CG-A--T--C--A--AC-AG-C--GT-AACA-TT--CC-A--AGGA--T--TA-  1907
HIV2D205   -C--AG-G-----C--TG-C--G--C--A--AT-AA-A--AC-AGCC-AC--CC-A--CAAA--C--CC-  1907
SIVMM251   -A--GG-G-----T--TA-C--A--G--A--AC-AG-A--AT-AGTC-TC--TC-A--GAAG--C--TA-  1910
SIVMM142   -A--GG-A-----T--CA-C--A--A--A--AT-AG-A--AT-AGTC-TC--TC-A--GAAG--C--TA-  1910
SIVMM239   -A--GG-A-----T--TA-C--A--A--A--GC-AG-A--AT-AGTC-TC--TC-A--GAAG--C--TA-  1922
SIVMNE     -A--GA-A-----T--TA-C--A--A--A--AC-AG-A--AT-AGTC-TC--TC-G--AAAG--C--TA-  1910
SIVSMMH4   -A--AG-T-----C--TG-G--A--A--T--CT-AG-C--AT-AGTC-TC--CC-A--AAAA--G--TA-  1799
SIVSMMPBJ  -A--AG-A-----C--TG-G--A--A--T--CT-AG-C--AC-AGTT-TC--CC-A--AAAA--G--TA-  1808

CONSENSUS  acaaGG?gcAGAaACcTttTAcacaGATGGaTCaTGcAatAgacagTCaAaaGAaGGaAAaGCAgGaTAt  1898
HIV2ROD    ACCA--TGC---G--C-TC--CACA-----A--C--C-AT-GGCAA--A-AA--A--A--A---G-A--T  1920
HIV2NIHZ   ACCA--CGC---A--T-TC--CACA-----A--C--C-AT-GGCAG--A-AA--A--G--A---G-A--T  1917
HIV2ISY    ACCA--CGC---G--C-TC--CACG-----A--T--C-AT-GGCAA--A-AA--G--A--A---G-A--T  1917
HIV2ST     ACTA--CGC---A--C-TC--CACA-----A--C--C-AT-AGCAA--A-GA--A--A--A---G-A--C  2031
HIV2BEN    ACCA--CGC---G--C-TC--CACA-----A--A--C-AT-GACAG--A-AA--G--A--A---G-A--T  1977
HIV2D194   ACCA--CAC---G--C-TT--CACA-----A--A--C-AT-GACAG--A-AA--A--A--A---G-A--T  2031
HIV2GH1    ACCA--CGC---G--C-TC--CACA-----A--A--C-AT-GACAG--A-AG--A--A--A---C-A--T  1977
HIV2D205   AGAA--GAG---A--C-AC--CACA-----G--C--C-AT-GAACC--A-AG--A--A--A---G-A--T  1977
SIVMM251   AGAG--AGA---A--C-AT--TACA-----A--A--T-AT-AACAG--A-AA--A--G--A---G-A--T  1980
SIVMM142   AGAG--AGA---A--C-AT--TGTA-----A--A--T-GT-AACAG--A-AA--A--A--A---G-A--T  1980
SIVMM239   AGAG--AGA---A--C-AT--TACA-----A--A--T-AT-AACAG--A-AA--A--G--A---G-A--T  1992
SIVMNE     AAAG--AGA---A--C-AT--TGTA-----A--A--T-AT-AACAG--A-AA--A--G--A---G-A--T  1980
SIVSMMH4   ACAG--AGC---A--A-TT--TGTA-----A--C--T-AT-GGCAG--A-GA--A--A--A---G-C--T  1869
SIVSMMPBJ  ACAG--AGC---A--A-TT--TGTA-----A--C--T-AT-GGCAG--A-GA--A--A--G---G-C--T  1878
```

```
CONSENSUS  aT?ACaGAtAGaGG?AaAgAcAagGtaAaagt?tTaGAaCagACtaCcAAtCAaCAaGCAGAatTaGAAG  1965
HIV2ROD    G-A--A--T--A--G-A-G-C-AG-TA-AGAAAC-A--G-AA--TA-C--T--G--A-----AC-A----  1990
HIV2NIHZ   A-A--A--T--A--G-G-G-C-GG-TA-AAGTAT-A--G-AA--TA-C--T--G--A-----AT-A----  1987
HIV2ISY    A-A--A--T--A--A-A-G-C-AA-TA-GGATAT-A--G-AA--TA-C--C--A--A-----AT-A----  1987
HIV2ST     A-A--A--T--A--A-G-G-C-AG-TG-GGCTAT-A--G-AA--CA-C--T--G--A-----AT-A----  2101
HIV2BEN    G-A--A--T--A--A-A-G-C-AA-TA-AAGTAT-A--A-AA--TA-C--T--G--G-----AT-A----  2047
HIV2D194   G-A--A--T--A--G-G-G-C-GG-TA-GAGTAT-A--G-AA--AT-C--T--G--A-----AC-A----  2101
HIV2GH1    G-G--A--T--A--A-G-G-C-AG-TA-GAGTAC-A--A-GA--TA-C--T--G--A-----AT-A----  2047
HIV2D205   G-C--T--C--G--A-A-G-T-AG-TT-AAGTGT-A--A-AG--AA-A--C--A--A-----AC-T----  2047
SIVMM251   A-C--A--T--G--C-A-G-C-AA-TA-AAGTGT-A--A-AG--TA-T--T--A--A-----AT-A----  2050
SIVMM142   A-C--A--C--G--C-A-G-C-AG-TA-AAGTGT-A--A-AG--TA-T--T--A--A-----AT-G----  2050
SIVMM239   A-C--A--T--G--C-A-G-C-AA-TA-AAGTGT-A--A-AG--TA-T--T--A--A-----AT-G----  2062
SIVMNE     A-C--A--T--G--C-A-G-C-AA-TA-AAGTCT-A--A-AG--TA-T--T--A--A-----AT-G----  2050
SIVSMMH4   G-G--G--T--G--C-G-G-C-AA-CA-AACTTT-A--A-AG--TA-C--C--A--A-----GT-G----  1939
SIVSMMPBJ  G-G--A--T--A--C-G-G-C-AG-CA-AACTGC-G--A-AG--TA-C--T--A--A-----GT-G----  1948

CONSENSUS  ccTTtgc?aTggCatTgaCaGAcTCaGg?CCaaAagc?AAtATtaTaGTAGAcTCaCAaTATGT?ATGGG  2031
HIV2ROD    CC--TGCGA-GG-AC-AA-A--C--G-GT--AA-AGTT--T--TA-A-----C--A--G-----A-----  2060
HIV2NIHZ   CC--TGCAA-GG-AC-AA-A--C--A-GT--AA-AGCT--T--TA-A-----C--A--A-----A-----  2057
HIV2ISY    CC--TGCAA-GG-AG-AA-A--C--A-GT--AA-AGTC--T--TG-A-----C--A--G-----A-----  2057
HIV2ST     CC--TGCGA-GG-AG-AA-A--C--A-GT--AA-GGCC--C--TA-A-----C--A--A-----A-----  2171
HIV2BEN    TC--TCGGA-GG-AC-GG-A--C--A-GC--AA-GGTT--T--CA-A-----T--A--G-----A-----  2117
HIV2D194   CC--TGCGA-GG-AC-GG-A--C--A-GT--CA-GGTT--T--CA-A-----C--A--G-----A-----  2171
HIV2GH1    CC--TGCGA-GA-AC-GA-A--C--G-GT--AA-AGTT--T--CA-A-----C--A--G-----A-----  2117
HIV2D205   CA--TGCAT-AG-AT-AA-A--C--A-AA--AC-AGTT--C--CA-A-----T--A--A-----C-----  2117
SIVMM251   CA--TCTCA-GG-AT-GA-A--C--A-GG--AA-GACA--T--TA-A-----T--A--A-----T-----  2120
SIVMM142   CA--TCTCA-GG-AT-GA-A--C--A-GG--AA-GGCA--T--TA-A-----C--A--A-----T-----  2120
SIVMM239   CA--TCTCA-GG-AT-GA-A--C--A-GG--AA-GGCA--T--TA-A-----T--A--A-----T-----  2132
SIVMNE     CA--TCTCA-GG-AT-GG-A--C--A-GG--AA-GGCA--T--TA-A-----T--A--A-----T-----  2120
SIVSMMH4   CC--CTATC-AG-CT-AG-A--T--G-GA--AA-AGCA--T--TA-A-----T--C--A-----T-----  2009
SIVSMMPBJ  CC--CTACC-AG-CT-AG-A--T--A-GA--AA-AGCA--T--TA-A-----T--C--A-----T-----  2018

CONSENSUS  aATaatagcagg?ca?CC?aCaGAatCaGAgagtagaaTAGTaaaccAaATaATaGAaGAaATGAT?AAa  2097
HIV2ROD    G--CAGTGCAAGCCAA--AA-A--GT-A--AAGTAAAA----GAACC-G--C--A--A--A-----A--A  2130
HIV2NIHZ   G--AGTAGCAGGCCAA--AA-A--GT-A--GAATAGAA----AAATC-G--C--A--A--A-----A--G  2127
HIV2ISY    A--AGTAACAGGCCAA--GG-T--AT-A--GAGTAGAA----AAATA-A--T--A--A--G-----A--A  2127
HIV2ST     A--AGTAGCAGGCCAA--AA-A--GT-A--GAGTAAAA----AAATC-A--C--A--A--A-----A--A  2241
HIV2BEN    G--AGTAGCAGGCCAG--AA-A--GT-A--AAATAGAA----GAACC-G--C--A--A--A-----A--G  2187
HIV2D194   G--AGTAGCAGGCCAA--AA-A--GT-A--AAATAGAA----AAACC-A--C--T--G--C-----A--G  2241
HIV2GH1    A--AGTAGTAGGCCAA--AA-A--GT-A--AAGTAGAA----AAACC-A--C--T--G--C-----A--G  2187
HIV2D205   A--AATAGCTGCACAG--AA-A--AA-A--ATCACCAA----AGCAA-A--A--T--A--A-----C--A  2187
SIVMM251   A--AATAACAGGATGC--TA-A--AT-A--GAGCAGGC----TAACC-A--A--A--A--A-----T--A  2190
SIVMM142   A--AATAACAGGATGC--TA-A--AT-A--GAGCAGGC----TAACC-A--A--A--A--A-----C--A  2190
SIVMM239   A--AATAACAGGATGC--TA-A--AT-A--GAGCAGGC----TAATC-A--A--A--A--A-----T--A  2202
SIVMNE     A--AATAACAGGATGC--TA-A--AT-A--GAGCAGGC----TAACC-A--A--A--A--A-----T--A  2190
SIVSMMH4   C--AATAGCAGGTCAG--CA-T--AT-A--AAGTAGAT----AAACC-G--A--A--G--A-----T--A  2079
SIVSMMPBJ  C--AGTAGCAGGTCAA--CA-T--AT-A--AAGTAGAT----AAACC-G--A--A--A--A-----T--A  2088
```

HIV2/SIV POL

```
CONSENSUS  AAggaaGcAaT?TATGT?GcaTGGGT?CCaGC?CAcAAaGG?aTaGGaGGaAAcCAgGAagTaGAcCAct  2162
HIV2ROD    --GGAA-C-A-C-----T-CA-----C--A--C--C--A--CA-A--G--A--C--G--AG-A--T--TT  2200
HIV2NIHZ   --AGAA-C-A-C-----T-CG-----C--A--C--C--A--CA-A--A--A--C--G--AG-A--T--TT  2197
HIV2ISY    --GGAA-C-A-C-----T-CA-----C--G--C--C--A--CA-A--A--A--T--A--AA-T--C--CT  2197
HIV2ST     --GGAA-C-A-C-----T-CA-----C--A--C--T--A--CA-A--A--A--T--G--GG-A--T--CT  2311
HIV2BEN    --GGAA-C-G-C-----T-CA-----C--A--C--T--A--CA-A--A--A--C--G--AG-A--C--TT  2257
HIV2D194   --AGAA-C-G-C-----T-CA-----C--A--C--T--A--CA-A--A--A--C--G--AG-A--C--TT  2311
HIV2GH1    --AGAA-C-G-C-----C-CG-----C--A--C--T--A--CA-A--A--A--C--G--AG-A--C--TT  2257
HIV2D205   --AGAG-C-G-A-----A-GA-----A--A--T--C--G--AC-G--T--T--T--G--AG-A--C--CC  2257
SIVMM251   --GTCA-A-A-T-----A-CA-----A--A--A--C--A--TA-A--A--A--C--A--AA-A--C--CC  2260
SIVMM142   --GACA-A-A-T-----G-CA-----A--A--A--C--A--TA-A--A--A--C--A--AA-A--C--CC  2260
SIVMM239   --GTCA-A-A-T-----A-CA-----A--A--A--C--A--TA-A--A--A--C--A--AA-A--C--CC  2272
SIVMNE     --GACA-A-A-T-----A-CA-----G--A--A--C--A--TA-A--A--A--C--A--AA-A--C--CC  2260
SIVSMMH4   --GGAA-C-A-T-----A-CA-----A--T--A--T--A--AA-A--A--C--T--A--AG-A--T--CC  2149
SIVSMMPBJ  --GGAA-C-A-T-----A-CA-----A--T--A--T--A--AA-A--A--A--T--A--AG-A--T--TC  2158

CONSENSUS  TaGTaAGtCAgGG?ATtAGACAaGT?tTaTTCcTagAaAaaATAGAgCCaGC?CAaGAaGAaCAtgaaAA  2229
HIV2ROD    -A--G--T--G--T--C-----A--GT-G---C-GG-A-AA-----G--C--T--G--A--A--TGAA--  2270
HIV2NIHZ   -A--A--T--G--C--T-----A--AT-A---C-AG-G-AA-----A--C--G--G--A--A--TGAA--  2267
HIV2ISY    -A--A--T--G--C--C-----A--AT-A---C-AG-G-GA-----G--C--T--G--A--A--TGGA--  2267
HIV2ST     -A--A--T--G--C--C-----A--AT-A---C-AG-G-AA-----A--C--T--G--G--A--TGAA--  2381
HIV2BEN    -A--A--T--A--C--C-----A--AT-A---C-GG-A-AG-----G--C--T--A--G--A--TGAA--  2327
HIV2D194   -A--A--T--G--C--C-----A--AT-A---C-GG-A-AG-----G--C--T--A--A--A--CGAA--  2381
HIV2GH1    -A--A--T--G--C--C-----A--AT-G---C-AG-A-GG-----G--C--C--A--A--A--TGAA--  2327
HIV2D205   -A--A--T--A--A--C-----G--CT-G---C-AG-A-AA-----A--A--C--G--A--G--TGAA--  2327
SIVMM251   -A--T--T--G--G--T-----A--TC-C---T-GG-A-AG-----G--A--A--A--A--A--TGAT--  2330
SIVMM142   -A--T--T--A--G--T-----A--TC-C---T-GG-A-AG-----G--A--A--A--A--A--TAGT--  2330
SIVMM239   -A--T--T--A--G--T-----A--TC-C---T-GG-A-AG-----G--A--A--A--A--A--TGAT--  2342
SIVMNE     -A--T--T--A--G--T-----A--TC-C---T-GG-A-AG-----G--A--A--A--A--A--TGAT--  2330
SIVSMMH4   -G--T--C--G--A--T-----A--CC-A---T-AA-A-AA-----A--A--A--A--A--G--TGAA--  2219
SIVSMMPBJ  -G--T--T--A--A--T-----A--CC-A---T-AG-A-AA-----A--A--G--G--A--G--TGAA--  2228

CONSENSUS  aTAtCATaGcAatgTaAAAGAacTggtccatAAaTTtGGa?TaCCca?acTaGTgGCaAgaCAgATAGTa  2297
HIV2ROD    A--T---A-C-ATG-A-----AC-GTCTCAT--A--T--AA-A--CAATT-A--G--A-GG--A-----A  2340
HIV2NIHZ   A--T---A-C-ATA-A-----AC-GTCCCAT--A--T--AA-A--CAAGC-A--G--A-GA--A-----A  2337
HIV2ISY    A--T---A-C-ATG-A-----AC-AGCCCAT--G--T--AT-A--CAACC-G--G--A-GA--A-----A  2337
HIV2ST     A--T---A-C-ATG-A-----AC-ATCCCAT--A--T--AC-G--CAAAT-A--G--A-GA--A-----A  2451
HIV2BEN    A--T---A-C-TTA-A-----AC-AACCCAT--A--T--AA-A--CCTTC-A--A--A-GA--G-----A  2397
HIV2D194   A--T---A-C-ATA-A-----AC-AACCCAT--A--T--AA-A--CCAAC-A--G--A-GA--G-----A  2451
HIV2GH1    A--T---A-C-ATA-G-----AC-AACCCAT--A--T--AA-A--CCAAC-A--A--A-GA--G-----A  2397
HIV2D205   A--T---G-C-ATG-A-----AC-GGTCCAT--A--C--AA-T--ACAAT-A--G--A-AA--G-----A  2397
SIVMM251   A--C---A-T-ATG-A-----AT-GGTATTC--A--T--AT-A--CAGAA-A--G--C-GA--G-----A  2400
SIVMM142   A--C---A-T-ACA-A-----AT-GGTATTC--A--T--AT-A--CAGAC-A--G--C-AA--G-----A  2400
SIVMM239   A--C---A-T-ATG-A-----AT-GGTATTC--A--T--AT-A--CAGAA-A--G--C-GA--G-----A  2412
SIVMNE     A--C---A-T-ATG-A-----AT-GGTATTC--A--T--AT-A--CAGAC-A--G--C-AA--G-----A  2400
SIVSMMH4   G--C---A-C-ATG-A-----AT-GGTATTC--A--T--TT-A--TAGGC-A--A--A-AA--G-----A  2289
SIVSMMPBJ  A--C---A-C-ATG-A-----AT-GGTATTC--A--T--TT-A--TAGGC-A--A--A-AA--G-----A  2298
```

```
CONSENSUS  aAcaCaTGtgacaAatgcCA?CAgAAAGGaGAaGCtkTaCATGGgCAaGtAAATqCAGAacTAGGgactT  2366
HIV2ROD    A-CT-A--TGCCC-ATGT--A--G-----G--A--T--A-----G--A-T----G----AC----CACT-  2410
HIV2NIHZ   A-CA-A--TGCCC-TGTG--A--G-----G--G--T--A-----G--A-T----G----AC----CACT-  2407
HIV2ISY    A-CA-A--TGCCC-GTGC--A--A-----G--A--T--A-----G--A-T----G----AC----CACC-  2407
HIV2ST     A-CA-A--CACCC-ATGT--G--G-----G--G--T--A-----G--A-T----G----AT----CACT-  2521
HIV2BEN    A-CT-A--TGCCC-ATGC--A--G-----A--A--C--A-----G--A-T----G----AA----CGTT-  2467
HIV2D194   A-CA-A--TGCCC-ATGC--A--G-----A--A--C--A-----G--A-T----G----AA----CGTT-  2521
HIV2GH1    A-CA-A--TGCCC-ATGC--A--G-----A--G--C--A-----G--G-T----G----AA----TGTC-  2467
HIV2D205   A-TT-C--TGATA-ATGC--A--A-----G--A--T--T-----A--G-T----G----CC----GACA-  2467
SIVMM251   G-CA-C--TGATA-ATGT--T--G-----A--A--T--A-----G--G-T----T----TC----GACT-  2470
SIVMM142   G-CA-A--TGATA-ATGT--T--A-----A--A--T--A-----G--G-T----T----CC----GACT-  2470
SIVMM239   G-CA-C--TGATA-ATGT--T--G-----A--G--T--A-----G--G-C----T----TC----GACT-  2482
SIVMNE     G-CA-A--TGATA-ATGT--T--G-----A--A--T--A-----G--G-T----T----TC----GACT-  2470
SIVSMMH4   G-CA-A--TGATA-ATGC--C--G-----A--A--C--A-----G--A-T----G----AC----GACT-  2359
SIVSMMPBJ  G-CA-A--TGATA-ATGC--C--G-----A--A--C--A-----A--A-T----G----AC----GACT-  2368

CONSENSUS  GGCAaATGGAcTgtACaCAttTAGAaGGaAAaaT?aTtATAGTaGCAGTaCATGT?GC?AGTGGaTTcAT  2433
HIV2ROD    ----A-----C-GC--A--TT----A--A--GA-CA-T-----A-----A-----T--A-----A--T--  2480
HIV2NIHZ   ----A-----C-GC--A--TT----A--A--AG-CA-T-----A-----A-----T--A-----A--T--  2477
HIV2ISY    ----A-----C-GC--A--CT----A--A--AA-CA-T-----A-----A-----T--A-----A--T--  2477
HIV2ST     ----A-----C-GC--A--CT----A--A--AA-CA-T-----A-----A-----T--A-----A--T--  2591
HIV2BEN    ----A-----C-AC--A--CT----A--A--AA-CA-T-----A-----A-----T--A-----A--C--  2537
HIV2D194   ----A-----C-GC--A--CT----A--A--AA-CA-T-----A-----G-----T--A-----A--C--  2591
HIV2GH1    ----A-----C-GC--A--CT----A--A--GA-CA-T-----A-----A-----T--A-----A--C--  2537
HIV2D205   ----G-----C-GT--A--TT----A--A--AA-TA-A-----G-----C-----A--C-----G--T--  2537
SIVMM251   ----A-----C-GT--C--TC----A--A--AA-AG-C-----T-----A-----A--T-----A--C--  2540
SIVMM142   ----A-----T-GT--C--TC----G--A--AA-AG-C-----T-----A-----A--T-----A--C--  2540
SIVMM239   ----A-----T-GT--C--TC----G--A--AA-AA-C-----T-----A-----A--T-----A--C--  2552
SIVMNE     ----A-----T-GT--C--TC----G--A--AA-AA-C-----T-----A-----A--T-----A--C--  2540
SIVSMMH4   ----A-----C-GT--G--CC----A--C--AA-AA-T-----T-----A-----G--T-----A--T--  2429
SIVSMMPBJ  ----A-----C-GT--A--CC----G--C--AA-AA-T-----T-----A-----G--T-----A--C--  2438

CONSENSUS  AGAaGCAGAaGTaATcCCaCAqGAaaCAGGaAGaCAqACaGCaoT?TTooTatTaAAatTgGCcaGcAGa  2502
HIV2ROD    ---A-----A--C--C--A--G--AT----A--A--A--A--AC-C--CC-AT-G--AC-G--AA-T--G  2550
HIV2NIHZ   ---G-----A--T--C--A--G--AT----A--G--A--A--AC-G--CT-AC-G--AC-G--CA-T--G  2547
HIV2ISY    ---A-----A--C--C--A--G--AT----A--G--A--A--AC-C--CC-AT-A--AC-G--CA-T--G  2547
HIV2ST     ---A-----A--C--C--A--G--AT----A--G--A--G--AC-C--CC-AC-A--AC-G--CA-T--G  2661
HIV2BEN    ---A-----A--C--C--A--G--AT----A--G--G--A--AC-C--CC-AT-A--AC-G--CA-T--G  2607
HIV2D194   ---A-----A--C--C--A--G--AT----A--G--G--A--AC-C--CC-AT-A--AC-G--CA-T--G  2661
HIV2GH1    ---A-----A--T--C--A--G--AT----A--G--G--A--GC-C--CC-AT-A--AC-G--CA-T--G  2607
HIV2D205   ---A-----G--A--A--C--A--GA----A--A--G--A--TC-C--CC-AC-A--GT-G--CA-C--A  2607
SIVMM251   ---A-----A--A--T--A--A--GA----A--A--G--A--AC-A--TC-GT-A--AT-G--AG-C--A  2610
SIVMM142   ---A-----A--A--T--A--A--AA----A--A--G--A--AC-A--TC-GT-A--AT-G--AA-C--A  2610
SIVMM239   ---A-----G--A--T--A--A--GA----A--A--G--A--AC-A--TC-GT-A--AT-G--AG-C--A  2622
SIVMNE     ---A-----A--A--T--A--A--GA----A--A--G--A--AC-A--TC-GT-A--AT-G--AA-C--A  2610
SIVSMMH4   ---G-----A--A--C--G--G--AA----A--A--A--A--AC-G--TC-GT-A--GT-A--TG-C--A  2499
SIVSMMPBJ  ---G-----A--A--T--G--G--AA----A--A--A--A--AC-G--TC-GC-A--AT-A--CA-C--A  2508
```

HIV2/SIV POL

```
CONSENSUS  TGGCC?aT?ACaCAccTgCAcACaGAtAAtGGtgCcAAcTTcaCtTCaCaggaaGTaAAgATGGT?GCaT  2569
HIV2ROD    -----AA-A--A--CT-G--T--A--T--T--TG-C--C--CA-T--A-AGGAG--G--G-----A--A-  2620
HIV2NIHZ   -----AA-A--A--CC-G--C--A--T--T--TG-C--C--CA-T--A-AGGAA--G--A-----A--A-  2617
HIV2ISY    -----AA-A--A--CT-G--C--A--T--T--TG-C--C--CA-T--A-AGGAG--A--G-----A--A-  2617
HIV2ST     -----AA-A--A--TT-G--C--A--C--T--TG-C--C--CA-T--A-AGGAA--A--G-----G--A-  2731
HIV2BEN    -----AA-A--G--CT-G--C--A--C--T--CC-C--C--CA-T--A-AGGAA--G--G-----G--A-  2677
HIV2D194   -----AA-A--A--CT-G--C--A--C--T--CC-C--C--CA-T--A-AGGAA--G--G-----G--A-  2731
HIV2GH1    -----AA-A--A--CT-A--C--A--C--T--CT-C--C--CA-T--A-AGGAA--G--G-----G--A-  2677
HIV2D205   -----TA-C--A--CC-A--C--A--C--C--TG-C--C--CA-C--A-CAAGT--A--G-----A--C-  2677
SIVMM251   -----TA-T--A--TC-A--C--A--T--T--TG-T--C--TG-C--G-AAGAA--A--G-----T--A-  2680
SIVMM142   -----TA-T--A--TC-G--C--A--T--T--TG-T--C--TG-T--G-AAGAA--A--G-----T--A-  2680
SIVMM239   -----TA-T--A--TC-A--C--A--T--T--TG-T--C--TG-T--G-AAGAA--A--G-----T--A-  2692
SIVMNE     -----TA-T--G--TC-A--C--A--T--T--TG-C--C--TG-T--G-AAGAA--A--G-----T--A-  2680
SIVSMMH4   -----CA-C--A--TC-G--T--T--T--T--TG-C--T--CA-A--A-AAGAA--G--A-----T--C-  2569
SIVSMMPBJ  -----CA-C--A--TC-G--T--T--T--T--TG-C--T--CA-A--A-AAGAA--G--A-----T--T-

CONSENSUS  GGTGGgtaGG?ATaGAaCAaaCcTTtGGaGTaCCtTAcAAtCCACAgAGcCAgGGaGTAGTgGAAGCaAT  2638
HIV2ROD    -----ATA--T--A--A--AT-C--T--A--A--T--C--T-----G--C--A--A-----A-----A--  2690
HIV2NIHZ   -----GTA--T--A--A--AA-C--T--A--G--T--C--T-----A--C--A--A-----A-----A--  2687
HIV2ISY    -----GTA--C--A--A--AT-C--T--A--A--T--C--T-----A--C--G--A-----A-----A--  2687
HIV2ST     -----ATA--T--A--A--AT-C--C--A--A--T--C--T-----A--C--A--A-----G-----A--  2801
HIV2BEN    -----GTA--T--A--A--AT-C--T--A--A--T--C--C-----A--C--G--A-----A-----A--  2747
HIV2D194   -----ATA--T--A--G--AT-C--T--A--A--T--C--T-----A--C--G--A-----A-----A--  2801
HIV2GH1    -----ATA--T--A--G--AT-C--T--A--A--T--C--C-----A--C--G--A-----A-----A--  2747
HIV2D205   -----GTA--A--A--A--AA-T--T--A--A--C--T--C-----A--T--A--A-----G-----A--  2747
SIVMM251   -----GCA--G--A--G--CA-C--T--G--A--A--C--T-----G--T--G--A-----G-----A--  2750
SIVMM142   -----GCA--G--A--G--CA-C--T--G--A--A--C--T-----G--T--G--A-----G-----A--  2750
SIVMM239   -----GCA--G--A--G--CA-C--T--G--A--A--C--T-----G--T--G--A-----G-----A--  2762
SIVMNE     -----GCA--G--A--G--CA-C--T--G--A--A--C--T-----G--T--G--A-----G-----A--  2750
SIVSMMH4   -----GCA--G--T--A--GA-C--T--G--G--T--T--T-----G--C--A--A-----G-----A--  2639
SIVSMMPBJ  -----GCA--G--T--A--GA-C--T--G--A--T--T--T-----G--C--A--G-----G-----G--  2648

CONSENSUS  GAAtCAcCAccTaAAaAatCAgATAgatAGAaTtAGaGAaCAgGCAaataCAaTaGAaAC?aTaGTAcTa  2707
HIV2ROD    ---T--C--TC-A--A-AC--A---AGT---A-C--A--A--G---AATA--A-A--A--AA-A---C-A  2760
HIV2NIHZ   ---T--T--TC-A--A-AT--G---GAC---A-T--A--G--G---AATA--G-A--A--AA-A---T-A  2757
HIV2ISY    ---T--C--CC-G--A-AT--G---GAA---A-T--A--G--G---AATA--A-G--A--AA-A---C-A  2757
HIV2ST     ---T--C--CC-A--A-AT--G---AGC---A-T--A--G--G---AACA--G-A--A--AA-A---C-A  2871
HIV2BEN    ---T--C--CC-A--G-AT--G---AGT---A-T--A--A--G---AATA--A-A--A--AA-A---C-G  2817
HIV2D194   ---T--C--CC-A--A-AT--G---AGT---A-T--A--A--G---AATA--A-A--A--AA-A---C-A  2871
HIV2GH1    ---T--C--CT-A--G-AT--G---AGT---A-T--A--A--G---AATA--A-A--A--CA-A---C-A  2817
HIV2D205   ---C--T--CC-G--A-AT--A---GAC---C-C--A--C--A---GTAT--A-A--G--AG-A---C-A  2817
SIVMM251   ---T--C--CC-G--A-AT--A---GAT---A-C--G--A--A---AATT--G-A--A--CA-A---T-A  2820
SIVMM142   ---T--C--CC-G--A-AT--A---GAT---A-C--G--A--A---AATT--G-A--A--CA-A---T-A  2820
SIVMM239   ---T--C--CC-G--A-AT--A---GAT---A-C--G--A--A---AATT--G-A--A--CA-A---T-A  2832
SIVMNE     ---T--C--TC-A--A-AT--A---GAT---A-C--G--A--A---AATT--A-G--A--CA-A---T-A  2820
SIVSMMH4   ---C--T--TT-A--A-CC--G---GAT---A-T--A--A--A---AACT--A-A--G--TA-A---C-A  2709
SIVSMMPBJ  ---C--T--TC-A--A-CC--A---GAT---A-T--A--A--G---AATT--A-A--A--TA-A---C-A  2718
```

```
CONSENSUS  ATGGCAgtTCAtTGcATGAATTTTAAaAgaAGGGGaGGAaTAGGGGATATGACcCCagcAGAAAGAtTAa  2777
HIV2ROD    ------AT---T--C----------A-GA-----G---A----------T--ATC-------T--A  2830
HIV2NIHZ   ------GT---T--C----------A-GA-----A---A----------C--AGC-------A--A  2827
HIV2ISY    ------GT---T--C----------A-GA-----A---A----------C--AGT-------C--G  2827
HIV2ST     ------GT---T--C----------A-GG-----A---A----------C--AGC-------C--A  2941
HIV2BEN    ------GT---T--C----------A-GA-----A---A----------C--AGC-------C--A  2887
HIV2D194   ------GT---T--C----------A-GA-----A---A----------C--AGC-------C--A  2941
HIV2GH1    ------GT---T--C----------A-GA-----A---A----------C--AGC-------C--A  2887
HIV2D205   ------AC---C--C----------A-GA-----A---A----------C--TGC-------C--G  2887
SIVMM251   ------GT---T--C----------A-GA-----A---A----------T--AGC-------T--A  2890
SIVMM142   ------GT---T--C----------A-GA-----A---A----------T--AGC-------T--A  2890
SIVMM239   ------GT---T--C----------A-GA-----A---A----------T--AGC-------T--A  2902
SIVMNE     ------GT---T--C----------A-GA-----A---A----------T--AGC-------T--C  2890
SIVSMMH4   ------GT---T--T----------A-GA-----A---A----------T--AGC-------T--G  2779
SIVSMMPBJ  ------GT---T--C----------A-GA-----A---A----------T--AGC-------T--G  2788

CONSENSUS  TcAAcATGaTcACcaCAGAaCAaGAaATACAaTTcctcCAA?caAAAAAtTcAAAaTTtaAAaAtTTtCg  2846
HIV2ROD    -C--T---A-C--CA----A--A--G-----A--CCTC---GCC-----T-C---A--AA--G-T--T-G  2900
HIV2NIHZ   -C--T---A-C--CA----A--A--A-----A--CCTC---GCA-----T-C---A--AA--A-T--T-G  2897
HIV2ISY    -C--T---A-C--CA----A--A--A-----A--CCTC---GCA-----T-C---A--AA--A-T--T-G  2897
HIV2ST     -C--T---G-C--TG----A--G--A-----A--CCTC---GCA-----T-C---A--AC--A-T--T-G  3011
HIV2BEN    -C--C---A-T--CA----A--A--A-----A--CCTC---AGA-----T-C---T--TA--A-T--C-A  2957
HIV2D194   -T--C---A-C--CA----A--A--A-----A--CCTC---AGA-----T-C---T--TA--A-A--C-A  3011
HIV2GH1    -C--C---A-T--CA----A--A--A---- A--CCTC---AGA-----T-C---T--TA--A-C--C-A  2957
HIV2D205   -T--C---A-A--CA----G--A--A-----G--CTTC---GCA-----T-T---A--TC--A-T--C-A  2957
SIVMM251   -T--C---A-C--TA----A--A--A-----A--TCAA---TCA-----C-C---A--TA--A-T--T-G  2960
SIVMM142   -T--C---A-C--TA----A--A--A-----A--TCAA---TCA-----C-C---A--TA--A-T--T-G  2960
SIVMM239   -T--C---A-C--TA----A--A--G-----A--TCAA---TCA-----C-C---A--TA--A-T--T-G  2972
SIVMNE     -T--C---A-C--TA----A--A--A-----A--CCAA---TCA-----C-C---A--TA--A-T--T-G  2960
SIVSMMH4   -C--T---A-C--CA----A--A--A-----A--CCAA---TCA-----T-C---A--TA--A-T--T-G  2849
SIVSMMPBJ  -C--T---A-C--CA----A--A--A-----A--CCAA---TCA-----T-C---A--TA--A-T--T-G  2858

                     \/3' sj (rare)                    \/ 5' sj (rare)
CONSENSUS  GGTCTATTaCAGaGAAGGCAGAgAt \gcTgTGGaAaGGACCtGGtGAgCTatTGTGGAAaGGgGAaGGA  2916
HIV2ROD    --------T---A---------G-T--GT-G---A-A-----T--G--A--AC-------A--A--A---  2970
HIV2NIHZ   --------T---A---------G-T--GC-G---A-A-----T--G--A--GC-------G--A--C---  2967
HIV2ISY    --------T---A---------A-T--AC-G---C-A-----T--G--G--AC-------A--G--C---  2967
HIV2ST     --------T---A---------G-T--GC-G---A-A-----T--G--A--AC-------G--G--C---  3081
HIV2BEN    --------A---A---------G-T--GC-G---A-A-----T--T--A--AC-------G--A--A---  3027
HIV2D194   --------A---A---------G-T--GC-G---A-A-----T--A--G--AC-------G--A--C---  3081
HIV2GH1    --------A---A---------G-T--GC-G---A-A-----T--A--A--AC-------G--A--C---  3027
HIV2D205   --------A---A---------G-T--AC-C---A-G-----T--T--A--AT-------A--G--A---  3027
SIVMM251   --------A---A---------G-T--AC-G---A-G-----C--T--G--AT-------A--G--A---  3030
SIVMM142   --------A---A---------G-T--GC-G---A-G-----C--T--G--AT-------A--G--A---  3030
SIVMM239   --------A---A---------G-T--AC-G---A-G-----C--T--G--AT-------A--G--A---  3042
SIVMNE     --------A---A---------G-T--GC-G---A-A-----T--T--G--AT-------A--G--A---  3030
SIVSMMH4   --------A---A---------G-C--GC-G---A-A-----C--T--G--AT-------A--G--A---  2919
SIVSMMPBJ  --------A---A---------G-C--GC-G---A-A-----C--T--G--AT-------A--G--A---  2928
```

HIV2/SIV POL

```
CONSENSUS  GCAGTCaTcgTaAAGGTAGGgaCaGAcAT?AAagTAgTACCaAGaAGgAAgGCtAAgATtATcAaagAcT   2985
HIV2ROD    ------C-AG-C--------AA-A--C--A--AA--A----A--A--G--A--C--G--C--C-GAG-C-   3040
HIV2NIHZ   ------A-AG-C--------GA-A--A--A--AG--G----A--A--G--A--C--G--C--T-AAG-C-   3037
HIV2ISY    ------A-AG-C--------GA-A--T--A--AG--A----A--A--A--G--C--G--C--C-GAG-C-   3037
HIV2ST     ------A-AG-C---  ----GG-T--C--A--AA--A----A--A--G--A--T--G--C--C-AAG-C-   3151
HIV2BEN    ------A-AG-C--------GA-A--C--A--AG--G----A--A--G--G--C--G--T--C-GGG-C-   3097
HIV2D194   ------A-AG-C--------GG-G--C--A--AG--G----A--A--G--G--C--G--T--C-GGG-C-   3151
HIV2GH1    ------A-AG-C--------GG-A--C--A--AG--A----A--A--G--G--C--G--T--C-GAG-C-   3097
HIV2D205   ------A-CA-A--------GA-A--A--C--AG--G----C--G--A--A--A--A--T--A-GGC-C-   3097
SIVMM251   ------A-CT-A--------GA-A--C--T--GG--G----C--A--A--G--T--A--T--C-AAG-T-   3100
SIVMM142   ------A-CT-A--------AA-A--C--T--GG--G----C--G--A--G--T--A--T--C-AAG-T-   3100
SIVMM239   ------A-CT-A--------GA-A--C--T--GG--G----C--A--A--G--T--A--T--C-AAG-T-   3112
SIVMNE     ------G-CT-A--------GA-A--C--T--GG--G----C--A--A--G--T--G--T--C-AAG-T-   3100
SIVSMMH4   ------A-CC-A--------GA-A--G--C--GG--G----A--G--G--A--T--A--T--C-AAG-C-   2989
SIVSMMPBJ  ------A-CC-A--------GA-A--G--C--GG--G----A--G--G--A--T--A--T--C-AAG-C-   2998

   vif cds ->
CONSENSUS  ATGGAggAgGaaAaGag?TGGATaGt?gttcCcac?tGGagGataccagggAGGcTgGAGAggTGGCAta   3052
HIV2kOD    -----GG-A-AC-A-AGA-----A-TGGTTC-CACCT--AG-GTGCCAGGG---A-G----AA-----TA   3110
HIV2NIHZ   -----GG-A-GC-A-AGA-----A-TGGTTC-CATTT--AG-GTGCCAGGG---A-G----GA-----***  3108
HIV2ISY    -----CC-A-GC-A-AGA-----A-CGGTTC-CACCT--AG-GTGCCAGGG---A-G----AA-----TA   3107
HIV2ST     -----GG-A-GC-A-AGA-----A-CGGTTC-AACTT--AG-GTGCCAGGG---A-G----GG-----TA   3221
HIV2BEN    -----GG-A-AC-G-AAC-----A-TAGTCC-CACCT--AG-GTGCCAGGG---A-G----AA-----TG   3167
HIV2D194   -----GG-A-GC-A-AAC-----A-TAGTTC-CACCT--AG-GTGCCAGGG---A-G----GG-----TA   3221
HIV2GH1    -----GG-A-GC-A-AAC-----A-TTCCCA-CTGGA--GT-CCAGGGAGG---A-G----GG-----TA   3167
HIV2D205   -----GG-G-AA-A-GAT-----T-TAGTGC-GACAT--AG-ATACCAGGC---C-A----GA-----CA   3167
SIVMM251   -----GG-G-AA-A-AGG-----A-CAGTTC-CACAT--AG-ATACCGGAG---C-A----GG-----TA   3170
SIVMM142   -----GG-G-AA-A-AGA-----A-TAGTTC-CACAT--AG-ATACCGGAG---C-A----GG-----TA   3170
SIVMM239   -----GG-G-AA-A-AGG-----A-CAGTTC-CACAT--AG-ATACCGGAG---C-A----GG-----TA   3182
SIVMNE     -----GG-G-AA-A-AGG-----A-CAGTTC-CACAT--AG-ATACCGGAG---C-A----GG-----TA   3170
SIVSMMH4   -----GG-G-AA-A-AAT-----A-TGGTTC-CACTT--AG-ATACCGGAG---C-A----GG-----TA   3059
SIVSMMPBJ  -----GG-G-AA-A-AAT-----A-TGGTTC-CACTT--AG-ATACCGGGG---C-A----AG-----TA   3068

CONSENSUS  g?CT??T?AA???      3060
HIV2ROD    G***               3114
HIV2ISY    G***               3111
HIV2ST     G***               3225
HIV2BEN    CC--TG-CAAGTA >    3430
HIV2D194   G***               3225
HIV2GH1    G***               3171
HIV2D205   GT--GA-T--***      3180
SIVMM251   G***               3174
SIVMM142   G***               3174
SIVMM239   G***               3186
SIVMNE     G***               3174
SIVSMMH4   G***               3063
SIVSMMPBJ  G***               3072
```

Splice sites have been determined for SIV only.

```
CONSENSUS  ATGGAggAgGaaAaGag?TGGATagt?GTtCCcAc?TGGAGGaTaCCaGggAGG???cTgGAGAggTGGC    64
HIV2ROD    -----GG-A-AC-A-AGA-----AGTG--T--C-CC------G-G--A-GG---...A-G----AA----    67
HIV2NIHZ   -----GG-A-GC-A-AGA-----AGTG--T--C-TT------G-G--A-GG---...A-G----GA----    67
HIV2ISY    -----CC-A-GC-A-AGA-----AGCG--T--C-CC------G-G--A-GG---...A-G----AA----    67
HIV2ST     -----GG-A-GC-A-AGA-----AGCG--T--A-CT------G-G--A-GG---...A-G----GG----    67
HIV2BEN    -----GG-A-AC-G-AAC-----AGTA--C--C-CC------G-G--A-GG---...A-G----AA----    67
HIV2D194   -----GG-A-GC-A-AAC-----AGTA--T--C-CC------G-G--A-GG---...A-G----GG----    67
HIV2GH1    -----GG-A-GC-A-AAC-----A...--T--C-CC------G-G--A-GG---AGGA-G----GG----    67
HIV2D205   -----GG-G-AA-A-GAT-----TGTA--G--G-CA------A-A--A-GC---...C-A----GA----    67
SIVMM251   -----GG-G-AA-A-AGG-----AGCA--T--C-CA------A-A--G-AG---...C-A----GG----    67
SIVMM142   -----GG-G-AA-A-AGA-----AGTA--T--C-CA------A-A--G-AG---...C-A----GG----    67
SIVMM239   -----GG-G-AA-A-AGG-----AGCA--T--C-CA------A-A--G-AG---...C-A----GG----    67
SIVMNE     -----GG-G-AA-A-AGG-----AGCA--T--C-CA------A-A--G-AG---...C-A----GG----    67
SIVSMMH4   -----GG-G-AA-A-AAT-----AGTG--T--C-CT------A-A--G-AG---...C-A----GG----    67
SIVSMMPBJ  -----GG-G-AA-A-AAT-----AGTG--T--C-CT------A-A--G-GG---...C-A----AG----    67

             <- pol cds
CONSENSUS  AtagccT?aTcAAgtAtcTgAAaTAtAaaACaaaaGA?cTa?a?aaGGt?tgcTATGT?CCcCAtcatAA   128
HIV2ROD    -TAGCC-TG-C--GT-TC-A--A--C-AA--AAAG--TC-AGAAAA--TGTGC-----T--C--CCAT--   137
HIV2NIHZ   -TAGCT-GG-C--GT-CT-G--A--C-GA--AAAA--CT-AGAGAA--TATGC-----T--C--TCAT--   137
HIV2ISY    -TAGCC-TA-C--GT-TC-A--A--C-GA--AAAA--TC-AGAACA--TGCGC-----T--C--CCAT--   137
HIV2ST     -TAGCC-TA-C--GT-TC-A--A--C-GA--AGGA--TC-AGAGAA--TGTGC-----T--C--CCAT--   137
HIV2BEN    -TGCCC-TG-C--GT-CC-G--A--C-GA--AAAA--CC-AGAGGA--TGCGC-----T--C--CCAC--   137
HIV2D194   -TAGCC-TG-C--GC-CC-G--G--C-GA--AAAA--CT-AGAGGA--TGCGC-----T--C--TCAC--   137
HIV2GH1    -TAGTC-TG-C--GT-CC-A--A--C-GA--AAGA--CC-AGAGGA--TGCGC-----T--C--CCAC--   137
HIV2D205   -CAGTC-GA-T--GT-TC-T--G--T-GA--AGGA--GT-GCAACA--TCTCT-----C--T--CCAC--   137
SIVMM251   -TAGCC-CA-A--AT-TC-G--A--T-AA--TA? \--TC-ACAAAA--TTTGC-----G--C--TTTT--   137
SIVMM142   -TAGCC-CA-A--AT-TT-G--A--T-AA--TAAA -TC-ACAAAA--CTTGC-----G--C--TCAT--   137
SIVMM239   -TAGCC-CA-A--AT-TC-G--A--T-AA--TAAA--TC-ACAAAA--TTTGC-----G--C--TTTT--   137
SIVMNE     -TAGCC-CA-A--AT-TC-G--A--T-AA--TAAA--TC-ACAAAA--TTTGC-----G--C--TCAT--   137
SIVSMMH4   -TAGCC-CA-C--AC-CC-G--A--T-AC--TAAA--CC-ACAGAT--CTTGT-----G--C--TCAT--   137
SIVSMMPBJ  -TAGCC-CA-C--AC-CC-G--A--T-AC--TAAA--CC-ACAGAA--CTTGT-----A--C--TCAT--   137

CONSENSUS  gGT?GGaTGGGCaTGGTGGACtTGCAGcAGaqTaATaTTcCCacTaaaaqaaqaaaqtcatcTaGAaqTa   197
HIV2ROD    G--G--A-----A--------T-----C--GG-A--A--C--AT-AAAAGGAAACAGTCATC-A--GA-A   207
HIV2NIHZ   G--G--G-----A--------T-----C--GG-A--A--C--AC-GAAAGAGAACAGTCATC-G--GA-A   207
HIV2ISY    G--G--G-----A--------T-----C--GG-A--A--C--AT-AAAAGGAAACAGTCATC-A--GA-A   207
HIV2ST     G--G--A-----G--------T-----C--GG-A--A--C--AT-AAAAGGAGAAAGTCATC-G--GA-A   207
HIV2BEN    G--G--A-----A--------T-----C--GG-A--A--C--AC-ACAAGGAAAAAGTCATC-A--AA-A   207
HIV2D194   G--A--A-----A--------T-----C--GG-A--A--C--AC-AGAAGGAGAAAGTCATC-A--GA-A   207
HIV2GH1    G--G--A-----G--------T-----C--GG-A--A--C--TC-AAAAGGAGAAAGTCATC-A--GA-A   207
HIV2D205   G--A--A-----T--------T-----T--AA-A--A--T--CC-AAACAAAGGAGCATGGC-A--AG-C   207
SIVMM251   G--C--A-----A--------C-----C--AG-A--C--C--CC-ACAGGAAGGAAGCCATT-A--AG-A   207
SIVMM142   G--C--A-----A-----  --C-----C--AG-A--C--C--AC-ACAGGAAGGAAGCCATT-A--AG-A   207
SIVMM239   G--C--A-----A--------C-----C--AG-A--C--C--AC-ACAGGAAGGAAGCCATT-A--AG-A   207
SIVMNE     G--C--A-----A--------C-----C--AG-A--C--C--AC-ACAAGAAAAAAGCCAAT-A--AG-A   207
SIVSMMH4   A--T--A-----A--------T-----C--AG-A--T--C--AT-AAGGGATGAGACTCATT-G--AG-A   207
SIVSMMPBJ  A--T--A-----A--------T-----C--AG-A--C--C--AT-AAAGGATGAAGCTCATT-G--AG-A   207
```

HIV2/SIV VIF

```
CONSENSUS  CAaGgaTAtTGGaActTaaCaCCAGAaAaaGGaTggcTcagctcttAtgCaGTgAGaaTaACcTGGTAc?  266
HIV2ROD    --G-CA--T---A-CT-AA-A-----A-AA--A-GGC-CTCCTCTT-TT-A--A--AA-A--T-----CA  277
HIV2NIHZ   --A-CA--T---A-CT-AA-A-----A-AA--A-GGC-CTCCTCTC-CT-A--G--AA-A--T-----TA  277
HIV2ISY    --G-CA--T---A-CC-AA-A-----A-AA--A-GGC-CTCCTCTT-TT-A--A--AA-G--T-----CT  277
HIV2ST     --G-CA--C---A-CC-AA-A-----A-AA--A-GGC-CTCCTCCT-TT-A--A--AC-A--T-----TA  277
HIV2BEN    --G-CA--T---A-CC-AA-A-----A-AA--A-GGC-CTCCTCTC-TG-A--A--AT-A--C-----TA  277
HIV2D194   --G-CA--T---A-CC-AA-A-----A-AA--A-GGC-CTCCTCTC-TT-A--A--GT-A--C-----TA  277
HIV2GH1    --G-CA--T---A-CC-AA-G-----A-AA--A-GGC-CTCCTCTC-TT-A--G--AA-A--C-----TA  277
HIV2D205   --A-GA--T---A-CC-AA-C-----A-GG--A-TCT-GAGCTCCT-TG-T--A--AC-A--A-----TG  277
SIVMM251   --A-GG--T---C-TT-GA-A-----A-GA--G-GGC-CAGTACTT-TG-A--G--GA-A--C-----CT  277
SIVMM142   --A-GG--T---A-TT-GA-A-----A-GA--G-GGC-CAGTACTT-TG-A--G--GA-A--C-----CT  277
SIVMM239   --A-GG--T---C-TT-GA-A-----A-AA--G-GGC-CAGTACTT-TG-A--G--GA-A--C-----CT  277
SIVMNE     --A-GG--T---A-TT-GA-A-----A-GA--G-GGC-CAGTACTC-TG-A--G--AA-A--C-----CT  277
SIVSMMH4   --A-GA--T---A-TT-GG-A-----A-AA--A-GGC-CAGTACTT-TG-A--A--AA-A--C-----CT  277
SIVSMMPBJ  --G-GA--T---A-TT-GA-A-----G-AA--G-GGC-CAGTACTT-TG-A--G--AA-A--C-----CT  277

CONSENSUS  caaaaaa?TTcTggaCAGATGT?ACaCCagActgtGCaGAca??cTAcTgCAtaGcaCTTATTTCcCTTG  332
HIV2ROD    CAGAAAAG--C-GGA-------T--C--AG-CTGT--A--TGTCC--A-A--TA-CA--------C----  347
HIV2NIHZ   CAGAAAAG--C-GGA-------T--C--AG-CTGT--A--TACCC--A-A--TA-CA--------T----  347
HIV2ISY    CAGAAGGG--C-GGA-------T--C--AG-CTGT--A--CACCC--A-A--CA-CA--------T----  347
HIV2ST     CAGAAAAA--C-GGA-------T--C--AG-CTGT--G--CTCCC--A-A--TA-CA--------T----  347
HIV2BEN    CAGAAAAG--C-GGA-------T--C--AG-CTGT--A--CATCC--A-A--TA-CA--------T----  347
HIV2D194   CAGAAAAG--C-GGA-------T--C--AG-CTGT--A--CTCCC--A-A--CA-CA--------T----  347
HIV2GH1    CAGAAAGG--C-GGA-------T--C--AG-CTAT--A--CATCC--A-A--TA-CA--------T----  347
HIV2D205   AGAGGAAC--T-ATA-------A--A--TG-TGTG--A--CCAGC--C-G--TG-GT--------T----  347
SIVMM251   CAAGGAAC--T-GGA-------A--A--AG-CTAT--A--CATTT--C-G--TA-CA--------C----  347
SIVMM142   CAAAGGAC--T-GGA-------A--A--AG-ATAT--A--TATTT--C-G--TA-CA--------C----  347
SIVMM239   CAAAGAAC--T-GGA-------A--A--AA-CTAT--A--CATTT--C-G--TA-CA--------C----  347
SIVMNE     CAAGGAAC--T-GGA-------A--A--AG-CTGT--A--CATTT--C-G--TA-CA--------C----  347
SIVSMMH4   CCAGAAAT--C-GGA-------A--A--AG-TTAT--A--CACTT--C-G--TA-CA--------C----  347
SIVSMMPBJ  CCAGAAAT--C-GGA-------G--A--AG-TTAT--A--CACTT--C-G--TG-CA--------C----  347

                \/J'aj
CONSENSUS  CTTTaCaGc?gg?GAAGTaaGaAgaGCCATCAGgCGaGAaaaqttatTgTCcTgCTGCAa?TaccC??aA  397
HIV2ROD    ----A-A-CAGGT-----AA-A-GA--------A--G--AAAGTTAT-G--C-G-----AT-ATC-CCG-  417
HIV2NIHZ   ----A-G-CAGGT-----AA-A-GA--------A--G--AAAATTAT-G--C-G-----AA-ACC-CCG-  417
HIV2ISY    ----A-G-CAGGT-----AA-A-GA--------G--A--AAAGTCAT-G--C-G-----AT-ATC-CCA-  417
HIV2ST     ----A-G-CAGGC-----AA-A-GA--------A--G--AAAGCTAT-A--C-G-----AC-ACC-CCA-  417
HIV2BEN    ----A-G-CAGGT-----AA-A-GA--------A--G--AAAGTTAT-G--C-G-----AC-ATC-CCA-  417
HIV2D194   ----A-G-CAGGT-----AA-A-GA--------A--G--AAAGTTAT-G--C-G-----AC-ACC-CCA-  417
HIV2GH1    ----A-G-CAGGT-----AA-A-GA--------A--G--AAAGTTAT-G--C-G-----AC-ATC-CCA-  417
HIV2D205   ----T-A-CCAAT-----AA-G-GA--------G--A--AAAGATAT-G--C-A-----AC-ATC-ATC-  417
SIVMM251   ----A-A-CGGGA-----GA-A-GG--------G--A--ACAACTGC-G--T-G-----AG-TCC-GAG-  417
SIVMM142   ----A-A-CGGGA-----GA-A-GG--------G--A--ACGACTGC-G--T-G-----GG-TCC-AAG-  417
SIVMM239   ----A-A-CGGGA-----GA-A-GG--------G--A--ACAACTGC-G--T-G-----GG-TCC-GAG-  417
SIVMNE     ----A-A-CGGGA-----GA-A-GG--------G--A--ACAACTGC-G--T-G-----GG-TCC-GAG-  417
SIVSMMH4   ----T-A-AGGGA-----AC-G-AG--------G--A--GAAATTGC-G--T-G-----AG-TCC-GAA-  417
SIVSMMPBJ  ----T-A-AGGGA-----AC-A-GG--------G--A--AAAATTGC-G--T-G-----AG-TCC-GAA-  417
```

```
                5'sj (rare)\/                                          vpX cds ->
CONSENSUS  GCtCAtaag??ccAGGTACCaag?cTaCAgTaccTaGCccTaa?aGTaGTaca??aa?atg?c???agat    456
HIV2ROD    --T--TAGAGCCC--------GTCAC-T--A-TTC-G--CT-AGTG--A--GCAACAAAATGAC...AGAC  484
HIV2NIHZ   --G--TAGATCCC--------GTCAC-T--A-TTC-G--CT-AGTG--A--GCAACAAAATGAC...AGAT  484
HIV2ISY    --C--TAAGTCCA--------GTCAC-C--A-TTC-G--CT-AGTA--A--ACAGCAAAATGAC...AAAC  484
HIV2ST     --C--TAAGTACC--------GTCAC-C--G-TTC-G--CT-AGTG--A--GCAACAAAATGGC...AGGC  484
HIV2BEN    --T--TAAAGCAC--------ATCAC-T--A-ACC-A--CC-AGTA--A--ACAACAAAATGAC...AGAC  484
HIV2D194   --T--TAAAGCAC--------ATCAC-T--A-ACC-A--CC-AGTG--A--GCAACAAAATGGC...AGAC  484
HIV2GH1    --T--TAAAGTAC--------ATCAC-T--A-ACC-A--CC-AGTG--A--GCAACAAAATGAC...AGAC  484
HIV2D205   --T--CGAAGGGC--------AAGCT-A--G-TTC-A--CC-AAGG--C--ACAGGAAGGAAAAAATGGAT  487
SIVMM251   --T--TAGGTACC--------AAGCC-A--G-ACT-A--AC-AAAA--A--A...AGCGATGTC...AGAT  481
SIVMM142   --T--TAAGCACC--------AAGTC-A--G-ACT-A--AC-GAGA--A--A...AGTCATGTC...AGAT  481
SIVMM239   --T--TAAGTACC--------AAGCC-A--G-ACT-A--AC-GAAA--A--A...AGCGATGTC...AGAT  481
SIVMNE     --T--TAAGACCC--------AAGTC-A--G-ACT-A--AC-GAGA--A--A...AGTTATGTC...AGAT  481
SIVSMMH4   --T--TAAAAATC--------AAGCC-A--G-ATC-A--AC-AACA--A--A...AGTCATGTC...AGAT  481
SIVSMMPBJ  --T--TAAGAATC--------AAGCC-A--G-ATC-A--AC-AACA--A--G...AGCCATGTC...AGAT  481

CONSENSUS  CCCAGaGAgAgaaT?CCaCCaGGAAACAGtGgaGAaGAgaCaaTaGgagAgGCcTTCGa?TGGCTAgAcA  524
HIV2ROD    -----A--G-CAG-A--A--A---------C-GC--A--GA-TA-C-GAG-G--C----CC------A-C-  554
HIV2NIHZ   -----G--A-CAG-G--A--A---------C-GC--G--GA-TA-C-AAG-G--C----CA------G-C-  554
HIV2ISY    -----A--G-CAA-A--A--A---------T-GC--A--AA-TA-C-AAG-G--C----AT------G-C-  554
HIV2ST     -----A--G-CAA-A--A--A---------T-GC--A--AA-TA-C-GAG-G--C----AG------G-C-  554
HIV2BEN    -----A--A-GGG-A--G--A---------T-GA--A--GA-CA-T-GAG-G--C----AG------G-G-  554
HIV2D194   -----A--A-GGG-G--G--A---------T-GA--A--GA-CA-T-GAG-G--C----AG------G-C-  554
HIV2GH1    -----A--A-GGG-A--G--A---------T-GA--A--GA-CA-T-GAG-G--C----AG------G-C-  554
HIV2D205   -----G--G-GAG-G--A--A---------C-AC--A--AA-AG-A-GAG-A--A----CT------G-A-  557
SIVMM251   -----G--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-G--C----AA------A-C-  551
SIVMM142   -----G--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-A--C----AG------A-C-  551
SIVMM239   -----G--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-G--C----AA------A-C-  551
SIVMNE     -----A--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-A--C----AA------A-C-  551
SIVSMMH4   -----G--G-GGA-C--A--T---------T-GA--A--AA-AA-A-GAA-G--C----AA------A-C-  551
SIVSMMPBJ  -----G--G-GAA-C--A--T---------T-GA--A--AA-AA-A-GCG-G--C----AC------G-C-  551

CONSENSUS  GaACagTAgaaGa?aTaAACaGaGagGCaGTqAAcCAcctqCC?aGgGAqcT?ATTTTCCAGGT?TGGCa  590
HIV2ROD    -G--AG--GAA-CCA-A---A-A-AA--A--G--T--CCTA--CC-A--AC-T----------G----A  624
HIV2NIHZ   -G--AG--GAG-CTA-A---A-A-AG--A--G--T--CCAC--AC-A--GC-C----------C----A  624
HIV2ISY    -G--GG--GAA-CCA-A---A-A-AG--A--G--C--CCTG--CA-G--GC-T----------G----A  624
HIV2ST     -G--GG--GAA-CCA-A---A-A-AG--A--G--C--CCTG--CC-A--GC-T----------G----A  624
HIV2BEN    -G--CA--GAA-CCT-A---A-G-AG--A--G--C--TCTG--CC-A--GC-C----------G----A  624
HIV2D194   -G--TA--GAA-CCT-A---A-G-AG--A--G--C--TCTG--CC-A--GC-C----------G----A  624
HIV2GH1    -G--TA--GAA-CCT-A---A-A-AG--A--G--C--TCTG--CC-A--GC-C----------G----A  624
HIV2D205   -A--AA--ACA-AGC-C---A-G-TA--G--C--C--TTTG--CC-A--AC-T----------C----A  627
SIVMM251   -A--AG--GAG-AGA-A---A-A-AG--A--A--C--CCTA--AA-G--GC-G----------T----A  621
SIVMM142   -A--AG--GAG-AGA-A---A-A-AG--G--A--C--CCTA--GA-G--GC-A----------T----A  621
SIVMM239   -A--AG--GAG-AGA-A---A-A-AG--G--A--C--CCTA--AA-G--GC-A----------T----A  621
SIVMNE     -A--AG--GAG-AGA-A---A-A-AG--A--A--C--CCTA--AA-G--GT-G----------T----A  621
SIVSMMH4   -A--AG--GAA-AAA-A---A-G-CA--A--A--T--CCTG--GA-G--GC-A----------T----G  621
SIVSMMPBJ  -A--AG--GAA-AAA-A---A-G-CA--A--G--T--TTTG--GA-G--AC-A----------C----G  621

CONSENSUS  aAGGTCcTGGq?ATAcTGGCa***    613
HIV2ROD    G-----C---AG---C----A---    648
HIV2NIHZ   G-----C---AG---C----A---    648
HIV2ISY    A-----C---AG---C----A---    648
HIV2ST     A-----C---AG---C----A---    648
HIV2BEN    A-----C---AG---T----A---    648
HIV2D194   A-----C---GC---T----A---    648
HIV2GH1    A-----C---AG---C----A---    648
HIV2D205   G-----T---GC---C----G---    651
SIVMM251   A-----T---GA---C----A---    645
SIVMM142   A-----T---GA---C----A---    645
SIVMM239   A-----T---GA---C----A---    645
SIVMNE     A-----T---GA---C----A---    645
SIVSMMH4   A-----C---GA---C----A---    645
SIVSMMPBJ  A-----C---GA---C----A---    645
```

HIV2/SIV VPX

Splice sites have been determined for SIV only.

```
CONSENSUS  ATG?cagatCCCAGaGAgAgaaT?CCaCCaGGAAACAGtGgaGAaGAgaCaaTaGgagAgGCcTTCGa?T   67
HIV2ROD    ---ACAGAC-----A--G-CAG-A--A--A---------C-GC--A--GA-TA-C-GAG-G--C----CC-   70
HIV2NIHZ   ---ACAGAT-----G--A-CAG-G--A--A---------C-GC--G--GA-TA-C-AAG-G--C----CA-   70
HIV2ISY    ---ACAAAC-----A--G-CAA-A--A--A---------T-GC--A--AA-TA-C-AAG-G--C----AT-   70
HIV2ST     ---GCAGGC-----A--G-CAA-A--A--A---------T-GC--A--AA-TA-C-GAG-G--C----AG-   70
HIV2BEN    ---ACAGAC-----A--A-GGG-A--G--A---------T-GA--A--GA-CA-T-GAG-G--C----AG-   70
HIV2D194   ---GCAGAC-----A--A-GGG-G--G--A---------T-GA--A--GA-CA-T-GAG-G--C----AG-   70
HIV2GH1    ---ACAGAC-----A--A-GGG-A--G--A---------T-GA--A--GA-CA-T-GAG-G--C----A G-   70
HIV2D205   ---...GAT-----G--G-GAG-G--A--A---------C-AC--A--AA-AG-A-GAG-A--A----CT-   67
SIVMM251   ---TCAGAT-----G--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-G--C----AA-   70
SIVMM142   ---TCAGAT-----G--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-A--C----AG-   70
SIVMM239   ---TCAGAT-----G--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-G--C----AA-   70
SIVMNE     ---TCAGAT-----A--G-GAA-C--A--T---------T-GA--A--GA-AA-A-GAG-A--C----AA-   70
SIVSMMH4   ---TCAGAT-----G--G-GGA-C--A--T---------T-GA--A--AA-AA-A-GAA-G--C----AA-   70
SIVSMMPBJ  ---TCAGAT-----G--G-GAA-C--A--T---------T-GA--A--AA-AA-A-GCG-G--C----AC-   70

CONSENSUS  GGCTAgAcAGaACagTAgaaGa?aTaAACaGaGagGCaGTgAAcCAcctgCC?aGgGAgcT?ATTTTCCA  134
HIV2ROD    -----A-C--G--AG--GAA-CCA-A---A-A-AA--A--G--T--CCTA--CC-A--AC-T--------  140
HIV2NIHZ   -----G-C--G--AG--GAG-CTA-A---A-A-AG--A--G--T--CCAC--AC-A--GC-C--------  140
HIV2ISY    -----G-C--G--GG--GAA-CCA-A---A-A-AG--A--G--C--CCTG--CA-G--GC-T--------  140
HIV2ST     -----G-C--G--GG--GAA-CCA-A---A-A-AG--A--G--C--CCTG--CC-A--GC-T--------  140
HIV2BEN    -----G-G--G--CA--GAA-CCT-A---A-G-AG--A--G--C--TCTG--CC-A--GC-C--------  140
HIV2D194   -----G-C--G--TA--GAA-CCT-A---A-G-AG--A--G--C--TCTG--CC-A--GC-C--------  140
HIV2GH1    -----G-C--G--TA--GAA-CCT-A---A-A-AG--A--G--C--TCTG--CC-A--GC-C--------  140
HIV2D205   -----G-A--A--AA--ACA-AGC-C---A-G-TA--G--C--C--TTTG--CC-A--AC-T--------  137
SIVMM251   -----A-C--A--AG--GAG-AGA-A---A-A-AG--A--A--C--CCTA--AA-G--GC-G--------  140
SIVMM142   -----A-C--A--AG--GAG-AGA-A---A-A-AG--G--A--C--CCTA--GA-G--GC-A--------  140
SIVMM239   -----A-C--A--AG--GAG-AGA-A---A-A-AG--G--A--C--CCTA--AA-G--GC-A--------  140
SIVMNE     -----A-C--A--AG--GAG-AGA-A---A-A-AG--A--A--C--CCTA--AA-G--GT-G--------  140
SIVSMMH4   -----A-C--A--AG--GAA-AAA-A---A-G-CA--A--A--T--CCTG--GA-G--GC-A--------  140
SIVSMMPBJ  -----G-C--A--AG--GAA-AAA-A---A-G-CA--A--G--T--TTTG--GA-G--AC-A--------  140

                                                    <- vif
CONSENSUS  GGT?TGGCaaAGGTCcTGGg?ATAcTGGCaTGAtGAacaaGGgATGtCa??aAGcTAcacaAAaTAtAGA  200
HIV2ROD    ---G----AG-----C---AG---C----A---T--ACAA--G---T-AGAA--T--CACA--G--T---  210
HIV2NIHZ   ---C----AG-----C---AG---C----A---T--ACAA--G---T-AACA--C--CACA--A--T---  210
HIV2ISY    ---G----AA-----C---AG---C----A---T--GCAA--G---T-ACGA--C--CACA--G--T---  210
HIV2ST     ---G----AA-----C---AG---C----A---T--ACAA--A---T-AATA--T--CACA--G--T---  210
HIV2BEN    ---G----AA-----C---AG---T----A---T--ACAA--G---T-AGCA--C--CACA--G--T---  210
HIV2D194   ---G----AA-----C---GC---T----A---T--ACAA--G---T-AACA--T--CACA--G--T---  210
HIV2GH1    ---G----AA-----C---AG---C----A---T--CCAA--G---T-ACCA--C--CACA--G--T---  210
HIV2D205   ---C----AG-----T---GC---C----G---G--ACAG--C---T-AATT--C--TACC--A--T---  207
SIVMM251   ---T----AA-----T---GA---C----A---T--ACAA--G---T-ACAA--C--TGTA--A--C---  210
SIVMM142   ---T----AA-----T---GA---C----A---T--ACAA--G---T-ACAA--C--TACA--A--C---  210
SIVMM239   ---T----AA-----T---GA---C----A---T--ACAA--G---T-ACCA--C--TGTA--A--C---  210
SIVMNE     ---T----AA-----T---GA---C- --A---T--ACAA--G---T-GCAA--C--TGTA--G--C---  210
SIVSMMH4   ---T----GA-----C---GA---C----A---T--AATG--A---T-AGAA--C--CACA--A--T---  210
SIVSMMPBJ  ---C----GA-----C---GA---C----A---T--AATG--G---T-AGTC--C--CACT--A--C---  210
```

```
CONSENSUS  TAcTTGTgctTaATaCAgAaaGCt?TgTttatgCAtt?cAagAaaGggtGtAg?TGcctgggggaaGgaC   267
HIV2ROD    --T----GCA-A--A--G-AA--AG-G-ACATG--TGTT-GG-AA-GGT-T-CT--CCTGGGGAGG-GA-   280
HIV2NIHZ   --T----GCT-A--A--G-TG--AA-G-ACATG--TGCT-AG-GA-ATG-C-CT--CCTGGGGGGA-G.-   279
HIV2ISY    --T----GCT-A--G--G-AA--TG-G-TCATG--TTTC-AG-AA-GGT-C-CT--CCGGGGGGAA-GA-   280
HIV2ST     --T----GCC-A--G--G-AA--TA-G-TCATA--TTCT-AG-GA-GGT-C-CT--CCTGGGGGGA-GA-   280
HIV2BEN    --T----GCC-A--G--A-AA--TA-A-TTACA--TTTC-AG-GA-GGT-C-CT--CTGGGGG.GA-GA-   279
HIV2D194   --T----GCA-A--G--G-AA--TG-G-ATATA--TTTC-AG-AG-GGT-C-CT--CCTGGGGAGA-GA-   280
HIV2GH1    --T----GCC-G--G--G-AA--TG-A-TTATA--TTTC-AG-GA-GGT-C-CT--CCTGGGGGGA-GA-   280
HIV2D205   --C----TGC-A--G--G-AA--AA-G-TTGTG--CTAT-CA-AG-GCT-T-GG--CCTGCAGGAG-GC-   277
SIVMM251   --C----GTT-A--G--A-AG--TT-A-TTATG--TTGC-AG-AA-GCT-T-GA--TCTAGGGGAA-GA-   280
SIVMM142   --C----GTT-A--A--A-AG--TT-A-TTATG--TTGC-AG-AA-GCT-T-GA--TCTAGGGGAA-GA-   280
SIVMM239   --C----GTT-A--A--A-AG--TT-A-TTATG--TTGC-AG-AA-GCT-T-GA--TCTAGGGGAA-GA-   280
SIVMNE     --C----GTT-A--A--A-AG--TT-A-TTATG--TTGC-AG-AA-GCT-T-GA--TCTAGGGGAA-GA-   280
SIVSMMH4   --C----GCT-G--A--G-AA--TC-G-TTGTG--TTGC-AG-AA-GGT-T-GG--CTTAGGAGAG-AG-   280
SIVSMMPBJ  --C----GCT-G--A--G-AA--TA-G-TTATG--TTGC-AG-AA-GGT-T-GG--CTTAGGAGGA-AG-   280

                                                                    vpr ->
                                                              \/3'sj
CONSENSUS  AtGGgcCaGga?GGa?TGG?AGAcCA?GGaCCTCCtCCTCCTCCcCCtCCAGG?cTaGcc***          325
HIV2ROD    -T--GC-A-GA.--G.---.---C--.--G-----T--------C--T-----TC-G-TC---          339
HIV2NIHZ   -T--GC-A-AA.--G.---.---T--G--A-----C--------A--T-----TT-A-TC---          339
HIV2ISY    -T--GC-A-GA.--G.---.---T--.--A-----C--------T--C-----TT-A-TC---          339
HIV2ST     -T--GC-G-GA.--A.---.---T--.--A-----C--------C--T-----TC-A-TC---          339
HIV2BEN    -T--GC-G-GAA--AT---A---C--A--A-----C--------C--T-----TC-A-TC---          342
HIV2D194   -T--CC-G-GA.--A.---.---C--.--A-----C--------C--T-----TC-A-TC---          339
HIV2GH1    -T--AC-G-GA.--A.---.---T--.--A-----C--------C--T-----TC-A-TC---          339
HIV2D205   -T--GC-A-GG.--A.---.---T--.--A-----T--------T--C-----CC-G-CC---          336
SIVMM251   -C--GG-A-GA.--A.---.---C--.--A-----T--------C--T-----AC-A-CA---          339
SIVMM142   -C--GG-A-GG.--A.---.---C--.--A-----T--------C--T-----AC-A-CA---          339
SIVMM239   -T--GG-A-GG.--A.---.---C--.--A-----T--------C--T-----AC-A-CA---          339
SIVMNE     -T--GG-A-GG.--A.---.---C--.--A-----T--------C--T-----AC-A-CA---          339
SIVSMMH4   -T--GG-A-GG.--A.---.---A--.--G-----T--------C--T-----AC-A-CA---          339
SIVSMMPBJ  -T--GG-A-GG.--A.---.---C--.--G-----T--------C--T-----AC-A-CA---          337
```

HIV2/SIV VPR

Splice sites have been determined for SIV only.

```
          <- vpX
CONSENSUS  ATGac?gaagca?c??cagA?a??CCtCCaGaagAtgagaccCCACagAGgGAaccatGGgatga?TGGg     62
HI"2ROD    ---GCTGAAGCACCAACAG-GCTC--C--G-TGG-TGGGACC----TG--G--GCCAG--GATGAG---A     70
HIV2NIHZ   ---ACTGAAGCACCAACAG-ACTC--C--G-AGG-TAGGACC----CG--G--GCCAG--GATGCA---G     70
HIV2ISY    ---ACTGAAGCACCAGCAG-GTTT--C--G-AGG-TGGGACC----CG--G--ACCAG--GATGAG---G     70
HIV2ST     ---ACTGAAGCACCAACAG-GTCT--C--G-AGG-TAGGACC----CG--G--GCCAG--GATGAG---G     70
HIV2BEN    ---ACTGAAGCACCAACAG-GTTT--C--A-AAG-TGGGACC----GG--G--CTTAG--AGTGAC---G     70
HIV2D194   ---ACTGAAGCACCAACAG-GTTT--C--A-AAG-TGGGACC----GG--A--GCTAG--AGTACC---G     70
HIV2GH1    ---ACTGAAGCACCAACAG-GTTT--C--A-AAG-TGGGACC----GG--G--ACTAG--GGTGAC---G     70
HIV2D205   ---GCAGAAGCAGCCCCAG-GATC--T--A-AGA-CGAGAAC----AA--A--ACCGT--GAAGAG---A     70
SIVMM251   ---...GAA.........G-AAGA--T--A-AAA-TGAAGGC----AA--G--ACCAT--GATGAA---G     58
SIVMM142   ---...GAA.........G-AAGA--T--A-AAA-TGAAGGC----AA--G--ACCAT--GATGAG---G     58
SIVMM239   ---...GAA.........G-AAGA--T--A-AAA-TGAAGGA----AA--G--ACCAT--GATGAA---G     58
SIVMNE     ---...GAA.........G-AAGA--T--A-AAG-TGAAGGC----AA--G--ACCAT--GATGAA---G     58
SIVSMMH4   ---GCA............G-AAGA--T--A-AAG-TGAAGCC----AG--G--ACCAT--GATGAA---G     58
SIVSMMPBJ  ---ACA............G-AAGA--T--A-AAG-TGAAGCC----AG--G--ACCGT--GATGAA---G     58

CONSENSUS  TAgtagaaattcTggaaGAaaTaaAagAAGAAGCttTaaagCAtTTTGAtCCtcGctTgCTAActgC?CT    131
HIV2ROD    --ATAGAAATCT-GAGA--AA-AA-AG-------TT-AAAG--T-----C--TC-CT-G----TTG-T--    140
HIV2NIHZ   --AT/ AAATCC-GAGA--AA-AG-AG-------TT-AAGG--T-----C--TC-CT-G----CTG-T--    140
HIV2ISY    --A1 .AAATTC-GAGA--AA-AA-GG-------TT-AAAG--T-----C--TC-CT-G----CTG-T--    140
HIV2ST     --ATAGAAACCC-GAGA--GA-AA-AT-------TT-AAAG--C-----C--TC-CT-G----TTA-T--    140
HIV2BEN    --ATAGAAACTC-GAGG--AA-AA-GG-------CT-AAGA--T-----T--CC-CT-G----TTG-T--    140
HIV2D194   --ATAGAAACTC-GAAG--AA-CA-GG-------CT-AAAA--T-----T--CT-CT-G----TTG-T--    140
HIV2GH1    --ATAAGAATTC-GGGA--AA-AA-GG-------CT-AAAG--T-----T--CC-CT-G----TCG-T--    140
HIV2D205   --GGGGAGATCC-GGAG--AA-AA-GC-------CT-AAAG--T-----T--TC-CT-G----CTG-G--    140
SIVMM251   --GTGGAGGTTC-GGAA--AT-GA-AG-------TT-AAAA--T-----T--TC-CT-G----CTG-A--    128
SIVMM142   --GTGGAAGTTC-GAAA--AC-GA-AG-------TT-AAAG--T-----T--TC-GC-T----CCG-A--    126
SIVMM239   --GTGGAGGTTC-GGAA--AC-GA-AG-------TT-AAAA--T-----T--TC-CT-G----CTG-A--    128
SIVMNE     --GTGGAGGTTC-GGAG--AC-GA-AG-------TT-AAAA--T-----T--TC-CT-G----CTG-G--    128
SIVSMMH4   --GTGGAAGTTC-AGAG--AT-AA-AG-------CC-GAAA--T-----T--TC-CT-G----CTG-G--    128
SIVSMMPBJ  --GTAGAAGTTC-GGAG--AA-AA-AG-------CC-GAAC--T-----T--TC-CT-G----CTG-G--    128

                      tat cds ->
                          \/J'sj
CONSENSUS  TGGtaactaTATctAtaaTAGaCATGGAGAcACcCtTGaaGGaGCagGAGagCTCATTAgaatCCT?CAa    200
HIV2ROD    ---CAAATA---CT-TAC---A--------C--C-T--AA--C--CA---AG-------AAGT---G--A    210
HIV2NIHZ   ---CAGATA---CT-TAC---G--------C--C-T--AA--C--CA---AG-------GAAT---G--A    210
HIV2ISY    ---CTACTA---CT-TAC---A--------C--C-T--AA--C--CA---AG-------GGGT---A--A    210
HIV2ST     ---CAACTA---CT-TGC---A--------C--C-T--AA--C--CA---GG-------GGAT---A--A    210
HIV2BEN    ---CTACTA---CC-TAA---A--------C--C-T--AA--C--CA---AG-------AAAC---A--A    210
HIV2D194   ---CAACTA---CT-TAA---A--------C--C-T--AA--A--CA---AG-------GAGT---A--A    210
HIV2GH1    ---CAACTA---CC-TAG---A--------C--A-C--AA--C--CA---AG-------GAAT---A--A    210
HIV2D205   ---TAACTT---CT-CAG---G--------T--C-T--CA--A--AG---AG-------AAAT---C--A    210
SIVMM251   ---TAATCA---CT-TAA---A--------C--C-T--AG--A--AG---AA-------GAAT---C--A    198
SIVMM142   ---TAATCA---CT-TAA---A--------C--C-T--AG--A--AG---AA-------GAAT---C--A    198
SIVMM239   ---TAATCA---CT-TAA---A--------C--C-T--AG--A--AG---AA-------GAAT---C--A    198
SIVMNE     ---TAATCA---CT-TAA---A--------C--C-T--AG--A--AG---AA-------AAAT---C--A    198
SIVSMMH4   ---TAACTA---TT-TGA---A--------C--C-T--AA--A--AG---AA-------GAAT---G--G    198
SIVSMMPBJ  ---TAACTA---CT-TGA---A--------C--C-T--AA--A--AG---AG-------GAAT---C--G    198
```

```
CONSENSUS  cGaGCcCTctTCatgCActTcAGAgc?GGaTGt??cCacTCaAGaAttgGccaa?Ca?GggGAgGAaaTC   265
HIV2ROD    C-A--C--TT--ACG--CT-C---GCA--A--TGGC-AC--A--A-TTG-C--GA-AA-GG--G--AA--   280
HIV2NIHZ   C-A--C--CT--GCG--CT-C---GCA--A--TGGC-AC--A--A-TTG-C--GA-AA-GG--G--AA--   280
HIV2ISY    C-A--C--CT--ACG--CT-C---GCA--A--TGGC-AC--A--A-TTG-C--AC-AA-GG--A--AA--   280
HIV2ST     C-A--C--CC--TTG--CT-C---GCA--A--CGGC-GC--A--G-TTG-T--GC-CA-GG--C--AA--   280
HIV2BEN    C-A--C--CT--GTG--CT-C---GCG--A--TAAC-GC--A--A-TTG-C***                   264
HIV2D194   C-A--C--CT--GTG--CA-C---GCG--A--TGAC-GC--A--A-AGG-C--AA-AA-GA--A--GC--   280
HIV2GH1    C-A--C--CT--GTG--CC-C---GCA--G--TAAC-GC--A--A-TTA-C--AA-AA-GA--A--AC--   280
HIV2D205   C-A--C--CT--CTC--CT-C---GCC--T--TCAA-AC--A--G-TTG-A--AT-AG-GG--G--AA--   280
SIVMM251   C-A--G--CT--ATG--TT-T---GGC--A--CAAC-AC--C--A-TCG-C--AC-TG-GG--G--AA--   268
SIVMM142   C-A--G--CT--ATA--TT-C---AGC--C--CAGC-AT--C--A-TCG-C--AC-TG-GG--G--AA--   268
SIVMM239   C-A--G--CT--ATG--TT-C---GGC--A--CATC-AC--C--A-TCG-C--AC-TG-GG--G--AA--   268
SIVMNE     C-G--G--CT--ATG--CT-C---GGC--C--CACC-AC--T--A-TCG-C--AT-TG-AG--G--AA--   268
SIVSMMH4   A-A--A--CT--ATC--TT-C---AGT--A--TGCC-AC--C--A-TCG-C--AT-TA-AG--G--AA--   268
SIVSMMPBJ  A-A--A--CT--ATC--TT-C---GGT--A--CAGA-AC--C--A-TCG-C--GT-TG-AG--G--AA--   268

CONSENSUS  CTctctCA?CTatACCgcCCcCTagaggcatgc?a***    303
HIV2ROD    --CTCT--G--AT---GA--C--AGAAACATGCAA---    318
HIV2NIHZ   --CTCT--G--AT---GA--C--AGAGGCATGCAC---    318
HIV2ISY    --CTCT--G--AT---GA--C--AGAAACATGCAA---    318
HIV2ST     --TTAT--G--AT---AA--C--AGAGGCATGCGA---    318
HIV2D194   --TGCC--G--GC---GA--C--AGAGGCATGCAC---    318
HIV2GH1    --TTCC--G--GC---GA--C--AGAGGCATGTAC---    318
HIV2D205   --CTCT--A--AT---GC--C--***                306
SIVMM251   --CTCT--A--AT---GC--T--***                294
SIVMM142   --CTCT--A--AT---GC--T--AGAAGCATGCTA---    306
SIVMM239   --CTCT--G--AT---GC--T--AGAAGCATGCTA---    306
SIVMNE     --CTCT--A--AT---GC--T--AGAAGAATGCTA---    306
SIVSMMH4   --CTCT--A--AT---GC--T--AGAGGCGTGCTA---    306
SIVSMMPBJ  --CTCT--A--AT---GC--T--AGAGGCGTGTTA---    306
```

:V2/SIV TAT

Splice sites have been determined for SIV only.

```
CONSENSUS  ATGGAGAcACccCtTGaaGGaGCagGAGagCTCATTAgaatCCTcCAacGaGCcCTctTCatgCActTcAG    70
HIV2ROD    -------C--C-T--AA--C--CA---AG-------AAGT---G--AC-A--C--TT--ACG--CT-C--    70
HIV2NIHZ   -------C--C-T--AA--C--CA---AG-------GAAT---G--AC-A--C--CT--GCG--CT-C--    70
HIV2ISY    -------C--C-T--AA--C--CA---AG-------GGGT---A--AC-A--C--CT--ACG--CT-C--    70
HIV2ST     -------C--C-T--AA--C--CA---GG-------GGAT---A--AC-A--C--CC--TTG--CT-C--    70
HIV2BEN    -------C--C-T--AA--C--CA---AG-------AAAC---A--AC-A--C--CT--GTG--CT-C--    70
HIV2D194   -------C--C-T--AA--A--CA---AG-------GAGT---A--AC-A--C--CT--GTG--CA-C--    70
HIV2GH1    -------C--A-C--AA--C--CA---AG-------GAAT---A--AC-A--C--CT--GTG--CC-C--    70
HIV2D205   -------T--C-T--CA--A--AG---AG-------AAAT---C--AC-A--C--CT--CTC--CT-C--    70
SIVMM251   -------C--C-T--AG--A--AG---AA-------GAAT---C--AC-A--G--CT--ATG--TT-T--    70
SIVMMP11   -------C--C-T--AG--A--AG---AA-------GAAT---C--AC-A--G--CT--ATG--TT-T--    70
SIVMM142   -------C--C-T--AG--A--AG---AA-------GAAT---C--AC-A--G--CT--ATA--TT-C--    70
SIVMM239   -------C--C-T--AG--A--AG---AA-------GAAT---C--AC-A--G--CT--ATG--TT-C--    70
SIVMNE     -------C--C-T--AG--A--AG---AA-------AAAT---C--AC-G--G--CT--ATG--CT-C--    70
SIVSMMH4   -------C--C-T--AA--A--AG---AA-------GAAT---G--GA-A--A--CT--ATC--TT-C--    70
SIVSMMPBJ  -------C--C-T--AA--A--AG---AG-------GAAT---C--GA-A--A--CT--ATC--TT-C--    70

CONSENSUS  Ag??GGaTG?a?cCa TCaAGaAttgGccAa?Ca?GggGAgGAaaTCCTctctCA?CTatACCgcCCcCT   133
HIV2ROD    -GCA--A--TGGC-AC--A--A-TTG-CC-GA-AA-GG--G--AA----CTCT--G--AT---GA--C--   140
HIV2NIHZ   -GCA--A--TGGC-AC--A--A-TTG-CC-GA-AA-GG--G--AA----CTCT--G--AT---GA--C--   140
HIV2ISY    -GCA--A--TGGC-AC--A--A-TTG-CC-AC-AA-GG--A--AA----CTCT--G--AT---GA--C--   140
HIV2ST     -GCA--A--CGGC-GC--A--G-TTG-TC-GC-CA-GG--C--AA----TTAT--G--AT---AA--C--   140
HIV2BEN    -GCG--A--TAAC-GC--A--A-TTG-CT-AA-AA-GA--A--AC----TGCC--G--GC---GA--C--   140
HIV2D194   -GCG--A--TGAC-GC--A--A-AGG-CC-AA-AA-GA--A--GC----TGCC--G--GC---GA--C--   140
HIV2GH1    -GCA--G--TAAC-GC--A--A-TTA-CC-AA-AA-GA--A--AC----TTCC--G--GC---GA--C--   140
HIV2D205   -GCC--T--TCAA-AC--A--G-TTG-AC-AT-AG-GG--G--AA----CTCT--A--AT---GC--C--   140
SIVMM251   -GGC--A--CAAC-AC--C--A-TCG-CC-AC-TG-GG--G--AA----CTCT--A--AT---GC--T--   140
SIVMMP11   -GGC--A--CAAC-AC--C--A-TCG-CC-AC-TG-GG--G--AA----CTCT--A--AT---GC--T--   140
SIVMM142   -AGC--C--CAGC-AT--C--A-TCG-CC-AC-TG-GG--G--AA----CTCT--A--AT---GC--T--   140
SIVMM239   -GGC--A--CATC-AC--C--A-TCG-CC-AC-TG-GG--G--AA----CTCT--G--AT---GC--T--   140
SIVMNE     -GGC--C--CACC-AC--T--A-TCG-CC-AT-TG-AG--G--AA----CTCT--A--AT---GC--T--   140
SIVSMMH4   -AGT--A--TGCC-AC--C--A-TCG-CC-AT-TA-AG--G--AA----CTCT--A--AT---GC--T--   140
SIVSMMPBJ  -GGT--A--CAGA-AC--C--A-TCG-CC-GT-TG-AG--G--AA----CTCT--A--AT---GC--T--   140

                    <- vpR                                                  \/3'sj
CONSENSUS  agAggcaTGctaTAAcacaTGcTatTGtAA?aagTG?tGcTaccAtTGcCAGct?TGTTTtcttaaaAAg   200
HIV2ROD    AG-AACA--CAA---CTCA--C-AT--T--GCGA--CT-C-ACC-T--T---ATG-----TCTAAAC--G   210
HIV2NIHZ   AG-GGCA--CAC---CTCC--C-AT--C--GAAA--CT-C-ACG-T--C---CTG-----TTTACAA--G   210
HIV2ISY    AG-AACA--CAA---CAAA--C-TT--T--GGGG--CT-C-TCC-T--C---CTG-----TTTAAAC--G   210
HIV2ST     AG-GGCA--CGA---CAAA--T-AC--T--AAAG--CT-C-ACC-T--C---ATG-----TTTAAAC--G   210
HIV2BEN    AG-GCCA--CAC---CAAA--C-AT--T--GCGA--CA-T-TCC-T--C---CTG-----CTCGAAA--G   210
HIV2D194   AG-GGCA--CAC---CTCA--C-AT--T--GCAG--CA-T-ACC-T--C---CTG-----CTCGAAA--G   210
HIV2GH1    AG-GGCA--TAC---CTCA--C-AT--T--GCAG--CA-C-TCC-T--C---CTG-----CTTGAAA--A   210
HIV2D205   TA-GGCA--CGA---TACA--C-AC--T--GAAA--CT-C-ACC-T--C---CTT-----TCTTAAA--G   210
SIVMM251   TG-GGCG--CTA---CACA--C-AT--T--AAAG--TT-C-ACC-T--C---TTT-----TCTTAAA--G   210
SIVMMP11   TG-GGCG--CTA---CACA--C-AT--T--AAAG--TT-C-ACC-T--C---TTT-----TCTTAAA--G   210
SIVMM142   AG-AGCA--CTA---CACA--C-AT--C--AAAG--TT-C-ACC-T--C---TTT-----TCTTAAA--G   210
SIVMM239   AG-AGCA--CTA---CACA--C-AT--T--AAAG--TT-C-ACC-T--C---TTT-----TCTTAAA--A   210
SIVMNE     AG-AGAA--CTA---CACA--C-AT--C--AAAG--TT-C-ACC-T--C---TTT-----TCTTAAA--G   210
SIVSMMH4   AG-GGCG--CTA---CAAG--C-AC--T--GAGG--TT-C-ATC-T--T---CAT-----TCTTAAA--G   210
SIVSMMPBJ  AG-GGCG--TTA---CAAG--C-AC--T--AAGG--TT-C-ATC-T--T---CAT-----TCTTAAA--G   210
```

```
                  rev cds ->
CONSENSUS  GG?cT?GGgATAtgtTATGagcga?aa?gcaga?GaAGaAGaacTccgAagAagaCTAAGgCTaaT?Cat    264
HIV2ROD    --GC-C--G---TGT----AACGAAAGGGCAGAC-A--A--GAC-CCA-AG-AAA-----A--CA-C-GT    280
HIV2NIHZ   --GC-C--G---TGG----ACCGAAAGGGCAGAC-A--A--GAC-CCA-AG-AAA-----G--CA-C-GT    280
HIV2ISY    --GC-C--G---TGT----ACCGAAAGGGCAGAC-A--A--GAG-CCG-AG-AAA-----G--CA-T-GT    280
HIV2ST     --GC-C--G---TGG----AACGAAAGGGCAGAA-A--A--AAC-CCG-AG-AAA-----G--CA-T-GT    280
HIV2BEN    --GC-C--A---TCA----AGCGAAAGGGCAGAC-A--A--GAC-CCA-GG-AAA-----A--CC-T-GC    280
HIV2D194   --GC-C--G---TGG----CGCGACAGGGCAGAC-A--A--GAC-CCA-GA-AAA-----A--CA-C-GC    280
HIV2GH1    --GC-C--G---TGG----CACGAAAAAAGCAGAC-G--A--AAC-CCA-GA-AGA-----A--CA-T-GT    280
HIV2D205   --TC-T--G---TGT----ACCGCTCG......A-A--G--ATC-GCA-AA-GAG-----A--AC-G-AC    274
SIVMM251   --AT-G--G---TGT----AGCAGTCACGAAAGA-A--A--AAC-CCG-AA-AGG-----G--AA-A-AT    280
SIVMMP11   --AT-G--G---TGT----AGCAGTCACGAAAGA-A--A--AAC-CCG-AA-AGG-----G--AA-A-AT    280
SIVMM142   --CT-G--G---AGT----AGAAGTCACACAGGA-A--A--AAC-CCG-AG-AGG-----G--AA-A-AT    280
SIVMM239   --CT-G--G---TGT----AGCAATCACGAAAGA-A--A--AAC-CCG-AA-AGG-----G--AA-A-AT    280
SIVMNE     --CT-G--G---TGT----AGCAGTCACGCAGAA-A--A--AAC-CCG-AG-AGG-----G--AA-A-AT    280
SIVSMMH4   --TT-G--A---TGC----AGCAG...CAC...C-A--A--AAC-CCG-AG-AGA-----G--AA-C-AT    274
SIVSMMPBJ  --CT-G--A---TGC----AGCAGCAA......C-A--A--AAC-CCG-AG-AGA-----G--AA-A-AT    274

                       exon 2\/exon 3
CONSENSUS  cTtCTgCAtCAgAc???aaAtcCaTAtCCacaaGGacc?GG?aCaGCCAacCAaaGAAGgaacAgAAgaA    329
HIV2ROD    C-C--A--C--G-C...AA-TC-A--T--ACAA--ACCG--GA-A----GC--AC----AAAC-G--GA-    347
HIV2NIHZ   C-T--G--T--G-C...AA-TC-A--T--ACAA--ACCA--AA-A----GC--GA----AAAC-G--GA-    347
HIV2ISY    C-C--G--T--G-C...AA-TC-A--T--ACAA--ACTG--AA-A----GA--GA----AAAC-G--GA-    347
HIV2ST     C-T--G--T--G-C...AA-TC-A--T--ACAA--ACCG--AA-A----GC--GA----AAAC-G--GA-    347
HIV2BEN    C-T--G--C--G-C...AA-TC-A--T--ACAA--ACCG--GA-A----GC--AC----GAAC-G--GA-    347
HIV2D194   C-C--G--T--G-T...AA-TC-A--T--ACAC--ACAG--GA-A----GC--AC----AAAC-G--GA-    347
HIV2GH1    C-T--G--T--G-C...AA-TC-A--T--ACAC--ACCG--GA-A----GC--AC----GAAC-G--GA-    347
HIV2D205   C-T--G--C--***                                                           288
SIVMM251   C-T--G--T--A-C...AA-CT-A--C--AACA--ACCC--CA-T----AC--GA----GCAA-G--GG-    347
SIVMMP11   C-T--G--T--A-C...AA-CT-A--C--AACA--ACCC--CA-T----AC--GA----GCAA-G--GG-    347
SIVMM142   C-T--G--T--A-C...GA-CC-A--C--AACA--ATCC--CT-T----AC--AA----GCAA-A--GG-    347
SIVMM239   C-T--G--T--A-C...AA-CC-A--T--AACA--ACCC--CA-T----AC--GA----GCAA-G--AG-    347
SIVMNE     C-T--G--T--A-C...AA-CC-A--T--GACA--ACCA--CA-T----AC--AA----GAAC-G--GG-    347
SIVSMMH4   T-C--G--T--A-C...AA-TC-C--T--ACAA--GCCA--AA-C----AC--AA----GAGA-G--GG-    341
SIVSMMPBJ  C-T--G--T--G-C...AA-TC-C--T--AGAC--GCCA--AA-T----AC--AA----GAGA-G--GG-    341

CONSENSUS  GAcGgtgGagacagcggTGG?aaCagcTccTgGCCtTGGCaga***     372
HIV2ROD    -AC-GTG-AAGCAACGG---AGA-AGA-AC-G---C----CGA---     393
HIV2NIHZ   -AC-TTG-AAGCAACGG---AGA-AGA-CT-G---T----CGA---     393
HIV2ISY    -AC-TTG-GAACGACGT---AAG-AGA-CC-G---T----CGA---     393
HIV2ST     -AC-TTG-AAACAGCGT---AGA-AAT-GG-G---C----CGA---     393
HIV2BEN    -AC-TCG-AGCCCACAG---TTA-GAC-TG-G---T----CAA---     393
HIV2D194   -AC-CCG-AGACGACAG---TTT-GGC-TG-G---T----CAC---     393
HIV2GH1    -AC-ACA-AGACGACGA---TTA-GAC-TG-A---T----CGA---     393
SIVMM251   -AC-GTG-AGAAGGCGG---CAA-AGC-CC-G---T----AGA---     393
SIVMMP11   -AC-GTG-AGAAGGCGG---CAA-AGC-CC-G---T----AGA---     393
SIVMM142   -AC-GTG-AGGCAGCGG---CAA-AGC-CC-G---T----AGA---     393
SIVMM239   -AC-GTG-AGAAGGCGG---CAA-AGC-CC-G---T----AGA---     393
SIVMNE     -AC-GTG-AGGCAGCGG---CAA-AGC-CC-G---T----AGA---     393
SIVSMMH4   -AC-GTG-AGACAGAGG---CAG-AGA-CT-G---T----AGA---     387
SIVSMMPBJ  -AC-GTG-AGGCAGAGG---CAA-AGA-CT-G---T----AGA---     387
```

```
                                                                     exon 2\/exon 3
CONSENSUS   ATGagcga?aa?gcaga?GaAGAAGaacTccgAagAAgaCTAAGgCT?aT?CatcTtCTgCAtCAgAcaA     65
HIV2ROD     ---AACGAAAGGGCAGAC-A-----GAC-CCA-AG--AA-----A--CA-C-GTC-C--A--C--G-CA-     70
HIV2NIHZ    ---ACCGAAAGGGCAGAC-A-----GAC-CCA-AG--AA-----G--CA-C-GTC-T--G--T--G-CA-     70
HIV2ISY     ---ACCGAAAGGGCAGAC-A-----GAG-CCG-AG--AA-----G--CA-T-GTC-C--G--T--G-CA-     70
HIV2ST      ---AACGAAAGGGCAGAA-A-----AAC-CCG-AG--AA-----G--CA-T-GTC-T--G--T--G-CA-     70
HIV2BEN     ---AGCGAAAGGGCAGAC-A-----GAC-CCA-GG--AA-----A--CC-T-GCC-T--G--C--G-CA-     70
HIV2D194    ---CGCGACAGGGCAGAC-A-----GAC-CCA-GA--AA-----A--CA-C-GCC-C--G--T--G-TA-     70
HIV2GH1     ---CACGAAAAAGCAGAC-G-----AAC-CCA-GA--GA-----A--CA-T-GTC-T--G--T--G-CA-     70
SIVMM251    ---AGCAGTCACGAAAGA-A-----AAC-CCG-AA--GG-----G--AA-A-ATC-T--G--T--A-CA-     70
SIVMMP11    ---AGCAGTCACGAAAGA-A-----AAC-CCG-AA--GG-----G--AA-A-ATC-T--G--T--A-CA-     70
SIVMM142    ---AGAAGTCACACAGGA-A-----AAC-CCG-AG--GG-----G--AA-A-ATC-T--G--T--A-CG-     70
SIVMM239    ---AGCAATCACGAAAGA-A-----AAC-CCG-AA--GG-----G--AA-A-ATC-T--G--T--A-CA-     70
SIVMNE      ---AGCAGTCACGCAGAA-A-----AAC-CCG-AG--GG-----G--AA-A-ATC-T--G--T--A-CA-     70
SIVSMMH4    ---AGCAGCACC......-A-----AAC-CCG-AG--GA-----G--AA-C-ATT-C--G--T--A-CA-     64
SIVSMMPBJ   ---AGCAGCAAC......-A-----AAC-CCG-AG--GA-----G--AA-A-ATC-T--G--T--G-CA-     64

CONSENSUS   ???AtcCaTAtCCacaaGGacc?GG?aCaGCCAacCAaaGAAGgaacAgAAgaAGacGgtgGagacagcg    130
HIV2ROD     ...-TC-A--T--ACAA--ACCG--GA-A----GC--AC----AAAC-G--GA--AC-GTG-AAGCAACG    137
HIV2NIHZ    ...-TC-A--T--ACAA--ACCA--AA-A----GC--GA----AAAC-G--GA--AC-TTG-AAGCAACG    137
HIV2ISY     ...-TC-A--T--ACAA--ACTG--AA-A----GA--GA----AAAC-G--GA--AC-TTG-GAACGACG    137
HIV2ST      ...-TC-A--T--ACAA--ACCG--AA-A----GC--GA----AAAC-G--GA--AC-TTG-AAACAGCG    137
HIV2BEN     ...-TC-A--T--ACAA--ACCG--GA-A----GC--AC----GAAC-G--GA--AC-TCG-AGGCGACA    137
HIV2D194    ...-TC-A--T--ACAC--ACAG--GA-A----GC--AC----AAAC-G--GA--AC-CCG-AGACGACA    137
HIV2GH1     ...-TC-A--T--ACAC--ACCG--GA-A----GC--AC----GAAC-G--GA--AC-ACA-AGACGACG    137
SIVMM251    ...-CT-A--C--AACA--ACCC--CA-T----AC--GA----GCAA-G--GG--AC-GTG-AGAAGGCG    137
SIVMMP11    ...-CT-A--C--AACA--ACCC--CA-T----AC--GA----GCAA-G--GG--AC-GTG-AGAAGGCG    137
SIVMM142    ...-CC-A--C--AACA--ATCC--CT-T----AC--AA----GCAA-A--GG--AC-GTG-AGGCAGCG    137
SIVMM239    ...-CC-A--T--AACA--ACCC--CA-T----AC--GA----GCAA-G--AG--AC-GTG-AGAAGGCG    137
SIVMNE      ...-CC-A--T--GACA--ACCA--CA-T----AC--AA----GAAC-G--GG--AC-GTG-AGGCAGCG    137
SIVSMMH4    ...-TC-C--T--ACAA--GCCA--AA-C----AC--AA----GAGA-G--GG--AC-GTG-AGACAGAG    131
SIVSMMPBJ   ...-TC-C--T--AGAC--GCCA--AA-T----AC--AA----GAGA-G--GG--AC-GTG-AGGCAGAG    131

                                      <- tat
CONSENSUS   gTGG?aaCagcTccTgGCCtTGGCagaTAgaaTAtattCAtTTCCTGATCCgCCaacTgataCGCCTCTT    199
HIV2ROD     G---AGA-AGA-AC-G---C----CGA--GCA--TATA--T----------G--AGC-GATT--------    207
HIV2NIHZ    G---AGA-AGA-CT-G---T----CGA--GCA--TATA--T----------G--TGC-GATT--------    207
HIV2ISY     T---AAG-AGA-CC-G---T----CGA--GAA--TATA--T----------G--TGC-GATC--------    207
HIV2ST      T---AGA-AAT-GG-G---C----CGA--AGA--TATA--T----------G--AGC-GATT--------    207
HIV2BEN     G---TTA-GAC-TG-G---T----CAA--AAC--TGTG--G----------A--TAC-GACT--------    207
HIV2D194    G---TTT-GGC-TG-G---T----CAC--AAC--CATA--A----------A--TAC-GACT--------    207
HIV2GH1     A---TTA-GAC-TG-A---T----CGA--AAC--TATA--T----------A--TAC-GACT--------    207
SIVMM251    G---CAA-AGC-CC-G---T----AGA--GAA--TATT--T----------G--AAC-GATA--------    207
SIVMMP11    G---CAA-AGC-CC-G---T----AGA--GAA--TATT--T----------G--AAC-GATA--------    207
SIVMM142    G---CAA-AGC-CC-G---T----AGA--GAA--TATT--T----------G--AAC-GATA--------    207
SIVMM239    G---CAA-AGC-CC-G---T----AGA--GAA--TATT--T----------G--AAC-GATA--------    207
SIVMNE      G---CAA-AGC-CC-G---T----AGA--GAA--TATT--T----------G--AAC-AATA--------    207
SIVSMMH4    G---CAG-AGA-CT-G---T----AGA--GAA--TATT--T----------G--AGT-GATA--------    201
SIVSMMPBJ   G---CAA-AGA-CT-G---T----AGA--GAA--TATT--T----------G--AGT-GATA--------    201
```

```
CONSENSUS  GActtGgcTaTaCAgcA?cTgCAGagacTt?CTATccAgga?cTaCCagAtCCtCCaaCcaat?ttccag   265
HIV2ROD    --CCA-AC-A-A--GC-TC-G---GGAC-TA----CC-GGAGC-T--TG-C--T--AA-TCATCTACCAG   277
HIV2NIHZ   --CCG-GC-A-A--AC-TC-G---GGAC-TA----CC-GGATC-C--CG-T--T--AA-CAATCTTCCAG   277
HIV2ISY    --CCA-AC-A-A--AC-GC-G---GGAC-TA----CC-GACTC-A--TG-T--T--AA-CACT.......   270
HIV2ST     --ACA-AC-A-A--AC-TC-G---GGAC-TA----CC-GGAGC-T--AG-C--T--AA-TAATCTCCCAG   277
HIV2BEN    --TCG-GC-A-A--AC-TC-G---AGAC-TA----CC-AGAAC-C--CG-C--G--GA-TGATCTCCCAG   277
HIV2D194   --CCG-GC-A-A--AC-GC-G---GGGC-TA----CC-AGAAC-C--CG-C--G--GA-TGATCTCCCAG   277
HIV2GH1    --CCG-GC-A-A--AG-TC-G---AGAC-TA----CC-CGAAC-C--CG-C--A--GA-TGATCTCCCAG   277
SIVMM251   --CTT-GC-A-T--GC-AC-G---AACC-TG----CG-GAGCA-A--AG-T--T--AA-CAATACTCCAG   277
SIVMMP11   --CTT-GC-A-T--GC-AC-G---AACC-TG----CG-GAGCA-A--AG-T--T--AA-CAATACTCCAG   277
SIVMM142   --CTT-GC-A-T--GC-AC-G---AACC-TG----CG-GAGCA-A--AG-T--T--AA-CAATATTCCAG   277
SIVMM239   --CTT-GC-A-T--GC-AC-G---AACC-TG----CG-GAGTA-A--AG-T--T--AA-CAATACTCCAG   277
SIVMNE     --CTT-GT-A-T--GC-AC-G---AACC-TG----CG-GAGCA-A--AG-T--T--AA-CAATATTCCAG   277
SIVSMMH4   --CTT-GC-A-T--GC-GC-G---GGAT-GG----TG-GGAGC-A--AG-T--T--AA-CAGTGCTCCAG   271
SIVSMMPBJ  --CTT-GC-A-T--GC-GT-G---AGAT-GG----TG-GGAAC-G--AA-C--T--AG-CAGTGCTCCAG   271

CONSENSUS  ag?ctctc??????????c??ag??????a??ag???tcagggactgGC_GAGacT***             300
HIV2ROD    AATCT.............................CAGAGACTG--T---AC----                 303
HIV2NIHZ   AGTCT............CCAGAGAGCACTAACAGCAATCAGAGACTG--T---GC----             324
HIV2ISY    ..................................CAGAGACTG--T---AC----                 291
HIV2ST     AGTCT............TCGGAGAGCATTGACAGCAGTCAGAGACTG--T---AT----             324
HIV2BEN    AGTCT.....................AACAGCAATCAGGGACTG--T---AC----                 312
HIV2D194   AGTCT.....................AACAGCAATCAGGGACTG--T---AC----                 312
HIV2GH1    AATCT.....................AACAGCAATCAGGGACTG--T---AC----                 312
SIVMM251   AGGCTCTCTGCGACCCTACGAAGGGTTCGAGAAGTCCTCAGGAC***                          324
SIVMMP11   AGGCTCTCTGCGACCCTACGAAGGATTCGAGAAGTCCTCAGGAC***                          324
SIVMM142   AGGCTCTCTGCGACCCTACGGAGAATTCGAGAAGTCCTCAGGCT***                          324
SIVMM239   AGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGAC***                          324
SIVMNE     AGATTCTCCACGACCCTACAGAGAATCCGAGAAGTCC  AGGAC***                          324
SIVSMMH4   AGCCTCTCAACGACGTTGCAAAG.........AGTCCG***                               303
SIVSMMPBJ  AGCCTCTCAAGGACGCTGCAGAG.........AGCCCG***                          303
```

HIV2/SIV ENV

Splice sites have been determined for SIV only.

```
CONSENSUS  ATGgg?tgt??taggAatCagCTgcttaTtGccatctTgcTa?taaGTgctt?c?tgat?tatTGca??c   61
HIV2ROD    ---.........ATG-AT-AG--GCTTA-T-CCATTT-AT-AGCTA--GCTTGCTTAGTATAT--CACCC   61
HIV2NIHZ   ---AAGGGTAGTAAG-AT-AA--GCTGA-T-CTATTG-AC-AGCTA--GCTTACCTAATACAT--CAAGC   70
HIV2ISY    ---AG...TGGTAAA-TT-AG--GCTTG-T-CCTTTC-GC-AACTA--GCTTGCTTAATATAT--CACCA   67
HIV2ST     ---TG...TGGTAGG-AT-AA--ATTTG-T-CCAGCT-GC-AGCTA--GCTTGCTTAATATAT--CGTCC   67
HIV2BEN    ---GAGCCTGGTAGG-AT-AG--GTTTG-T-TCATTT-AC-AACAA--GCTTGCTTAGTATAT--TAGCC   70
HIV2D194   ---GAGCCTGGTAGG-AT-AG--GCTTG-T-CCATTT-AT-AACTA--GCTTGCTTAATATAT--CAAAC   70
HIV2GH1    ---TG...TGGTAAG-GT-TA--ATGTG-T-CCAGCT-GC-AGCTA--GCTTACTTAGTATAT--CACCC   67
HIV2D205   ---GCATATTTTAGC-GC-GC--GCCTA-T-CGCTCC-GC-TATAG--ATCAGTGGGTTTGTA--TAAAC   70
SIVMM251   ---GGATGTCTTGGG-AT-AG--GCTTA-C-CCATCT-GC-TTTAA--GTCTATGGGATCTAT--TACTC   70
SIVMM142   ---GGTTGTCTTGGA-AT-AG--GCTTA-C-CCATCT-GC-TCTAA--GTCTATGGGATTTAT--TATTC   70
SIVMM239   ---GGATGTCTTGGG-AT-AG--GCTTA-C-CCATCT-GC-TTTAA--GTCTATGGGATCTAT--TACTC   70
SIVMNE     ---GGATGTCTTGGG-AT-AG--GCTTA-C-CCATCT-GT-TCTAA--GCCTATGGGATCTAT--CATTC   70
SIVSMMH4   ---GGATGTCTTGGG-AT-AG--GCTTA-C-CGCTCT-GC-AGTAA--GTTTTAGAGATTTGT--TGTTC   70
SIVSMMPBJ  ---GGATGTCTTGGG-AT-AG--GCTTA-C-CGCTCT-GC-ATTAA--GCTTCAGGGATTTAT--TGTTC   70

CONSENSUS  aaTaTGT?ACtGT?TTcTAtGG?gTaCCcgc?TGGAggAAtGCaaC?aTTCCCcTctTtTGTGcAACcAa  126
HIV2ROD    AA-A---A--T--T--C--T--CG-A--CACG----AA--T--AA-CA-----C-CT-T----C---C-G  131
HIV2NIHZ   AA-T---G--T--T--C--C--CA-A--CGCG----GG--T--AT-CA-----C-GT-T----C---C-A  140
HIV2ISY    AA-A---G--T--T--C--T--AG-A--CGTG----AA--T--AT-CA-----C-CT-T----C---T-A  137
HIV2ST     AA-A---G--T--T--C--T--CG-G--CGTG----GA--T--AT-CA-----C-CT-T----C---T-A  137
HIV2BEN    AG-A---G--T--T--C--T--CA-A--CGCG----AA--T--AT-TA-----T-AT-T----C---T-A  140
HIV2D194   AA-A---G--T--T--C--T--CA-A--CGCG----GA--T--AT-TA-----C-AT-T----C---C-A  140
HIV2GH1    AA-A---G--T--T--C--T--CG-G--CGTG----GA--T--AT-CA-----C-CT-T----C---T-A  137
HIV2D205   AA-A---T--T--C--C--T--CA-A--CGCA----GG--C--AA-AG-----C-CA-T----C---C-C  140
SIVMM251   AA-A---C--A--C--T--T--TG-A--AGCT----GG--T--GA-AA-----C-CT-C----C---C-A  140
SIVMM142   AA-A---C--A--C--T--T--TG-A--AGCT----GG--T--GA-AA-----C-CT-C----C---C-A  140
SIVMM239   TA-A---C--A--C--T--T--TG-A--AGCT----GG--T--GA-AA-----C-CT-T----C---C-A  140
SIVMNE     AA-A---C--A--C--T--T--TG-A--AGCT----GG--T--GA-AA-----C-CT-C----T---C-G  140
SIVSMMH4   AA-A---A--A--A--C--T--TG-A--AGCA----AG--T--GA-AA-----C-CT-C----C---C-A  140
SIVSMMPBJ  AA-A---A--A--A--C--T--TA-A--AGCA----GG--T--GA-AG-----C-CT-C----C---C-A  140

CONSENSUS  aAAtAGgGAtACtTGGGGaACaataCAgTGcttgCCaGAcAATGaTGAtTAt?caGAaaT?gc??T?AAt  191
HIV2ROD    A--T--G--T--T-----A--CATA--G--CTTG--T--C----A---T--TCAG--AA-AACTT-G--T  201
HIV2NIHZ   A--T--A--T--T-----A--AATA--G--CTTA--A--C----A---T--TCAG--AA-AACTC-A--T  210
HIV2ISY    A--T--A--T--T-----A--CATA--G--CTTG--A--C----A---T--TCAA--GA-ACCTT-G--T  207
HIV2ST     A--T--A--T--T-----A--CATA--G--CTTG--A--C----A---C--TCAG--AA-AGCTT-A--T  207
HIV2BEN    A--T--A--C--T-----G--CATA--G--CTTG--A--C----A---T--TCAG--AA-AATTT-A--T  210
HIV2D194   A--T--A--T--T-----G--CATC--G--CTTG--A--C----A---T--TCAG--AA-AACCT-A--T  210
HIV2GH1    A--C--A--T--T-----A--CATA--G--CAAG--A--C----A---T--TCAA--AA-AACCT-G--T  207
HIV2D205   A--C--A--C--C-----A--TGTA--G--TCTC--A--C----G---C--CACT--GA-CAGGC-A--C  210
SIVMM251   G--T--G--T--T-----A--AACT--G--CCTA--A--T----G---T--TTCA--AT-GGCCC-T--T  210
SIVMM142   G--T--G--T--T-----A--AACT--G--CCTA--A--T----A---T--TTCA--AT-GGCCC-T--T  210
SIVMM239   G--T--G--T--T-----A--AACT--G--CCTA--A--T----G---T--TTCA--AG-GGCCC-T--T  210
SIVMNE     G--T--G--T--T-----A--AACT--G--CCTA--A--T----A---T--TTCA--AT-GGCCC-T--T  210
SIVSMMH4   G--T--G--C--T-----A--AACA--A--CTTG--A--T----A---T--CTCA--AT-GGCAA-C--T  210
SIVSMMPBJ  G--T--G--C--T-----A--AACA--A--CTTG--A--T----G---T--CTCA--AT-GGCAA-T--T  210
```

```
CONSENSUS  gT?ACAGAggctTTtGAtGCaTGGgAtAATAcAGT?ACAgAACAaGCAaTAGAgGAtGTcTGGaa?CT?T  257
HIV2ROD    G-A-----GGCT--T--T--A---A-T----C---A---G----A---A----A--T--C---CAT--A-  271
HIV2NIHZ   G-G-----GGCT--C--T--A---A-T----C---A---G----A---G----G--T--C---AAT--A-  280
HIV2ISY    G-A-----GGCT--T--C--A---G-T----T---A---G----A---G----A--T--C---AAT--A-  277
HIV2ST     G-G-----GGCC--C--C--A---A-T----C---A---G----A---G----A--T--C---AGT--A-  277
HIV2BEN    G-G-----GGCT--T--T--A---A-T----C---G---G----A---G----A--T--C---CAT--A-  280
HIV2D194   G-G-----AGCT--T--T--A---G-T----C---A---G----A---A----A--T--C---AGA--G-  280
HIV2GH1    G-G-----GGCT--C--C--A---G-T----C---A---G----A---G----A--T--C---AGT--A-  277
HIV2D205   A-A-----GGCT--T--T--A---G-T----C---G---C----G---G----T--T--G---AGA--C-  280
SIVMM251   G-T-----AAGC--T--T--T---G-G----C---C---G----G---A----G--C--A---CAA--C-  280
SIVMM142   G-T-----AAGC--T--T--T---G-G----C---C---G----G---A----G--C--A---CAA--C-  280
SIVMM239   G-T-----AAGC--T--T--C---A-T----C---C---G----G---A----G--T--A---CAA--C-  280
SIVMNE     A-T-----AAGC--T--T--T---G-G----C---C---G----G---A----G--T--A---CAT--C-  280
SIVSMMH4   G-C-----GGCT--T--T--T---G-T----C---C---G----A---A----G--T--G---AAC--C-  280
SIVSMMPBJ  G-C-----GGCT--T--T--T---G-T----C---C---G----A---A----G--T--A---AAC--C-  280

CONSENSUS  TtGAgACaTCaATaAAaCCaTGtGTcAAatTaaC?CCatTaTGtgT?gCaATGag?TGt???????????  313
HIV2ROD    -C--G--A--A--A--A--A--T--C--AC-AA-A--TT-A--TG-AG-A---AAA--C.........  330
HIV2NIHZ   -T--G--A--A--A--A--A--T--C--AT-AA-A--CT-A--TG-AG-A---AAC--TACAAGG.....  345
HIV2ISY    -T--G--A--A--A--A--A--T--C--AT-AA-G--TT-A--TG-AA-A---AAC--T.........  336
HIV2ST     -T--G--A--A--A--A--A--C--C--AC-AA-A--CT-A--TG-AG-A---CGT--T.........  336
HIV2BEN    -T--G--A--A--A--A--A--T--C--GC-AA-A--TC-A--TG-GG-A---AAT--TAGCAGGGTTCA  350
HIV2D194   -T--G--A--A--A--A--A--T--C--GT-GA-G--CC-A--TG-GG-G---AAT--T.........  339
HIV2GH1    -T--G--A--A--A--A--A--C--C--AC-AA-A--TC-A--TG-AG-A---AGT--T.........  336
HIV2D205   -T--A--C--C--A--A--A--T--C--AC-AA-C--AC-G--TG-GG-A---AAC--T.........  339
SIVMM251   -T--G--C--A--A--G--T--T--A--AT-AT-C--AT-A--CA-TA-T---AGA--C.........  339
SIVMM142   -T--G--C--A--A--G--T--T--A--AT-AT-C--AT-A--CA-TA-T---AGA--C.........  339
SIVMM239   -T--G--C--A--A--G--T--T--A--AT-AT-C--AT-A--CA-TA-T---AGA--C.........  339
SIVMNE     -T--G--C--A--A--G--T--T--A--AT-AA-C--AT-A--CA-TA-T---AAA--C.........  339
SIVSMMH4   -T--A--A--C--T--G--C--T--A--AC-CA-C--AC-A--TA-AG-A---AGA--T.........  339
SIVSMMPBJ  -T--A--A--C--T--G--C--T--A--AC-CA-C--AC-A--TA-AA-A---AGG--T.........  339

CONSENSUS  ????AataaaActga?acaga?a?????qg????ac????aca?caac?????????????????????  341
HIV2ROD    ....-GCAGC-CA.....................GAGAGCAGCACAGGGAACAAC......ACAACC  366
HIV2NIHZ   ....-ACATG-CC.....................................................  354
HIV2ISY    ....-ACGCA-GT.....................ACAGAGAGC.....................  354
HIV2ST     ....-ACAGC-CA...ACTGCAAAA.....AACACAACCTCC.....................  366
HIV2BEN    AGGG-ATACC-CG...ACCCCGAATCCC.....AGGACCTCGAGT.....................  387
HIV2D194   ....-ATATA-CT.....................................................  348
HIV2GH1    ....-ACAGC-CC.....................ACTAACAACACAACC.....................  360
HIV2D205   ....-GTAAA-CCGAAACAAAC...CCAGGG...AATGCCAGTAGTACT.....................  378
SIVMM251   ....-ATAAA-GTGAGACAGATAGATGGGGATTGACAAA..TCATCAACA.....................  384
SIVMM142   ....-ATAAA-GTGAGACAGATAAATGGGGATTGACAAAATCATCAACA.............ACAACA  390
SIVMM239   ....-ATAAA-GTGAGACAGATAGATGGGGATTGACAAAATCAATAACA.............ACAACA  390
SIVMNE     ....-ACAAA-GTGAGACAGATAAATGGGGATTGACAAAATCATCAACA.............ACAACA  390
SIVSMMH4   ....-ATAAA-CTGAGACAGATAGGTGGGGTTTGACAC..AAACGCAGGG.............ACAACA  390
SIVSMMPBJ  ....-ATAAA-GTGAGACAGACAGATGGCGTTTAACAGGAACACCAGCACCAACAACAACACAAACAACA  405
```

```
CONSENSUS  ??????????caacaacaac?acaacaaca?ca????????????a?a??a?a?at?aTaaatGA?act?  379
HIV2ROD    TCAAAGAGCACAAGCACAACCACAACCACACCCACAGACCAG......GAGCAAGAGA-AAGT--GGATA  430
HIV2NIHZ   ..............ACATGGACAGGGAGGACAGACACCCAA...AATATCACAATTA-AAAT--CACTT  406
HIV2ISY    ............GCAGTTGCAACTACAAGCCCATCT.........GGACCTGATATGA-AAAT--TACTG  403
HIV2ST     .........ACACCAACAACCACCACAACAGCA............AACACAACA...A-AGGA--GAATT  412
HIV2BEN    ............TCCACAACCTCG...AGACCACCCACATCC...GCAGCCTCCATAA-AAAT--AACTT  439
HIV2D194   ......TCAGGGACTACCGCGACCCCGAGTCCACCA.........AACATTACAATAA-AGAT--AAATT  403
HIV2GH1    ............ACAACAGGGAGCACAACA..............GGGATGTCAGAGA-AAAT--AACTT  403
HIV2D205   ...ACCACCACTAAGCCTACTACCACCTCTCGT............GGGCTGAAAACGA-TAAC--AACAG  433
SIVMM251   ...ACAATAACAACAGCAGCACCAACATCAGCACCAGTATCAGAAAAAATAGACATGG-CAAT--GACTA  451
SIVMM142   GCA...TCAACAACAACAACAACAACAGCAAAATCAGTA...GAGACAACAGACATAG-CAAT--GACTA  454
SIVMM239   GCA...TCAACAACATCAACGACAGCATCAGCA............AAAGTACACATGG-CAAT--GACTA  445
SIVMNE     GCA......CCAACAGCAATACCAACAAAAGCA.........GAGGCAATAAAAGTGG-CAAT--GAATA  445
SIVSMMH4   ACA...ACAGCAATAACAACAACAGCAACACCAAGTGTA......GCAGAAAATGTTA-AAAT--AAGTA  451
SIVSMMPBJ  ACAACACAAGCATCAACAACACCAACATCACCAATA.........ACAGCAAAGGTTG-AAAT--CAGTG  466

CONSENSUS  ?tcc?tg?ata????a?gAcAa?TGcaCAGG?tTggaa?AaGAggaaatgaT??a?TGtaA?TTcaatAT  434
HIV2ROD    CTCCATGCGCACGCGCAG-C-AC--CT----AT-GGGAG-G--AGAAACGA-CAAT--CC-G--CAAT--  500
HIV2NIHZ   C.CCAT..GCGCGTGCAG-C-AC--CA----AT-AAAGG-G--AGAAATGA-CGAC--TC-G--TAGT--  473
HIV2ISY    ATCCATGCATACAATTGA-C-AT--CT----AC-GAGGG-G--AGACATGG-CGAG--TC-G--CAAT--  473
HIV2ST     CTTCATGCATACGCACAG-C-AC--CA----GT-GGGAG-A--AGAGATGG-CGAC--TC-G--CAAT--  482
HIV2BEN    CTAACTGCATAGAAAACA-C-CA--CG----AT-AGGGT-T--GGAGATGA-GCAA--TG-G--CAAT--  509
HIV2D194   CTACCTGTATAGGCGACA-C-AC--CA----AT-AGGGA-A--AGAGGTGG-TGAG--TG-G--CAAT--  473
HIV2GH1    CT...CCATCATACTCAG-C-AC--CA----AT-GGGGA-G--AGAGATAG-CAAT--TC-G--CTAT--  470
HIV2D205   ACCCATGCATAAAAAATG-C-GC--CA----AC-AGGAG-A--GGAAATAA-GCAA--TA-T--TAGT--  503
SIVMM251   GTTCTTGTATAGCTCAGA-T-AT--CA----CT-GGAAC-A--GCAAATGA-AAGC--TA-A--CACC--  521
SIVMM142   GTCCTTGTGTAGTTCATG-T-AT--CA----CT-GGAAC-A--GCCAATGA-AAGC--TA-A--CAAC--  524
SIVMM239   GTTCTTGTATAGCCCAGG-T-AT--CA----CT-GGAAC-A--GCAAATGA-AAGC--TA-A--CAAC--  515
SIVMNE     GTCCTTGTATAAATCATG-T-AT--CA----CT-GGAAC-A--GCCAATGA-AAGC--TA-A--CAAC--  515
SIVSMMH4   ATCCTTGCATAAAAAATA-T-GT--TG----CT-GGAAC-G--GCCCATGA-AGGT--TA-A--TAAC--  521
SIVSMMPBJ  ATCCTTGTATAAAAATTA-T-AT--TA----CT-GGAAC-G--GCCCATGG-AAGT--TA-A--TAAC--  536

CONSENSUS  GAcaGGgtTAaaaagaGAtaAgaaaAaa?agTAtAatGAaaCaTGGTAcTcaa?aGAT?TggttTGTgaa  501
HIV2ROD    --CA--AT--GAAAGA--TA-GAAA-AACAG--T-AT--AA-A-----C-CAAAA---G-GGTT---GAG  570
HIV2NIHZ   --CA--AT--GAGAGA--CA-GAGA-AACAG--T-CT--AG-A-----C-CAAAA---G-GGTA---GAC  543
HIV2ISY    --CA--AC--GAGTTA--TA-GAAA-AACAG--T-GT--AA-C-----C-CAAAA---G-GGTT---GAA  543
HIV2ST     --CA--AT--GAGAGG--TA-GAAA-AACTA--T-AT--AA-A-----C-CAAAA---G-AGTC---GAA  552
HIV2BEN    --AG--GT--GAACAA--TA-GAAA-GGAGG--T-AG--CA-A-----T-TAGAA---G-GGTT---GAC  579
HIV2D194   --CG--GC--GAACAA--TA-GAAA-GGAAG--T-AT--CG-A-----C-CAAGA---G-GGTT---GAC  543
HIV2GH1    --CA--AT--GAGAGA--TA-GAAA-AACAG--T-AT--AA-A-----C-CAAAA---G-AGTT---GAA  540
HIV2D205   --CG--AC--AGAAGA--TG-GCTA-AACAA--T-AA--CA-C-----C-CAGAA---T-AGAG---...  570
SIVMM251   --CA--GT--AAAAGA--CA-GACA-AGGAG--C-AT--AA-T-----C-CTACA---T-GGTT---GAA  591
SIVMM142   --CA--GT--AAAAGA--CA-GAAA-AGGAG--C-AT--AA-T-----C-CTGCA---T-GGTT---GAA  594
SIVMM239   --CA--GT--AAAAGA--CA-GAAA-AAGAG--C-AT--AA-T-----C-CTGCA---T-GGTA---GAA  585
SIVMNE     --CA--GT--AAAAGA--CA-GAGA-GAGAG--C-AT--AA-T-----C-CTGCA---T-GGTT---GAA  585
SIVSMMH4   --CA--GT--AATAGG--CA-AAAG-AAGAA--T-AT--AA-A-----T-CAAGA---T-AATC---GAG  591
SIVSMMPBJ  --CA--GT--AAAAGA--CA-AAAG-GAGAA--T-AT--AA-A-----T-CAAGA---T-AGTT---GAA  606
```

```
CONSENSUS  ?????????AAta?cAc???????????ga?ag?agaTGtTAcATgaa?catTGcAAcACatcagT?ATcc  547
HIV2ROD    ...ACAAAT--TAGC-CAAATCAGACCCAG.....--T--C--GAACCAT--C--C--ATCAG-C--CA  631
HIV2NIHZ   .........--TAAC-CC...TCAAGTCAGAGCAAG--T--C--GAACCAT--C--C--ATCAG-C--CA  601
HIV2ISY    ...TCAGAT--CAGC-CAGACCGA......AAAAGA--T--C--GAACCAT--C--C--ATCAG-C--CA  604
HIV2ST     ...TCA...--TGAC-CCAAGAAA...GAGAAAACA--T--C--GAACCAC--C--C--ATCAG-C--CA  613
HIV2BEN    .........--CACA-CA.........GCTGGCACA--T--C--GAGACAT--C--C--ATCAA-C--CA  631
HIV2D194   ...AAGACA--CGGA-CA...........GGCACA--T--C--GAGACAT--C--C--ATCAG-C--CA  598
HIV2GH1    ...TCAAAT--TACC-AAGATGGGAAA...AACAGA--T--C--GAACCAT--C--C--ATCAG-C--CA  604
HIV2D205   .........--TAAT-CCAGGAAGTATACCAGCAGA--C--T--AAGAACC--C--C--AACAA-T--CC  631
SIVMM251   CAA...GGG--TAGC-CT...GATAATGAAAGCAGA--C--C--GAATCAC--T--C--TTCTG-T--CC  655
SIVMM142   CAA...GGG--TAGC-CT...GGTAATGAAAGTA A--T--C--GAATCAC--T--T--TTCTG-T--CC  658
SIVMM239   CAA...GGG--TAAC-CT...GGTAATGAAAGTAGA--T--C--GAACCAC--T--C--TTCTG-T--CC  649
SIVMNE     CAA...GGT--TAGC-CT...GAAAATGAAAGTAGA--T--C--GAATCAC--T--C--TTCTG-T--TC  649
SIVSMMH4   CAGTCAGCG--TGAA-GT.........GAGAGTAAA--T--C--GCATCAT--T--C--CAGTG-T--TC  652
SIVSMMPBJ  CAAAATAGC--TGAA-CT.........GACAGTAAA--C--T--GAACCAT--T--C--CAGTG-T--TC  667

CONSENSUS  aAGAgTcaTGTGAcAAgcAtTATTGGCAtgctaT?AggTtTAGaTAtTGTGCaCCaCCgGGtTaTgc  T  616
HIV2ROD    C---A-CA-----C--GC-C--------TGCTA-A-GG-T---A--C-----A--A--G--T-A-GCCC-  701
HIV2NIHZ   C---G-CA-----T--GC-C- ------TGCTA-G-GG-T---A--C-----A--A--G--T-T-GCCT-  671
HIV2ISY    C---G-CA-----C--GC-C--------TGCTA-G-GA-T---A--C-----A--A--G--T-T-GTCT-  674
HIV2ST     C---G-CA-----C--GC-C--------TACTA-G-GG-T---A--T-----A--A--G--T-T-GCCC-  683
HIV2BEN    A---G-CA-----T--GC-C--------TGCTA-G-GG-T---A--C-----A--A--G--C-T-GCCC-  701
HIV2D194   A---G-CA-----C--GC-C--------TGCTA-G-AG-T---A--C-----A--A--G--T-T-GCCC-  668
HIV2GH1    C---G-CA-----T--GC-C--------TGCTA-A-AG-T---A--C-----A--A--G--T-A-GCCC-  674
HIV2D205   A---G-CA-----C--AC-T--------CAGCT-A-GG-T---G--T-----T--C--G--G-T-TTTC-  701
SIVMM251   A---A-CT-----C--AC-T--------TACTA-T-GA-T---G--T-----A--T--A--T-A-GCTT-  725
SIVMM142   A---G-GT-----C--AG-T--------TGCTA-T-GA-G---A--T-----A--T--A--T-A-GCTT-  728
SIVMM239   A---G-CT-----C--AC-T--------TGCTA-T-GA-T---G--T-----A--T--A--T-A-GCTT-  719
SIVMNE     A---A-CT-----C--AC-T--------TGCTA-T-GA-T---G--T-----A--T--A--T-A-GCTT-  719
SIVSMMH4   A---A-CC-----C--GC-T--------TGCTA-T-GA-T---A--C-----A--G--A--T-A-GCTT-  722
SIVSMMPBJ  A---A-CC-----C--AC-T--------TGCTA-T-GA-T---A--T-----A--G--A--T-A-GCTT-  737

CONSENSUS  gcTaAGaTGtAATGAtaCcAAtTAttcAGGCTTtg?gCCtAA?TGttaTAAgGTaGTaGatttCttCaTGc  684
HIV2ROD    AT-A--A--T-----TA-C--T--TTC------TGCA--C--C--TTC---A--A--A-CTT-TA-A--C  771
HIV2NIHZ   AC-A--A--C-----TA-C--T--TTC------TGCA--T--C--CTC---A--A--A-CTG-TA-A--C  741
HIV2ISY    GC-A--G--C-----TA-C--T--CTC------TGAG--C--T--CTC---A--A--A-CTT-TA-A--T  744
HIV2ST     GC-A--A--C-----TA-C--T--TTC------TGAG--C--T--TTC---G--A--A-CTG-TA-A--T  753
HIV2BEN    AT-A--A--T-----TA-C--C--TTC------TGAA--T--G--CAC---A--A--A-CTG-TT-A--C  771
HIV2D194   AC-A--A--C-----TA-C--C--TTC------TGAA--T--G--CTC---A--A--A-CTG-TT-A--C  738
HIV2GH1    AC-A--A--C-----TA-C--T--TTC------CGAG--A--G--TTC---A--A--A-CTT-TA-A--C  744
HIV2D205   AC-A--A--T-----TA-C--C--TTC------CATG--C--C--CAG---G--A--A-CGT-CT-C--C  771
SIVMM251   GC-T--A--T-----CA-A--T--TTC------TATG--T--A--TTC---G--G--G-TCT-TT-A--C  795
SIVMM142   GC-T--A--T-----CA-A--T--TTC------TATG--T--C--TTC---G--A--G-TCT-TT-A--C  798
SIVMM239   GC-T--A--T-----CA-A--T--TTC------TATG--T--A--TTC---G--G--G-TCT-TT-A--C  789
SIVMNE     GC-T--A--T-----CA-A--T--TTC------TATG--T--C--TTC---G--G--G-TCT-TT-A--C  789
SIVSMMH4   GC-T--G--T-----TT-A--T--TTT------TGCT--T--C--TTC---G--A--G-TTT-TT-A--C  792
SIVSMMPBJ  GC-T--G--T-----TT-A--T--TTC------TGCT--T--C--TAC---G--A--A-TTT-TT-A--C  807
```

HIV2/SIV ENV

```
CONSENSUS  ACaAGgATGATGGAaACgCAaaCtTCtACtTGGtTTgGcTTtAATGG?ACtAGaGCAGAaAAtAGaACaT   753
HIV2ROD    --C--G--------A--G--AA-T--C--A---T--G-C--T-----C--T--A-----G--T--A--A-   841
HIV2NIHZ   --C--A--------A--G--AA-T--T--A---T--G-C--T-----C--T--A-----A--T--A--A-   811
HIV2ISY    --A--A--------A--G--AC-T--T--T---C--G-C--T-----C--T--G-----A--T--A--A-   814
HIV2ST     --A--G--------A--G--AA-C--C--T---T--G-C--T-----C--T--G-----A--T--A--A-   823
HIV2BEN    --A--G--------A--G--AA-T--T--T---T--G-C--T-----C--T--A-----A--T--A--A-   841
HIV2D194   --A--G--------A--G--AA-T--T--T---T--G-C--T-----C--T--A-----A--T--A--A-   808
HIV2GH1    --A--G--------A--G--AA-C--C--T---T--G-C--T-----C--T--G-----G--T--A--A-   814
HIV2D205   --A--A--------A--A--GT-C--T--A---T--G-C--C-----T--A--G-----G--C--G--A-   841
SIVMM251   --A--G--------G--A--GA-T--T--T---T--G-C--T-----A--T--A-----A--T--A--T-   865
SIVMM142   --A--G--------G--A--GA-T--T--T---T--C-G--T-----A--T--A-----A--T--A--C-   868
SIVMM239   --A--G--------G--A--GA-T--T--T---T--G-C--T-----A--T--A-----A--T--A--T-   859
SIVMNE     --A--A--------G--A--GA-T--T--T---T--G-C--T-----A--T--A-----A--T--A--T-   859
SIVSMMH4   --A--A--------G--G--AA-C--T--T---T--G-C--C-----T--T--G-----A--T--A--A-   862
SIVSMMPBJ  --A--A--------A--A--AA-C--T--T---T--G-T--C-----T--T--A-----A--T--A--A-   877

CONSENSUS  AtATcTAtTGGCATgg?AgagatAAtAGgACtAT?ATtAGctTaAAtAagtatTAtAAtcT?acaaTgca   820
HIV2ROD    -T--C--T------GGC-GAGAT--T--A--T--C--C--CT-A--C-AATAT--T--TC-CAGTT-GCA   911
HIV2NIHZ   -T--C--T------GGT-AAGAT--C--A--T--T--C--CT-A--T-ACTTT--T--TC-CACTA-GCA   881
HIV2ISY    -T--C--T------GGT-GGGAT--C--A--T--T--C--CT-A--C-AATAT--T--TC-CACCA-ACT   884
HIV2ST     -T--C--T------GGT-GGGAT--T--A--T--C--T--CT-A--C-AGTTT--T--TC-CACCG-ACA   893
HIV2BEN    -T--C--T------GGC-GAGAT--T--G--T--C--T--CT-A--C-AGTAT--T--TC-CACAA-GCG   911
HIV2D194   -T--C--T------GGT-AGGAT--T--G--T--C--T--CT-A--C-AGTAT--T--TC-CACAA-GCA   878
HIV2GH1    -T--C--T------GGC-GAGAT--T--G--T--C--C--CT-A--C-AGTAT--T--CC-TTCTA-ACA   884
HIV2D205   -T--A--T------GAA-AAGAC--T--G--C--C--A--CT-A--T-CATAC--T--TT-GTCAA-ACA   911
SIVMM251   -T--T--C------GGT-GGGAT--T--G--T--A--T--TT-A--T-AGTAT--T--TC-AACAA-GAA   935
SIVMM142   -T--T--C------GGT-GAGAT--T--G--T--A--T--TC-A--T-AGCAT--T--TC-AACAA-GAA   938
SIVMM239   -T--T--C------GGT-GGGAT--T--G--T--A--T--TT-A--T-AGTAT--T--TC-AACAA-GAA   929
SIVMNE     -T--T--C------AGC-AAGAT--T--G--T--A--T--TT-G--T-AGTAT--T--TC-AACAA-GAA   929
SIVSMMH4   -C--T--T------GGC-AAAGT--T--A--C--A--T--CT-G--T-AGTAT--T--TC-AACAA-GAG   932
SIVSMMPBJ  -C--T--T------GGC-GAAGC--T--A--C--A--T--CT-A--T-AGTAT--T--TC-AACAA-GAG   947

CONSENSUS  ?TGTAggAGaCCaGgaAAtAAgAcaGTtttACcAaTaAc?attaTGTCaGg?tt?gt?TTcCAcTC????   881
HIV2ROD    T----AG--G--A-GG--T--G-TA--GAA--A-A-A-TGCTTA----A-GACATGTG--T--C--CCAC   981
HIV2NIHZ   T----AG--G--G-GA--T--G-CA--GTT--C-A-A-CATTTA----A-GGTTTAAG--T--C--C...   948
HIV2ISY    T----GG--A--A-AA--T--A-CA--TGT--C-A-A-CACTCA----A-GCCGCAGA--T--C--C...   951
HIV2ST     T----AG--G--A-GA--C--G-CA--TGT--C-A-A-CACTCA----A-GGTTAGTG--T--C--C...   960
HIV2BEN    T----AG--A--A-GA--T--G-CA--TTT--C-A-A-CACTTA----A-GATTAGTG--T--C--T...   978
HIV2D194   T----AG--A--A-GA--T--G-CA--TGT--C-A-A-CACTTA----A-GGCGAAGG--T--C--T...   945
HIV2GH1    T----AG--A--A-GA--C--G-CA--TGT--C-A-A-CACTCA----A-GGTTAGTG--T--C--C...   951
HIV2D205   C----AG--G--A-GA--C--G-CG--TGT--C-A-A-GAACCG----A-GACTACTT--C--T--A...   978
SIVMM251   A----GA--A--A-GA--T--G-CA--TTT--C-G-C-CCATTA----T-GATTGGTT--C--C--A...  1002
SIVMM142   A----GA--A--A-GA--T--G-CA--TTT--C-G-C-CCATTA----T-CATTGGTT--C--C--A...  1005
SIVMM239   A----GA--A--A-GA--T--G-CA--TTT--C-G-C-CCATTA----T-GATTGGTT--C--C--A...   996
SIVMNE     A----GA--A--A-GA--T--G-CA--TTT--C-G-C-CCATCA----T-GATTGGTT--C--C--A...   996
SIVSMMH4   A----GA--A--A-AA--T--G-CA--TTT--C-G-C-CCATTA----A-GGTTGGTC--C--T--G...   999
SIVSMMPBJ  A----GA--A--A-GA--T--G-CA--CTT--C-G-C-CCATTA----A-GGTTGGTC--C--T--G...  1014
```

```
CONSENSUS  ???CagccaaTc???AAtaa?AgaCC?aggCAgGCaTGGTGctGGTTtgaaGGaaa?TGGAaggaaGCcA   942
HIV2ROD    TAC-AGCCGA-C...--TAAA-GA--CAGA--A--A-----CT----CAAA--CAAA----AAGAC--C-  1048
HIV2NIHZ   ...-AGCCAG-CATC--TAAA-AA--CAGG--A--A-----TT----CGAA--CCAA----AGGAA--C-  1015
HIV2ISY    ...-AGAAGA-CATC--TAAA-AA--CAGG--A--A-----CC----CAAA--CGAG----GGGAA--C-  1018
HIV2ST     ...-AGCCAA-C...--TAGA-GA--CAGG--A--A-----CT----CAAA--CGAG----AGGAA--C-  1024
HIV2BEN    ...-AGCCAA-C...--CACA-GG--TAGG--G--A-----CC----TGGA--CAGA----GGGAA--C-  1042
HIV2D194   ...-GGCCAG-CTAC--CAAA-AA--TGGG--G--A-----TT----TCAA--CAAC----TAGAA--C-  1012
HIV2GH1    ...-AGCCAA-C...--TACA-GA--TAGG--A--A-----CT----CAAA--CAAG----GGGAA--C-  1015
HIV2D205   ...-AGCCTA-C...--TAAG-GA--CAGA--A--T-----CT----[illegible]AG--AAAC----CAGAA--C-  1042
SIVMM251   ...-AACCAA-C...--TGAT-GG--AAAG--G--A-----TT----TGGA--AAAA----AGGAT--A-  1066
SIVMM142   ...-AACCAG-C...--TGAG-GG--AAAG--G--A-----TA----TGGA--AAAT----AGGAG--A-  1069
SIVMM239   ...-AACCAA-C...--TGAT-GG--AAAG--G--A-----TT----TGGA--AAAA----AGGAT--A-  1060
SIVMNE     ...-AACCAA-C...--TGAT-GG--AAAA--G--A-----TA----TGAA--AAAT----AGGAG--A-  1060
SIVSMMH4   ...-AGCCCA-A...--TGAG-GA--AAAA--G--C-----CT----TGAA--AAGC----AAAAG--C-  1063
SIVSMMPBJ  ...-AACCCA-A...--TGAG-GA--AAAA--G--C-----CT----TGGA--AGAA----AAAAG--C-  1078

CONSENSUS  T?cagGAgGTgAAgcaaACC?Ttgt?aAACATCCCAGgTAta?aGGa??????Ac?Aatga?Acaga?aa   999
HIV2ROD    -GCAG--G--G--GGAA---C-TGCAA----------G--TAGA--A......-CC-ATGAC-CAAGGAA  1112
HIV2NIHZ   -GCAG--G--G--GGAG---C-TGCGA----------A--TAAA--G...AAC-GG-GCCGC-CAGAGAA  1082
HIV2ISY    -GCAG--G--G--ACAA---C-TGTAA----------G--TAAA--A......-CC-ATGAC-CAAATAA  1082
HIV2ST     -GAAG--G--G--GCTA---C-TGCAA----------G--TAAA--A......-CC-ACGAC-CAGAAAA  1088
HIV2BEN    -GCAG--G--G--GCAA---C-TGTAC----------A--CAAA--A......-TC-ATGAT-CAGGGAA  1106
HIV2D194   -GCGG--G--G--GCAA---C-TGCAA----------G--CGGA--A......-CA-ATGAT-CAGGAAA  1076
HIV2GH1    -GCAG--G--A--GCAA---C-TATAA----------G--TAAA--A......-CC-ACGAC-CAAAGAA  1079
HIV2D205   -AAAG--G--G--AAGG---A-CATAA----------G--TAAA--AGGTGCA-AA-ATATC-CAAGCGT  1112
SIVMM251   -AAAA--G--G--ACAG---A-TGTCA----------G--TACT--A......-CT-ACAAT-CTGATAA  1130
SIVMM142   -AAAA--G--G--GCAG---A-TGTCA----------G--TACT--A......-CT-ACAAT-GTGATAA  1133
SIVMM239   -AAAA--G--G--GCAG---A-TGTCA----------G--TACT--A......-CT-ACAAT-CTGATAA  1124
SIVMNE     -AAAA--G--A--GCAG---A-TGTCA----------G--TACT--A......-CT-ACAAT-CTGATAA  1124
SIVSMMH4   -CCAG--A--G--GGAA---T-GGTCA----------G--TACG--A......-CT-ATGAT-CTAGGAA  1127
SIVSMMPBJ  -CCAG--A--G--GGAA---T-GGTCA----------G--TACG--A......-CT-ATAAG-CTGAACA  1142

CONSENSUS  aAttaa?tT?aca-jc?CcaggaggAgg?tcaGAtCCaGAaGT?aCaTtcATGTGGAC?AA?TGcaGaGGA  1062
HIV2ROD    T-TTAGCT-TGCAGCG-CAGGAAA-GGCTCA--C--A--A--AG-A-AC--------T--C--CA-A---  1182
HIV2NIHZ   T-TTAAAT-TAAAGCA-CAGGGAG-GGCTCA--C--A--A--AA-A-AC--------T--C--CA-A---  1152
HIV2ISY    A-TTAACT-TACAGCA-CAGAAAA-GACTCA--C--A--A--AG-A-AT--------T--C--CA-A---  1152
HIV2ST     A-TTCGTT-TATAGCG-CAGGAGA-CGCTCA--C--A--A--GG-A-AC--------T--C--CA-A---  1158
HIV2BEN    A-TTAACT-TACGAAA-CGGGAGC-GGCTCA--C--G--A--GG-A-TT--------T--C--CA-A---  1176
HIV2D194   A-TTAACT-TACGAAG-CAGGAAT-GGTTCA--C--A--A--GA-A-AC--------T--C--CA-A---  1146
HIV2GH1    T-TTAACT-TACAAAA-CAGGAAG-GGCTCA--C--A--A--GG-A-AC--------T--C--CA-A---  1149
HIV2D205   A-AGTTAG-ATCAGAA-ATGGAAA-GGTTCA--T-                                     1147
SIVMM251   A-TCAATT-AACGGCT-CTGGAGG-GGA...--T--G--A--TA-C-TC--------A--T--CA-A---  1197
SIVMM142   A-TCAATT-GACGGCT-CTAGAGG-GGA...--T--G--A--TA-C-TC--------A--T--CA-A---  1200
SIVMM239   A-TCAATT-GACGGCT-CTGGAGG-GGA...--T--G--A--TA-C-TC--------A--T--CA-A---  1191
SIVMNE     A-TCAATT-GACGGC[illegible]-CTGGAGG-GGA...--T--G--A--TA-C-TC--------A--T--CA-A---  1191
SIVSMMH4   A-TTAATC-AACAGCT-CAGCAGG-GGA...--T--A--A--CA-T-TT--------A--T--TC-A---  1194
SIVSMMPBJ  A-TTAAGC-AACAGCT-CAGGAGG-GGA...--T--A--A--TA-T-TC--------A--T--TC-A---  1209

CONSENSUS  GAaTttcTcTAcTGcAAaATGA?TTGGTTccT?AATTGGgTAGAaaA?aqg??????????????????  1110
HIV2ROD    --G-TTC-C--C--C--C----C------CC-C------A----GA-TAAG.....................  1233
HIV2NIHZ   --A-CTC-C--C--C--C----C------CC-C------G----GA-CAGG.....................  1203
HIV2ISY    --A-TCC-C--T--C--C----C------CC-T------G----AA-C.......................  1200
HIV2ST     --A-TTC-C--C--C--T----C------CC-C------G----AA-CAGA.....................  1209
HIV2BEN    --A-TTC-C--C--T--C----C------CC-C------G----AG-C.......................  1224
HIV2D194   --A-TTC-C--C--T--T----C------CC-C------G----AA-TAAG.....................  1197
HIV2GH1    --A-TTC-C--C--C--C----C------CC-C------G----AA-CAGA.....................  1200
SIVMM251   --G-TCC-C--C--T--A----A------TC-A------G----GG-TAGGGAT.................G  1252
SIVMM142   --G-TTC-C--C--T--A----A------TC-A------G----AG-TAGGAGT.................C  1255
SIVMM239   --G-TCC-C--C--T--A----A------TC-A------G----AG-TAGGAAT.................A  1246
SIVMNE     --G-TTC-C--C--T--A----A------TC-A------G----AG-TAAGAATCTG.........ACTG  1252
SIVSMMH4   --A-TCT-A--T--C--A----A------TC-T------G---GG-CAGAGACCAAAAGGGTGGCAGAT  1264
SIVSMMPBJ  --A-TCT-A--T--C--A----A------TC-T------G---AA-TATA...CAAAATGGTTCTAGAT  1276
```

```
CONSENSUS  ????????????????aagga?cagcaa?agag?AAtTAtgtgccaTGcCatATaag?CAaaTAATtAA  1159
HIV2ROD    ........................ACACACCGC--T--TGCACCG--C-AT--AAAG--AA----T--  1277
HIV2NIHZ   ................ACGGGTCAGAAACAGCGC--?--TGCACCG--C-GT--AAGG--AA----T--  1256
HIV2ISY    ................AAGACGGGTCAACAGCAT--C--TGTGCCG--C-AT--AGAG--AA----T--  1253
HIV2ST     ................ACGAATCAGACACAGCAC--T--TGTGCCA--C-AT--AAAG--AA----T--  1262
HIV2BEN    ................AAGAACCAAACACGGCGC--C--T......--C-AT--AAAG--GA----T--  1271
HIV2D194   ................ACGAACCAAACACACCGC--C--TGCGCCA--C-AT--AAGG--GA----T--  1250
HIV2GH1    ................CCGAATCAGACACAGCAC--T--TGCGCCG--C-AT--AAGG--AA----T--  1253
SIVMM251   TAACTACCCAGAGGCCAAAGGAACGGCATAGAAGG--T--CGTGCCG--T-AT--TAGA--AA----C--  1322
SIVMM142   TAACTACCCAGAAGCCAAAGGAACGGCATAAAAGG--T--CGTACCA--T-AT--TAGA--AA----C--  1325
SIVMM239   CAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGG--T--CGTGCCA--T-AT--TAGA--AA----C--  1316
SIVMNE     GAACTACCCAGAAGCCACAGGAACGGCATAAAAGG--T--CGTGCCA--T-AT--TAGA--AA----C--  1322
SIVSMMH4   GGAAACAACAAAATAGGAAAGAGCAACAGAAGAAA--T--TGTGCCA--T-AT--TAGA--AA----C--  1334
SIVSMMPBJ  GGACAAGTCAAAACCAGAAAGAGCGACAAAGGAGA--T--TGTGCCA--T-AT--TAGA--GA----C--  1346

CONSENSUS  cAC?TGGCAtAaaGTAGGgAaAAATgTaTATTTGCCTCCtAGggAaGGagA?cTgaCcTG?aAcTCcACa  1226
HIV2ROD    C--A-----T-AG-----G-G----G-A----------C--GG-A--GG-GC-GT-C--CA-C--A--A  1347
HIV2NIHZ   T--C-----C-GG-----G-A----C-A----------C--GG-A--GG-GT-GA-C--CA-C--A--G  1326
HIV2ISY    T--C-----T-AG-----G-A----G-A----------T--GG-A--AG-GT-GT-C--CG-A--A--A  1323
HIV2ST     T--C-----C-AG-----G-A----G-A----------T--GG-A--AC-GT-AA-C--CA-C--T--A  1332
HIV2BEN    T--C-----T-AA-----G-A----G-A----------T--GG-A--GG-GT-GG-C--TG-A--A--A  1341
HIV2D194   C--C-----T-AG-----G-C----G-A----------T--GG-A--GG-GT-GA-C--CA-T--A--A  1320
HIV2GH1    T--C-----C-AG-----G-A----G-A----------T--AG-A--AC-GT-GA-C--CA-C--A--A  1323
SIVMM251   C--T-----T-AA-----C-A----G-T----------A--AG-G--AG-CC-CA-G--TA-C--C--A  1392
SIVMM142   C--T-----T-AA-----C-A----G-T----------A--AG-G--AG-CC-CA-G--TA-C--C--A  1395
SIVMM239   C--T-----T-AA-----C-A----G-T----------A--AG-G--AG-CC-CA-G--TA-C--C--A  1386
SIVMNE     C--T-----T-AA-----C-G----G-T----------A--AG-G--AG-CC-CA-G--TA-T--C--A  1392
SIVSMMH4   C--G-----C-AA-----C-A----G-A----------T--GG-A--AG-CC-GA-A--CA-T--C--T  1404
SIVSMMPBJ  C--G-----C-AA-----C-A----G-G----------T--GG-A--AG-CC-GA-A--TA-T--C--T  1416

CONSENSUS  GTgACcAG?aT?AT?GCaaAcATtgAttggat?gatggaaAtcag??????AC?AatAT?ACctttAGTG  1284
HIV2ROD    --A--C--CA-A--T--TA-C--TG-CTGGCAAAACAATA-TCAG.......--A-AC--T--CTTT----  1411
HIV2NIHZ   --G--C--CA-A--T--CA-C--TG-T......GCGGGAG-TCAA.......--A-AT--T--CTTT----  1384
HIV2ISY    --G--C--TA-C--T--TA-C--TG-T...GTTGATGGAG-TAACCGG....--A-AT--T--CTTT----  1387
HIV2ST     --G--C--CA-A--T--TA-C--TG-C...GGAGGAGAGA-CCAG.......--A-AT--T--CTTT----  1393
HIV2BEN    --A--C--CA-A--T--TA-C--TG-C...ATAGATAAAA-TCGGACTCAT--C-AC--T--CTTT----  1408
HIV2D194   --A--C--CA-A--T--TA-C--TG-C...TCAGATGGAA-TCAG.......--C-AC--T--CTTT----  1381
HIV2GH1    --G--C--CA-A--T--TA-C--TG-C...GTAAATAGTA-CCAG.......--A-AT--T--CTTT----  1384
SIVMM251   --G--C--TC-C--A--AA-C--AG-TTGGACTGATGGAA-CCAA.......--T-GT--C--CATG----  1456
SIVMM142   --G--C--TC-C--A--AA-C--AA-TTGGACTGATGGAA-CCAA.......--T-GT--C--CATG----  1459
SIVMM239   --G--C--TC-C--A--AA-C--AG-TTGGATTGATGGAA-CCAA.......--T-AT--C--CATG----  1450
SIVMNE     --G--C--TC-C--A--AA-C--AG-TTGGATTGATGGAA-CCAA.......--T-AT--C--CATG----  1456
SIVSMMH4   --A--T--TC-C--A--AG-G--AG-TTGGATCAATAGCA-TGAG.......--C-AT--C--CATG----  1468
SIVSMMPBJ  --A--T--CC-C--A--AG-A--AG-TTGGATCAATGGCA-TGAG.......--C-AT--C--CATG----  1480

CONSENSUS  CAGAqGtGGCAGAACTgTA?cGATTgGAatTGGGaGAtTAtAAATTaqTAGAaaTaAC?CCaATTqqcTT  1352
HIV2ROD    ----G-T---------A--CA----G--GT----A--T--T-----GG----AA-A--A--A---GGC--  1481
HIV2NIHZ   ----G-C---------A--CC----G--AT----G--T--C-----AG----AA-C--A--A---GGC--  1454
HIV2ISY    ----G-T---------A--CC----G--AT----G--T--T-----AG----AG-A--A--A---GGC--  1457
HIV2ST     ----G-T---------A--CC----A--AT----G--T--T-----GA----AG-A--A--A---GGC--  1463
HIV2BEN    ----A-T---------G--CC----A--AC----A--C--C-----AA----AA-A--A--A---GGC--  1478
HIV2D194   ----A-T---------G--CC----A--AT----G--C--C-----GA----AG-A--A--A---CCG--  1451
HIV2GH1    ----G-T---------A--CC----A--AT----A--T--T-----AA----AG-A--A--A---GGC--  1454
SIVMM251   ----G-T---------G--TC----G--GT----A--T--T-----AG----GA-C--T--G---GGC--  1526
SIVMM142   ----G-T---------G--TC----G--AT----A--T--T-----AG----AA-C--T--A---GGC--  1529
SIVMM239   ----G-T---------G--TC----G--AT----A--T--T-----AG----GA-C--T--A---GGC--  1520
SIVMNE     ----G-T---------G--TC----G--AT----A--T--T-----AG----GA-C--T--A---GGC--  1526
SIVSMMH4   ----G-T---------G--TC----G--GT----A--T--C-----AA----GA-T--T--A---GGC--  1538
SIVSMMPBJ  ----G-T---------G--TC----G--GT----A--T--C-----AG----GA-T--T--A---GGC--  1550
```

```
CONSENSUS  cGCcCCtACaaatgtgAaaAG?TAc?Cc?Ct???g?tccctc?AGaaatAaaAGAGG?GTctTtGTGCTA  1413
HIV2ROD    C--A--T--AAAAGAA-AA--A--CT-CT-T...GCTCACGGG--ACAT-CA-----T--GT-C------  1548
HIV2NIHZ   C--A--T--ATCAGTA-AG--A--CT-CT-T...GCTCACCAG--ACAT-CA-----T--GT-T------  1521
HIV2ISY    C--C--T--AGCAGAA-A --A--CT-CT-T...GCTCCAGGG--ACAT-AG-----T--GC-T------  1524
HIV2ST     T--A--T--ACCAGTA-AA--A--CT-CT-T...GCTCCAGTG--GAAT-AA-----T--AT-C------  1530
HIV2BEN    C--A--T--AGATCAG-GA--G--CT-CT-A...ACTCCAGTG--GAAC-AA-----T--GT-C------  1545
HIV2D194   C--A--T--AAAAGAG-AA--A--TT-CT-G...GCTCCAGTG--GAAC-AA-----T--GT-C------  1518
HIV2GH1    C--A--T--GAGAGAG-AA--A--TT-CT-G...GCTCCAGTG--GAAC-AA-----T--GT-T------  1521
SIVMM251   G--C--C--AGATGTG-AG--G--CA-TA-TGGTGGCACCTCA--AAAT-AA-----G--CT-T------  1596
SIVMM142   G--C--C--AAATGTG-AG--G--CA-TA-TGGTGGCACCTCA--AAAT-AA-----G--CT-T------  1599
SIVMM239   G--C--C--AGATGTG-AG--G--CA-TA-TGGTGGCACCTCA--AAAT-AA-----G--CT-T------  1590
SIVMNE     G--C--C--AAATGTG-AG--G--CA-TA-TGGTGGCACCTCA--AAAT-AA-----G--CT-T------  1596
SIVSMMH4   G--C--C--AAGTGTA-GA--G--CA-CA-AACTGGTGCCTCA--AAAT-AG-----G--CT-T------  1608
SIVSMMPBJ  C--C--C--AAGTGTA-AA--G--CA-CA-AACTGGTGCCTCA--AAAT-AA-----G--CT-T------  1620

CONSENSUS  GGGTTCtTgGGTTTTCTCgCgACaGCAGGttCTGCaATGgGcGcggcGTCctTGACgcTGtCgGCtCAGT  1483
HIV2ROD    ------T-G---------G-A--A-----TT----A---G-C-CGGC---CC----CG--T-G--T----  1618
HIV2NIHZ   ------C-A---------G-A--A-----TT----C---G-C-CGGC---CT----GC--T-A--T----  1591
HIV2ISY    ------C-A---------A-G--A-----TG----A---G-G-CGGC---TC----GC--T-G--T----  1594
HIV2ST     ------T-A---------A-G--A-----AG----A---G-C-CGGC---CT----GC--T-G--T----  1600
HIV2BEN    ------T-G---------G-G--A-----TT----A---G-C-CGCG---CC----GC--T-A--C----  1615
HIV2D194   ------T-G---------G-G--A-----TT----A---G-C-GCGC---CT----GC--T-G--T----  1588
HIV2GH1    ------T-G---------G-G--A-----TT----A---G-C-CGGC---CC----GC--T-G--T----  1591
SIVMM251   ------T-G---------G-A--G-----TT----A---G-C-CGGC---GT----GC--A-C--T----  1666
SIVMM142   ------T-G---------G-A--G-----TT----A---G-C-CGGC---GT----CG--A-C--T----  1669
SIVMM239   ------T-G---------G-A--G-----TT----A---G-C-CGGC---GT----GC--A-C--T----  1660
SIVMNE     ------T-A---------G-A--G-----TT----A---G-C-CGGC---GT----GC--A-C--T----  1666
SIVSMMH4   ------T-G---------G-G--A-----TT----A---G-C-CGGC---CG----GC--T-G--T----  1678
SIVSMMPBJ  ------T-G---------G-G--A-----TT----A---A-C-CGGC---CG----GC--T-G--T----  1690

CONSENSUS  CcCGgACTTTatTggctGGGATAGTGCAGCAACAGCAACAGCTGTTGGAcGTqGTCAAGAGACAACAAGA  1553
HIV2ROD    -C--G-----AC-GGCC----------------------------------C--G----------------  1688
HIV2NIHZ   -T--G-----AT-GGCC----------------------------------T--G----------------  1661
HIV2ISY    -T--G-----AT-CCGT----------------------------------C--G----------------  1664
HIV2ST     -T--G-----AT-GGCC----------------------------------C--G----------------  1670
HIV2BEN    -C--G-----AC-GGCC----------------------------------C--A----------------  1685
HIV2D194   -C--G-----AC-GGCC----------------------------------C--G----------------  1658
HIV2GH1    -C--G-----AC-GGCC----------------------------------C--A----------------  1661
SIVMM251   -C--G-----AT-GGCT----------------------------------C--G----------------  1736
SIVMM142   -C--G-----AT-GGCT----------------------------------C--G----------------  1739
SIVMM239   -C--A-----AT-GGCT----------------------------------C--G----------------  1730
SIVMNE     -C--G-----AT-GGCT----------------------------------C--G----------------  1736
SIVSMMH4   -C--G-----GT-GGCT----------------------------------T--G----------------  1748
SIVSMMPBJ  -C--G-----GT-GGCT----------------------------------T--G----------------  1760

CONSENSUS  AtTGTTGCGAcTGACCGTCTGGGGAaC?AA?AAcCTCCAG?C?AGaGTCaCTGCtATcGAGAAqTAC?Ta  1618
HIV2ROD    -C---------C---------------A-G--A--C------G-A--A---A----T--A-----G---C-A  1758
HIV2NIHZ   -A---------C---------------A-A--A--T------G-A--A---A----T--A-----G---C-A  1731
HIV2ISY    -A---------C---------------A-T--A--C------G-A--A---A----T--T-----G---C-A  1734
HIV2ST     -A---------C---------------A-A--A--T------G-A--A---A----T--C-----A---T-A  1740
HIV2BEN    -A---------C---------------A-G--A--C------G-A--A---A----T--C-----G---C-A  1755
HIV2D194   -A---------T---------------A-G--A--T------G-A--A---A----T--C-----A---T-A  1728
HIV2GH1    -A---------C---------------A-G--A--T------G-A--A---A----T--C-----G---T-A  1731
SIVMM251   -T---------C---------------A-A--G--C------A-T--G---A----C--C-----G---T-A  1806
SIVMM142   -T---------C---------------A-A--G--C------A-T--G---T----C--C-----G---T-A  1809
SIVMM239   -T---------C---------------A-A--G--C------A-T--G---A----C--C-----G---T-A  1800
SIVMNE     -T---------C---------------A-A--G--C------A-T--A---A----C--C-----G---T-A  1806
SIVSMMH4   -T---------C---------------A-T--G--C------A-T--A---A----T--C-----G---C-G  1818
SIVSMMPBJ  -T---------C---------------G-T--G--C------A-T--A---A----T--C-----G---C-A  1810
```

HIV2/SIV ENV

```
CONSENSUS  aaygAcCAGGCgCag?TaAATtCaTGGGGaTGTGCgTTTAGaCAaGTCTGcCAcACtaCTGT/CCatGG?  1687
HIV2ROD    CAGG-C-----G-GGC-A---T-A-----A-----G-----A--A-----C--C--TA-------AT--G  1828
HIV2NIHZ   AAGG-C-----G-AGC-A---T-A-----A-----G-----A--A-----C--C--TT-------AT--G  1801
HIV2ISY    GCAG-C-----G-GAC-A---T-A-----A-----G-----A--A-----C--C--TA-------AT--G  1804
HIV2ST     AAGG-C-----G-AAC-A---T-A-----A-----G-----A--A-----C--C--TA-------AT--G  1810
HIV2BEN    AAGC-T-----A-AGC-A---T-A-----A-----G-----A--G-----C--C--TA-------GT--G  1825
HIV2D194   AAGG-C-----A-AGC-A---T-A-----A-----G-----G--G-----C--C--TA-------AT--G  1798
HIV2GH1    AAGG-C-----A-AGT-A---T-A-----G-----G-----G--G-----C--C--TA-------AT--G  1801
SIVMM251   AAGG-C-----G-AGC-A---G-T-----A-----G-----A--A-----C--C--TA-------AT--C  1876
SIVMM142   AAGG-C-----G-AGC-A---G-T-----A-----G-----A--A-----T--C--TA-------AT--C  1879
SIVMM239   AAGG-C-----G-AGC-G---G-T-----A-----G-----A--A-----C--C--TA-------AT--C  1870
SIVMNE     AAGG-C-----G-AGC-A---G-T-----A-----A-----A--A-----C--T--TA-------AT--C  1876
SIVSMMH4   AAGG-T-----G-AGC-A---T-A-----A-----T-----G--A-----T--C--TA-------AT--C  1888
SIVSMMPBJ  AAGG-T-----G-AGC-A---T-A-----A-----T-----G--G-----C--C--TA-------AA--C  1900

CONSENSUS  ?aAATGa?acttTaacaCCtgA?TGGaAcAAtatgAC?TGGCA?GA?TGGGA??aa?a?gTccaCttccT  1747
HIV2ROD    TT----ATTCCT-AGCA--TG-C---G-C--TATG--G-----G--A-----AAAACAAG-CCG-TACC-  1898
HIV2NIHZ   TA----ATACCT-GACA--TG-T---A-C--TATG--G-----G--A-----ACAAAAAG-CCG-TACC-  1871
HIV2ISY    TA----ACACCT-AACA--TG-G---A-C--CATG--A-----A--A-----ACACAAAA-CCG-TTCC-  1874
HIV2ST     TA----ACACCT-AACG--TG-T---A-C--CATG--A-----G--A-----GCAACGAA-CCG-AACC-  1880
HIV2BEN    TA----ACTCTT-ATCG--TG-C---A-A--TATG--A-----G--G-----GAAACAAG-CCG-TACC-  1895
HIV2D194   TA----ACTCCT-AACA--TG-C---A-C--TATG--A-----G--A-----AAAACGAG-CCA-TACC-  1868
HIV2GH1    TA----ATTCTT-ATCG--TG-C---A-C--TATG--A-----G--A-----AAAACAAG-CCG-TACC-  1871
SIVMM251   CA----CAAGTC-AACA--AG-C---A-C--TGAT--T-----A--G-----GCGAAAGG-TGA-TTCT-  1946
SIVMM142   CA----CAAGTC-AACA--AG-T---A-C--TGAG--T-----A--G-----GCGGAAGG-TGA-TTCT-  1949
SIVMM239   CA----CAAGTC-AACA--AA-G---A-C--TGAG--T-----A--G-----GCGAAAGG-TGA-TTCT-  1940
SIVMNE     CA----CAAATC-AACA--AA-T---A-C--TGAG--T-----A--G-----GCGAAAGG-TGA-TTCT-  1946
SIVSMMH4   CA----AAACAT-GGTG--TA-T---A-C--TATG--T-----A--G-----AAGACAGG-TGA-TTCC-  1958
SIVSMMPBJ  CA----ACACAT-GACA--TA-C---A-C--TATG--T-----A--G-----AAAACAGG-GAA-TTCC-  1970

CONSENSUS  aGAgGcAAATAT?A?tcaaagttTaGAagAgGCaCAAaTtCAaCAaGAaAAgAAtAtGTATGAatTaCAA  1815
HIV2ROD    G--G-C------C-GTAAAAGTT-A--AC-G--A---A-T--G--A--G--A--T-T------AC-A---  1968
HIV2NIHZ   G--G-C------C-GTCAAAGTC-A--AC-A--A---A-T--G--A--A--G--T-T------GC-A---  1941
HIV2ISY    A--G-C------C-GTGAGAGTT-A--AC-G--A---A-C--G--A--A--G--T-T------GC-G---  1944
HIV2ST     A--G-C------C-GTGAAAGTT-A--AC-G--A---A-C--G--A--A--G--C-T------AC-A---  1950
HIV2BEN    A--G-C------C-GTCAAAGTT-A--AG-A--C---A-T--A--A--A--G--T-T------AT-A---  1965
HIV2D194   A--G-C------C-GTCAAAGTT-A--AC-G--A---A-T--A--A--A--G--T-T------AC-A---  1938
HIV2GH1    A--G-C------C-GTCAAAGTT-A--AC-G--A---A-T--A--A--A--G--T-T------AT-A---  1941
SIVMM251   G--G-A------A-CAGCCCTCC-A--AG-G--A---A-T--A--A--G--G--C-T------AT-A---  2016
SIVMM142   G--G-C------A-CGGCCCTCC-A--AG-G--A---A-T--A--A--G--G--C-T------AT-A---  2019
SIVMM239   G--A-A------A-CAGCCCTCC-A--GG-G--A---A-T--A--A--G--G--C-T------AT-A---  2010
SIVMNE     G--G-A------A-CGGCCCTTT-A--AG-G--A---A-T--A--A--A--G--C-T------AT-A---  2016
SIVSMMH4   A--G-C------A-CTCAATTAT-A--AG-A--A---A-T--G--A--A--G--T-T------AT-G---  2028
SIVSMMPBJ  A--G-C------A-CTCAATCAT-G--AG-A--A---A-T--G--A--A--G--T-C------AT-G---  2040

CONSENSUS  AAatTaAATAGCTGGGATgT?tTTgqCAAtTGGtTTGAc?TtaCcTC?TGGaT?AagTAtAT?CAaTAtG  1880
HIV2ROD    --AT-A-----------A-TT--GG---T---T----CT-AA-C--C---G-C-AG--T--T--A--T-  2038
HIV2NIHZ   --AT-A----------G-TT--AC---T---C----CT-CA-C--C---G-C-GG--T--T--A--T-  2011
HIV2ISY    --GC-A----------G-TT--GG---T---T----CT-AA-C--C---A-C-AG--T--T--A--T-  2014
HIV2ST     --AT-A----------G-TT--GG---C---T----TT-AA-C--C---A-C-AA--T--T--G--T-  2020
HIV2BEN    --AT-A----------A-TC--GG---C---T----CT-AA-C--C---G-C-AG--T--T--A--T-  2035
HIV2D194   --AC-A----------G-CT--GG---C---T----TT-GA-C--C---A-C-AA--T--T--A--T-  2008
HIV2GH1    --GC-A----------G-CT--GG---C---T----TT-AA-C--C---A-C-AG--T--T--A--T-  2011
SIVMM251   --GT-G----------G-GT--GG---T---T----CC-TG-T--T---A-A-AG--T--A--A--T-  2086
SIVMM142   --GT-G----------G-GT--GG---T---T----CC-TA-T--T---A-A-AG--T--A--A--T-  2089
SIVMM239   --GT-G----------G-GT--GG---T---T----CC-TG-T--T---A-A-AG--T--A--A--T-  2080
SIVMNE     --GT-G----------G-GT--GG---T---T----CC-TG-T--T---A-A-GG--T--A--A--C-  2086
SIVSMMH4   --AT-A----------A-CT--GG---T---T----CC-TA-T--T---A-A-GA--T--A--A--T-  2098
SIVSMMPBJ  --AT-A----------A-TT--GG---T---T----CC-TA-T--T---A-A-AA--T--A--A--T-  2110
```

```
CONSENSUS  GagTttataTAgT?gTaGgAgTaaTAgtttTaAGAaTAGtaATaTATaTAgTaCAaaTGtTA??TAgg?T  1946
HIV2ROD    -AG-GCTTA--A-AG-A-C-G-AA--GCTT-A---A---TG--A---G--G-A--AA--T--AG--GGC-  2108
HIV2NIHZ   -AG-TTATG--G-AG-A-G-A-AG--GCTT-A---A---TA--A---A--G-A--GA--T--AG--GAC-  2081
HIV2ISY    -AG-CATGA--G-AG-A-G-A-AG--GCTC-C---A---TA--A---G--G-A--AA--C--AG--GAC-  2084
HIV2ST     -AG-TTATA--G-AG-A-G-A-AA--GTTT-A---A---TA--A---G--G-A--AA--T--AG--GAC-  2090
HIV2BEN    -AG-GCATA--G-AG-G-G-A-AA--GCTT-A---A---CA--C---G--G-G--AT--T--AG--GAT-  2105
HIV2D194   -AG-TTATA--G-AG-A-G-A-AA--GGTT-A---A---CC--A---A--G-G--AT--T--AG--GAC-  2078
HIV2GH1    -AG-TTATA--G-AG-A-G-G-GA--GTTT-A---A---CA--A---A--G-G--AT--T--AG--GAC-  2081
SIVMM251   -AA-TTATG--G-TG-A-G-G-AA--CTGT-A---A---TG--C---A--G-A--AA--C--GC--AGT-  2156
SIVMM142   -AA-TTATA--A-TG-A-G-G-AA--CTGT-A---A---TG--C---A--G-A--AA--C--GC--GGT-  2159
SIVMM239   -AG-TTATA--G-TG-A-G-G-AA--CTGT-A---A---TG--C---A--G-A--AA--C--GC--AGT-  2150
SIVMNE     -AG-TTATA--G-TG-A-G-G-AA--CTGT-A---A---TG--C---A--G-A--AA--C--GC--AGT-  2156
SIVSMMH4   -TG-ACTTA--G-TC-A-G-G-AA--GGGT-A---A---TA--A---G--G-G--AA--T--GC--GGT-  2168
SIVSMMPBJ  -TG-ACTGA--G-TC-A-G-G-AA--GGAT-A---A---TG--A---G--G-G--GA--T--GC--GGT-  2180

                                                          <- stop in some SIVs
                                                        \/ \/3'sja       \/3'sj
CONSENSUS  tAGaaAgGGcTATaGGCCtGT?TTCTCtTCCCC?CCcgcTTat?tcCA?cAGAtcCaTAtCCacaaGGac  2012
HIV2ROD    T--AA-G--C---A----T--T-----T-----C--CGG--ATATC--AC---TC-A--T--ACAA--AC  2178
HIV2NIHZ   T--AA-G--C---A----T--T-----T-----C--CGG--ACATC--AC---TC-A--T--ACAA--AC  2151
HIV2ISY    T--AA-G--C---A----T--T-----T-----C--CGG--A*ATT--AC---TC-A--T--ACAA--AC  2154
HIV2ST     T--AA-G--C---A----T--T-----T-----C--CGC--ACTTC--AC---TC-A--T--ACAA--AC  2160
HIV2BEN    T--AA-G--C---A----T--T-----T-----C--CGG--ATCTC--AC---TC-A--T--ACAA--AC  2175
HIV2D194   T--AA-G--C---A----T--T-----C---  --C--CGG--ATCTC--AC---TC-A--T--ACAC--AC  2148
HIV2GH1    T--AA-G--C---A----T--T-----T-    ---T--CGG--ATCTC--AC---TC-A--T--ACAC--AC  2151
SIVMM251   A--GC-G--G---A----A--G-----T-----A--CTC--ATTTC--GT---CT-A--C--AACA--AC  2226
SIVMM142   A--AC-G--G---A----A--G-----T-----A--CTC--ATTTC--GT---CC-A--C--AACA--AT  2229
SIVMM239   A--GC-G--G---A----A--G-----T-    -A--CTC--ATTTC--GC---CC-A--T--AACA--AC  2220
SIVMNE     A--GC-A--G---A----A--G-----T-----A--TTC--ATTTC--GT---CC-A--T--GACA--AC  2226
SIVSMMH4   A--AC-G--T---A----A--G-----T-----T--CGC--ATGTT--GC---TC-C--T--ACAA--GC  2238
SIVSMMPBJ  A--AC-G--T---A----A--G-----T-----T--CGC--ATGTT--GC---TC-C--T--AGAC--GC  2250

                                                                  <- early stop
                                                                     in GH1
CONSENSUS  c?GG?aCaGCCA?cCAaaGAAGgaacAgAAgaAGacGgtgGagacagcggTGG?aaCagcTccTgGCCtT  2078
HIV2ROD    CG--GA-A----GC--AC----AAAC-G--GA--AC-GTG-AAGCAACGG---AGA-AGA-AC-G---C-  2248
HIV2NIHZ   CA--AA-A----GC--GA----AAAC-G--GA--AC-TTG-AAGCAACGG---AGA-AGA-CT-G---T-  2221
HIV2ISY    TG--AA-A----GA--GA----AAAC-G--GA--AC-TTG-GAACGACGT---AAG-AGA-CC-G---T-  2224
HIV2ST     CG--AA-A----GC--GA----AAAC-G--GA--AC-TTG-AAACAGCGT---AGA-AAT-GG-G---C-  2230
HIV2BEN    CG--GA-A----GC--AC----GAAC-G--GA--AC-TCG-AGGCGACAG---TTA-GAC-TG-G---T-  2245
HIV2D194   AG--GA-A----GC--AC----AAAC-G--GA--AC-CCG-AGACGACAG---TTT-GGC-TG-G---T-  2218
HIV2GH1    CG--GA-A----GC--AC----GAAC-G--GA--AC-ACA-AGACGACGA---TTA-GAC-TG***--T-  2221
SIVMM251   CC--CA-T----AC--GA----GCAA-G--GG--AC-GTG-AGAAGGCGG---CAA-AGC-CC-G---T-  2296
SIVMM142   CC--CT-T----AC--AA----GCAA-A--GG--AC-GTG-AGGCAGCGG---CAA-AGC-CC-G---T-  2299
SIVMM239   CC--CA-T-   --AC--GA----GCAA-G--AG--AC-GTG-AGAAGGCGG---CAA-AGC-CC-G---T-  2290
SIVMNE     CA--CA-T----AC--AA----GAAC-G--GG--AC-GTG-AGGCAGCGG---CAA-AGC-CC-G---T-  2296
SIVSMMH4   CA--AA-C----AC--AA----GAGA-G--GG--AC-GTG-AGACAGAGG---CAG-AGA-CT-G---T-  2308
SIVSMMPBJ  CA--AA-T----AC--AA----GAGA-G--GG--AC-GTG-AGGCAGAGG---CAA-AGA-CT-G---T-  2320
```

```
                <- tat
CONSENSUS  GGC?gaTAgaaTAtattCAtTTCCTGATCCgCCa?cTgataCGCCTCTTGAc?tGgcTaTaCA?cA?cTg   2143
HIV2ROD    ---CGA--GCA--TATA--T----------G--AGC-GATT----------CCA-AC-A-A--GC-TC-G   2318
HIV2NIHZ   ---CGA--GCA--TATA--T----------G--TGC-GATT----------CCG-GC-A-A--AC-TC-G   2291
HIV2ISY    ---CGA--GAA--TATA--T----------G--TGC-GATC----------CCA-AC-A-A--AC-GC-G   2294
HIV2ST     ---CGA--AGA--TATA--T----------G--AGC-GATT----------ACA-AC-A-A--AC-TC-G   2300
HIV2BEN    ---CAA--AAC--TGTG--G----------A--TAC-GACT----------TCG-GC-A-A--AC-TC-G   2315
HIV2D194   ---CAC--AAC--CATA--A----------A--TAC-GACT----------CCG-GC-A-A--AC-GC-G   2288
HIV2GH1    ---CGA--AAC--TATA--T----------A--TAC-GACT----------CCG-GC-A-A--AG-TC-G   2291
SIVMM251   ---AGA--GAA--TATT--T----------G--AAC-GATA----------CTT-GC-A-T--GC-AC-G   2366
SIVMM142   ---AGA--GAA--TATT--T----------G--AAC-GATA----------CTT-GC-A-T--GC-AC-G   2369
SIVMM239   ---AGA--GAA--TATT--T----------G--AAC-GATA----------CTT-GC-A-T--GC-AC-G   2360
SIVMNE     ---AGA--GAA--TATT--T----------G--AAC-AATA----------CTT-GT-A-T--GC-AC-G   2366
SIVSMMH4   ---AGA--GAA--TATT--T----------G--AGT-GATA----------CTT-GC-A-T--GC-GC-G   2378
SIVSMMPBJ  ---AGA--GAA--TATT--T----------G--AGT-GATA----------CTT-GC-A-T--GC-GT-G   2390

CONSENSUS  CAG?gacTtaCTATccAgga?ct?ccagaccCtCcaaCcaatcttCCAgag?cTctc???gac?ct?c?a   2203
HIV2ROD    ---GGAC-TA----CC-GGAGCTTCCTGACC-T-CAA-TCATCTA---GAATC-C.............   2373
HIV2NIHZ   ---GGAC-TA----CC-GGATCTCCCCGATC-T-CAA-CAATCTT---GAGTC-CCAGAGAGCACTAACA   2361
HIV2ISY    ---GGAC-TA----CC-GA............T-TAC-TGATCCT---ACCAC-C.............   2337
HIV2ST     ---GGAC-TA----CC-GGAGCTTCCAGACC-T-CAA-TAATCTC---GAGTC-TCGGAGAGCATTGACA   2370
HIV2BEN    ---AGAC-TA----CC-AGAACTCCCCGACC-G-CGA-TGATCTC---GAGTC-A............ACA   2373
HIV2D194   ---GGGC-TA----CC-AGAACTCCCCGACC-G-CGA-TGATCTC---GAGTC-A............ACA   2346
HIV2GH1    ---AGAC-TA----CC-CGAACTCCCCGACC-A-CGA-TGATCTC---GAATC-A............ACA   2349
SIVMM251   ---AACC-TG----CG-GAGCATACCAGATC-T-CAA-CAATACT---GAGGC-CTCTGCGACCCTACGA   2436
SIVMM142   ---AACC-TG----CG-GAGCATACCAGATC-T-CAA-CAATATT---GAGGC-CTCTGCGACCCTACGG   2439
SIVMM239   ---AACC-TG----CG-GAGTATACCAGATC-T-CAA-CAATACT---GAGGC-CTCTGCGACCCTACAG   2430
SIVMNE     ---AACC-TG----CG-GAGCATACCAGATC-T-CAA-CAATATT---GAGAT-CTCCACGACCCTACAG   2436
SIVSMMH4   ---GGAT-GG----TG-GGAGCTACCAGATC-T-CAA-CAGTGCT---GAGCC-CTCAACGACGTTGCAA   2448
SIVSMMPBJ  ---AGAT-GG----TG-GGAACTGCCAAACC-T-CAG-CAGTGCT---GAGCC-CTCAAGGACGCTGCAG   2460

                                   <- rev          nef ->
CONSENSUS  agaatcaGaGA????cT?AGacttaaa?cagCCtaccT?CAaTATGGGtGcgagTggaTCCAaGAaGCg?   2265
HIV2ROD    ......A-A--CTGGC-G--ACTTAGAACAG--TTCT-G--A------T-CGAG-GGA----A--A--AT   2437
HIV2NIHZ   GCAATCA-A--CTGGC-G--GCTTAAAGCAG--TACC-G--G------T-CGAG-GGA----A--A--GT   2431
HIV2ISY    ......A-A--CTGGC-G--ACTCAAGGCAG--TACC-G--G------T-CGAG-GGA----A--A--GT   2401
HIV2ST     GCAGTCA-A--CTGGC-G--ATTTAACACAG--TACC-G--A------G-CGAG-GGA----A--A--GT   2440
HIV2BEN    GCAATCA-G--CTGGC-G--ACTTAAGGCGG--CAAC-G--A------T-CGAG-GGA----A--A--TT   2443
HIV2D194   GCAATCA-G--CTGGC-G--ACTTAAGGCGG--TACC-G--A------T-CGAG-GGA----A--A--GT   2416
HIV2GH1    GCAATCA-G--CTGGC-G--ACTTAAAGCGG--TACC-G--A------G-CGAG-GGA----A--A--AT   2419
SIVMM251   AGGGTTC-A--AGTCC-C--GACTGAACTGA--TACC-A--A------T-GAGC-ATT----T--G--GG   2506
SIVMM142   AGAATTC-A--AGTCC-C--GCTTGAACTGA--TACC-A--A------T-GAGC-ATT----A--A--GG   2509
SIVMM239   AGGATTC-A--AGTCC-C--GACTGAACTGA--TACC-A--A------T-GAGC-ATT----T--G--GG   2500
SIVMNE     AGAATCC-A--AGTCC-C--GACTGAACTAA--TACC-A--A------T-GAGC-ACT----A--G--GG   2506
SIVSMMH4   AGAGTCC-T--AGTCA-C--AATTGAAATAG--TACC-A--A------T-GCGC-ATT----A--A--AG   2518
SIVSMMPBJ  AGAGCCC-T--AGTCA-C--AGTTCAGATAG--TACC-A--G------T-GCGT-ACC----A--A--AG   2530

CONSENSUS  tcCaaGcc?tcgcgAg??cTgcgagAGAgACTcTTgCgggCGCgtGgaGagactTaTGggag?ctctggg   2331
HIV2ROD    TC-AG-CCGCCGCG-GGGC-ACAAG---G---C--G-GGG---GT-CA-GGGCT-G--GAGGGTATTGGA   2507
HIV2NIHZ   TC-AA-CCCTTGCA-GGAC-ACAAG---G---C--G-AGG---GG-GA-AGACT-G--GAGGGCGCTGCA   2501
HIV2ISY    TC-AG-CCCTCGCG-GGGT-ACAAG---G---C--A-GAG---GG-GA-GAGCT-G--GGGGGCTCTGGG   2471
HIV2ST     TC-GA-CCTTCGCG-GGGC-ACGGG---G---C--A-AAA---CT-GA-AGGCT-C--GGGGACACTGGG   2510
HIV2BEN    TC-AA-CATTCGCG-GGAC-ACGAG---G---C--G-GGG---GT-GG-ATGGT-A--GGAAGCAGCGCG   2513
HIV2D194   TC-GA-CATTCGCA-GGAC-GCGAG---G---A--G-GGG---GT-GA-GGGGT-A--TGAAGCAGCGCA   2486
HIV2GH1    TC-AA-CATTCGCA-AGAC-ACGAG---G---C--G-GAG---AT-GG-GGGGT-A--TGCAGCAGTGCA   2489
SIVMM251   TC-AA-CCGGCTGG-GATC-GCGAC---A---C--G-GGG---GT-GA-AGACT-A--GGAGACTCTTAG   2576
SIVMM142   TC-AA-CCGCCCAG-GATC-GCGAC---G---C--G-GGG---GT-GG-AGAAT-A--GGAGGCTCTTCA   2579
SIVMM239   TC-AG-CCGTCTGG-GATC-GCGAC---G---C--G-GGG---GT-GG-AGACT-A--GGAGACTCTTAG   2570
SIVMNE     TC-AA-TCGCCTGG-GATC-GCGAC---G---C--G-GGG---GT-GG-AGACT-A--GGAGACTCTGGG   2576
SIVSMMH4   TA-AA-CGTGGTGG-AACT-GCGCG---G---C--G-AAG---GT-GG-AGACA-A--GGAGACTCTGGG   2588
SIVSMMPBJ  CG-AG-CGTAGTGG-AATT-GTACG---G---C--G-AAG---GT-GA-AGACT-A--GGAGACTCTGGG   2600
```

HIV2/SIV NEF

HIV-2/SIV NEF CDS

```
CONSENSUS  ATGGGtG ???TggaTCCAaGAaGCggtcCaaGcc??c??gAga?cTgcgagAGAgACTcTTgCgggCGC    62
HIV2ROD    -----T-CGAG-GGA----A--A--ATTC-AG-CCGCCGCG-GGGC-ACAAG---G---C--G-GGG---    70
HIV2NIHZ   -----T-CGAG-GGA----A--A--GTTC-AA-CCCTTGCA-GGAC-ACAAG---G---C--G-AGG---    70
HIV2ISY    -----T-CGAG-GGA----A--A--GTTC-AG-CCCTCGCG-GGGT-ACAAG---G---C--A-GAG---    70
HIV2ST     -----G-CGAG-GGA----A--A--GTTC-GA-CCTTCGCG-GGGC-ACGGG---G---C--A-AAA---    70
HIV2BEN    -----T-CGAG-GGA----A--A--TTTC-AA-CATTCGCG-GGAC-ACGAG---G---C--G-GGG---    70
HIV2D194   -----T-CGAG-GGA----A--A--GTTC-GA-CATTCGCA-GGAC-GCGAG---G---A--G-GGG---    70
HIV2GH1    -----G-CGAG-GGA----A--A--ATTC-AA-CATTCGCA-AGAC-ACGAG---G---C--G-GAG---    70
SIVMM251   -----T-GAGC-ATT----T--G--GGTC-AA-CCGGCTGG-GATC-GCGAC---A---C--G-GGG---    70
SIVMMP11   -----T-GAGC-ATT----T--G--GGTC-AA-CCGACTGG-GATC-GCGAC---A---C--G-GGG---    70
SIVMM142   -----T-GAGC-ATT----A--A--GGTC-AA-CCGCCCAG-GATC-GCGAC---G---C--G-GGG---    70
SIVMM239   -----T-GAGC-ATT----T--G--GGTC-AG-CCGTCTGG-GATC-GCGAC---G---C--G-GGG---    70
SIVMNE     -----T-GAGC-ACT ---A--G--GGTC-AA-TCGCCTGG-GATC-GCGAC---G---C--G-GGG---    70
SIVSMMH4   -----T-GCGC-ATT----A--A--AGTA-AA-CGTGGTGG-AACT-GCGCG---G---C--G-AAG---    70
SIVSMMPBJ  -----T-GCGT-ACC----A--A--AGCG-AG-CGTGGTGG-AATT-GTACG---G---C--G-AAG---    70

CONSENSUS  qtGgaGagactTaTGggagactctg?gaagagttGGaAGggGgATactcGCaaTcCCaaGaaGgATcAGg   131
HIV2ROD    GT-CA-GGGCT-G--GAGGGTATTGGAACGAATC--G--GG-A--ACTC--GG-T--AA-AA-G--C--A   140
HIV2NIHZ   GG-GA-AGACT-G--GAGGGCGCTGCAACGAATC--G--GG-G--ACTT--AG-C--AC-AA-G--C--G   140
HIV2ISY    GG-GA-GAGCT-G--GGGGGCTCTGGGACGAATC--A--GG-G--ACTC--AG-T--AC-AA-G--C--G   140
HIV2ST     CT-GA-AGGCT-C--GGGGACACTGGGACAAATT--G--GG-A--ACTT--AG-C--AA-AA-G--C--G   140
HIV2BEN    GT-GG-ATGGT-A--GGAAGCAGCGCGACGCATC--G--GG-A--ACTC--AG-T--AA-AA-A--C--G   140
HIV2D194   GT-GA-GGGGT-A--TGAAGCAGCGCAACGCATC--G--GG-A--CCTC--AG-C--AA-AA-G--C--G   140
HIV2GH1    AT-GG-GGGGT-A--TGCAGCAGTGCAACGCGTC--G--GG-A--ACTC--AG-C--AA-AA-G--C--G   140
SIVMM251   GT-GA-AGACT-A--GGAGACTCTTAGGAGAGGT--A--AT-G--CCTC--AA-C--TA-GA-G--T--G   140
SIVMMP11   GT-GG-AGACT-A--GGAGACTCTTAGGAGAGGT--A--AT-G--CCTC--AA-C--TA-GA-G--T--G   140
SIVMM142   GT-GG-AGAAT-A--GGAGGCTCTTCAAAGGGGT--A--AT-G--CCTC--AA-C--TA-GA-G--T--A   140
SIVMM239   GT-GG-AGACT-A--GGAGACTCTTAGGAGAGGT--A--AT-G--ACTC--AA-C--CA-GA-G--T--A   140
SIVMNE     GT-GG-AGACT-A--GGAGACTCTGGGAAGAGTT--A--AT-G--ACTC--AA-C--TA-GA-G--C--A   140
SIVSMMH4   GT-GG-AGACA-A--GGAGACTCTGGGAAGGGTT--A--AG-G--ACTC--AA-C--TA-GC-C--C--G   140
SIVSMMPBJ  GT-GA-AGACT-A--GGAGACTCTGGGAAGGGTT--A--GG-A--ACTC--AA-C--AA-AC-C--C--G   140

                                        <- env
CONSENSUS  CAaGGgct?GAacTcaC?CTCtTGTGAGggaCagaaaTAt?atcAgGGagA?TttaTgAAtAccCCaTGG   197
HIV2ROD    --G--AGCA--AA-CG-C---C------GGA-GGCAG--TCAGC-G--AG-C-TTA-G--T-CT--A---   210
HIV2NIHZ   --G--AGCA--AC-CG-C---C------GGA-AGCGA--TCAGC-G--AG-T-TTG-A--T-CC--A---   210
HIV2ISY    --G--AGCA--AA-TG-C---C------GGA-AGAGA--TCAGC-A--AG-C-TTA-G--T-CC--A---   210
HIV2ST     --G--GGCA--AA-CG-C---C------GGA-GGCGG--TCAAC-G--AG-T-TTA-G--T-CC--A---   210
HIV2BEN    --G--AGCA--AC-CG-C---C------GGA-AGCAG--TCAGC-G--AG-G-ACA-G--C-GC--A---   210
HIV2D194   --G--AGCA--AA-CG-C---C------GGA-AGCAG--TCAGC-G--AG-G-TCA-G--C-CC--A---   210
HIV2GH1    --A--GGCA--AA-CG-C---C------GGA-AGCAG--TCGGC-G--AG-T-TCA-G--C-CC--A---   210
SIVMM251   --A--GCTT--GC-CA-G---T------GGA-AGAAA--CAATC-G--GC-G-ATA-G--T-CT--A---   210
SIVMMP11   --A--GCTT--GC-CA-G---T------GGA-AGAAA- CAATC-G--GC-G-ATA-G--T-CT--A---   210
SIVMM142   --A--GCTT--GC-CA-T---T------GGC-AAAAA--CAATC-G--AG-A-ACA-G--T-CT--A---   210
SIVMM239   --A--GCTT--GC-CA-T---T------GGA-AGAAA--CAATC-G--AC-G-ATA-G--T-CT--A---   210
SIVMNE     --A--GCTC--GC-TA-T---T------GGA-AGAAA--CAATC-G--AC-G-TTA-G--T-CT--A---   210
SIVSMMH4   --A--GCTT--GC-CA-T---T------CCT-AGAAA--TAGTG-G--AC-A-TTA-G--T-CC--C---   210
SIVSMMPBJ  --A--GCTT--GC-CA-T---T------CCT-AGAAA--TTGTG-G--TC-A-TTA-G--T-CC--T---   210
```

```
CONSENSUS  AgaaaCCCaGCagcagaaaggcaaAAa?ta?caTAcAgacAaCAAAAtATgGATGATgTaGAT???ga?g   261
HIV2ROD    -AGGA---A--AGCAGAAAGGGAG--AAATTTG--C-GGC-A-----T--G------G-A---TCAGATG   280
HIV2NIHZ   -GAAC---A--AGCAGAAAGGGAG--AGAATTG--C-AAC-G-----T--G------G-A---CTAGATG   280
HIV2ISY    -GAAC---A--AACAGAAAAGGAA--AGAATCG--C-GGC-A-----T--G------G-A---TCAGATG   280
HIV2ST     -GAGC---A--A...GAAGGGGAG--AGGCTCG--C-AGC-A-----T--G------G-A---TCAGATG   277
HIV2BEN    -GAAA---A--AACAGAAAGACAG--AGATTTG--T-GGC-G-----T--G------G-A---TCTGATG   280
HIV2D194   -GAAC---A--AGCAATAGGGCAG--AAATTCA--T-AGC-G-----T--G------G-A---TCTGATG   280
HIV2GH1    -GAAC---G--AATAGAAGGACAG--AAAGTTA--C-AGC-G-----T--G------A-A---TCTAGTG   280
SIVMM251   -GAAA---A--TGAAGAAAAAGAA--ATTAGCA--C-GAA-A-----T--G------A-A---GAGGAAG   280
SIVMMP11   -GAAA---A--TGAAGAAAAGGAA--ATTAGCA--C-GAA-A-----T--G------A-A---GAGGAGG   280
SIVMM142   -GAAA---A--TGAAGAGAGGAAA--ATTACCA--C-GAA-A-----T--A------A-A---GAGGAAG   280
SIVMM239   -GAAA---A--TGAAGAGAGAGAA--ATTAGCA--C-GAA-A-----T--G------A-A---GAGTAAG   280
SIVMNE     -AAAA---A--TGGAGAGAGGGAA--ATTAGCA--C-GAA-A-----T--A------A-A---GAAGAAG   280
SIVSMMH4   -GGAA---A--AACAGAAAGAGCA--ATTAGGT--T-GAC-A-----C--G------G-G---GATGAAG   280
SIVSMMPBJ  -GGAA---A--AACAGAAAGAGCA--ATTAGAT--T-GAC-A-----C--G------G-G---.....A   274

CONSENSUS  aTgaTgAcctagTagG?gt?cCaGt?a??cCaagAGT?CC?ctaAGagcaATgAc?tacAaatTGGCAat   323
HIV2ROD    A-GA-G-CCAAG-AA-AGTTT-T-TCACAC-AAA---A--ACTA--ACCA--G-CACAT-GAT-----AT   350
HIV2NIHZ   A-GA-G-CCAAG-AG-ATTCC-T-TCACAC-AAG---A--ATTA--ACCA--G-CATTC-AAT-----GT   350
HIV2ISY    A-GA-G-CCTAG-AG-GGTCT-T-ACACAT-AAG---A--ATTG--AGCA--G-CATAT-GAA-----GT   350
HIV2ST     A-GA-G-CCTAG-AG-GGTCC-T-TCACAC-AAG---A--ATTA--AGAA--G-CATAT-GGT-----AG   347
HIV2BEN    A-GA-G-CCTAA-AG-AGTTC-T-TTACAC-AAG---A--ACGG--AGAA--G-CCTAT-AAT-----AT   350
HIV2D194   A-GA-G-CCTAG-GG-AGTTC-T-TTATGC-AAG---A--GCTG--AGAA--G-CCTAT-AAC-----AT   350
HIV2GH1    A-GA-G-CCTAG-GG-AGTTC-T-TTACAC-AAG---A--ACTA--AGCA--G-CCTAT-AAT-----GT   350
SIVMM251   A-GA-G-CTTGG-AG-GGTAT-A-TGAGGC-AAA---T--CCTA--AGCA--G-CTTAC-AAT-----AT   350
SIVMMP11   A-GA-G-CTTGG-AG-GGTAT-A-TGAGGC-AAA---T--CCTA--AGCA--G-CTTAC-AAT-----AT   350
SIVMM142   A-GA-G-CTTGG-AG-GATAC-A-TTGAGG-CAG---T--CCTA--AACA--G-GTTAC-AAT-----AT   350
SIVMM239   A-GA-G-CTTGG-AG-GGTAT-A-TGAGGC-AAA---T--CCTA--AACA--G-GTTAC-AAT-----AT   350
SIVMNE     A-AA-G-CTTGG-AG-GGTAC-A-TGAGGC-ACG---T--CTTA--AATA--A-GTTAC-AAT-----GT   350
SIVSMMH4   A-GA-G-CTTAA-AG-TGTCT-A-TGCACC-AAG---C--CTTA--GGCC--G-CATAC-AAT-----AT   350
SIVSMMPBJ  G-GC-G-CTTAG-AG-TTGTC-A-TGAGCC-AAG---C--CGTA--AATC--G-CATAC-AAT-----AT   344

                                                  LTR -> (for further information, see the LTR)
CONSENSUS  AGAtaTGTCtcATTTtATAAAagaaAagggggGGACTGgAaGGGaTtTaTTAcAgTga?agaAGACaTAgA   392
HIV2ROD    ---TA----AC----A-----AACA-GGGGG------G-A---A-G-T---C-G-GAAAGA----A--A-   420
HIV2NIHZ   ---TA----TC----T-----AGAA-AAGGG------G-A---C-G-T---T-G-CAGAGA----A--G-   420
HIV2ISY    ---CA----AG----A-----AGAT-AGGGG------G-A---A-G-A---C-G-GAGAGA----A--G-   420
HIV2ST     ---TA----AC----G-----AGAA-AGGGG------G-A---C-G-A---C-G-GATAGG----G--G-   417
HIV2BEN    ---TA----AC----T-----AGAA-AAGGG------C-A---A-G-T---C-G-AGGAGG----A--G-   420
HIV2D194   ---TA----AC----T-----AGAA-AAGGA------G-A---A-A-T---C-G-AGGGAG----A--G-   420
HIV2GH1    ---TA----AC----T-----GAAA-GG...------G-T---A-G-T---C-G-AGGGAT----A--G-   417
SIVMM251   ---TA----TC----T-----AGAA-AGGGG------G-A---A-T-A---C-G-GCAAGA----A--G-   420
SIVMMP11   ---TA----TC----T-----AGAA-AGGGG------G-A---A-T-A---C-G-GCAAGA----A--G-   420
SIVMM142   ---TA----TC----T-----AGAA-AGGGG------G-A---A-T-A---C-G-GCAAGA----A--G-   420
SIVMM239   ---CA----TC----T-----AGAA-AGGGG------G-A---A-T-A---C-G-GCAAGA----A--G-   420
SIVMNE     ---TA----TC----T-----AGAA-AGGGG------G-A---A-T-A---C-G-GAAAGA----A--A-   420
SIVSMMH4   ---CA----TC----T ----AGAA-AGGGG------G-A---A-T-A---C-A-GAGAGA----A--G-   420
SIVSMMPBJ  ---CA----TC----T-----AGAA-AGGGG------G-A---A-T-A---C-G-GATAGA----A--A-   414
```

HIV2/SIV NEF

```
CONSENSUS  aTctTagAcaTgTaCtTaGAaaaggaagaaGG?ATaaTaccAGAtTGGCAgaatTAcA?ctCa?GG?CcA  458
HIV2ROD    A-CT-AA-TA-A-A-T-A--AAAGGAAGAA--G--AA-TGC---T-----GAAC--C-.CT-AT--G-C-  489
HIV2NIHZ   A-CT-AG-CT-A-A-T-A--CAAG.....................................................  444
HIV2ISY    A-CC-AG-CA-A-A-T-A--AAAGGAAGAA--G--AA-TCC---T-----GAAC--T-.CT-AT--G-T-  489
HIV2ST     G-CC-AG-CA-A-A-T-A--AAAGGAAGAG--A--AA-TGG---C-----GAAC--T-.CT-AT--A-C-  486
HIV2BEN    A-CC-AG-CA-A-A-C-A--AAAAGAGGAA--G--AA-ACC---T-----GAAT--T-.CT-AT--G-C-  489
HIV2D194   A-CC-AG-CT-G-T-C-A--AAAGGAGGAA--G--AA-ACC---T-----GAAT--T-.CT-AT--G-C-  489
HIV2GH1    A-CC-AG-CT-G-A-C-A--AAAGGAGGAA--G--AA-ACC---T-----GAAT--T-.CT-AT--G-C-  486
SIVMM251   A-CT-AG-CA-G-A-T-A--AAAGGAAGAA--C--CA-ACC---T-----GGAT--C-CCT-A.--A-C-  489
SIVMMP11   A-CT-AG-CA-G-A-T-A--A.........--C--CA-ACC---T-----GGAT--C-CCT-A.--A-C-  480
SIVMM142   A-CT-AG-CA-A-A-T-A--AAAGGAAGAA--C--CA-ACC---T-----GA..--C-.CT-C.--A-C-  486
SIVMM239   A-CT-AG-CA-A-A-T-A--AAAGGAAGAA--C--CA-ACC---T-----GGAT--C-CCT-A.--A-C-  489
SIVMNE     A-CT-AG-CA-G-A-T-A--AAAGGAAGAA--C--CA-GCC---T-----GAAT--C-CCT-G.--A-C-  489
SIVSMMH4   A-AT-AG-TA-G-A-T-G--AAAGGAAGAA--A--AA-ACC---T-----AAAT--C-CAT-A.--G-C-  489
SIVSMMPBJ  A-AT-AG-TC-G-A-T-A--AAAGGAAGAA--A--AA-ACC---T-----AAAT--C-CAG-A.--G-C-  483

CONSENSUS  gGaataAGaTAcCcaA?Gt?cTttGGgTGGcTaTGGAAa?TAgT?cCaGTA?atgTctCAgAtGAqGcac  523
HIV2ROD    G-AGTA--A--C-CA-T-TTC-TT--G---C-A-----GC--G-AC-A---GATG-CC--C-A--A-GGG  559
HIV2NIHZ   .....................................................................  444
HIV2ISY    G-AGTA--G--C-CA-T-TTC-TT--G---C-A-----GC--G-AC-A---ACTG-CC--C-A--A-GGG  559
HIV2ST     G-AGTA--G--T-CA-A-TTC-TT--G---T-A-----GC--G-AC-A---GATG-CC--C-A--G-GAG  556
HIV2BEN    G-AGTA--G--C-CA-T-TAC-TC--G---C-G-----GC--G-AT-A---GAAC-CT--C-A--G-CAG  559
HIV2D194   G-AACA--G--C-CA-T-TAC-TC--G---C-G-----AC--G-AC-A---GACA-CT--C-A--G-CAG  559
HIV2GH1    G-AGTA--G--C-CA-T-TGT-TC--G---C-A-----GC--G-AC-A---GATG-CT--C-A--G-CAG  556
SIVMM251   G-AATT--A--C-CA-A-ACA-TT--C---C-A-----AT--G-CC-T---AATG-AT--G-T--G-C C  559
SIVMMP11   A-AATT--A--C-CA-A-ACA-TT--C---C-A-----AT--G-CC-T---AATG-AT--G-T--G-CAC  550
SIVMM142   G-AATT--A--C-TA-A-ATG-TT--C---C-A-----AT--A-CC-T---AATG-AT--G-T--G-CAC  556
SIVMM239   G-AATT--A--C-CA-A-ACA-TT--C---C-A-----AT--G-CC-T---AATG-AT--G-T--G-CAC  559
SIVMNE     G-ACCT--A--C-CA-A-ACA-TT--C---C-A-----AT--G-CC-T---AATG-AT--G-T--G-CAC  559
SIVSMMH4   G-AATT--A--C-CT-T-CAC-AT--G---C-C-----AT--G-CC-A---GATG-CT--G-T--A-CTC  559
SIVSMMPBJ  G-AATC--A--C-CT-T-TTC-TT--A---C-C-----AT--G-CC-G---AATG-CT--G-T--A-CTC  553

CONSENSUS  AgGA?gatGAg???ac?catTactTagtgCAtcCaGCaCAaac?t?caagtggGATGAccc?tggGGaGA  585
HIV2ROD    -G--CACT--G...ACTCAC-GCT-AGTA--TC-A--A--AACAAGCAAGTTT-----CCCGCAT--G--  626
HIV2NIHZ   .....................................................................  444
HIV2ISY    -G--CACT--G...ACTCTC-GCT-AATG--CT-A--A--AGTAAGCAGATTT-----CCCGCAT--G--  626
HIV2ST     -T--CAGT--G...ACTCAC-GCT-AGTG--TC-A--A--AACAAGCAGGTTT-----CCCGCAT--A--  623
HIV2BEN    -G--AGAT--G...GCCAAC-GCT-AGTA--CC-A--A--AACAAGCAGACAT-----TGAGCAT--G--  626
HIV2D194   -G--AGTA--G...ACCAAC-GCT-AGTA--CC-A--A--AACAAGCAGATAT-----CGAGCAT--G--  626
HIV2GH1    -G--CGAT--G...ACCAAC-ACT-AACA--CC-A--A--AACAAGCAGACAT-----CGAGCAT--A--  623
SIVMM251   -G--GGAT--GAGG...CAT-ATT-AATG--GC-A--T--AACTTCCAAGTGG-----CCCTTGG--A--  626
SIVMMP11   -G--GGAT--GAGG...CAT-ATT-AATG--GC-A--T--AACTTCCAAGTGG-----CCCTTGG--A--  617
SIVMM142   -G--GGAT--GGAG...CAT-ATT-AGTG--CC-A--T--AACTTCCCAGTGG-----CCCTTGG--A--  623
SIVMM239   -G--GGAT--GGAG...CAT-ATT-AATG--TC-A--T--AACTTCCCAGTGG-----CCCTTGG--A--  626
SIVMNE     -G--GGGT--GGAG...AAT-ATT-ACTG--TC-A--T--AACTTCCCAGTGG-----CCCTTGG--A--  626
SIVSMMH4   -G--AGAC--G...ACACAT-GCC-GGTG--TC-G--A--GACTTATCAGTGG-----CCCCTGG--A--  626
SIVSMMPBJ  -G--AGAC--G...ACACAT-ACT-GATG--TC-A--A--AACTTCTCAGTGG-----CCCCTGG--A--  620
```

```
CONSENSUS  ggtacTagt?TGGaagTTTGAtcCaacgcTaGCctata?tTAtgaGgCaTttatT?gatAcCCAGAaGAg   652
HIV2ROD    GACAC-AGTC---GAG-----TC-CTTGC-G--TTATAGT--CGA-G-T-TTAT-CGGT-C-----G--A   696
HIV2NIHZ   ..........................................G-T-TTAC-CTGT-C-----G--A   468
HIV2ISY    GACAC-AGTC---AAG-----CC-CATGC-G--TCATGAG--CAC-A-C-TTAT-CTAT-C-----G--A   696
HIV2ST     AACAT-AGTT---AGG-----CC-CACGC-A--TTTTAGC--CGA-G-C-TTAT-CGAT-C-----G--G   693
HIV2BEN    GACAT-AGTG---CAG-----CT-CATGC-G--CTATAAC--CAA-G-C-TCAC-CTGT-C-----A--G   696
HIV2D194   GACAC-AGTT---CGG-----CC-CATGC-G--CTATAGT--CAA-G-C-TCAT-CTGC-C-----A--A   696
HIV2GH1    AACAC-ACTT---CGG-----CC-CACGC-G--CTATGAT--CAA-G-T-TCAT-CTGC-T-----A--G   693
SIVMM251   GGTTC-AGCG---AAG-----TC-AACTC-A--CTACACT--TGA-G-A-ATGC-AGAT-C-----A--G   696
SIVMMP11   GGTTC-AGCG---AAG-----TC-AACTC-A--CTACACT--TGA-G-A-ATGT-AGAT-C-----A--G   687
SIVMM142   GGTTC-AGCA---AAG-----TC-AACTC-A--CTACACT--TGA-G-A-ATAT-AGAT-C-----A--G   693
SIVMM239   GGTTC-AGCA---AAG-----TC-AACTC-G--CTACACT--TGA-G-A-ATGT-AGAT-C-----A--G   696
SIVMNE     GGTTC-AGTA---AAG-----TC-AACTC-A--CTACACT--TGA-G-A-ATAT-AGAT-C-----A--G   696
SIVSMMH4   GGTAC-GGCA---AAG-----TC-AGAAT-A--TTATAGC--TAA-G-A-TTAT-AAGT-C-----A--G   696
SIVSMMPBJ  GGTAC-GGCA---AAG-----TC-AAAGT-A--TTATAAC--TAA-G-A-TTGT-GAGC-C-----A--G   690

CONSENSUS  TTtGG???caAgTCAGGccTG?cAgAggAaGAggtgaagagaAggCt?a??GCAAGaGGc?Tac?atatA   713
HIV2ROD    --T--GCACA-G-----CC--CC-G-GG-A--GTGGAAGGCG-GA-TGAAA-----A--AA-ACCATTT-   766
HIV2NIHZ   --T--GCATA-T-----AC--CC-G-GA-A--GTGGAAGGCG-GA-TGAAA-----G--AA-ACCATTT-   538
HIV2ISY    --T--GCACA-G-----AA--GA-G-AG-T--CTGGAAGGCA-AA-TGAAA-----A--GA-ACCATTT-   766
HIV2ST     --T--GTACA-G-----CC--CC-G-GG-T--ATGGAAGGCA-GA-TGAAA-----A--GA-ACCGTTT-   763
HIV2BEN    --T--GCACA-G-----AT--CC-G-GA-A--ATGGAAGGCA-AA-TGAAA-----A--GA-ACCATAT-   766
HIV2D194   --T--GCACA-G-----AT--CC-G-GA-A--GTGGAAGGCA-AA-TGAAA-----A--GA-ACCATAT-   766
HIV2GH1    --T--GCACA-G-----AT--CC-G-GA-A--GTGGAAGGCA-AA-TGAAA-----A--GA-ACCATAT-   763
SIVMM251   --.--AAGCA-G-----CC--TC-G-GG-A--GGTTAGAAGA-GG-***                        744
SIVMMP11   --T--AAGCA-G-----CC--TC-G-GG-A--GGTTAGAAGA-GG-TAACC-----A--CC-TCTTAAC-   757
SIVMM142   --T--AAGCA-G-----CC--TC-G-GA-A--GGTTAAAAGA-GG-TAGCC-----A--CC-TCTTGAA-   763
SIVMM239   --T--AAGCA-G-----CC--TC-G-GG-A--GGTTAGAAGA-GG-TAACC-----A--CC-TCTTAAC-   766
SIVMNE     --T--AAGCA-G-----CC--TC-G-GG-A--GGTTAGAAGA-GG-TAACC-----A--CC-CTTAAAA-   766
SIVSMMH4   --T--TAGTA-G-----CT--TC-G-GG-A--GGTAAAGAGA-GG-TAACC-----A--CC-TTT.AAA-   765
SIVSMMPBJ  --T--TAGTC-G-----CT--TC-A-GG-A--GGTGCAGAGA-GG-TAACC-----A--CC-TTTAAAA-   760

CONSENSUS  ?ggctgACAaGAaGgAAACtaGC???    737
HIV2ROD    GT***                         771
HIV2NIHZ   GT***                         543
HIV2ISY    GT***                         771
HIV2ST     GC***                         768
HIV2BEN    GTGAA***                      774
HIV2D194   GTGAA***                      774
HIV2GH1    GT***                         768
SIVMMP11   TGGCTG---A--G-G----TC--***     783
SIVMM142   TGGCTG---G--A-G----TA--***     789
SIVMM239   TGGCTG---A--A-G----TC--***     792
SIVMNE     TGGCTG---A--G-G----TA--***     792
SIVSMMH4   TGGCTG---A--A-G----AA--TGA >  929
SIVSMMPBJ  TGGCTG---A--A-A----AA--***     786
```

```
LOCUS       HIV2D194     9472 bp ss-RNA
DEFINITION  Human immunodeficiency virus type 2 (HIV-2), isolate D194 from
            a Gambian person; complete proviral genome.
ACCESSION   J04542 X52223
SOURCE      Human immunodeficiency virus type 2, isolate D194, clone 10;
            derived from circular proviral DNA; biologically active clone.
REFERENCE   1  (bases 1 to 9399)
  AUTHORS   Kuhnel,H., von Briesen,H., Dietrich,U., Adamski,M., Mix,D.,
            Biesert,L., Kreutz,R., Immelmann,A., Henco,K., Meichsner,C.,
            Andreesen,R., Gelderblom,H. and Rubsamen-Waigmann,H.
  TITLE     Molecular cloning of two West African human immunodeficiency virus
            type 2 isolates that replicate well in macrophages: a Gambian
            isolate, from a patient with neurologic acquired immunodeficiency
            syndrome, and a highly divergent Ghanian isolate
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 2383-2387 (1989)
REFERENCE   2  (bases 1 to 9472; revises [1])
  AUTHORS   Kuhnel,H., von Briesen,H., Dietrich,U., Adamski,M., Mix,D.,
            Biesert,L., Kreutz,R., Immelmann,A., Henco,K., Meichsner,C.,
            Andreesen,R., Gelderblom,H. and Rubsamen-Waigmann,H.
  TITLE     ;
  JOURNAL   Unpublished (1991)
COMMENT     The revision in reference [2] corrected a possible frameshift
            mutation in gag and a premature stop in env observed in the
            original copy of the sequence.
            The D194 isolate was taken from a Gambian male in 1986; the
            patient suffered from severe neurological symptoms up to his
            death in 1987.
BASE COUNT     3220 a   1946 c   2382 g   1924 t
ORIGIN
        1 agtcgctctg cggagaggct ggcagattga gccctgggag gttctctcca gcactagcag
       61 gtagagcctg ggtgttccct gctagactct caccagtgct tggccggcac tgggcagacg
      121 gctccacgct tgcttgctta aagacctctt AATAAAGctg ccagttagaa gcaagttaag
                                    signal ->
      181 tgtgtgttcc catctctcct agtcgccgcc tggtcattcg gtgttcatct gagtaacaag
      241 accctggtct gttaggaccc ttcccgcttt gagaatccaa ggcaggaaaa tccctagcag
          <- lys-tRNA pbs
      301 gttggcgccc gaacagggac ttgaaagagg actgagaagc cctggaacac ggctgagtga
      361 aggcagtaag ggcggcagga acaaaccacg acggagtgct cctagaaaag cgcgggccga
      421 ggtaccgaag cggcgtgtgg agcgggagtg aaagaggcct ccgggtgaag gtaagtacct
      481 acaccgaaaa ctgtagccag aaaaggcttg ttatcctacc tttagacagg tagaagattg
      541 tgggagATGg gcgcgagaaa ctccgtcttg agagggaaaa aagcagacga attagaaaaa
      gag start ->
      601 gttaggttac ggcccaacgg aaagaaaaga tacaggttaa aacatgttgt gtgggcagcg
      661 aatgaattag acagattcgg attggcagag agcctgttgg aatcaaaaga aggttgccaa
      721 aagattctta aagttttaga accattagta ccaacagggt cagaaaattt aaaaagcctt
      781 tttaataccg tctgcgtcat ttggtgcttg cacgcagaag agaaagtgaa agatactgaa
      841 gaagcaaaga aactagcaca gagacatcta gtggcagaaa caggaactgc agagaaaatg
      901 ccaaatataa gtagaccaac agcaccacct agtgggaaag gaggaaactt ccccgtgcaa
      961 caggcaggcg gcaactatat ccatgtgccg ctgagccccc gaactctaaa tgcttgggta
     1021 aaattagtag aggaaaagac gttcggggca gaagtagtgc caggatttca ggcactctca
     1081 gaaggctgca cgcctatga tatcaatcaa atgcttaatt gtgtgggcga tcaccaagca
     1141 gctatgcaaa taatcagaga aattattaat gaggaagcag cagattggga tgcgcagcac
     1201 ccaataccag gccccttacc agcagggcag cttagagacc caaggggctc tgacatagca
     1261 ggaacaacaa gcacagtaga tgaacagatc cagtggatgt ataggcaacc aaatcccgtg
     1321 ccggtagggа acatctacag gagatggatc cagatagggc tacagaaatg tgtcaggatg
     1381 tacaacccaa ctaacatctt agatgtgaag cagggaccaa aagaatcgtt ccagagctat
     1441 gtagacagat tctacaaaag cctaagggca gaacaaacag acccggctgt aaaaaattgg
     1501 atgacccaaa cgctgctaat acagaatgcc aacccagact gcaagttagt attaaaaggа
     1561 ctaggatgga atcccacact agagtagatg ctgactgcct gcccagggct aggcggacca
     1621 agccagaaag ccagactaat ggctgaagcc ctaaagtagt ctttgacgcc agccctatc
     1681 ccatttgcag cagcccaaca aagaagggca atTAGgtgtt ggaattgtgg aaagtagggа
                                    pol orf ->
     1741 cactcggcga aacagtgccg agcacccaga agacagggct gctggaagtg tggcaagtca
     1801 ggacacatca tggcaaactg cccgtgaaga caggcaggtt tttagggat ggggcccacg
```

```
1861 ggaaagcagc cccgcaactt ccccgcggcc caagctcctc aggggctgat accaacagca
1921 cccccaatag atccagcagt ggacctgttg gagaaatata tgcagcaagg gagaaagcag
1981 agagagcaga gggagagacc atacaaggag gtgacggagg acttactgca cctcgagcag
2041 ggagagacgc cccacagagg ggcgacagag gacttgctac acctcaattc tctctttgga
2101 aaagaccagt agtcacagca ttcatcgagg atcagccggt agaagtctta ctagacacag
2161 gagctgatga ctcaatagta gcaggaatag agttagggga caattacact ccaaaaatag
2221 tgggggaat aggggattc ataaatacca aagaatataa aaatgtagaa ataaaggtac
2281 taaataaaag agtaagagcc accataatga caggagatac cccaatcaac atttttggca
2341 gaaatattct ggcaaccttа ggcatgtcat taaacctacc agtcgccaag ttagacccaa
2401 taaaagtaac attgaagcca gggaaagatg gaccaaggct gaaacaatgg cccctaacaa
2461 aagaaaaaat agaagcacta aaagaaattt gtgaaaaaat ggaaagggag ggccaactag
2521 aagaagcacc tccaactaat ccttataata ccccccacatt tgcaattaag aaaaaggaca
2581 agaacaaatg gagaatgcta atagatttta gagaactaaa cagggtgact caagatttca
2641 cagaaattca gctaggaatt ccacacccgg caggattagc caaaaagaaa aggattactg
2701 tactagatgt aggggatgcc tacttttcca taccactaca tgaagatttt aggcaatata
2761 ctgcatttac cctaccatca gtaaacaatg cagagccaga aaaaagatat gtatataagg
2821 tcttaccaca aggatggaaa ggatcaccag caatctttca attcatgatg aggcaaatct
2881 tagaaccttt cagaaaagca aacccagacg tcattctcat ccaatacatg gatgatatct
2941 taatagctag tgacaggacg ggtttagagc atgacaaagt agtcctgcaa ctaaaagaac
3001 ttctgaatgg cctagggttc tctaccccag atgagaagtt ccaaaaggac cctccgtttc
3061 aatggatggg ctatgaattg tggccaacta aatggaaact gcagaaaata caattacctc
3121 agaaagaaat atggacagtc aatgacatcc aaaaactagt aggagttttg aactgggcgg
3181 cgcagatcta tccagggata aaaaccaagc atttatgtaa attgattaga ggaaaaatga
3241 cactcacaga ggaagtacag tggacagagt tagcagaggc agaactagaa gaaaacaaaa
3301 ttatcttaag tcaggaacaa gagggatcct actatcagga agaagaagaa ctagaagcaa
3361 cagtcatcaa aagccaagac aatcagtggg catacaaaat acaccaggga gagagggttc
3421 taaaagtagg aaagtatgcg aagataaaaa atactcatac caatggggtc agactactag
3481 cacaagtagt ccaaaaaata ggaaaggaag cactggtcat ttggggacga gtgccaaaat
3541 ttcacctacc ggtagagaga gacacctggg agcaatggtg ggataactac tygcaagtaa
3601 catgggtccc agagtgggac ttcgtatcta ccccaccact ggtcaggttg acatttaact
3661 tggtaggaga tcctatacca ggcacagaga ccttttacac agatggatca tgcaataçac
3721 agtcaaaaga aggaaaagca ggatatgtaa cagatagagg gagagacagg gtaagagtat
3781 tagagcaaac atccaatcag caagcagaac tagaagcctt tgcgatggca ctggcagact
3841 caggtcccaa ggttaatatc atagtagact cacagtatgt aatggggata gtagcaggcc
3901 aaccaacaga gtcagaaaat agaatagtaa accaaatcat tgaggacatg ataaagaaag
3961 aagcagtcta tgttgcatgg gtcccagccc ataaaggcat aggaggaaac caggaagtag
4021 accatttagt aagtcagggc atcagacaag tattattcct ggaaaagata gagcccgctc
4081 aagaagaaca cgaaaaatat catagcaata taaaagaact aacccataaa tttggaatac
4141 cccaactagt ggcaagacag atagtaaaca catgtgccca atgccaacag aaaggagaag
4201 ccatacatgg gcaagtaaat gcagaaatag gcgtttggca aatggactgc acacacttag
4261 aaggaaaaat cattatagta gcagtgcatg ttgcaagtgg attcatagaa gcagaagtca
4321 tcccacagga atcaggaagg cagacagcac tcttcctatt aaaactggcc agtaggtggc
4381 caataacaca cttgcacaca gacaatggcc ccaacttcac ttcacaggaa gtgaagatgg
4441 tggcatggtg gataggtata gagcaatcct ttggagtacc ttacaatcca caaagccagg
4501 gagtagtaga agcaatgaat caccacctaa aaaatcagat aagtagaatt agagaacagg
4561 caaatacaat agaaacaata gtactaatgg cagttcattg catgaatttt aaaagaaggg
4621 gaggaatagg ggatatgacc ccagcagaaa gactaattaa catgatcacc acagaacaag
4681 aaatacaatt cctccaaaga aaaaattcaa attttaaaaa attccaggtc tattacagag
4741 aaggcagaga tcagctgtgg aaaggacctg gagagctact gtggaaggga gacggagcag
4801 tcatagtcaa ggtaggggcg gacataaaag tagtaccaag aaggaaggcc aagattatca
4861 gggactATGg aggaaggcaa gaactggata gtagttccca cctggagggt gccagggagg
 vif start ->
4921 atggagaggt ggcaTAGcct tgtcaagcac ctgaagtaca gaacaaaaga cttagaggag
                        <- pol
4981 gtgcgctatg ttccccatca caaggtagga tgggcatggt ggacttgcag cagggtaata
5041 ttcccactag aaggagaaag tcatctagag atacaggcat attggaacct aacaccagaa
5101 aaagaatggc tctcctctca ttcagtaagg ttaacctggt atacagaaaa gttctggaca
5161 gatgttaccc cagactgtgc agactcccta atacacagca cttatttctc ttgctttacg
5221 gcaggtgaag taagaagagc catcagaggg gaaaagttat tgtcctgctg caactaccca
5281 caagctcata aagcacaggt accatcactt caatacctag ccctagtggt agtgcaacaa
5341 aATGgcagac cccagagaaa gggtgcccgc agggaacagt ggagaagaga ccattggaga
vpX start ->
5401 ggcctcgag tggctagaca ggactatata agccttaaac agggaggcag tgaaccgatt
5461 gcccgagag ctcatttcc aggtgtggca aaggtcctgg gcatattggc aTGAtgaaca
```

```
                                                                  <- vif
5521 agggatgtca acaagttaca caaagtatag atatttgtgc ataatgcaga aagctgtgta
5581 tatacatttc aagaaggggt gcacttgcct ggggagagga catggcccgg gaggatggag
5641 accaggacct ccccctcctc cccctccagg tctagtcTAA TGactgaagc accaacagag
                                          <- vpX
                                       vpR ->
5701 tttcccccag aagatgggac cccacggaga gagctaggga gtacctgggt aatagaaact
5761 ctgaaggaaa tcaaggaaga agccttaaaa cattttgatc cctgcttgct aattgctctt
5821 ggcaactata tctataatag acATGgagac acccttgaag gagccagaga gctcattaga
                tat start ->
5881 gtcctacaac gagccctctt cgtgcacatc agagcgggat gtgaccgctc aagaaagggc
5941 caaacaagga gaagagctcc ttgcccagct gcaccgaccc ctagaggcat gcacTAActc
                                                                  <- vpR
6001 atgctattgt aagcagtgca gttaccattg ccagctgtgt ttcttgaaaa aagggctcgg
6061 gatatggtAT Gcgcgacagg gcagacgaag aaggactcca agaaaaacta agactcatcc
   rev start ->
6121 gcctcctgca tcagataaGT aagtATGgag cctggtagga atcagctgct tgttgccatt
                5' sj /\  env ->
6181 ttattaacta gtgcttgctt aatatattgc aaacaatatg tgactgtttt ctatggcata
6241 cccgcgtgga gaaatgcatc tattcccctа ttttgtgcaa ccaaaaatag agatacttgg
6301 gggaccatcc agtgcttgcc agacaatgat gattatcagg aaataacctt aaatgtgaca
6361 gaagcttttg atgcatggga taatacagta acagaacaag caatagaaga tgtctggaga
6421 ctgtttgaga catcaataaa accatgtgtc aagttgacgc ccctatgtgt ggcgatgaat
6481 tgtaatataa cttcagggac taccgcgacc ccgagtccac caaacattac aataatagat
6541 gaaaattcta cctgtatagg cgacaacaac tgcacaggat tagggaaaga agaggtggtt
6601 gagtgtgagt tcaatatgac ggggctagaa caagataaga aaaggaagta taatgacgca
6661 tggtactcaa gagatgtggt ttgtgacaag acaaacggaa caggcacatg ttacatgaga
6721 cattgcaaca catcagtcat caaagagtca tgtgacaagc actattggga tgctatgaag
6781 tttagatact gtgcaccacc gggttttgcc ctactaagat gcaatgatac caactattca
6841 ggctttgaac ctaagtgctc taaagtagta gctgcttcat gcacaaggat gatggaaacg
6901 caaacttcta cttggtttgg ctttaatggc actagagcag aaaatagaac atatatctat
6961 tggcatggta aggataatag gactatcatt agcttaaaca agtattataa tctcacaatg
7021 cattgtaaga gaccaggaaa taagacagtt gtaccaataa cacttatgtc agggcgaagg
7081 tttcactctc ggccagtcta caacaaaaaa cctcggcagg catggtgttg gtttcaaggc
7141 aactggatag aagccatgcg ggaggtgaag caaacccttg caaaacatcc caggtacgga
7201 ggaacaaatg atacaggaaa aattaacttt acgaagccag gaataggttc agacccagaa
7261 gtgacatata tgtggactaa ctgcagagga gaatttctct actgtaatat gacttggttc
7321 ctcaattggg tagaaaataa gacgaaccaa acacacggca actatgcgcc atgccatata
7381 aggcagataa ttaacacctg gcataaggta gggacaaatg tatatttgcc tcctagggaa
7441 ggggagttga cctgcaattc aacagtaacc agcataattg ctaacattga ctcagatgga
7501 aatcagacca acattacctt tagtgcagaa gtggcagaac tgtaccgatt agaattgggg
7561 gactacaaat tgatagaagt aacaccaatt ccgttcgcac ctacaaaaga gaaaagatat
7621 tcctcggctc cagtgaggaa caaaagaggt gtgttcgtgc tagggttctt gggttttctc
7681 gcgacagcag gttctgcaat gggcggcgcg tccttgacgc tgtcggctca gtcccggact
7741 ttactggccg ggatagtgca gcaacagcaa cagctgttgg acgtggtcaa gagacaacaa
7801 gaaatgttgc gattgaccgt ctgggaacg aaaaatctcc aggcaagagt cactgctatc
7861 gagaaatact taaaggacca ggcacagcta aattcatggg gatgtgcgtt taggcaggtc
7921 tgccacacta ctgtaccatg ggtaaatgac tccttaacac ctgactggaa caatatgaca
7981 tggcaggaat gggaaaaacg agtccactac ctagaggcaa atatcagtca aagtttagaa
8041 caggcacaaa ttcaacaaga aaagaatatg tatgaactac aaaaactaaa tagctgggat
8101 gtctttggca actggtttga tttgacctcc tggatcaaat atattcaata tggagtttat
8161 atagtagtag gaataatagg tttaagacta gccatatata tagtgcaatt gttaagtaga
8221 cttagaaagg gctataggcc tgttttctcc tccccccccg gttatctcca acAGatccat
                                                                  /\ 3'sj
8281 atccacacgg acaggggaca gccagccaac gcagaaacag aagaagacgc cggagacgac
8341 agtggtttcg gcttgtggcc ttggccacTA Aactacatac aattcctgat ccacctactg
                                       <- tat
8401 actcgcctct tgaccgggct atacaacagc tgcaggggct tactatccaa gaactccccg
8461 acccgccgac tgatctccca gagtctaaca gcaatcaggg actggctgag acttAAgggg
                                                                  <- rev
8521 gcctacctgc aatATGggtg cgagtcgtat cgaagaagct tccgagcatt cgcaaggact
        nef start ->
8581 gcgagagaga ctattgcggg cgcgtcgtag ggggttatgt gaagcagcgc acgcatcggg
8641 agggaatccc tcgcagtccc aagaagcgat cagcagggag cagaaatccg cctcctgTGA
```

```
                                                                    <-env
8701 gggacagcag tatcagcagg gagagttcat gaacacccca tggagaaccc cagcagcaat
8761 agggcagaaa aattcatata agcagcaaaa tatggatgat gtagattctg atgatgatga
8821 cctagtggga gttcctgtta tgccaagagt accgctgaga gaaatgacct ataaactggc
8881 aatagatatg tcacattttt taaaagaaaa aggaggactg gaagggatat tttacagtag
                                    3'LTR ->
8941 ggagagacat agaatcctag acttgttcct agaaaaggag gaagggataa taccagattg
9001 gcagaattat actcatgggc caggaacaag gtacccaatg tacttcgggt ggctgtggaa
9061 actagtacca gtagacatct cacaagaggc agaggaagta gagaccaact gcttagtaca
9121 cccagcacaa acaagcagat atgatgacga gcatggggag acactagttt ggcggtttga
9181 ccccatgctg gcctatagtt acaaggcctt cattctgcac ccagaagaat ttgggcacaa
9241 gtcaggattg ccagagaaag agtggaaggc aaaactgaaa gcaagaggga taccatatag
9301 tgaaTAAcag gaacaaccat acttggtcag ggcaggaaat agctactaag aacagctgag
        <- nef
9361 actgcaggga ctttccagaa ggggctgtaa ccaagggagg gacatgggag gagctggtgg
9421 ggaacgccct catattctct gtataaatgt acccgcttct tgcattgtat tc
```

```
CONSENSUS  TGGATGGGATTTATTACTCCGAaAGgAG?aAtaagATccTgAA?cTgTAtGCccT?AAtGAaTGGGGaAT   67
SIVAGMTYO  ----------------------T--G--AA-TAAG--CC-G--TC-G--T--TC-T--T--A-----G--   70
SIVAGM155  ----------------------T--G--AA-TAAG--CC-T--TC-G--T--CC-C--T--A-----A--   70
SIVAGM3    ----------------------C--G--GA-TCAG--CC-A--CC-G--C--CC-C--T--G-----A--   70
SIVAGM677  ----------------------A--A--AG-AAAG--CC-G--TT-G--T--CT-G--C--G-----A--   70
SIVAGM692  ----------------------A--A--GG-AAAG--TT-G--CC-G--T--AT-A--T--A-----A--   70
SIVAGM9    ----------------------A--G--GA-TAGA--CC-C--CC-A--T--TC-T--T--A-----A--   70

CONSENSUS  cATtGATGATTGGaAtGCtTggtCa??aGGaCCaGG?ATaAGaTacCC?agatgCTTtGGcTTcTGcTTt  133
SIVAGMTYO  A--T---------A-T--C-GGT-GAAG--A--A--A--A--A-TC--TAAATG---T--G--C--C--T  140
SIVAGM155  C--T---------A-C--A-GGT-AAAA--A--T--G--A--A-AC--GAGGTG---T--C--C--C--C  140
SIVAGM3    C--T---------A-T--T-GGT-AGAA--A--A--A--C--A-AC--AAGATG---C--C--C--C--T  140
SIVAGM677  A--A---------C-A--T-ACT-ACCA--C--G--G--A--G-AC--GAGAGT---T--C--C--C--T  140
SIVAGM692  C--A---------C-A--T-ATA-TCCA--T--A--C--C--A-AT--AAGATG---T--G--C--T--T  140
SIVAGM9    C--T---------A-T--A-GGT-AGCA--A--A--C--A--A-AT--CCGCTG---T--C--T--C--C  140

CONSENSUS  aAg?TaGTaCCaGTgGaccTgcaTGAgGAaGCa?aaacaTGTGA?AGaCAtTGccTggTgCAtCCaGCaC  200
SIVAGMTYO  A-GC-A--G--A--G-ACT-ACA---G--A--ACAAACA-----A--A--T--CC-AG-C--T--A--G-  210
SIVAGM155  A-GC-A--A--G--T-CCC-GCA---G--A--AGAAACA-----A--G--T--CT-GG-A--C--A--A-  210
SIVAGM3    A-AT-G--A--A--A-ACC-GCA---G--A--AGAGACT-----G--A--T--CC-GG-G--T--A--A-  210
SIVAGM677  A-GC-A--C--A--G-ACC-GCA---G--G--ACGCAAC-----G--A--C--TC-GA-G--T--A--A-  210
SIVAGM692  G-AT-A--G--A--G-ACC-TAG---G--A--GCAAGGA-----A--G--C--TC-GG-C--T--T--T-  210
SIVAGM9    A-GT-A--A--G--A-AGA-GCA---A--G--AGAAACC-----G--A--T--CT-GG-G--T--T--A-  210

CONSENSUS  Aa?Tg??gGAaGA?CCtGATGGtATca?tCATGGAGAaatcTTGGtcTGGAagTTtgAtCC??AtGtT?G  262
SIVAGMTYO  -GA-GGGA--A--T--A-----T--CAGC--------GATC----TG----AG--TG-T--.T-T-T-G-  279
SIVAGM155  -AC-GCAT--A--C--T-----T--AAAT--------AATA----CA----AG--TG-T--.A-T-T-G-  279
SIVAGM3    -AG-GAGG--A--C--T-----A--CAAC--------AGTC----TC----AG--TG-T--.C-T-T-A-  279
SIVAGM677  -GA-GGGG--A--T--T-----A--AGAT--------AGTC----TC----AG--TG-C--.G-A-T-G-  279
SIVAGM692  -AT-ACAG--G--T--A-----T--CTGG--------AACA----TC----GA--CA-T--.C-T-C-A-  279
SIVAGM9    -AG-AAAA--G--C--C-----C--CAGT--------GACC----TC----AG--TG-C--CT-T-T-A-  280

                                                                          <- nef
CONSENSUS  CagTgcAc?acgaccCcaa?agggAgta?tTta?tGACaTgcAtgcact?Gtga?agaGgaag?TAGc?t  325
SIVAGMTYO  -AA-AC-GTACGACC-CAATCGGG-GTACT-TACT---A-GC-TGGGCTG-TGA.AGA-GAAG.---CCA  347
SIVAGM155  -TG-TC-GTACGACC-CTCAAGGG-GTACT-TACA---T-AT-TTCAACA-TGGTACA-GAAAC---CC.  348
SIVAGM3    -AG-CC-ATATGACC-TAACAGAA-ATATC-CACT---A-GC-TGATCTT-GCA.AGA-GAAG.---C.T  346
SIVAGM677  -GG-GG-GTACCGCC-......GG-CATGT-TAAG---A-GC-CGAACAT-CAA.AGC-C....---..T  336
SIVAGM692  -A.-GC-AGGCCATG-CA....GG-G.TGT-CAAT---A-GC-TGCAACA-TGG....-GAAG.---C..  336
SIVAGM9    -AG-GC-GTATGACC-AAACAGAC-GTATT-AGAA---A-GC-TGCACTG-TGA.AGA-GAAG.---C.T  347

CONSENSUS  gaccgcaggctTGcGGTTa?gc?cgTc?cc?aggaGatgACatt?????????TGCTgAc????  378
SIVAGMTYO  GACCGCAAGCC--C----A.GAACA-CACCATGGA-ATG--ATTAAAAAC.........----G-C....  403
SIVAGM155  GACCACAGGCT--C----T.CCTGG-TGCCTAGGA-ATG--ATTAAGAAC.........----G-C....  404
SIVAGM3    AACCGCAGGCT--T----AAGCACA-CACCATGGT-ATG--ATTAAGAAC.........----G-C....  403
SIVAGM677  GTCAGCA.CTT--C----GGGACTT-CCGCCA.....................................  367
SIVAGM692  ..........T--C----A.GCGCG-CCGGGACCT-TGT--CAACCAGCATAGCAACCA----A-TGAGC  395
SIVAGM9    AACCGCAGGCT--T----AAC..CG-TGCCGGGGA-ATG--ATTTGAAAC.........----G-CAA..  405

           enhancer                    enhancer
CONSENSUS  ??GGGACTTTCCAgc?a?GGGacttTCcaaGGCGGgaCAtGGGCGGTacgGGGAGTGGcTTtaCCCTCAG  444
SIVAGMTYO  ..----------GCGAA---ACTT--CAA-----GA--T-------CCG--------C--TA-------  471
SIVAGM155  ..----------GC.AC---ACTT--CAA-----GA--T-------ACG--------C--TA-------  471
SIVAGM3    ..----------GC.AA---ACTT--CAG-----GT--T-------ACG--------C--TA-------  470
SIVAGM677  ..----------CA.GT---TGGA--GGA-----TA--G-------ACT--------C--TC-------  434
SIVAGM692  TA----------G.............................AAG--------T--AA-------  432
SIVAGM9    ..----------A....---ACTT--CAG-----GC--T-------ACG--------T--TA-------  469
```

AGM LTR

```
          signal ->          putative mRNA start ->
CONSENSUS  A?gcTGcATAaAAGCAGATGCTCgctGGCTTGTAACTCAGT?CTCTTACTAGGAGaCCAGCTtGAGCCTG    512
SIVAGMTYO  -.GC--C---A----------GCT--------------.-------------A------T-------    539
SIVAGM155  -.GC--C---A----------GCT--------------.-------------A------T-------    539
SIVAGM3    -.GC--C---A----------GCT--------------.-------------A------T-------    538
SIVAGM677  -.GC--C---A----------GCT--------------.-------------A------A-------    502
SIVAGM692  -TAT--T---T----------TTG--------------G-------------.------A-------    501
SIVAGM9    -.GC--C---A----------GCT--------------.-------------A------T-------    537

CONSENSUS  GGTGTTCGCTGGTtAGCCTAACCtGGt?TGGCC?aCCaGGGGTAAG?GaCTCCttGGCTT?agAaaGCT?    577
SIVAGMTYO  ----------T---------T--T.-----.A--A--------.-A----TT-----.AG-AA---.    604
SIVAGM155  ----------T---------T--T.-----.A--A--------.-A----TT-----.AG-AA---.    604
SIVAGM3    ----------T---------T--T.-----.A--A--------.-A----TT-----.AG-AA---.    603
SIVAGM677  ----------T---------C--T.-----.A--G--------.-A----TT-----CAT-TA---C    569
SIVAGM692  ----------.---------T--AC-----CT--A--------A-C----AC-----..G-AT---T    568
SIVAGM9    ----------T---------T--T.-----.A--A--------.-A----TT-----.GG-AA---.    602

    signal ->
CONSENSUS  AATAAAccTtgcctGCaTTAGagcttaTct??GAGTCaaGTGcccctCATTga??cGCCtc?acTC??ttg    640
SIVAGMTYO  ------C.-TGCCT--A----AGCTTA-CT..-----AA---CCCT----GA..C---TC.AC--TCTTG    668
SIVAGM155  ------C.-TGCCT--A----AGCTTA-CT..-----AA---CCCT----AA..T---TC.AC--..TTG    666
SIVAGM3    ------TC-TCGCT--A----AGCTTC-CT..-----AA---CCCT----GA..C---TC.AC--..TTG    666
SIVAGM677  ------CC-.GCTC--.----TCGCTA-ATTG-----AA---C..T----GCTGC---GAGCC--..TAG    633
SIVAGM692  ------CC-TGCCT--A----AAGTAC-TC..-----GT---GTCC----GC..C---TCCGT--.....    629
SIVAGM9    ------CA-TGCCT--A----AGCTTA-CC..-----AA---CCCT----GA..C---TC.AC--..AAG    665

CONSENSUS  aAcGgG??aA??cTtcctTACTGGGTTCTCtCTct?aCCCAGGcGaGAGAAACTCCAGCA              695
SIVAGMTYO  A-C-G-..A-T.C-TCCT----------T--CTGA------C-A-------------              725
SIVAGM155  A-C-G-AGA-G..-TCCT----------T--CAAA------C-A-------------              724
SIVAGM3    A-C-G-TAA-A.C-TCCT----------T--CA.A------C-A-------------              724
SIVAGM677  .-G-T-...-ACC-CTCT----------.--GT.A------T-G-------------              687
SIVAGM692  .-C-G-..A-TCC-C.AA----------T--.T.G------G-A-------.-------              683
SIVAGM9    C-G-G-..A-CCG-TCCT----------T--CTGA------C-A-------------              723
```

AGM GAG CDS

```
CONSENSUS  ATGGGggcGGcTacCTCAGCACTgaataGGAGAcaatTaGAc?aaTTtGAgcAtATacGaCTtCGcCCGA   69
SIVAGMTYO  -----GGC--C-AC---------AAATA-----CAAT-A--CCAA--T--GA-A--AC-A--T--C----   70
SIVAGM155  -----GGC--C-AC---------GAATA-----CAAT-A--TGAA--T--GC-T--AC-A--T--C----   70
SIVAGM3    -----GGC--C-AC---------AAATA-----CAAT-A--CAAA--T--GC-T--AC-A--T--C----   70
SIVAGM692  -----TTC--G-TC---------GTCAG-----AAAT-A--CCAA--T--AC-T--AC-T--T--C----   70
SIVAGM677  -----CGG--G-CA---------GTCAG-----AGCC-C--CACG--C--GA-G--TA-G--A--T----   70

CONSENSUS  aCGGaAAgAAaAAGTAcCAAaTtAAACATTTAATATGGGCAGGcAAggAaATGGAgCGcTTcGGccTcCA  139
SIVAGMTYO  A---A--G--A-----C---A-T-------------------C--GG-G-----G--C--C--CC-C--  140
SIVAGM155  A---A--G--A-----T---A-T-------------------C--GA-G-----C--C--C--CC-C--  140
SIVAGM3    C---A--G--A-----C---A-T-------------------C--GG-A-----G--C--C--CC-C--  140
SIVAGM692  A---A--G--A-----C---T-G-------------------C--GG-A-----G--C--T--CC-C--  140
SIVAGM677  A---G--A--G-----C---A-T-------------------A--AG-A-----A--A--T--GT-A--  140

CONSENSUS  TGAgAagtT?tTaGAaaCAgAaGAaGG?TGTaAaAagATCATAGAAGT?cTct?cCCccTaGAACCaACa  205
SIVAGMTYO  ---G-GGT-GT-G--GA--G-G--G--G---A-A-GA----------CC-CTAC--CC-A-----A--A  210
SIVAGM155  ---G-AGT-AT-G--GA--G-G--A--T---A-A-AG----------TC-CTCT--CC-A-----A--A  210
SIVAGM3    ---G-GAT-AC-A--AT--G-A--A--A---A-G-AG----------AC-CTAC--GC-A-----A--A  210
SIVAGM692  ---A-AGT-GT-A--AA--G-A--G--G---A-A-AG----------AT-GCTT--CT-A-----A--C  210
SIVAGM677  ---G-AAC-TT-A--AA--A-A--A--C---C-A-AA----------TT-AACC--GT-G-----G--A  210

CONSENSUS  GGgTCgGAgGG?tTaAAAagtCTGTttAAtcTggtgTGCGTa?T?t?tTGC?T?CAc?aaGAa?aGaAAG  267
SIVAGMTYO  --A--G--G--CT-A---AGT----TC--TC-TGTG-----AC-ATAT---T-G--CAAG--ACA-A---  280
SIVAGM155  --G--G--A--AA-G---AGT----AT--TC-GGTG-----AT-GCTT---G-C--CCAA--AAA-A---  280
SIVAGM3    --G--G--G--CT-A---AGT----TT--CC-TGTG-----AT-GTTT---G-A--CAAA--TAA-G---  280
SIVAGM692  --G--G--A--TT-A---AGC----TC--TT-GACC-----CA-TTGC---A-T--TCAG--AGC-A---  280
SIVAGM677  --C--C--G--GC-A---GCT----TT--TT-GTGC-----CA-TTGG---A-T--CGCA--ACA-A---  280

CONSENSUS  TGAAAGACACAGAgGAAGCagTAgcAA?AgTaAgaCAAcacTgCCAtCTAGTGGAcAAAgaaaaaAatGC  336
SIVAGMTYO  -------------G-----AG--GC--C-G-A-GA---CAC-G---T--------A---GAAAAA-GT--  350
SIVAGM155  -------------G-----TT--GC--T-G-A-GA---TGC-G---C--------C---GAAAAA-CT--  350
SIVAGM3    -------------A-----AG--GC--T-G-A-GA---TGC-G---T--------G---GAAAGA-AT--  350
SIVAGM692  -------------G-----AG--AT--G-A-A-AG---CAG-G---T--------C---GGTGAG-AT--  350
SIVAGM677  -------------G-----TG--GT--C-G-T-AG---CAC-A---T--------C---AATGAG-AA--  350

CONSENSUS  Ag??a?a??????ac?????ca?ct?gtggacaaaAgAAaaatgacAcaggA??a?CagcgcCaccTGGt  387
SIVAGMTYO  -............ACAGAGACATCTAGTGGACAAA-G--AAATGAC-AGGG-ATAG-AGCGC-ACC---G  408
SIVAGM155  -GTT.........ACG...CCACCTGGTGGACAGC-G--AAATAAC-CAGG-GGAA-AGCGA-ACC---T  408
SIVAGM3    -GAAAGAAATACAACAGAGACATCTAGTGGACAAA-G--AAATGAC-AGGG-GTAA-AGTGC-ACC---T  420
SIVAGM692  -GCCAAA.....................GGAATAG-T--G.......-CAAC-...C-GACAC-TAG---T  390
SIVAGM677  -GCTAAA..........................A-G--AAATGAG-CAAC-...G-GCCAC-TGG---C  390

CONSENSUS  ggcagtcaaAATTtcCCaGcacaAcA?CAggg?AATGcaTGGgTaCAtgtgCC?tTgtcaCC?cGcACct  453
SIVAGMTYO  TGCAGTCAG----TT--A-CGCA-C-A--AGGA----CC---G-A--TGTA--CT-GTCA--GC-C--CT  478
SIVAGM155  GGCAGCCAA----TT--C-CACA-C-G--AGGG----CA---G-G--TGTA--AC-TTCA--TC-C--CC  478
SIVAGM3    GGCAGTCAA----TC--A-CACA-C-A--GGGA----CA---A-A--TGTG--CT-GTCA--AC-C--CT  490
SIVAGM692  AGGAGTCAA----AC--G-CACA-C-G--GAAT----TA---G-A--TGTG--AC-TAGC--CA-A--AT  460
SIVAGM677  GAATCAAGA----AC--A-TAGT-A-T--GAAT----CC---G-A--CCAG--TT-GTCT--GC-C--GT  460

CONSENSUS  TAAATGCgTGGGTaAAAgcagTaGAgGAgAAaAaaTttGGaGCAGAaaTAGT?CCCATGTTcCAaGCccT  522
SIVAGMTYO  -------G-----A---GCAG-A--G--G--A-AA-TT--A-----AA----A-----T--A--CC-  548
SIVAGM155  -------A-----A---GCAG-A--A--G--A-AA-TT--G-----AA----A-----C--A--CC-  548
SIVAGM3    -------G-----A---GCAG-A--G--G--A-AA-TC--A-----AA----G-----C--G--TT-  560
SIVAGM692  -------T-----A---GTAA-T--A--A--G-AA-TT--A-----CA----T-----T--G--CC-  530
SIVAGM677  -------G-----C---TGCG-G--G--A--A-GG-GG--A-----AG----C-----C--A--AC-  530
```

AGM GAG

```
CONSENSUS  ?TCAGAa3G?TGcacccCcTATGAcaT?AAtCAgATGcTtAATGT?cTaGGAGAcCATCAgGGGGC?tTa   587
SIVAGMTYO  A-----A--C--CACAC-C-----CA-T--T--G---C-T-----GC-A-----T-----A-----AT-A   618
SIVAGM155  C-----A--C--CACCC-A-----CA-C--T--G---C-T-----CT-A-----T-----G-----CT-G   618
SIVAGM3    A-----A--G--CACAC-C-----CA-C--T--A---C-T-----CC-G-----C-----A-----GC-A   630
SIVAGM692  G-----A--A--TACCC-A-----TG-G--C--A---T-G-----TC-A-----C-----G-----CC-G   600
SIVAGM677  C-----G--A--TCTCT-C-----TG-A--T--G---C-C-----AA-A-----C-----G-----AT-A   600

CONSENSUS  CAaATagTgAAaGAaaTcATtAATGAaGAAGCaGCccAGTGGGAtataaCaCAcccaCcaCCaGCaGG?C   656
SIVAGMTYO  --A--AG-G--A--GA-C--T-----A-----A--CC-------TGTAA-A--CCCA-TA--C--A--A-   688
SIVAGM155  --A--AG-G--A--AA-A--T-----G-----A--CC-------TGTAA-C--CCCA-CG--G--A--C-   688
SIVAGM3    --A--AG-A--A--AA-C--C-----G-----A--CC-------TATAG-T--CCCA-CA--A--A--A-   700
SIVAGM692  --G--AG-G--A--GG-C--C-----A-----T--CC-------CATTA-A--TCCC-CA--A--A--G-   670
SIVAGM677  --A--TC-T--G--AG-C--T-----A-----A--AG-------CAGGA-A--CAGA-CA--A--T--C-   670

CONSENSUS  C?tTaCCAGCaGGaCAgcTcAGaGAcCC?ag?GG?TCaGAtATAGCAGGgACcACcAGcaCagT?CAaGA   721
SIVAGMTYO  -CC-A-----A--A--GC-C--G--C--TCGC--C--A--T--------G--C--C--CT-AG-A--A--   758
SIVAGM155  -CT-G-----G--A--GC-C--G--T--GGGG--A--A--T--------G--C--T--TA-AG-G--A--   758
SIVAGM3    -AT-A-----A--A--AC-C--A--T--TAGA--C--T--C--------A--C--C--CA-AG-G--A--   770
SIVAGM692  -GC-C-----A--G--AT-G--A--T--AAGG--G--A--C--------G--T--T--TA-CA-T--A--   740
SIVAGM677  -GT-A-----A--G--GC-A--A--C--GACA--G--A--T--------A--T--C--CT-AA-T--G--   740

CONSENSUS  aCAgcTaGAaTGGAtcTacActGCcAAcCCaaGagTAGAtGTaGG?GCcatcTAtaGaAgaTGGaTtAT?   789
SIVAGMTYO  A--GT-A--A----TC-AT-CT--T--C--CC-GG----T--A--T--CATC--CC-G-GA---A-T--T   828
SIVAGM155  G--GC-A--G----TC-AT-CT--T--C--AA-GG----T--A--G--CATC--TC-A-GA---A-C--C   828
SIVAGM3    A--GC-G--A----TA-AC-CA--C--T--CA-AG----T--G--T--CATC--TA-A-GG---A-T--C   840
SIVAGM692  A--AC-A--A----TT-AC-CA--C--C--AA-AA----C--G--A--TATC--TA-G-GA---G-A--A   810
SIVAGM677  A--AA-A--G----CC-TC-AT--C--T--AA-AA----C--A--G--ACAA--CA-A-AA---G-T--T   810

CONSENSUS  ctaGGgtT?CAAAAatgtGT?aaaATGTAcAAtCCagtatc?GTccTaGAcAT?aGaCAgGGaCC?AaAG   854
SIVAGMTYO  CTA--AC-T-----GTGT--CAAA-----C--C--AGTATCA--CC-A--C--TA-G--G--A--T-A--   898
SIVAGM155  CTA--GT-A-----ATGT--AAAA-----C--T--AGTGTCT--TT-A--T--CA-A--A--G--C-A--   898
SIVAGM3    CTG--GT-G-----ATGT--AAAA-----C--C--AGTGTCT--CT-A--C--AA-A--G--G--C-A--   910
SIVAGM692  GCA--GC-G-----ATGT--CAGA-----T--T--AACAGGG--TC-G--T--AA-A--A--A--A-G--   880
SIVAGM677  TTG--CT-A-----GGTA--GCAG-----C--T--CCAAAAG--CC-A--C--TC-A--G--A--T-A--   880

CONSENSUS  AacC?TTcaa?GAtTAtGTaGAcAGaTTcTAcAAaGCaaT?AGAGCaGAaCAaGCctCagggGAaGT?AA   920
SIVAGMTYO  -GC-C--CAAG--T--T--G--C--A--T--C--G--AA-T-----A--A--A--CT-AGGG--A--G--   968
SIVAGM155  -AC-A--CAAA--T--T--A--C--A--C--T--A--AA-A-----A--A--A--TT-AGGA--A--C--   968
SIVAGM3    -AG-A--CAAA--C--C--A--T--G--C--C--A--AA-A-----T--G--G--CT-AGGA--A--A--   980
SIVAGM692  -AT-T--TAGC--T--T--A--T--A--C--C--G--CC-G-----A--A--A--CT-TCAG--T--T--   950
SIVAGM677  -AC-C--CCAG--C--T--A--C--A--C--T--A--CC-G-----A--A--A--AC-ACAG--T--T--   950

CONSENSUS  acA?TGGATGACAgAatCttTgcTcATtCAgAATGCtAAtCCaGAtTGtAAagTcATtcTgAAgGGccTa   989
SIVAGMTYO  AC-A---------G-AT-AT-AC-C--T--A-----T--T--A--T--T--GG-C--CC-G--G--CC-A  1038
SIVAGM155  AC-A---------G-AT-TT-GC-C--T--G-----C--C--A--T--C--AG-A--TT-G--G--CC-A  1038
SIVAGM3    AC-G---------G-AT-AT-AC-C--T--G-----T--T--A--C--T--AG-C--CC-A--G--CC-G  1050
SIVAGM692  GA-T---------G-CA-TC-GT-G--T--A-----T--C--A--G--T--GG-C--TC-G--A--GC-A  1020
SIVAGM677  AA-T---------C-AA-TT-GC-T--C--G-----C--T--G--T--T--AT-G--TC-G--A--AT-G  1020

CONSENSUS  GGaATGcAcCCcACccT?GAaGAAATGcTaAc?GC?TGtCAgGGaGTaGGaGG?CCaag?tAcAAaGC?A  1053
SIVAGMTYO  --A---C-C--C--CC-T--A------T-A-CG--T--T--G--G--A--A--C--AAGCT-C--A--A-  1108
SIVAGM155  --G---C-C--C--TC-T--A------C-G-CA--C--T--A--G--G--A--C--AAGTT-C--A--C-  1108
SIVAGM3    --A---C-T--C--TC-A--A------T-A-CT--C--T--A--A--G--A--A--AAGTT-C--A--A-  1120
SIVAGM692  --C---C-C--T--CT-G--A------C-T-CG--A--C--G--A--A--G--A--CCAAT-C--A--C-  1090
SIVAGM677  --A---A-T--A--CT-G--G------C-A-TA--T--C--G--A--A--A--G--ACAAC-T--G--T-  1090
```

```
CONSENSUS  AagT?ATGGcAGAAATGATGca?aa?atgca?ag?cagAAcATGgT?CAgcagGgaggtc?aagagg?aG  1115
SIVAGMTYO  -AG-A----C----------CAGACCATGCAAAATCAA--C---G-G--GCAG-GAGGTCCAAAA...A-  1175
SIVAGM155  -AG-C----C----------CAGAACCTGCAGAGTCAG--C---G-A--GCAG-GAGGTGGAAGGGGAA-  1178
SIVAGM3    -AG-G----C----------CAAAATATGCAAAGCCAG--C---A-G--ACAG-GCGGTCAGAGAGGAA-  1190
SIVAGM692  -AT-G----T----------AATCAAATGCAGGGGGTC--C---G-A--GCAA-CAGGAATAGGAGGTA-  1160
SIVAGM677  -GC-A----T----------AGTAAT......GGACAG--T---G-C--AGTG-GACCTCAGAAAAAGG-  1154

CONSENSUS  accaaGAccc???CCactaAaaTGtTacAAttTGTGGaAAATTTGGcCAtaTGCAaAgacAaTGccctgaa  1182
SIVAGMTYO  ACAAA--CCC...--ACTA-GA--T-AT--T-----A--------C--TA----A-GAC-A--TCCGGAA  1242
SIVAGM155  ACCAA--CCC...--GCCA-AG--T-AC--C-----A--------C--CA----G-GGC-G--TCCTGAG  1245
SIVAGM3    ACCAA--CCC...--AGTA-AG--T-AC--T-----A--------C--TA----A-GAC-A--CCCTGAA  1257
SIVAGM692  AGGGA--GGAAGA--AGTT-AA--C-AC--A-----A--------G--TG----G-AAA-T--CACTCAA  1230
SIVAGM677  CCCCC--GGG...--GCTA-AA--C-TT--T-----C--------A--TA----A-GGG-A--CAAGGCA  1221

CONSENSUS  ccAaGaaaaataaaATGccT?AAaTGtGGaAAacc?GG?CAttTaGCaaaAGAcTGCAgg???GGAcagG  1246
SIVAGMTYO  CC-A-GAAAACAAA---TC-A--G--T--A--ATTG--A--CC-A--AAA---C----GG...---CAG-  1309
SIVAGM155  CC-A-AAAAATAAA---TC-T--A--T--A--GCCA--G--CT-A--AAA---C----GG...---CAG-  1312
SIVAGM3    CC-A-AAAGATGAG---CT-G--A--T--G--ACCA--G--TT-A--AAA---T----GA...---CAG-  1324
SIVAGM692  AA-G-GCCAGTAGT---CC-G--A--T--A--ACCT--C--TT-T--TCG---T----GA...---GCA-  1297
SIVAGM677  CC-A-ACAGATCAA---CT-T--G--C--C--AATT--C--TA-G--AAA---C----AGAAT---CAG-  1291

      pol orf ->
CONSENSUS  taAATTTTTTAGGGTATGGccggTGGatgGGagCaAAACCaAgAAATTTT?????????cccGcaGc?ac  1306
SIVAGMTYO  TG----------------ACGG---ATG--GG-A-----G-G-------.........CCC-CC-CTAC  1370
SIVAGM155  TG----------------CCGG---ATG--GA-A-----A-G-------.........CCC-CA-CCAC  1373
SIVAGM3    TA----------------CCGG---ATG--AG-G-----C-G-------.........CCC-CC-CTAC  1385
SIVAGM692  TA----------------CAGG---ATG--AG-A-----A-A-------TTAGAACACAGA-CA-CA..  1365
SIVAGM677  CA----------------CCAT---GGA--AG-G-----A-G-------GTGCAATACAGA-GA-ACAC  1361

CONSENSUS  tcTtgg?g?ggaaCCaactGCgCCcCCtccaccga?ca?c?c?ac?cC?TACGAcccaGCaAagaaGCTc  1368
SIVAGMTYO  TC-TGGAGCGGAA--GAGT--G--T--TCCACCGAGCGGCACCACCC-A-----CCCA--A-AGAA---C  1440
SIVAGM155  TC-TGGGGCGGAA--AAGT--G--C--TCCACCGAACAACTCTACAC-T-----CCCA--A-AGAA---C  1443
SIVAGM3    TC-TGGGGTGGAG--AACT--G--C--TCCACCGAGT.........C-A-----CCCT--A-AGAA---C  1446
SIVAGM692  .G-C.........--CTCC--C--T--ACCGCCGCACAACCCA?GGG-G-----CGAA--C-CTCG---T  1425
SIVAGM677  AG-TGGTCTGGAA--AACA--C--C--A.........ATGGAAACAG-T-----TCCA--A-AGAA---C  1422

CONSENSUS  CTgcAGcAaTATgCagAgaAgGGgaa?CaactgAGgga?cAaa??aaaaA?ccacc?gca?agaatccgg  1431
SIVAGMTYO  --GC--C-A---G-AG-GA-A--GAAA-AACTG--GGAGC-AAAGAGGA-TCCACCGGCAATGAATCCGG  1510
SIVAGM155  --GC--C-G---G-AG-GA-A--GAAA-AAATG--AAATC-GAACAGAA-CCCCCCAGCGAACAATCCAG  1513
SIVAGM3    --GC--C-A---G-AG-CA-G--GAAG-AGTTG--GGAAC-AAGGAAAA-ACCACCAGCAGTGAATCCCG  1516
SIVAGM692  --GG--A-A---A-CC-AG-G--AGCC-AACAA--GAGAA-AGTAGAGA-GAGC...TCCCAAGCGGGGA  1492
SIVAGM677  --CC--C-G---G-AG-GA-G--ACAG-GCCTG--AGAGG-GAGAGAAC-GACAAGGAAACAGAAGGAGA  1492

CONSENSUS  attgga??GAGGgataTTCtTTGAacTCCCTCTTTGGAGaAGACCAA***    1479
SIVAGMTYO  ATTGGACC----GATA---T----AC-------------A----------  1560
SIVAGM155  ATTGGAAC----GATA---T----AC-------------A----------  1563
SIVAGM3    ATTGGACA----GATA---T----AC-------------A----------  1566
SIVAGM692  GGGAGGAA----ATTA---C----AA-------------A----------  1542
SIVAGM677  AAGAAGTG----ATGT---C----GC-------------G----------  1542
```

AGM POL

AGM POL CDS

```
 gag cds ->
CONSENSUS  ?TG?GG?AAA?TTGGCCATATG??AAGACAATGTCCGGAACCAAG?AA?A??A?ATG??T?AAaTGTGGa   23
SIVAGMTYO  T--T--A---T-----------CA--------------------G--A-CA-A---TC-A--G-----A   70
SIVAGM155                                                              A-----A     7
SIVAGM3                                          ------A--G-TG-G---CT-G--A-----G   31
SIVAGM677  G--C--C---A-----------GC.....................................           24

                                                 pol orf ->
CONSENSUS  AAaccaGGgCActTAGCAAAAGAcTGCAg???gGGACAGGt?AATTTTTTAGGGTATGGcCggTGGatgG   89
SIVAGMTYO  --ATTG--A--CC----------C----G...G-------TG----------------A-GG---ATG-  137
SIVAGM155  --GCCA--G--CT----------C----G...G-------TG----------------C-GG---ATG-   74
SIVAGM3    --ACCA--G--TT----------T----G...A-------TA----------------C-GG---ATG-   98
SIVAGM677  ..................------C----AGAAT-------CA----------------C-AT---GGA-   77

CONSENSUS  G?gC?AAACCaAGAAATTTTcccgc?gc?ActcttGgggcgGaaccaagtGcgCCccCtcCaCCgagca?  154
SIVAGMTYO  -GG-A-----G---------CCCGCCGCT-CTCTT-GAGCG-AACCGAGT-CG--TC-TC-A--GAGCGG  207
SIVAGM155  -GA-A-----A---------CCCGCAGCC-CTCTT-GGGCG-AACCAAGT-CG--CC-TC-A--GAACAA  144
SIVAGM3    -AG-G-----C---------CCCGCCGCT-CTCTT-GGGTG-AGCCAACT-CG--CC-TC-A--GAG...  165
SIVAGM677  -AG-G-----A---------GTGCAATAC-GAGGA-ACACA-TTGGTCTG-AA--AA-AG-C--CCCAAT  147

CONSENSUS  c?c?acacC?TACGAcCCaGCAAAGAAGCTCCTgCAGCA?TATGCAGAgAA?GGgaA?CaacTGAG?gAg  217
SIVAGMTYO  CACCACCC-A-----C--A--------------G-----A--------G--A--GA-A-AAC----GG-G  277
SIVAGM155  CTCTACAC-T-----C--A--------------G-----G--------G--A--GA-A-AAA----AA-T  214
SIVAGM3    .....TC-A-----C--T--------------G-----A--------C--G--GA-C AGT----GG-A  229
SIVAGM677  GGAAACAG-T-----T--A--------------C-----G--------G--G--AC-G-GCC----AG-G  217

CONSENSUS  cA?A?ga?aaA?cCacc?gcaa?gAAtccggAttggaccGAGGgataTTCtTTGAaCTCCCTCTTTGGAG  281
SIVAGMTYO  C-A-AGAGGA-TC-ACCGGCAATG--TCCGG-TTGGACC----GATA---T----A--------------  347
SIVAGM155  C-G-ACAGAA-CC-CCCAGCGAAC--TCCAG-TTGGAAC----GATA---T----A--------------  284
SIVAGM3    C-A-GGAAAA-AC-ACCAGCAGTG--TCCCG-TTGGACA----GATA---T----A--------------  299
SIVAGM677  G-G-GAGAAC-GA-AAGGAAACAG--GGAGA-AGAAGTG----ATGT---C----G--------------  287

                 <- gag
CONSENSUS  aAGACCAATaAa?acagtt??cATAGaAGG???ccCc?Ttaa?GCatT?tTAGAtACaGGgGCAGATGAc  342
SIVAGMTYO  A--------A-AGACAGTGTAT----A---GGTCC-CA-TAAG--AC-GC----C--A--G--------C  417
SIVAGM155  A--------A-GGACCTGTATA----G---AACTG-CG-TAAG--AT-AT----T--A--G--------C  354
SIVAGM3    A--------A-AAACAGTTTAC----A---GGTCC-CA-CAGA--AT-AT----T--G--G--------T  369
SIVAGM677  G--------G-AACGAGTCATC----A---AACGC-AG-GCAA--CT-GT----T--A--A--------C  357

CONSENSUS  AC?ATaAT?aAaGAaa?aGAtTTaCAaTTatcagga??????cCATGGagacC?AAaaTa?TAGG?GGcA  400
SIVAGMTYO  --C--A--TA-A--AAAT--T--A--A--ATCAGGT......C-----AGAC-C--AA-TA----G--C-  481
SIVAGM155  --T--A--AA-G--TACA--T--A--A--AAGGGGA......T-----AGAC-A--AA-AG----A--A-  418
SIVAGM3    --C--T--AA-A--AGCA--T--A--A--ATCAGGA......A-----AAAC-A--AA-AA----G--C-  433
SIVAGM677  --T--A--TC-A--AAAG--C--G--C--TCCCCCACATAAAC-----CGTT-C--GG-AG ---A--T-  427

CONSENSUS  T?GGaGGaGGgcTtaAtGTaAAAGAaTATaatgat???gAAGTA?aAtTgGAaGA?AAaATttTgAgaGG  464
SIVAGMTYO  -A--A--A--CC-TA-T--A-----A---AACGACAGGG-----AA-A-A--A--T--A--TT-G-GA--  551
SIVAGM155  -T--G--A--GT-AA-C--A-----A---GATAATGTAG-----CA-T-G--A--C--G--AT-A-GA--  488
SIVAGM3    -T--A--G--AC-CA-T--A-----G---AGTGATAGGG-----AG-T-G--A--C--A--TT-G-GA--  503
SIVAGM677  -A--A--A--GA-TC-T--C-----A---CAGGGGGTAC-----CA-T-G--G--T--A--CA-C-CC--  497

CONSENSUS  aaCaaTa?TgaTAGGA??aACtCCcATcAAtATaATaGGaAGAAAt?TttTaGCcC?gGCAGgagccaaa  529
SIVAGMTYO  AA-AA-AT-GT-----GCA--T--C--T--T--A--A--T-----TT-GC-G--C-CG----TGCCCCGG  621
SIVAGM155  AA-AG-CC-CA-----GCA--T--C--C--T--C--A--A-----CT-TT-A--C-AG----GAGCCAAA  558
SIVAGM3    GA-CA-AT-GA-----AGC--T--C--A--C--A--T--A-----TA-AT-A--A-CA----GAGCCAAA  573
SIVAGM677  CT-AA-TC-AA-----AGT--A--A--C--T--T--A--A-----TA-TT-A--T-AG----GCATGAAA  567
```

```
CONSENSUS  TTAGTaATGGGacaa?T?TCagaa?aaATtcc??tcAC?cctGTa?aatTaAA?GAAGGGgctagaGGAC  591
SIVAGMTYO  -----A-----ACAAT-A--AGAAAAA--TCCTGTC--ACCT--CAAAT-G--G------GCTCGG----  691
SIVAGM155  -----G-----GCAAT-G--GCAGACA--ACCAATC--CCCG--ACGCT-A--G------GCCAGA----  628
SIVAGM3    -----A-----TCAAC-G--AGAACAA--TCCCATT--CCCT--GAAAT-A--A------GCTAGA----  643
SIVAGM677  -----T-----AGTTC-A--TAGTCAG--TGAGGAA--AAAA--ACAAC-A--A------AAAGAT----  637

CONSENSUS  CttgatT?AaaCAATGGCCccTcTCtAaAGAaAAaAT??aAGC?tTaca?GAAATaTG?aa?caAtTaGA  654
SIVAGMTYO  -CTGTG-A-GA--------TC-C--T-A---G--G--TGA---TT-ACAG-----A--TTCCCA-T-A--  761
SIVAGM155  -ACGAT-G-AG--------AC-C--T-A---A--A--AAT---CC-GCAA-----T--CAAAAC-T-A--  690
SIVAGM3    -TTTCT-A-AA--------CC-C--C-A---A--A--AAA---CT-ACAG-----A--TGACCA-T-A--  713
SIVAGM677  -TAAAT-G-AA--------CT-A--A-G---A--A--TGA---TT-AACA-----A--CAAACA-A-G--  707

CONSENSUS  ggA?GAaGGAAAA?Taag?Agg?TAGGaGGAGAaAATgCaTAcAAtACaCCAgT?TTttgcATAAaGAAa  719
SIVAGMTYO  GC-G--A------A-CAGT-GAG----A-----A---G-A--C--T--C---A-A--TTGC----A---G  831
SIVAGM155  GG-A--A------T-AAGC-GGG----G-----C---G-A--C--T--A---G-A--CTGT----G---A  768
SIVAGM3    GA-A--A------A-TAGC-AGA----A-----G---G-A--C--C--T---G-G--TTGC----A---A  783
SIVAGM677  AG-G--G------T-ATCT-GGA----A-----A---C-T--T--T--A---G-G--TGCC----A---A  777

CONSENSUS  AA?GAcAAatCaCA?TGGAGaATGcTaGTAGAtTT?AGgGAacTaAAcAAaGC?ACaCAAGA?TT?TTtG  783
SIVAGMTYO  --G--C--AT-C--G-----G---C-A-----C--T--A--GT-A--T--G--A--C-----T--C--T-  901
SIVAGM155  --A--C--AT-A--G-----A---C-G-----T--C--G--AC-C--C--A--T--A-----C--C--T-  838
SIVAGM3    --A--C--GT-A--A-----A---T-A-----T--T--G--AC-A--C--A--A--A-----T--T--C-  853
SIVAGM677  --G--T--AA-A--A-----A---C-T-----T--C--G--AC-A--C--A--T--T-----C--T--T-  847

CONSENSUS  AaGT?CA?tTaGG?ATaCC?CA?CCAgCaGGgtTaaa?AA?AtGAagcAaATaACagT??TAGA??TaGG  842
SIVAGMTYO  -A--G--AT-A--G--A--C--C---G-A--AT-AAGA--G-T--GAC-G--A--AG-TT----TG-A--  971
SIVAGM155  -A--C--AT-A--T--A--C--T---G-A--GT-AAAG--A-T--AGC-A--A--CA-TA----TG-G--  908
SIVAGM3    -A--A--GT-A--C--A--T--T---T-A--GT-CGAA--G-T--CGG-A--A--AG-AT----CA-A--  923
SIVAGM677  -G--T--GC-G--A--T--T--C---G-G--CC-TCAG--A-A--AGC-A--C--AG-AA----CA-A--  917

CONSENSUS  gGAtGCcTATTATtc?ATACCAtT?ga?ccagAgTTtAGaAAaTAtACaGCtTTtACcAT?CCt?CaGTa  907
SIVAGMTYO  A--C--C------TCC------T-GGATCCAA-T--T--G--A--T--T--T--T--T--T--TA-A--G  1041
SIVAGM155  G--T--A------AGC------C-GGATCCTG-G--T--A--A--C--A--T--C--C--C--TA-G--A  978
SIVAGM3    G--T--C------TCA------T-AGACCCAG-G--T--A--G--T--C--T--T--C--T--AT-A--A  993
SIVAGM677  G--T--C------TCA------T-ATGCAAGG-A--C--A--A--T--A--A--T--C--C--CT-A--A  987

CONSENSUS  AAtAATca?GG?CCaGGgAt?AG?TATCA?TTcAA?TGtCT?CCgCA?GGaTGGAA?GGATC?CCtACaA  967
SIVAGMTYO  --T---CAG--A--C--G-TT--G-----A--C--C--T--C--G--A--G-----A-----T--T--A-  1111
SIVAGM155  --C---GAG--A--A--C-TA--A-----A--T--T--C--A--G--G--C-----G-----C--G--A-  1048
SIVAGM3    --T---CAA--G--A--T-CT--A-----G--C--C--T--T--A--A--A-----G-----C--A--T-  1063
SIVAGM677  --T---ACA--G--A--G-TA--G-----G--C--T--T--G--T--G--A-----A-----T--T--A-  1057

CONSENSUS  TtTTcCA?AA?ACaGCAqCa??cATT?TaGA?GA?ATaAAaA?/gActta??a?cgcTaaccATTGTcCA  1025
SIVAGMTYO  -C--C--A--T--A---G-ATCC---T-G--G--G--A--A-GAA-CTTGCCAGCAC-AACC-----A--  1181
SIVAGM155  -T--C--A--C--A---T-AAAA---C-A--A--A--A--G-AAG-ATTAAAACAGC-GACG-----C--  1118
SIVAGM3    -T--T--G--C--A---G-TTCC---C-A--A--A--A--A-AGG-GTTAAAACCCC-AACC-----G--  1133
SIVAGM677  -T--C--G--T--G---G-AAAC---T-A--G--G--C--A-GGC-CACTCCTGGGT-AGAA-----C--  127

CONSENSUS  aTACATGGAtGA??TaTGGqTaGggTC?cAaqAaqAtGaac?cAcqCATgAtca??t?GTAqAa?aacTa  1087
SIVAGMTYO  A--------T--TT-A---G-A-GT--TC-AG-AA-T-AACAC-CC---G-CAAATTA---G-ACAGT-A  1251
SIVAGM155  G--------T--CC-C---G-A-GA--AC-AG-AG-G-GTCCA-AG---G-TCAGCTA---C-AACAC-T  1188
SIVAGM3    A--------T--CC-A---G-A-GG--TC-GG-AG-T-AATAC-CG---G-TCGGTTG---G-ACAAC-A  1203
SIVAGM677  A--------C--TT-G---T-G-CG--AG-CC-TG-T-AGACT-GA---A-TCAACAG---G-CATAG-A  1197

CONSENSUS  AGaAaqAaatT?caaG??tqqGGatTAGAAAC?CCAGAcAAqAA?qTqCAaA?AqAACCaCCtTatGAqT  1151
SIVAGMTYO  --A-CA-AAT-ACAA-CCTGG--CT-------C-----A--G--GA-G--A-A-G----A--T-AT--G-  1321
SIVAGM155  --G-AT-GAT-GCAA-AATGG--AT-------A-----G--A--GG-G--A-G-G----T--C-TT--G-  1258
SIVAGM3    --A-TG-AAT-AAGT-CCTGG--AT-------A-----C--G--AG-A--G-A-A----A--T-AT--G-  1273
SIVAGM677  --A-AG-TGC-GCTA-AAAAA--TC-------C-----C--G--AG-C--A-G-G----G--A-GG--A-  1267
```

```
CONSENSUS  GGATGGGaTA?AAAtTgtggCCtcA?AAaTGGaagaTaa?cAgaATAgAAtTagaggac???aAAGaagA  1215
SIVAGMTYO  -------A--C---C-TTGG--TC-C--A---GAAC-AAGC-GA---C--C-GGAGGAA...A---ATG-  1388
SIVAGM155  -------A--T---T-ATGG--TC-T--A---AAGT-ACAA-GT---G--T-AGAGAAG...A---AAC-  1325
SIVAGM3    -------A--C---T-GTGG--AC-C--G---CAGA-AAGC-GC---G--T-AGAAGAC...A---AAG-  1340
SIVAGM677  -------G--T---T-GCAT--GA-T--A---ACCA-TAAC-AA---G--T-ACCCCCCTTAG---GAG-  1337

CONSENSUS  ATGGAC?GTaAAtgAtaTaCAgAa?tTaGTaGGgaaaacTAAATTGGGCAgcaCAatTtTAtCCAGGacTt  1283
SIVAGMTYO  ------T--C--1G-CA-C--G-AGT-A--T--GAAAC----------GCA--AT-G--T-----TC-T  1458
SIVAGM155  ------A--G--TG-TC-T--G-AAT-G--A--GAAAT----------GCA--AT-A--T-----AT-G  1395
SIVAGM3    ------T--A--TG-TA-A--A-GAC-A--G--GAAAC----------GCA--GC-T--C-----AC-C  1410
SIVAGM677  ------A--A--CA-AA-A--G-AGG-A--A--AGTTC----------AGT--AA-T--T-----AA-T  1407

CONSENSUS  A?aACaAaAaatatcTGtaagtTaaTtAgA???GGaAA?AA?AA??Ta?TAGAagtagTAgcaTGGACa?  1343
SIVAGMTYO  -AG--A-G-...ATA--CAAGT-AA-T-C-GGA--A--G--A--TC-GT----GCTAG--GCT-----AC  1525
SIVAGM155  -GA--A-A-AATATC--TAAGC-AC-T-G-...--A--G--A--TT-AT----CGTGG--GAA-----CC  1462
SIVAGM3    -GA--T-A-AACTTG--TAAAT-AA-C-G-...--A--A--G--CT-AC----AACAG--ACC-----AG  1477
SIVAGM677  -AA--C-A-CATACC--TGCCA-GT-G-G-...--G--A--G--CC-CC----AGAAA--GTA-----AG  1474

CONSENSUS  ?aGAgGCAGAaGCaGAaTAtg?agAaAA?aaaGaGATccTaaAaacagaaCA?GAAGGaACcTA?TA??a  1406
SIVAGMTYO  CT--G-----A--T--A--TGCAG-A--TGCA-A---TC-TA-AACAGAA--G-----A--C--T--CAA  1595
SIVAGM155  CA--G-----A--A--G--CGAAG-A--CAAG-A---CC-AA-AACAGAG--A-----T--T--T--TGC  1532
SIVAGM3    AG--A-----A--A--A--TGCAG-A--CAAA-A---CT-AA-AACGGAA--G-----G--C--C--CAA  1547
SIVAGM677  AA--G-----G--A--A--TAAGA-C--TCAA-G---AG-GC-GGAAACA--A-----A--A--C--TGA  1544

CONSENSUS  aCCaggaAaAcccaTtAg?GCAgCAGTaCA?AAa?taG?AG?aGGgCAaTGGa??TACCAATTCAagCAa  1469
SIVAGMTYO  A--AGGA-T-CCTA-T-GG---G----A--G--ATTG-A--GA--A--G---AGT----------AA--A  1665
SIVAGM155  A--AGAA-A-CCCC-T-GG---G----A--G--ATTA-G--AT--G--A---TCA----------AG--G  1602
SIVAGM3    A--AGGA-G-CCCA-C-GA---G----G--A--ACTA-A--GA--T--A---AGT----------AG--A  1617
SIVAGM677  C--TCTC-A-GAAT-A-TA---A----T--A--GCAA-G--AA--G--A---ACA----------CC--A  1614

CONSENSUS  GAaGGacaAgT?TTaAA?GTaGGaAaaTac?CcAAgCA?Aaa?acACtCAtACcAATGAg?T?cGt??a?  1529
SIVAGMTYO  --A--ACA-G-C--G--A--A--A-AA-ACA-C--G--A-AGAAC--C--T--A-----AC-TC-CACAT  1735
SIVAGM155  --A--AAA-A-C--A--G--A--G-AG-TCG-C--A--G-AAGCT--T--C--C-----GT-GC-TGTAC  1672
SIVAGM3    --G--ACA-G-A--A--A--A--T-AA-ACA-A--G--G-AAAAC--T--T--C-----GT-CC-TGTAT  1687
SIVAGM677  --A--GCC-G-A--A--G--G--A-GA-ATG-C--G--A-GAGAA--T--T--T-----TC-AA-GACTC  1684

CONSENSUS  TaGCaggatTaGTaCA?AAaaTttGtAAaGAagCtcTagttATtTGGGGaaaatTaCCagttcT?gAACT  1597
SIVAGMTYO  -A--TGGTT-A--G--G--GA-TT-C--A--AG-TC-AGTT--T-----GATAT-A--AGTTC-AG----  1805
SIVAGM155  -A--AGGAG-A--A--G--AA-AG-G--A--GG-CC-AGTA--T-----ACAAT-A--CACTT-TG----  1742
SIVAGM3    -G--AGGAT-A--A--A--AC-TT-T--A--AT-TT-AGTT--A-----AGAGT-G--AGTCC-TG----  1757
SIVAGM677  -A--ACACC-T--C--A--AA-CT-T--G--AG-AC-TACC--T-----AAGAC-T--ACGAG-AC----  1754

CONSENSUS  CCCa?TaGAqAqqqAa??ATGGGAacaaTGGTGGqcgGA?TA?TGGCA?GTaaq?TGGAT?CC?GAaTGG  1658
SIVAGMTYO  ---GA-A--A-GAG-GGT------GCAA------GCG--T--C-----G--AAGC-----T--C--A---  1875
SIVAGM155  ---AG-G--G-GGG-CAC------ACAA------GCA--C--T-----A--CAGT-----A--C--A---  1812
SIVAGM3    ---AA-A--G-GGG-AGT------ACAA------GCT--T--C-----G--AAGT-----T--A--C---  1827
SIVAGM677  ---AG-A--C-AGA-AAC------TATG------CAG--C--T-----A--ATCC-----A--A--A---  1824

CONSENSUS  GA?TTTGTcAG?accCCaCcc?TaqTaAaa?T?TGGTAtaC??T?acaAAaGAACCcAT?cca???GA?G  1715
SIVAGMTYO  --T-----C--CACC--A-CTT-GC-C-AAC-A-----CA-AT-AACA--A-----C--ACCCAAG--G-  1945
SIVAGM155  --C-----C--TGTT--G-CCT-AG-A-CTT-G-----TA-AC-GACT--G-----C--CCCGGGA--G-  1882
SIVAGM3    --A-----C--TACC--A-CCC-AG-A-AAT-A-----TA-CC-GACA--A-----C--ACCAAAG--A-  1897
SIVAGM677  --G-----T--CACA--A-TCC-AG-A-AAC-G-----TT-CT-AGTA--A-----A--CAAAGGA--A-  1894

CONSENSUS  AtGT?TAcTATGT?qatGGaGCatqtaa?AqAaa?tCaAaaqaaGGaAAaGCAGGaTA?aTc?CacAata  1779
SIVAGMTYO  -C--T--C-----A...--A--ATGCAAC-G-AATT-A-AAGAA--A--A-----A--CA-CT-AC-ATA  2012
SIVAGM155  -T--C--?-----AGAT--A--CTGTAAT-G-CAGT-G-AAGAG--A--A-----C--CA-AA-CC-ACA  1952
SIVAGM3    -T--C--C-----GGAT--A--TTGTAAT-G-AATT-A-GGGAA--A--A-----A--TA-CA-AC-ATA  1967
SIVAGM677  -T--T--T---GGAT--G--AGCATCC-A-GTGA-C-AATTA--T--G-----A--TC-GT-AG-GAG  1964
```

```
CONSENSUS  aGGaAAAcaaAGagTagaaaaatTAGAAAAcACaACaAA?CA?CAAGCagAAtTaAcaGCcaTaAAaATG  1847
SIVAGMTYO  C--A---CAG--AG-AGAAACAT-------C--T--C--T--G-----AA--T-A-CA--TA-A--A---  2082
SIVAGM155  A--C---CAA--AG-ACAACAGC-------C--A--A--T--A-----TG--C-G-CA--CA-A--A---  2022
SIVAGM3    T--G---CAA--GG-GGAAAAAT-------T--A--A--C--G-----AG--T-A-TG--CA-A--A---  2037
SIVAGM677  A--A---AGT--AA-TAGGGAAT-------C--C--T--C--A-----AG--T-A-CA--AG-T--G---  2034

CONSENSUS  GCatTgGA?GA?AGtGGgcctAAtGTaAA?ATAGTaACAGAtTC?CAATATGcaATGgg?ATaTTgACaG  1912
SIVAGMTYO  --TT-G--A--C--T--GCCT--T--G--C-----A-----C--T-------CA---GGA--T--G--A-  2152
SIVAGM155  --CT-G--G--T--C--CCCT--A--C--T-----A-----T--A-------CG---GGC--A--G--A-  2092
SIVAGM3    --AC-A--A--T--T--GCCT--T--A--C-----A-----T--A-------CA---GGA--A--A--T-  2107
SIVAGM677  --AT-G--G--C--T--AGAA--T--A--T-----C-----T--T-------TA---AAC--C--G--A-  2104

CONSENSUS  CacaaCCcacacA?AGTgAcTCaCC?tTAgTaGAaCAaAT?ATAgcactaaTg?TacAaAA?caacagaT  1977
SIVAGMTYO  -ACAA--CACAC-A---G-T--A--AT--G-A--G--A--T---GCCTTAA-GA-AC-A--GCAACAAA-  2222
SIVAGM155  -ACAG--CACAC-G---G-C--C--AC--G-A--A--A--A---GCACAGA-GG-AC-G--AGAAGCCA-  2162
SIVAGM3    -CCAA--CACAC-G---G-C--A--CT--A-A--A--A--T---GCACTAA-GG-AC-A--ACATCAGA-  2177
SIVAGM677  -ATGT--ACAGG-A---A-C--A--CT--G-G--A--G--A---CAAGCCC-AA-GA-A--GAGGCAGG-  2174

CONSENSUS  ?TA?tTgCAaTGGGTACCaGC?cAtAAaGGqATAGGaGGcAATgaaGA?ATAGAtAAAtTAGTaAG?aAa  2042
SIVAGMTYO  A--TT-G--G--------A--AC-T--A--A-----A--A---GAG--G-----T---T----G--TA-A  2292
SIVAGM155  C--TC-G--A--------T--TC-T--A--T-----G--C---GAA--A-----C---T----A--CA-G  2232
SIVAGM3    A--CT-G--A--------A--AG-C--A--G-----A--C---GAA--G-----T---C----A--TC-A  2247
SIVAGM677  C--CT-A--A--------A--TC-T--G--G-----A--C---ACA--A-----T---T----A--CA-A  2244

CONSENSUS  GGaaTtAGaa?aaTttTaTTctTaGaaA?aATAGAAGAAGCaCAaGAagAaCATGaaAggTA?CAtAatA  2109
SIVAGMTYO  --CA-T--AAGAG-TT-A--CT-A-AA-AA----------T--A--AA-G----AA-GA--T--T-AT-  2362
SIVAGM155  --AG-T--AAGAA-AT-G--CA-T-GC-GG----------A--A--AG-A----AT-GG--T--C-GT-  2302
SIVAGM3    --GA-G--GAAAA-TT-A--TT-A-AA-AA----------C--G--GG-A----AA-GG--C--T-AT-  2317
SIVAGM677  --AA-A--ACAGA-CC-C--CT-A-AT-GA----------A--A--TG-C----CA-AG--C--T-AC-  2314

CONSENSUS  AtTGGAgaAa?cTaGcAgAcacaT?TGG??TaCCacAaATaGTaGCaAAAGA?ATaGTaGCcatgTGtCC  2174
SIVAGMTYO  -T----AA-ACC-A-C-G-TACA-A---GC-T--AC-A--A--A--A-----G--A--G--CATG--T--  2432
SIVAGM155  -C----GA-ATC-A-C-G-CACA-T---AT-G--AC-A--A--A--T-----A--T--A--AATG--C--  2372
SIVAGM3    -T----GG-ACT-A-C-G-CACT-A---GC-A--AC-A--T--G--A-----A--A--A--CATG--T--  2387
SIVAGM677  -T----GA-GTA-G-T-C-GGAA-T---AT-A--TA-T--A--A--A-----G--A--A--GGCA--T--  2384

CONSENSUS  aAAATGtCAaaTAAaaGG?GAACCa?tgCATGGaCAaGTaGAtGCcTCacc?Gga??aTGGCAgATGGAc  2239
SIVAGMTYO  A-----T--AA---AG--A-----AGTG-----A--A--G--T--C--ACCT-GAACA-----G-----T  2502
SIVAGM155  A-----T--AG---AA--G-----AATA-----A--A--A--T--T--ACCA-GAGTG-----G-----C  2442
SIVAGM3    A-----T--GA---AA--G-----AGTC-----G--A--A--T--C--GCCA-GGGTA-----A-----C  2457
SIVAGM677  C-----C--AA---GA--A-----TAAG-----A--G--A--C--C--CATT-AAACT-----G-----C  2454

CONSENSUS  TG?ACaCATtTAGAAgqaAAaqTa?TcATAGTaGC?GTcCATGTAGCCAGTGGaTTcATAGAAGCAGA?G  2305
SIVAGMTYO  --T--T---C-----AAA--AG-AG-C-----T--G--C--------------A--C----------A-  2572
SIVAGM155  --C--A---A-----GGA--AA-AG-G-----A--G--C--------------G--T----------G-  2512
SIVAGM3    --T--A---T-----GGC--GG-AA-C-----A--A--C--------------A--C----------A-  2527
SIVAGM677  --C--C---T-----GGA--AG-TA-A-----A--A--A--------------A--C----------G-  2524

CONSENSUS  TtAT?CCtAG?GAAACaGG?AAaGA?ACaGCAaAqTTt?T?tTaAAaaTacTa?g?AGATGGCCcaTaAc  2367
SIVAGMTYO  -C--A--T--G-----A--A--A--A--G---A-G--TC-AT-A--AA-AC-GAGT--------TA-A-C  2642
SIVAGM155  -T--C--T--G-----A--A--A--G--A---A-G--CT-GT-A--AA-AA-AGGA--------CA-C-C  2582
SIVAGM3    -T--A--T--A-----A--G--A--A--A---A-A--TT-AT-A--GA-AC-AAGT--------CA-A-C  2597
SIVAGM677  -G--C--A--A-----T--G--G--G--A---C-C--TC-GC-G--AC-GT-AGCA--------AG-G-A  2594

CONSENSUS  aCA?cTaCA?ACaGA?AATGGaCCaAA?TTtACcTCtCA?qAaGT?GCaGCaaTgTG?TGGTGGGGaAAa  2430
SIVAGMTYO  A--GT-A--C--A--C-----G--T--C--T--C--C--AG-A--G--A--AA-A--T--------A--A  2717
SIVAGM155  T--CC-C--T--A--T-----A--A--T--C--T--T--GG-A--A--T--TA-G--C--------A--G  2652
SIVAGM3    C--AC-G--T--A--C-----A--C--T--T--G--T--AG-A--A--A--AA-G--T--------A--A  2667
SIVAGM677  A--TC-A--C--T--T-----C--A--C--T--C--T--GA-T--G--A--GG-G--C--------T--T  2664
```

```
CONSENSUS  aTaGAaCAcAC?ACaGGt?TACCaTATAAcCCaCA?TCaCA?GGatctaTAGAaAG?ATGAAcAaaCA?t   2494
SIVAGMTYO  A-T--A--T--A--A--TA----A-----C--C--A--T--A--ATCAA----A--C-----C-AG--AT  2782
SIVAGM155  G-A--A--C--A--G--GG----A-----T--A--G--C--G--ATCTA----A--T-----C-AA--AT  2722
SIVAGM3    A-A--A--C--C--A--TG----C-----C--T--A--A--A--CTCTA----G--T-----T-AA--GT  2737
SIVAGM677  A-A--G--C--C--T--AA----T-----C--A--G--A--G--TAGTG----A--C-----C-GA--GC  2734

CONSENSUS  TaAAaGA?ATaATtgggaAAATAAGAGATGA?TGtcAatatacaGA?aCAGCAGTaCttATGGCttgcca  2561
SIVAGMTYO  -A--A--G--A--TGGGA-------------T--CC-ATATACA--GG-------A-TG-----TTGCAT  2852
SIVAGM155  -G--A--G--A--TGGAA-------------C--TC-ATATACA--AA-------A-TT-----CTGCCA  2792
SIVAGM3    -A--A--A--A--TGGGA-------------C--TC-ATACACA--AA-------A-TT-----ATGCCA  2807
SIVAGM677  -C--G--A--C--CTCTC-------------T--TG-GAGATTG--GA-------G-AA-----TACGCA  2804

CONSENSUS  caT?CACAATTTTAAAAGAAAGGGAGGAATAGGGGCgctaaC??ctGCAGAgAGA?TaaTtAATATgaTa  2627
SIVAGMTYO  AC-T-------------------------------ACAGA-TTCA-----G---C-AA-T-----AA-A  2922
SIVAGM155  CA-T-------------------------------GCTAA-AGCT-----G---C-AA-A-----GA-A  2862
SIVAGM3    CA-C-------------------------------GTTAA-ACCG-----G---T-AA-C-----GA-T  2877
SIVAGM677  TA-C-------------------------------TATCT-TAGT-----A---T-GG-T-----GC-A  2874

CONSENSUS  ACaACACAAtTAGAAaTa?Aa??tcTACAAAcCaAAAT?CAAAAAATTTT?AATTTTAgaGTCTACTACA  2692
SIVAGMTYO  --A------T-----A-AC-ACATT------C-A----T----------A-------GA----------  2992
SIVAGM155  --A------T-----A-CA-CACTC------C-A----C----------G-------GA----------  2932
SIVAGM3    --T------T-----T-AC-ACACC------C-A----T----------A-------GA----------  2947
SIVAGM677  --A------C-----C-AA-TACTC------A-C----C----------G-------AG----------  2944

CONSENSUS  GAGAAGGgAGAGAtCC?GTgTGGAAaGGACCaGcaC?a?TaATcTGGAAAGG?GAAGGcGC?GTgGTc?T  2756
SIVAGMTYO  -------G-----C--T--G-----A-----G-CA-AAT-A--C----------G-----A--A--G--CC-  3062
SIVAGM155  -------C-----T--A--G-----G-----T-CT-GCC-G--C--------A-----C--G--A--TC-  3002
SIVAGM3    -------G-----T--T--C-----A-----A-GA-AGT-A--T--------G-----T--A--G--CA-  3017
SIVAGM677  -------T-----T--A--G-----A-----A-CG-GAC-C--C--------A-----C--G--G--AA-  3014

                                                          vif cds ->
CONSENSUS  cAA?GaaGGtgaaGA?cTaAAgGtagtcCCaAGaAGgAAaGC?AAaATtATaAA?GAtTATGaaccaAgA  2822
SIVAGMTYO  C--G-AC--AAGT--CC-A--G-TTGTA--A--A--G--A--T--A--T--T--G--T----AACCC-A-  3132
SIVAGM155  C--G-AA--TGAA--AC-G--G-TAGTT--G--A--G--A--A--A--C--A--A--C----AGCCA-G-  3072
SIVAGM3    C--A-GA--TGTG--AT-A--A-AATAC--A--A--G--A--A--A--T--A--G--T----AACCA-G-  3087
SIVAGM677  T--A-AG--GGAA--CA-C--G-TAGTC--C--G--A--G--T--G--T--C--A--T----GAGAG-G-  3084

CONSENSUS  aAAA?AaTGGgT?aTGAG?gTaaa?TGGAaGGTgcc?GaGgagctgATaAa?A?ATGG??GGGGATAGT?  2870
SIVAGMTYO  C---G-G---G-AA----GG-GACG----A---ACCA-G-GATCTG--A-C***                  3186
SIVAGM155  A---C-T---G-GA----AC-CACC----A---GCAG-A-GAAGTG--C-CC-A----CA---------*** 3144
SIVAGM3    A---G-A---G-GA----AG-AACT----A---GCCG-A-GAGCTG--A-C***                  3141
SIVAGM677  A---C-A---A-AG----GG-AGTA----G---GTCA-A-AGGCAA--A-GC-G----AG---------G > 3174
```

AGM VIF CDS

```
 pol cds ->
CONSENSUS  ATGaaccaAgAaAAA?AaTGGgT?aTGAG?gTaac?TGGAaGGTgcc?GaGgagctgATaAc?AaATGGc     64
SIVAGMTYO  ---AACCC-A-C---G-G---G-AA----GG-GACG----A---ACCA-G-GATCTG--A-CT-A----C     70
SIVAGM155  ---AGCCA-G-A---C-T---G-GA----AC-CACC----A---GCAG-A-GAAGTG--C-CC-A----C     70
SIVAGM3    ---AACCA-G-A---G-A---G-GA----AG-AACT----A---GCCG-A-GAGCTG--A-CT-A----C     70
SIVAGM677  ---GAGAG-G-A---C-A---A-AG----GG-AGTA----G---GTCA-A-AGGCAA--A-GC-G----A     70

CONSENSUS  agGGgATAGTgAgaTA?tgGATgAggaa?AgaaA?tTgaaaTGG?AaTA?a?aatgCAtTacCAaaTtac    120
SIVAGMTYO  AG--A-----C-GA--TTG---G-GACAA-GAA-TT-GAAA---A-C--TTATATG--T-AC--AA-TAC    140
SIVAGM155  AG--G-----T-GA--TTG---G-ATAAA-GGA-TC-GAAA---G-A--CAAAATG--T-AT--AA-CAC    140
SIVAGM3    AA--G-----G-GG--CTG---G-GGACT-GAA-AT-AGAC---A-A--TCGAATG--C-AC--AA-TAC    140
SIVAGM677  GG--G-----G-CT--CAA---C-GGAAT-AAC-AT-GCCT---G-A--CAGACAT--T-GG--GG-GCA    140

CONSENSUS  aTGGgcaTggTacACcatgAG?agaTatgtaATaCCccTag?aaaAcaTGgagaaATacatGT?gAt?Ta    194
SIVAGMTYO  A---GCT-GG-AC--CATG--TAGA-ATGTA--A--AA-AGGAAA-CA--GGGAA--ATGT--AG-CC-A    210
SIVAGM155  T---GCA-GG-AC--TATG--CAGA-ATGTA--A--CC-CCCAGG-AG--GAGAA--CCAT--GG-TA-C    210
SIVAGM3    A---GCA-GG-AC--AATG--TAGA-ATGAG--A--CC-AGGGCA-CA--GAAGT--ACAT--AG-TC-A    210
SIVAGM677  A---CAG-TT-GG--CTAC--CCAG-TCATT--C--CT-ATCAAA-GA--ATTAC--AGAA--GA-TA-T    210

CONSENSUS  TATtgqcAt?TaaCaCCAgAa?aaGGATGGCT?TCaAc?tATGcaGtaGGgaTAcaaTATgtaagc?a?a    258
SIVAGMTYO  ---TGGC-TT-AA-A---G-GCAA---------A--C-CAT---CA-TA--TA--CAA---GTAAGCAATT    280
SIVAGM155  ---TGGC-TT-AG-T---A-ACAA---------C--A-CTT---CA-TA--AA--CAA---GTTAGCCTAG    280
SIVAGM3    ---TGGC-TC-GA-A---G-AAAG---------A--A-CAT---CT-AG--GA--CAG---CTAAGCAATA    280
SIVAGM677  ---CACA-CC-CA-C---G-AAGA---------C--A-GTC---GA-TA--GT--TCC---TACCATCAAA    280

CONSENSUS  t?ga??ctaaaTATAg?ACAGAAtTaGAtCCtgc?ACAGCAGA?agcaTgATACAt?gtcAtTATTTTA?    320
SIVAGMTYO  TAGAATCTAAA----GA------T-A--C--TGCT--------TAGTA-A-----TGGTC-C-------A    350
SIVAGM155  TAAATGATAAA----GA------T-A--T--CAAT--------CTCCA-G-----TTGTC-T-------C    350
SIVAGM3    GGGATCCTTGG----GG------T-G--T--TGCA--------TAGCC-G-----TACCC-T-------C    350
SIVAGM677  AGGGA......----AG------G-A--T--AGGA--------CAGAA-G-----CCTAT-T-------A    344

CONSENSUS  ?TGTTTTAcAGA?AGaGCCATCcaacA?GCtcTaaggGGacA?AggTt?a?cTtcTGtca?TT?ccaGaa    382
SIVAGMTYO  T-------A---A--A------CAAC-A--TC-GAGG--CC-C-GA-TTGTC-TC--TCAG--TCCA-AA    420
SIVAGM155  C-------C---T--A------CAAC-G--AC-AAGG--AA-C-GG-TCATC-TC--TCAA--TCCA-GA    420
SIVAGM3    T-------C---A--G------AGGA-A--CC-ATTG--AC-G-GG-TCACC-TC--TCAG--CCCC-AG    420
SIVAGM677  C-------C---T--A------CAAC-G--TA-CAGA--GG-G-AG-ATACG-GG--CACA--CAAG-AA    414

                                                                    vpX cds ->
CONSENSUS  GGaCAtAAaa?aacaGG?CAGGTACc?TCttTGCAata?tTaGC?cTacTtGCacAtcaAAATGGCcTCA    447
SIVAGMTYO  --G--T--AAGCACA--A-------CA--TT----GTACC-A--TC-GC-C--AC-TCA-------C---    490
SIVAGM155  --A--T--ACTAACA--T-------CC--CT----ATATT-A--AT-AC-A--CC-TCA-------C---    490
SIVAGM3    --A--C--GAAAACA--A-------CC--TT----ATACT-A--TC-CC-T--AC-CCA-------C---    490
SIVAGM677  --A--T--A.......--T-------AA--AC----ACTTT-G--AC-AG-T--AT-TAC-------A---    478

CONSENSUS  GGaAGAGATCC?AGAGA?qcaagACCaGGaq?acTaGAAATtTGGGATcTaaqCAGGGAGCCqTGGGACq    514
SIVAGMTYO  --G--------A-----GGCAAG---A--AGAAG-A-----T------C-AAG---------G------A    560
SIVAGM155  --A--------C-----GGAGAG---A--AGGAC-A-----T------C-CAG---------G------G    560
SIVAGM3    --C--------C-----AGCAAG---G--GGAAC-A-----A------T-GAG---------G------G    560
SIVAGM677  --A--------A-----ACCTTT---A--ATGGC-G-----C------C-AGA---------A------G    548

CONSENSUS  AATGGCTaaqAGA?ATGcTaqaaGAt?TcAAccAqGAaGCcA???tGCAtTTTGGq?qaqAacTcCT?tt    577
SIVAGMTYO  -------AAG---T---T-ACAA--TC-C--CC-G--G--G-GGCT---T-----GAGAG-AC-C--GTT    610
SIVAGM155  -------CAG---T---G-ACAA--AA-C--CA-C--A--C-AACT---T-----CCGAG-AC-C--ATA    610
SIVAGM3    -------AAG---C---C-AGAA--TA-C--TC-G--A--C-AGAT---T-----GCGCG-GC-C--GTT    610
SIVAGM677  -------ACA---C---C-CAGG--TC-A--CG-A--A--C-GAAG---C-----AATGA-CA-G--AAT    610
```

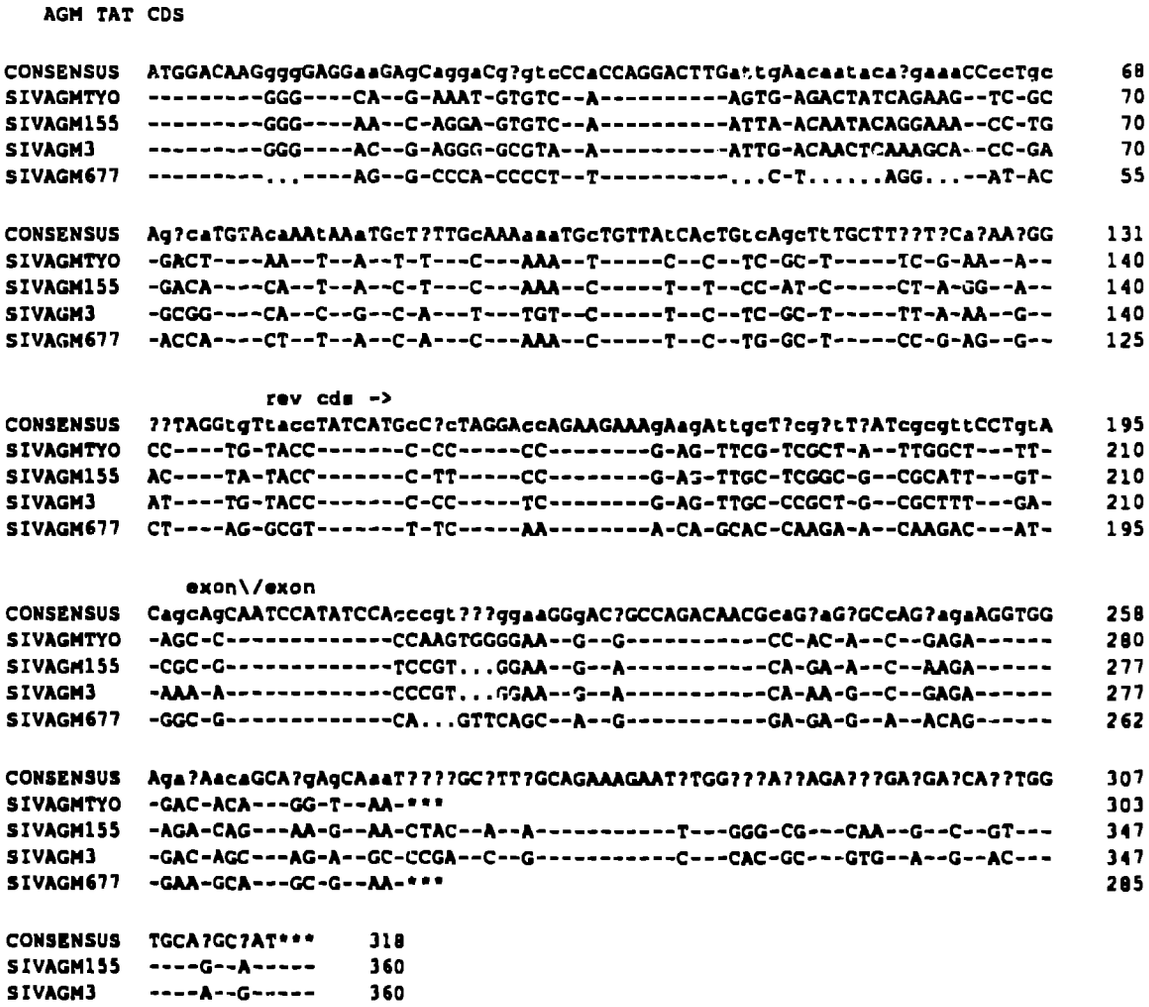
AGM VIF

```
CONSENSUS  CCaAGTATGGAA?TA?TGTcagGAGGA?GGaga?AGacAtgG?ActCCCATg?tgGAAAGGGC?TA???      638
SIVAGMTYO  --A---------C--C---CAG-----A--AGAA--AC-TG-T-CT-----GATG--------C--CAA >   708
SIVAGM155  --A---------T--T---CAG-----A--GGAG--AC-GG-A-GA-----AGCG--------A--***     699
SIVAGM3    --A---------C--T---CAG-----G--AGAA--GA-TC-C-CT-----GCTA--------T--***     699
SIVAGM677  --G---------T--C---GTA-----G--AAGG--AC-***                                660
```

AGM VPX CDS

```
 vif cds ->
CONSENSUS  ATGGCcTCAGGaAGAGATCC?AGAGA?gcaagACCaGGag?acTaG?AATttGGGATcTaagCAGGGAGC     67
SIVAGMTYO  -----C-----G--------A-----GGCAAG---A--AGAAG-A-----T------C-AAG--------     70
SIVAGM155  -----C-----A--------C-----GGAGAG---A--AGGAC-A- ---T------C-CAG--------     70
SIVAGM3    -----C-----C--------C-----AGCAAG---G--GGAAC-A-----A------T-GAG--------     70
SIVAGM677  -----A-----A--------A-----ACCTTT---A--ATGGC-G-----C------C-AGA--------     70

CONSENSUS  CgTGGGACgAATGGCTaagAGA?ATGcTagaaGAt?TcAAccAgGAaGCcA???tGCAtTTTGGg?gagA    131
SIVAGMTYO  -G------A-------AAG---T---T-ACAA--TC-C--CC-G--G--G-GGCT---T-----GAGAG-    140
SIVAGM155  -G------G-------CAG---T---G-AGAA--AA-C--CA-C--A--C-AACT---T-----CCGAG-    140
SIVAGM3    -G------G-------AAG---C---C-AGAA--TA-C--TC-G--A--C-AGAT---T-----GCGCG-    140
SIVAGM677  -A------G-------ACA---C---C-CAGG--TC-A--CG-A--A--C-GAAG---C-----AATGA-    140

CONSENSUS  acTcCT?ttCCaAGTATGGAA?TA?TGTcagGAGGA?GGaga?AGacAtggtActCCcatgatgGAaAgg    196
SIVAGMTYO  AC-C--GTT--A---------C--C---CAG-----A--AGAA--AC-TGGT-CT--CATGATG--A-GG    210
SIVAGM155  AC-C--ATA--A---------T--T---CAG-----A--GGAG--AC-GGGA-GA--CATAGCG--A-GG    210
SIVAGM3    GC-C--GTT--A---------C--T---CAG-----G--AGAA--GA-TCGC-CT--CATGCTA--A-GG    210
SIVAGM677  CA-G--AAT--G---------T--C---GTA-----G--AAGG--AC-TAAT-CC--ATGGAAT--G-TA    210

                     <- vif
CONSENSUS  GccTA?AA?TAtTATagatTaGTtCAaAA?gCTcTcTTTGTgCATTTccG?TGTGGaTG??GcAG?AGac    259
SIVAGMTYO  -CC--C--A--T---AGGC-A--A--A--GG--C-C-----G-----TC-A-----G--CA-G--A--GC    280
SIVAGM155  -CA--T--G--T---CGCT-A--T--G--AG--C-C-----G-----CC-G-----A--TC-C--G--AC    280
SIVAGM3    -CT--T--A--T---AAAT-G--G--A--AG--C-C-----G-----CC-G-----A--CC-C--A--AC    280
SIVAGM677  -GC--C--G--C---AGAA-T--T--A--GT--A-G-----A-----CA-A-----T--TA-A--G--AG    280

                              tat cds ->
CONSENSUS  aaCCcTTTgaaCCaTACGA?GAgAGGAGa?ATGGACAAGGgGGAGGaagagCaggaCg?gtcCCaCCAGG    326
SIVAGMTYO  AG--C---GAA--A-----G--G-----AG----------G-----CAGAG-AAAT-GTGTC--A-----    350
SIVAGM155  AA--C---GAG--A-----G--G-----AA----------G-----AAGAC-AGGA-GTGTC--A-----    350
SIVAGM3    AA--C---GAA--A-----A--A-----GG----------G-----ACGAG-AGGG-GCGTA--A-----    350
SIVAGM677  GA--T---TCC--T-----A--G-----AA----------A-----A...G-CCCA-CCCCT--T-----    347

CONSENSUS  ACTTGat***    336
SIVAGMTYO  -----AG---    360
SIVAGM155  -----AT---    360
SIVAGM3    -----AT---    360
SIVAGM677  -----CA---    357
```

AGM TAT

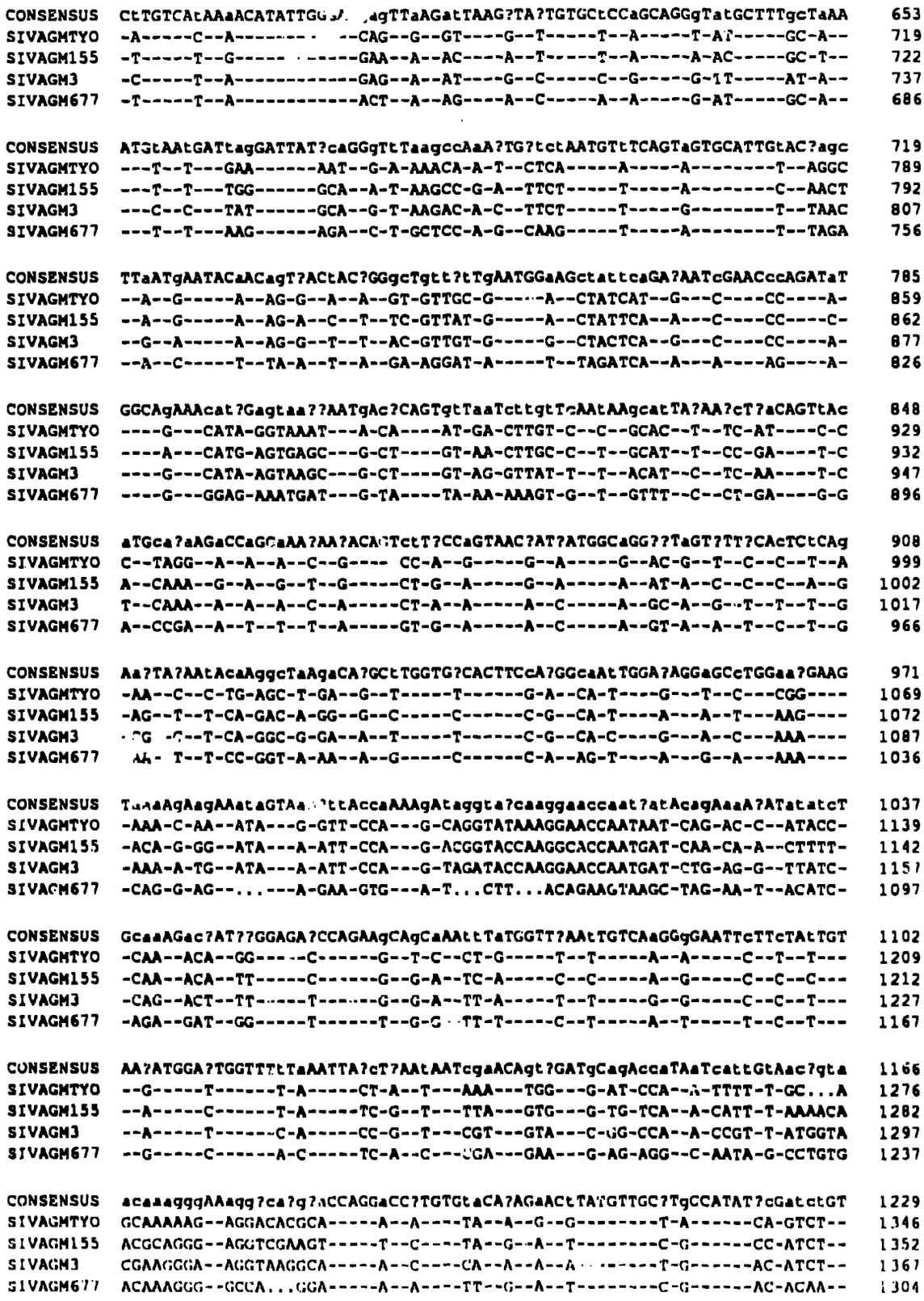

AGM TAT CDS
CONSENSUS ATGGACAAGgggGAGGaaGAgCaggaCg?gtcCCaCCAGGACTTGattgAacaataca?gaaaCCccTgc 68
SIVAGMTYO ---------GGG----CA--G-AAAT-GTGTC--A----------AGTG-AGACTATCAGAAG--TC-GC 70
SIVAGM155 ---------GGG----AA--C-AGGA-GTGTC--A----------ATTA-ACAATACAGGAAA--CC-TG 70
SIVAGM3 ---------GGG----AC--G-AGGG-GCGTA--A----------ATTG-ACAACTCAAAGCA--CC-GA 70
SIVAGM677 ---------...----AG--G-CCCA-CCCCT--T----------...C-T......AGG...--AT-AC 55
CONSENSUS Ag?caTGTAcaAAtAAaTGcT?TTGcAAAaaaTGcTGTTAtCAcTGtcAgcTtTGCTT??T?Ca?AA?GG 131
SIVAGMTYO -GACT----AA--T--A--T-T---C---AAA--T-----C--C--TC-GC-T-----TC-G-AA--A-- 140
SIVAGM155 -GACA----CA--T--A--C-T---C---AAA--C-----T--T--CC-AT-C-----CT-A-GG--A-- 140
SIVAGM3 -GCGG----CA--C--G--C-A---T---TGT--C-----T--C--TC-GC-T-----TT-A-AA--G-- 140
SIVAGM677 -ACCA----CT--T--A--C-A---C---AAA--C-----T--C--TG-GC-T-----CC-G-AG--G-- 125
rev cds ->
CONSENSUS ??TAGGtgTtaccTATCATGcC?cTAGGAccAGAAGAAAgAagAttgcT?cg?tT?ATcgcgttCCTgtA 195
SIVAGMTYO CC----TG-TACC-------C-CC-----CC--------G-AG-TTCG-TCGCT-A--TTGGCT---TT- 210
SIVAGM155 AC----TA-TACC-------C-TT-----CC--------G-AG-TTGC-TCGGC-G--CGCATT---GT- 210
SIVAGM3 AT----TG-TACC-------C-CC-----TC--------G-AG-TTGC-CCGCT-G--CGCTTT---GA- 210
SIVAGM677 CT----AG-GCGT-------T-TC-----AA--------A-CA-GCAC-CAAGA-A--CAAGAC---AT- 195
exon\/exon
CONSENSUS CagcAgCAATCCATATCCAcccgt???ggaaGGgAC?GCCAGACAACGcaG?aG?GCcAG?agaAGGTGG 258
SIVAGMTYO -AGC-C----------CCAAGTGGGGAA--G--G----------CC-AC-A--C--GAGA------ 280
SIVAGM155 -CGC-G----------TCCGT...GGAA--G--A----------CA-GA-A--C--AAGA------ 277
SIVAGM3 -AAA-A----------CCCGT...GGAA--G--A----------CA-AA-G--C--GAGA------ 277
SIVAGM677 -GGC-G----------CA...GTTCAGC--A--G----------GA-GA-G--A--ACAG------ 262
CONSENSUS Aga?AacaGCA?gAgCAaaT????GC?TT?GCAGAAAGAAT?TGG???A??AGA???GA?GA?CA??TGG 307
SIVAGMTYO -GAC-ACA---GG-T--AA-*** 303
SIVAGM155 -AGA-CAG---AA-G--AA-CTAC--A--A----------T---GGG-CG---CAA--G--C--GT--- 347
SIVAGM3 -GAC-AGC---AG-A--GC-CCGA--C--G----------C---CAC-GC---GTG--A--G--AC--- 347
SIVAGM677 -GAA-GCA---GC-G--AA-*** 285
CONSENSUS TGCA?GC?AT*** 318
SIVAGM155 ----G--A----- 360
SIVAGM3 ----A--G----- 360

AGM REV CDS

```
 tat cds ->                                                                exon\/exon
CONSENSUS  ATGcC?cTAGGAccAGAAGAAAgAagAttgcT?cg?tT?ATcgcgttCCTgtACagcAgCAATCCATATC      66
SIVAGMTYO  ---C-CC-----CC--------G-AG-TTCG-TCGCT-A--TTGGCT---TT--AGC-C-----------      70
SIVAGM155  ---C-TT-----CC--------G-AG-TTGC-TCGGC-G--CGCATT---GT--CGC-G-----------      70
SIVAGM3    ---C-CC-----TC--------G-AG-TTGC-CCGCT-G--CGCTTT---GA--AAA-A-----------      70
SIVAGM677  ---T-TC-----AA--------A-CA-GCAC-CAAGA-A--CAAGAC---AT--GGC-G-------  ---      70

CONSENSUS  CAcccgtg???gaaGGgAC?GCCAGACAACGcaG?aG?GCcAG?agaAGGTGGAga?AccaGCA?gAgCA     127
SIVAGMTYO  --CCAAGTGGGGAA--G--G-----------CC-AC-A--C--GAGA-------GAC-ACA---GG-T--     140
SIVAGM155  --TCCGTG...GAA--G--A-----------CA-GA-A--C--AAGA-------AGA-CAG---AA-G--     137
SIVAGM3    --CCCGTG...GAA--G--A-----------CA-AA-G--C--GAGA-------GAC-AGC---AG-A--     137
SIVAGM677  --CAGTTCAGC...--A--G-----------GA-GA-G--A--ACAG-------GAA-GCA---GC-G--     137

CONSENSUS  aaT???agcctTgGcAGaAAGAatc????acaccagagaggAagAgCA?ctgGTGcatgcaaTTgAccaa     189
SIVAGMTYO  AA-TAGAGTCT-G-T--A----...CTC............C-AG-G--G...---TATGCAG--G-CCGC     192
SIVAGM155  AA-CTACGCAT-A-C--A----ATTTGGGGGACGAGACAAG-GG-C--GTTG---CAGGCAA--G-CCAA     207
SIVAGM3    GC-CCGAGCCT-G-C--A----ATCTGGCACAGCAGAGTGG-AG-G--ACTG---CAAGCGA--G-CCAA     207
SIVAGM677  AA-TGATAAGA-T-C--G----GTCCTCAACACCTTCGAGG-TC-A--ACTG---GCTCAAC--C-AGAG     207

CONSENSUS  ?TGgt?ct?GAga?tcAgcA??TGG?tatACAACAg?TGCCTGACCCTCCTcaTtcagcT???           243
SIVAGMTYO  C--GCTGAC--GGCTC-AC-CT---CTAT------GT--------------CA-TCAGC-***           255
SIVAGM155  T--GTTCTT--CACTC-GC-TC---TTAC------GC--------------TC-CAAGC-***           270
SIVAGM3    T--GTGCTC--TCAGC-GC-TC---CTAT------GT--------------AG-TCATC-***           270
SIVAGM677  C--CAGCTT--GAATA-GG-CT---TGTT------TC--------------CA-ATTCA-CAGGACAGC >   441
```

AGM ENV CDS

```
CONSENSUS  ATGaggaag??atTaataa???Tag?aaTAaTAgtaATAGGgATAgGaaTAG?gaTaagtacaaagca?C    62
SIVAGMTYO  ---AGGTATACAA-AATAACCT-AGGAA--A--GTG-----A---G-GA---TGT-AAGT...AAG...-    64
SIVAGM155  ---ACAAAGTTCT-AGGAATTT-TATAG--T--GGA-----G---G-AA---GGA-AAGTACAAAACAG-    70
SIVAGM3    ---AAGCTGACAT-ACTG...A-AGGGA--C--TTA-----G---G-AG---TGC-TAATACAAGGCAA-    67
SIVAGM677  ---GGGAGATTGC-TATAAAAA-ACTAA--A--GCA-----G---A-TA---GAA-A...GGT...AAC-    64

CONSENSUS  a?Tgg?TaACAGTgTTtTATGGA?TaCCAGTATGGAAAAAcagctCaGT?CAgGCtTT?TGCATGACaCC   127
SIVAGMTYO  AA-GGA-A-----C--T------A-A--------------CAGCT-G--G--G--T--C--------T--   134
SIVAGM155  AG-GGA-A-----G--C------G-A--------------CAGCT-A--C--A--T--T--------A--   140
SIVAGM3    AA-GGG-C-----A--T------G-A--------------CAGCT-A--A--G--T--C--------A--   137
SIVAGM677  TG-ATG-G-----G--T------A-C--------------TTCAA-A--T--G--A--T--------G--   134

CONSENSUS  cAc?ACcAg?tT?TGGGCaACtACtAA?TgcATACCAGATGATCATGActAtACAGA?GTaCC?cTaAA?   190
SIVAGMTYO  C-CC--A-GCT-A-----T--T--T--T-GC-----------------CT-T-----A--A--TC-A--T   204
SIVAGM155  T-CT--T-GGT-G-----A--T--T--T-GC-----------------CT-T-----A--A--AC-G--T   210
SIVAGM3    C-CC--C-GAC-A-----A--T--T--C-CG-----------------CT-C-----G--A--AT-A--C   207
SIVAGM677  C-AT--C-ATA-G-----A--C--C--C-GC-----------------TA-T-----G--G--TC-A--C   204

CONSENSUS  ATcAC?GAacCaTTtGA?GCaTGGgctGA?agaAA?CCctTagTAgcACAAGCAGgaAGTAA?ATaCA?c   254
SIVAGMTYO  --C--T--AC-A--T--G--A---GGT--TAGA--C--AT-AA--GC--------CC-----C--C--TT   274
SIVAGM155  --A--A--GC-A--T--A--A---GCA--CAGA--T--CT-AG--GC--------GA-----C--T--CC   280
SIVAGM3    --C--T--AC-A--T--A--A---GCT--CAGA--C--CT-AG--GC--------GA-----T--A--CC   277
SIVAGM677  --T--A--AG-T--C--G--T---...--T...--T--GC-GG--AA--------AG-----T--A--TC   268

CONSENSUS  T?CTcTTTGA?CA?ACtaTgAagCCtTGTGTaAAg?TaTCaCCtcTaTG?AT?AAaATG???TGTGTAGA   315
SIVAGMTYO  -A--C-----G--A--TA-G-AA--T-----G--GT-A--A--AC-A--C--C--G---AAT--------   344
SIVAGM155  -G--G-----A--G--AT-A-AG--C-----A--GC-A--A--TC-A--T--C--A---AAT--------   350
SIVAGM3    -G--A-----G--G--TC-G-AG--A-----A--AT-A--A--TT-G--C--T--A---TCC--------   347
SIVAGM677  -A--C-----A--A--GA-G-GG--T-----T--GC-C--C--CA-A--T--T--A---TCC--------   338

CONSENSUS  gtT?AAt??c?Cagca???????????????cctacaaccacgcc?aaatctACc???acggca?caaca   361
SIVAGMTYO  GT-A--TTCCA-AAGAGAAAGGGCGACAACACCTACAACGACGCCGAAATCT--C.............   399
SIVAGM155  GT-A--AGGCT-CGCA......ACCTCTACCCCAGCAACC.........TCT--T...ACGGCAGGAACC   402
SIVAGM3    AT-G--CTCCT-TGAG..............CCTACCACCACTCCTAAAAGT--C...ACGGCCTCAACA   399
SIVAGM677  GC-G--TGGTA-AGCC.....................ACGACAAAGGCC--CACTACTGCAACTACA   384

CONSENSUS  aca???????????????????????accctaCCcTGTgt??g?aa?aagacaggt?c?aa?ctacAgtcaT   403
SIVAGMTYO  .....................GGCCTA--C---GTAGGCCCGACGTCAGGTGAAAATCTAC-GTCC-   445
SIVAGM155  AAA.....................CTA--C---GTTAGAAATAAAACAGACTCCAACCTAC-GTCA-   448
SIVAGM3    ACCAATATCACAGCCTCAACAACCACTTTG--G---GTCCAGAACAAGACAAGTACTGTGTTAG-ATCA-   469
SIVAGM677  ACA...ATGACT...........ACC...--C---.........................C-GAAT-   409

CONSENSUS  G?Aatg?aaccatcATAGAAa?gGA?aTg?atGA?GAgcCaGC?TCtAA?TG?ACaTTtGCaATgGC?GG   463
SIVAGMTYO  -T-ATGCAAGCATT------AGG--GA-GGAG--T--GC-C--C--T--T--T--A--C--A--G--T--   515
SIVAGM155  -C-ACGACACCATC------AAG--GA-GAAT--C--GG-A--G--A--C--C--C--T--T--G--T--   518
SIVAGM3    -T-ATGAAACAATC------AAG--AT-AAAT--A--GC-T--T--T--T--T--A--T--A--G--A--   539
SIVAGM677  -C-GTACAGAGCAG------GGA--AA-GGCA--G--AC-A--A--C--C--C--T--T--A--T--A--   479

CONSENSUS  gTAtgtaAGaGAtca?AA?AAgAAt???TAttcagTggtgTGGaATGATgcaGAaaTctttTG?AAg???   524
SIVAGMTYO  C--TGTA--A--TCAG--G--A--TTAT--TTCTG-GGTG---A-----GCA--AA-CTAT--C--AAAT   585
SIVAGM155  G--CATT--G--CCAA--G--G--T...--CTCAG-AGTA---A-----GCA--AA-CTTT--T--G...   582
SIVAGM3    G--TGTA--A--TCAG--A--G--G...--TTCAG-GGTG---A-----GCA--AA-CATG--T--G...   603
SIVAGM677  A--TCAA--A--TGTA--A--G--T...--TAGCA-GACC---T-----CAG--GT-AGTC--C--TAAT   546

CONSENSUS  aaga?taaaagt??????g?tAgcAaaGAgTGtTA?ATGATtCAtTGTAATGAtTCAGTtATAAAaGAAG   585
SIVAGMTYO  AAGACTAATAGC......ACT-GC-AA--G--T--C-----T--T--------C-----T-----A----   649
SIVAGM155  CGTAGTACATCGCATAATGGG-CA-AA--G--C--T-----C--C--------T-----T-----G----   652
SIVAGM3    AAGGGTAACAAT......TCT-AC-GA--A--T--T-----T--T--------T-----T-----A----   667
SIVAGM677  AAAACAGGAACTGAAAAGGGA-GT-AG--T--T--C-----A--T--------T-----G-----A----   616
```



```
CONSENSUS  ?aTaAATGATTggTA?ACtcTaTCaAagAAaacaTATGCtCC?CCAAGaGAAGGaCAtTTGcA?TGcAcg  1295
SIVAGMTYO  AA-A-------CC--T--CC-A--A-AG--AACT-----T--G-----A-----A--T---C-A--C-GG  1416
SIVAGM155  CA-A-------GG--T--TA-A--A-AG--AACA-----T--A-----A-----A--T---C-G--C-CG  1422
SIVAGM3    CA-T-------GG--C--AC-A--A-GG--AACC-----A--G-----A-----G--C---C-A--C-CA  1437
SIVAGM677  AG-A-------GG--C--TG-C--T-AA--GGTA-----T--A-----G-----T--T---G-G--T-AC  1374

CONSENSUS  TCcaCaGTcAC?GggaTg?caGTgGagcTaaAtTAtAAtA??aAgaacaGgaCa???AAtGT?ACa?TaA  1356
SIVAGMTYO  --CA-A--C--T-GGA-GACA--T-AGC-TA-T--T--T-GTA-AAACA-AA-C...--C--G--AC-A-  1483
SIVAGM155  --CA-A--T--T-GGA-GACA--A-AGC-AA-C--T--T-ACC-GAACA-GA-A...--T--A--AT-G-  1489
SIVAGM3    --CA-G--A--G-GTA-GTCA--G-AGC-AA-T--C--T-GTA-GAACA-GA-T...--T--A--AT-A-  1504
SIVAGM677  --AT-A--C--G-CAC-ATAC--G-CAA-AG-T--T--C-ACA-GTCTG-CC-AATA--T--G--CC-A-  1444

CONSENSUS  GTCCcCAGaTAgaaaccATcTGGGCggcaGAAtTGGGcagaTAcAAATTaGTaGA?ATtACACCAATTGG  1425
SIVAGMTYO  ----C---A--GAATCT--C-----GGCT---T----CAGA--C-----A--G--A--C----------  1553
SIVAGM155  ----C---A--GAAACC--C-----GGCA---T----CAGA--C-----G--A--G--T----------  1559
SIVAGM3    ----C---A--GAAACC--C-----AGCA---T----CAGG--C-----A--A--A--T----------  1574
SIVAGM677  ----T---G--CGCAGC--A-----GTAC---C----AGAC--T-----A--A--G--A----------  1514

CONSENSUS  cTTtGCaCCcACAGAaGTaAG??GaTAtACgGGaggcca?gagAG?caAAA??GaGTcCCgTTcGTGCTA  1489
SIVAGMTYO  C--T--A--C-----A--A--GC-T--T--G--AGGACATGAG--GCA---GA-A--C--G--T------  1623
SIVAGM155  A--T--A--C-----A--C--GC-A--C--G--AGGCCAAGAG--GCA---AC-A--C--G--C------  1629
SIVAGM3    C--C--A--C-----A--A--AA-G--T--G--AGGTCATGAC--AAC---GC-A--C--G--C------  1644
SIVAGM677  C--T--T--T-----T--A--AA-A--T--T--CCCCACA...--AGA---AA-G--G--A--C------  1581

CONSENSUS  GGGTTccTAGGCTTCTTgGGaGCTGCTGG?ACTGCAATGGGaGCAGCGGCgAca?ccCTGACgGTCCAGT  1557
SIVAGMTYO  -----CT----------G--G--------A-----------A--------G-GCAGC-----G-------  1693
SIVAGM155  -----CC----------G--A--------G-----------A--------G-CAGCC-----G-------  1699
SIVAGM3    -----CC----------A--A--------G-----------A--------G-CAGCC-----G-------  1714
SIVAGM677  -----TC----------G--A--------A-----------C--------A-CAACG-----A-------  1651

CONSENSUS  CtC?GCATTT?CTTGCTGGGATA?TGCAGCAGCAGAAGAAtcTGCTGGCGGCTGTGGa?gctCAACAgCA  1623
SIVAGMTYO  -C-G------G-----------C----------------TC---------------AGGCT-----G--  1763
SIVAGM155  -T-A------A-----------T----------------TC---------------GAGCT-----G--  1769
SIVAGM3    -T-A------A-----------C----------------TC---------------AGGCT-----G--  1784
SIVAGM677  -T-G------G-----------T----------------CT---------------AACAG-----A--  1721

CONSENSUS  GaTGTTGAAGCTGACCATTTGGGGTGTqAAAAACCTCAATGCCCGCGTCACAGCtCTtGAGAAGTACCTa  1693
SIVAGMTYO  -A-------------------------T-------------------------C--T----------A  1833
SIVAGM155  -A-------------------------G-------------------------T--T----------G  1839
SIVAGM3    -A-------------------------G-------------------------T--T----------A  1854
SIVAGM677  -T-------------------------G-------------------------T--C----------A  1791

CONSENSUS  GaGGAtCAGGCaCGq?TaAA??CtTGGGGqTGcGC?TGGAAaCAAGTaTGTCAtAC?ACaGT?ccgTGG?  1756
SIVAGMTYO  -A---T-----A--AC-A--CT-C-----G--C--A-----A-----A-----T--C--A--GGAG---C  1903
SIVAGM155  -C---T-----A--GT-A--CG-T-----G--C--G-----A-----A-----T--A--A--ACCC---A  1909
SIVAGM3    -A---C-----G--GT-G--TG-T-----G--C--A-----G-----C-----T--A--C--ACCG---C  1924
SIVAGM677  -A---T-----A--GC-A--TT-A-----A--T--G-----A-----A-----C--C--A--GCCA---A  1861

CONSENSUS  ?gTggAatAAt???ACtCCtgA?TGGaAtAATATGACtTGGtTGGAgTGGGAaAgACA?ATag?tgatTT  1819
SIVAGMTYO  CC-GG-CA--TCGG--T--GG-T---C-A--------T---T----G-----A-G---A--AGCTGAT--  1973
SIVAGM155  CG-GG-AT--T...--A--AG-G---A-T--------C---T----G-----A-A---G--AGAAGGA--  1976
SIVAGM3    AG-GG-AT--TAGG--C--TG-T---A-T--------T---C----A-----A-G---G--ATCGTAT--  1994
SIVAGM677  AG-AT-AT--C...--T--TA-G---G-C--------T---T----G-----G-G---A--TAATGCC--  1928

CONSENSUS  GGAaqGcAACATaACaaaACaATT?Ga?qA?GCaaqaqaaCAqGAqqaaAAGAAT?TgGATqc?TA?CA?  1882
SIVAGMTYO  ---AA-C-----T--GGG--A---A-TGA-G--TAGAGAA--G--GGAA------C-A---GCC--T--G  2041
SIVAGM155  ---GG-C-----A--AAA--A---G-AAC-G--AAGGGAA--A--GGAA------T-G---GCT--T--A  2046
SIVAGM3    ---AG-T-----A--AAC--A---A-AGG-A--CAGAGCA--G--GGAG------T-G---GCA--C--A  2064
SIVAGM677  ---AG-C-----A--TCA--T---G-AAG-A--ACAAAAT--G--ATCA------C-G---CTG--C--G  1998
```

```
CONSENSUS  AA?TT?act??tTGGTCag?tTTcTGGTCtTGGTTCgattTcTCAAaaTGGcTgaaCatTtTaAAaATaG  1947
SIVAGMTYO  --G--AACTAGT-----AGAT--C-----T------GATT-C----AA---C-TAA-AT-T-A--A--G-  2113
SIVAGM155  --G--GTCAGAC-----GAGT--T-----T------GATT-T----AA---C-GAA-AT-T-A--G--A-  2116
SIVAGM3    --A--AAGTAGT-----AGAT--C-----T------GATT-C----AG---C-GAA-AT-C-A--A--A-  2134
SIVAGM677  --A--GGATGAT-----AGGG--C-----A------TCAC-G----CT---T-AGG-TA-G-T--A--A-  2068

CONSENSUS  GaTTTTT?Gta?TAgTaggtaTtaTAGG?TTAAGATTgcttTacacA?TATatg??TG?ATagctAgggT  2010
SIVAGMTYO  -A-----A-TAA--G-AGGAA-AA----G--------ACTT-ACAC-G---ATGGA--T--AGTG-GGG-  2183
SIVAGM155  -C-----G-CAG--A-AGGCG-TA----G--------GCTT-ACAC-T---ATACT--C--AGCT-GGG-  2186
SIVAGM3    -A-----G-ATG--C-AGGTA-TA----A--------GCTT-ATAC-G---ATTCT--C--AGCT-GGG-  2204
SIVAGM677  -A-----A-TGA--G-GATTA-TC----A--------TGCA-GGGT-T---GGGGA--T--CAGA-ATA-  2138

CONSENSUS  TAGGCAGGG?TA?tcTCCTcTatctCCaCAGATCCATATCCAcccgtgg???aaGGgAC?GCCAGACAAC  2074
SIVAGMTYO  ---------A--TGT----C-ATCT--A--------------CCAAGTGGGGAA--C -G----------  2253
SIVAGM155  ---------T--CTC----T-ATCT--T--------------TCCGTGG...AA--G--A----------  2253
SIVAGM3    ---------T--CTC----C-TTCT--A--------------CCCGTGG...AA--G--A----------  2271
SIVAGM677  ---------A--TAA----C-C...--C--------------CAGT...TCAGC--A--G----------  2202

CONSENSUS  GcaG?aG?GCcAG?agaAGGTGGAga?AacaGCA?gAgCAaaT???agccttgGcAGaAAGA?TCc???A  2132
SIVAGMTYO  -CC-AC-A--C--GAGA-------GAC-ACA---GG-T--AA-TAGAGTCTTG-T--A--- C --C...-  2320
SIVAGM155  -CA-GA-A--C--AAGA-------AGA-CAG---AA-G--AA-CTACGCAT***                   2304
SIVAGM3    -CA-AA-G--C--GAGA-------GAC-AGC---AG-A--GC-CCGAGCCTTG-C--A----A--TGGC-  2341
SIVAGM677  -GA-GA-G--A--ACAG-------GAA-GCA---GC-G--AA-TGATAAGATT-C--G----G--CTCA-  2272

CONSENSUS  ca?c???gaGGa??a?caactggTG?atgca?TTgAcca?cTGg?gctcGAg?atcAgcActTGGctaTA  2190
SIVAGMTYO  AG...AGCA--............--TATGCAG--G-CCGCC--GCTGAC--GGCTC-AC-CT---CTA--  2375
SIVAGM3    CAGCAGAGT--AAGAGCAACTGG--CAAGCGA--G-CCAAT--GTGCTC--TCAGC-GC-TC---CTA--  2411
SIVAGM677  CACCTTCGA--ATCAACAACTGG--GCTCAAC--C-AGAGC--CAGCTT--GAATA-GG-CT---TGT--  2342

                                        <- rev            nef cds ->
CONSENSUS  CAACAgtTGCCTGACCCTCCTcaTtca?cTtAGGA?aGCttT?caaTACcTgCAaTATGGGCTcg?ggAA  2256
SIVAGMTYO  -----GT--------------CA-TCAGC-T----AA--TT-TCAA---C-G--A--------CGCAG--  2445
SIVAGM3    -----GT--------------AG-TCATC-T----GC--TT-CCAG---A-A--A--------TGGGG--  2481
SIVAGM677  -----TC--------------CA-ATTCA-C----CA--AG-GGGA---C-G--G--------CCAGC--  2412

CONSENSUS  CTCaAAgccGcaGCAcaAGaa?tagttcaagCTcTGGCa?GcgtTGCacaaaAcGC?tG??aCcagaTat  2321
SIVAGMTYO  ---A--ACC-GC---CA--AAATACTCCAAA--C----AG-CGT---ACAAA-C--AT-TCA-CAGA-AT  2515
SIVAGM3    ---A--GCC-CA---CA--AAGCAGTTGTCG--T----AC-CCT---ACAAA-C--GG-CTA-CAGA-AT  2551
SIVAGM677  ---C--GAG-CA---AC--GGCTTGCTCAAG--C----GA-GGC---GAGGG-A--CT-GGG-AGAC-GG  2482

CONSENSUS  GgctTgcTtgCaGATCCGCTTATaGGqcaaTCaTCAACaqTCCAAGAAGAGTGCGaCAAGGCCTTGAaqa  2391
SIVAGMTYO  -GCT-GC-TG-A----------A--AACA--G-----AG--------------A----------GGA  2585
SIVAGM3    -GCT-GC-TG-A----------A--GCAA--A-----TC--------------A----------AGG  2621
SIVAGM677  -TGC-AT-GT-C----------C--GCAG--A-----AG--------------G----------AAA  2552

CONSENSUS  AaTCCTtaat***   2404
SIVAGMTYO  -A----TAAT---   2598
SIVAGM3    -A----TAAT---   2634
SIVAGM677  -G----GGGG---   2565
```

AGM NEF

AGM NEF CDS

```
 env cds ->
CONSENSUS  ATGGGCT?ggcgAaCTCaAAgccGcaGCAcaAGaag?agttca?gcTcTGGC?aGcgtTGCac?aaAcGC    65
SIVAGMTYO  -------CGCAG-A---A--ACC-GC---CA--AAATACTCCAAAC-C----AG-CGT---ACAAA-C--    70
SIVAGM155  -------TGGGG-G---A--GCC-CA---CA--AAGCAGTTAACCA-C----GA-CTT---ACGCA-C--    70
SIVAGM3    -------TGGGG-A---A--GCC-CA---CA--AAGCAGTTGTCGC-T----AC-CCT---ACAAA-C--    70
SIVAGM677  -------CCAGC-A---C--GAG-CA---AC--GGCTTGCTCAAGC-C----GA-GGC---GAGGG-A--    70

CONSENSUS  g?g?caCcagaTatGgctTgcTtgCaGATCCGCTTAT?GGgcaaTCaTCAAC??TCCAAGAAGAGTGCGa   130
SIVAGMTYO  ATGTCA-CAGA-AT-GCT-GC-TG-A----------A--AACA--G-----AG--------------A   140
SIVAGM155  GGCACA-CAGA-AT-GCT-GC-TG-A----------C--GCAA--A-----TC--------------A   140
SIVAGM3    GGGCTA-CAGA-AT-GCT-GC-TG-A----------A--GCAA--A-----TC--------------A   140
SIVAGM677  CTGGGG-AGAC-GG-TGC-AT-GT-C----------C--GCAG--A-----AG--------------G   140

                                  <- env
CONSENSUS  CAAGGcCTTGA?gaA?TCCTtaattaGgAA?agaaatGG?AA?ATGACtccAGAaGgaaGa?g?cTaCAa   193
SIVAGMTYO  -----C-----GGA-A----TAATTA-G--ACAAAAT--C--C-----GGA---G-AAA-GAGGC-T--A   210
SIVAGM155  -----G-----GGA-G----TAATTA-G--GAGAAAT--C--C-----TCC---A-GAA-ACGTC-A--G   210
SIVAGM3    -----C-----AGG-A----TAATTA-G--GAGAAAT--T--A-----TCC---A-GAA-AAAAT-A--A   210
SIVAGM677  -----C-----AAA-G----GG...G-T--A.....--T--A-----TCC---C-GCC-CCGCC-G--A   201

CONSENSUS  GAaGGaGAca??TggGAtGA?TGGtctGATGA?GAaGA?GAAGT?GGaTTtCCaGT?agACCaaGAGTgC   256
SIVAGMTYO  --A--A--CACC-GG--A--G---TCG-----T--G--A-----G--A--T--A--GAG---AA----A-   280
SIVAGM155  --C--G--CCAA-GG--T--A---TCA-----A--A--T-----G--A--T--A--AAG---AA----G-   280
SIVAGM3    --A--A--TAAA-GG--T--A---TCT-----A--A--T-----A--A--T--A--AAG---AA----G-   280
SIVAGM677  --A--A--CACC-TT--T--G---GAT-----T--A--A-----A--C--C--T--GCA---TC----C-   271

                                                                        LTR ->
CONSENSUS  Cc?TAAGaCAaATgACcTAtAAA?T?GCaGTgGA?TTTTCgCACTTTTTAAAAgaAAAGGG?GGACTGGA   321
SIVAGMTYO  -CT----G--A--G--T--T---C-T--A--G--T-----G----------GA------A--------   350
SIVAGM155  -AC----A--A--A--A--C---C-T--A--A--T-----G----------GA------A--------   350
SIVAGM3    -GC----A--A--G--C--T---T-A--G--G--C-----G----------GA------G--------   350
SIVAGM677  -CT----A--G--G--C--T---T-A--A--G--C-----C----------TC------G--------   341

CONSENSUS  TGGGATtTATTACTCcGAtAGqAGaaAtaAGATCCTgAAtcTGTAtGCccTcAAtGA?TGGGGaAT?ATt   389
SIVAGMTYO  ------T--------C--T--G--AA-TA--------G--TC----T--TC-T--T--A-----G--A--T   420
SIVAGM155  ------T--------C--T--G--AA-TA--------T--TC----T--CC-C--T--A-----A--C--T   420
SIVAGM3    ------T--------C--C--G--GA-TC-------A--CC----C--CC-C--T--G-----A--C--T   420
SIVAGM677  ------A--------T--A--A--AG-AA-------G--TT----T--CT-G--C--G-----A--A--A   411

CONSENSUS  GATGATTGGaAtGCtTggTCaaaaGGaCCaGG?ATaAGaTaCCCgAgatgCTTtGGcTTCTGCTTtAAgc   458
SIVAGMTYO  ---------A-T--C-GG--GAAG--A--A--A--A--A-T---T-AATG---T--G--------T--GC   490
SIVAGM155  ---------A-C--A-GG--AAAA--A--T--G--A--A-A---G-GGTG---T--C--------C--GC   490
SIVAGM3    ---------A-T--T-GG--AGAA--A--A--A--C--A-A---A-GATG---C--C--------T--AT   490
SIVAGM677  ---------C-A--T-AC--ACCA--C--G--G--A--G-A---G-GAGT---T--C--------T--GC   481

CONSENSUS  TaGTaCCaGTgGaCcTgCATGAGGAaGCA?aaAcaTGTGA?AGaCAtTGccTggTgCAtCCAGCaCA?aT   525
SIVAGMTYO  -A--G--A--G-A-T-A--------A---CAA-CA-----A--A--T--CC-AG-C--T-----G--GA-   560
SIVAGM155  -A--A--G--T-C-C-G--------A---GAA-CA-----A--G--T--CT-GG-A--C-----A--AC-   560
SIVAGM3    -G--A--A--A-A-C-G--------A---GAG-CT-----G--A--T--CC-GG-G--T-----A--AG-   560
SIVAGM677  -A--C--A--G-A-C-G--------G---CGC-AC-----G--A--C--TC-GA-G--T-----A--GA-   551

CONSENSUS  GqqqGAAGA?CCtGATGG?AT?aa?CATGGAGAa?TcTTGGtcTGGAAGTTTGAtCC?AtGTTgGCagT?   588
SIVAGMTYO  -GGA-----T--A-----T--CAGC--------GA-C----TG----------T--T-T---G--AA-A   610
SIVAGM155  -CAT-----C--T-----T--AAAT--------AA-A----CA----------T--A-T---G--TG-T   630
SIVAGM3    -AGG-----C--T-----A--CAAC--------AG-C----TC----------T--C-T---A--AG-C   630
SIVAGM677  -GGG-----T--T-----A--AGAT--------AG-C----TC----------C--G-A---G--GG-G   621
```

```
CONSENSUS  cAgTAcgaCCCcaa?agggAgtacttTtActGACaTgcAtgaactt???gg?AagaGgaag***   646
SIVAGMTYO  C-G--CGA---CAATCGGG-GTACT-T-CT---A-GC-TGCGCTGGTG...-AGA-GAAG---   690
SIVAGM155  C-G--CGA---CTCAAGGG-GTACT-T-CA---T-AT-TTCAACAGTTGGT-CAG-AAAC---   693
SIVAGM3    C-A--TGA---TAACAGAA-ATATC-C-CT---A-GC-TGATCTT...GGC-AGA-GAAG---   690
SIVAGM677  G-G--CCG---G......G-CATGT-T-AG---A-GC-CGAACAT...GCA-AGC-C***     672
```

```
LOCUS       SIVAGM677    9623 bp ss-DNA
DEFINITION  Simian immunodeficiency virus from African Green Monkey; gri-1
            lambdaII or 677 clone (from a grivet); complete genome.
ACCESSION   M66437
SOURCE      Simian immunodeficiency virus from an African grivet (green
            monkey), gri-1, lambda clone II (biologically active clone);
            proviral DNA, also denoted 677.
REFERENCE   1  (bases 1 to 9623)
  AUTHORS   Fomsgaard,A., Hirsch,V.M., Allan,J.S. and Johnson,P.R.
  TITLE     A highly divergent proviral DNA clone of SIV from a distinct
            species of African green monkey.
  JOURNAL   Virology 182, 397-402 (1991)
COMMENT     Kindly submitted prior to publication and in a computer readable
            form by Phillip Johnson, NIAID/Georgetown University.
            The typical stop codon "tag" in rev (position 8574) is "cag" making
            rev in SIVgri-1 (lambdaII) longer than rev in most other sequences.
            The gri-1 sequence is the most divergent primate lentivirus reported
            to date.
BASE COUNT     3243 a   1807 c   2470 g   2103 t
ORIGIN      5' terminus of 5'LTR.
        1 tggatgggat atattactct gaaagaagag aaaagatcct gaatttgtat gccttgaacg
    5'LTR ->
       61 agtggggaat aatagatgat tggcaagctt actcaccagg cccggggata aggtacccga
      121 gagtctttgg cttctgcttt aagctagtcc cagtggacct gcatgaggag gcacgcaact
      181 gtgagagaca ctgtctgatg catccagcac agatggggga agatcctgat ggaatagatc
      241 atggagaagt cttggtctgg aagtttgacc cgaagttggc ggtggagtac cgcccggaca
      301 tgtttaagga catgcacgaa catgcaaagc gctagtgtca gcactttgcg gttgggactt
      361 tccgccaggg actttccaca gtgggtggat cggaggcggt acaggggcgg tactgggagt
      421 ggctttcccc tcagagctgc ataaaagcag atgctcgctg gcttgtaact cagtctctta
      481 ctaggagacc agctagagcc tgggtgttcg ctggttagcc taacccggtt ggccaccggg
      541 ggtaaggact ccttggcttc atatagctcA ATAAAcctgc tcgcttagtc gctatattgg
                                signal ->                <- R rpt
      601 agtcaagtgc tcattgctgc gccgagcctc tagaggtgaa cctctcttac tgggttctcc
      661 tgtacccagg tgggagaaac tccagcagtg gcgcccgaac agggacttga gaagaggcat
                                        <- lys-tRNA pbs
      721 cggcaccgac cgctgagttg ctgagcgtcg gagagggacg actcaggtag ggtgagagcc
      781 tacgagtttt ttgctaccta gtcagcgaga aaggctaggc cgcgacaggg gcgcgggtcc
      841 cattagtggc aaccaaccca gttggacgaa gggttggtag gggacgggtc ggagcaATGg
                                                          gag start ->
      901 gcgggggtca ctcagcactg tcagggagaa gcctcgacac gttcgagaag attaggctac
      961 gtccgaacgg gaaaaagaag taccaaatta aacatttaat atgggcagga aaagaaatgg
     1021 aacgatttgg gttacatgag aaacttttag aaacaaaaga aggctgtcaa aaaatcatag
     1081 aagttttaac cccgttggaa ccgacaggct ccgaggggct aaaagctctg tttaatttgt
     1141 gctgcgtcat ttggtgcatt cacgcagaac agaaagtgaa agacacagag gaagctgtag
     1201 taacagttaa gcaacactac catctagtgg acaaaaatga gaaagcagct aaaaagaaaa
     1261 atgagacaac agcgccacct ggtggcgaat caagaaatta cccagtagta aatcagaata
     1321 atgcctgggt acaccagcct ttgtctccgc gcacgttaaa tgcgtgggtc aaatgcgtgg
     1381 aggaaaaaag gtggggagca gaagtagtcc ccatgttcca agcactctca gagggatgtc
     1441 tctcctatga tgtaaatcag atgctcaatg taataggaga ccatcagggg gcattacaaa
     1501 ttcttaagga agtcattaat gaagaagcag cagagtggga caggacacac agaccaccag
     1561 ctggcccgtt accagcaggg cagctaagag acccgacagg gtcagatata gcaggaacta
     1621 ccagctcaat tcaggaacaa atagagtgga ccttcaatgc caatccaaga atagacgtag
     1681 gggcacaata cagaaaatgg gttattttgg gcttacaaaa ggtagtgcag atgtacaatc
     1741 cccaaaaggt cctagacatt cgacagggac ctaaagaacc cttccaggac tatgtagaca
     1801 gattctataa agccctgaga gcagaacaag caccacagga tgttaaaaat tggatgacac
     1861 aaactttgct tatccagaat gccaatccgg attgtaaatt gattctgaaa ggattgggaa
     1921 tgaatccaac cttggaggaa atgctaatag cttgccaggg agtaggaggg ccacaacata
     1981 aggctaagct aatggtagaa atgatgagta atggacagaa tatggtccaa gtgggacctc
     2041 agaaaaaggg cccccgaggg ccgctaaaat gctttaattg tggcaaattt ggacatatgc
     2101 aaagggaatg caaggcacca agacagatca aatgcttTAA gtgcggcaaa attggccata
                                            pol orf ->
     2161 tggcaaaaga ctgcaagaat ggacaggcaa atttttagg gtatggccat tggggaggag
     2221 cgaaaccaag aaattttgtg caatacagag gagacacagt tggtctggaa ccaacagccc
     2281 ccccaatgga aacagcttac gatccagcaa agaagctcct ccagcagtat gcagagaagg
```

```
2341 gacagcgcct gagagaggag agagaacaga caaggaaaca gaaggagaaa gaagtggagg
2401 atgtttcctt gagctccctc tttggaggag accaaTGAaa cgagtcatca tagaaggaac
                                          <- gag
2461 gccagtgcaa gccttgttag atacaggagc agatgacact ataattcaag aaaaggactt
2521 gcactttccc ccacataaac catggcgttc caaggtagta ggaggtatag gaggagggat
2581 tcatgtcaaa gaatatcagg gggtacaagt acaattggag gataaaatca tcaccggctc
2641 aattctaata ggaagtacac caatcaatat tataggaaga aatattttag ctcaggcagg
2701 catgaaatta gttatgggag ttctatctag tcagattgag gaaacaaaag tacaactaaa
2761 agaagggaaa gatggaccta aattgaaaca atggcccttа tcaagagaaa aaattgaagc
2821 tttaacagaa atatgcaaac aaatggaaga ggagggaaaa ttatctagga taggaggaga
2881 aaatccttat aatacaccag tgtttgccat aaagaaaaag gataaaacac aatggagaat
2941 gcttgtagat ttcagggaac taaacaaagc tactcaagac ttttttgagg ttcagctggg
3001 aattcctcac ccagcgggcc ttcagaaaaa gaagcaaatc acagtaatag acataggggа
3061 tgcctattat tcaataccat tatgcaagga attcagaaaa tatacagcat ttaccatccc
3121 ctcagtaaat aatacagggc cagggataag gtatcagttc aattgtctgc ctcagggatg
3181 gaaaggatct cctacaattt tccagaaтac ggcagcaaac attttagagg agatcaaaag
3241 gcacactcct gggttagaaa ttgtccaata catggacgat ttgtggttgg cgtcagacca
3301 tgatgagact agacataatc aacaggtaga catagtaaga aagatgctgc tagaaaaagg
3361 tctagaaacc ccagacaaga aagtccaaag agaaccgcca tgggaatgga tggggtataa
3421 attgcatccg aataaatgga ccattaacaa aatagaatta ccccccttag aaggagaatg
3481 gacagtaaac aaaatacaga aggtagtagg agttctaaat tgggcaagtc aaatttatcc
3541 aggaattaaa accaaacata cctgtgccat gttgagaggg aaaaagaacc tcctagaaga
3601 aatagtatgg acagaagagg cagaggcaga atataagaac aatcaaggga tagtgcagga
3661 aacacaagaa ggaacatact atgaccctct caaagaatta atagcaacag ttcaaaagca
3721 aggagaaggg caatggacat accaattcac ccaagaaggg gcagtattaa aggtgggaag
3781 atatgccaag caaagagaaa ctcatactaa tgatctaagg actctagcac accttgtcca
3841 aaaaatctgt aaggaagcac ttaccatttg gggaagactt ccacgagtac aactcccagt
3901 agacaagaaa acatgggata tgtggtggca ggactattgg caagtatcct ggataccaga
3961 atgggagttt gttagcacac cactcctagt aaaactgtgg tattccttag taaaagaacc
4021 aatcaaagga gaagatgttt attatgtgga tggggcagca tccaaagtga ccaaattagg
4081 taaggcagga tatctgtcag agagaggaaa aagtagaatt agggaattag aaaacaccac
4141 tcaccaacaa gcagaattaa cagcagttaa gatggcattg gaggacagtg gagaaaatgt
4201 aaatatagtc acagattctc aatatgtaat gaacatcttg acagcatgtc cacaggaaag
4261 taactcacca ttagtggaac agataaraca agccctaatg aaaaagaggc aggtctactt
4321 acaatgggta ccagctcata aggggatagg aggcaatica gaaatagata aattagtaag
4381 caaaggaata agacagatcc tcttcttaga tagaatagaa gaagcacaag atgaccatgc
4441 aaagtaccat aacaattgga gaagtatggt acaggaattt ggattaccta atatagtagc
4501 ataagagata gtagcggcat gtcccaaatg ccaaataaga ggagaaccta agcatggaca
4561 ggtagacgcc tccattgaaa cttggcagat ggactgcacc catttagaag gaaaagttat
4621 aatagtagca gtacatgtag ccagtggatt catagaagca gaggtgatcc caagaggaac
4681 tgggaaggag acagcacact ttctgctgaa actgttagca agatggccag tgaaacatct
1741 acacactgat aatggcccaa actttacctc tcagaatgtg gcagcggtgt gctggtgggg
4801 taatatagag cacaccactg gaatacctta taacccacag tcacagggta gtgtagaaag
4861 catgaacaga cagctcaagg aaatcatctc tcaaataaga gatgattgtg agagattgga
4921 gacagcagtg caaatggcta cgcatatcca caatttтaaa agaaagggag gaataggggg
4981 tatctctagt gcagaaagat tggttaatat gctaacaaca caactagaac taaatactct
5041 acaaaaccaa atccaaaaaa ttttgaattt taaggtctac tacagagaag gtagagatcc
5101 agtgtggaaa ggaccagcgc gaatcatctg gaaaggagaa ggcgcggtgg taatтaaaga
5161 ggggcaagac atcaaggtag tccccaggag aaaggctaag attatcaaag attATGgaga
                                                    vif start ->
5221 gagaaaaaca atggatagtg agggtagtat ggagggtgtc agagaggcaa ataagcagat
5281 ggagggggat agtgacttac aagatcagga aTAAacaatt gccttgggaa tacagacatc
                                          <- pol
5341 actggcaggt gcaatggcag ttttggacct acagccagtt cattatcccc ttatcaaaag
5401 atgattacat agaagtgaat atttatcaca acctcacccc agaaagagga tggctctcaa
5461 gtcatgggat agggttatcc tattaccatc aaaagggata taagacagaa gtagatccag
5521 gaacagcaga cagaatgata cacctatatt attttaactg ttttacagat agagccatcc
5581 aacagcctat cagagggagg aagtatacgt ggtgcacatt caaggaagga catcaaaggt c
5641 aggtacaatc actgcaactt ttggcactag ttgcatatac aaATGgcatc aggaagatat
                                          vpX start ->
5701 ccaagaagaa ctttaccagg atgggtgggg atctgggatc tagacagggа gccatgggac
5761 gaatggctac aagacatgct cagggatcta aacgaaggag ccaggaaggc ctttggaatg
5821 aacatgctaa tccgagtatg gaattactgt gtagagggag gaagggaaca TAAtaccccа
                                                    <- vif
```

```
5881 tggaatgaga taggctacaa gtactataga attgttcaaa agtctatgtt tgtacatttc
5941 agatgtggtt gtagaaggag aggacctttt tccccttacg aagagaggag aaATGgacaa
                                                    tat start ->
6001 ggaggaggag ccccaccccc tcctccagga cttgcaTAGg ccattacaac catgtactaa
                                              <- vpX end
6061 taaatgctat tgcaaaaaat gctgttatca ctgtgagctt tgcttcctgc agaagggctt
6121 aggagtgcgt tatcATGtct ctaggaaaag aagaaaaaca agcactcaag ataatcaaga
          rev start ->
6181 ccctatacgg cagcaGTaag tATGgggaga ttgcttataa aaatactaat aatagcaata
               5' sj /\  env ->
6241 gggataagta taggaatagg taacctgtat gtgacagtgt tttatggaat cccagtatgg
6301 aaaaattcaa cagttcaggc attttgcatg acgcccaata ccaatatgtg ggcaaccacc
6361 aactgcatac cagatgatca tgataataca gaggtgcctc taaacattac agaagctttc
6421 gaggcttggg ataatccgct ggtaaaacaa gcagagagta atatacatct actctttgaa
6481 caaacgatga ggccttgtgt taagctctcc cccatatgta ttaaaatgtc ctgtgtagag
6541 ctgaatggta cagccacgac aaaggccacc actactgcaa ctacaacaat gactacccca
6601 tgtcagaatt gcagtacaga gcagatagaa ggagaaatgg cagaggaacc agcatccaac
6661 tgcacttttg caattgcagg atatcaaaga gatgtaaaaa agaattatag catgacctgg
6721 tatgatcagg agttagtctg caataataaa acaggaagtg aaaagggaag taaggattgt
6781 tacatgatac attgtaatga ttcagtgata aaagaagctt gtgataaaac atattgggat
6841 actttaagag taagatactg tgcaccagca gggtatgctt tgctaaaatg taatgataag
6901 gattatagag gctttgctcc aaagtgcaag aatgtttcag tagtgcattg tactagatta
6961 atcaatacta ctataactac agggatagga ttaaatggta gtagatcaga aaatagaaca
7021 gagatatggc agaaaggagg aaatgataat gatacagtta taataaagtt gaataagttt
7081 tacaacttga cagtgagatg ccgaagacct ggtaataaaa cagtgttgcc agtaacaatc
7141 atggcagggt tagtatttca ctctcagaaa tataatacca ggttaaaaca agcgtggtgc
7201 cacttccaag gagattggaa aggggcatgg aaagaagtca gagaagaagt aaagaaagtg
7261 aaaaatctta cagaagtaag catagaaaat atacctctga gaaggatatg gggagatcca
7321 gaatcagcga atttttggtt caattgtcaa ggtgaatttt tctattgtaa gatggactgg
7381 tttatcaatt atctaaacaa tcgaacagaa gatgcagaag gtactaatag gacctgtgac
7441 aaagggaagc caggaccagg accatgtgtt cagagaactt atgttgcctg ccatatacga
7501 caagtagtaa atgattggta cactgtctct aaaaaggtat atgctccacc aagggaaggt
7561 catttggagt gtaactcatc agtcacggca ctatacgtgg caatagatta taacaacaag
7621 tctggcccaa taaatgtgac cctaagtccc caggtacgca gcatatgggc gtacgaactg
7681 ggagactata aattagtaga gataacacca attggctttg ctcctacaga tgtaagaaga
7741 tatactggcc ccacaagaga aaaaagggtg ccattcgtgc tagggtttct aggcttcttg
7801 ggagctgctg gaactgcaat gggcgcagcg gcaacaacgc tgacagtcca gtctcggcat
7861 ttgcttgctg ggatattgca gcagcagaag aacttgctgg cggctgtgga acagcaacaa
7921 cagttgttga agctgaccat ttggggtgtg aaaaacctca atgcccgcgt cacagctctc
7981 gagaagtacc tagaggatca ggcacggcta aattcatggg gatgtgcgtg gaaacaagta
8041 tgtcacacca cagtgccatg gaagtataat aacactccta agtgggacaa tatgacttgg
8101 ttggagtggg agagacaaat taatgccttg gaaggcaaca taactcaact atcggaagaa
8161 gcacaaaatc aggaatcaaa gaatctggat ctgtaccaga aattggatga ttggtcaggg
8221 ttctggtcat ggttctcact gtcaacttgg ttaggctatg ttaaaatagg atttttagtg
8281 atagtgatta ttctaggatt aagatttgca tgggtattat ggggatgtat cagaaatatt
8341 aggcagggat ataatcctct cccccAGatc catatccaca gttcagcgga acggccagac
                                  /\ 3' sj
8401 aacggaggag ggcaagacag aggtggagaa agcagcagca gcaaatTGAt aagattgcag
                                                        <- tat
8461 gaagagtcct caacaccttc gaggatcaac aactggtggc tcaacttcaa gagctgcagc
8521 ttgagaataa ggacttggtg ttacaacatc tgcctgacac tcatcatatt catcaggaca
8581 gcagtgggat acctgcagtA TGggctccag caactccaag aggcagcaac agggcttgct
            nef start ->
8641 caagctctgg cgagggctgc gagggaagcc tggggcagac tgggtgctat tgtccgatcc
8701 gcttatcggg cagtcatcaa cagtccaaga agagtgcggc aaggccttGA aaaagtcctg
                                                        <- rev
8761 gggTAAaggt aaaatgacta cagacggccg ccgcctgcaa gaaggagaca cctttgatga
        <- env
8821 gtgggatgat gatgaagaag aagtaggctt cctgtgcaa cctcgagtcc cctaagaca
8881 gatgacctat aaattagcag tggacttttc ccacttttta aaatcaaagg gggactgga
8941 tgggatatat tactctgaaa gaagagaaaa gatcctgaat ttgtatgcct tgaacgagtg
9001 gggaataata gatgattggc aagcttactc accaggccag gggataaggt accccgagag
9061 ctttggcttc tgctttaagc tagtcccagt ggacctgcat gaggagtcac gcaactgtga
9121 gagacactgt ctgatgcatc cagcacagat gggggaagat cctgatgtga tagatcatgg
```

```
9181 agaagtcttg gtctggaagt ttgacccgaa gttggcggtg gagtaccgcc cggacatgtt
9241 taaggacatg cacgaacatg caaagcgcTA Gtgtcagcac tttgcggttg ggactttccg
                                  <- nef end
9301 ccagggactt tccacagtgg gtggatcgga ggcggtacag gggcggtact gggagtggct
9361 ttcccctcag agctgcataa aagcagatgc tcgctggctt gtaactcagt ctcttactag
           signal ->                          R rpt ->
9421 gagaccagct agagcctggg tgttcgctgg ttagcctaac ccggttggcc accgggggta
9481 aggactcctt ggcttcatat agctcAATAA Acctgctcgc ttagtcgcta tattggagtc
                       signal ->
9541 aagtgctcat tgctgcgccg agcctctaga ggtgaacctc tcttactggg ttctcctgta
9601 cccaggtggg agaaactcca gca
```

```
LOCUS       SIVMND121      710 bp ss-RNA
DEFINITION  Simian immunodeficiency virus isolated from a mandrill, infectious
            clone pMD-121, tat and rev exons (first expressed exons).
ACCESSION   M62883
SEGMENT     1 of 2
REFERENCE   1  (bases 1 to 710)
  AUTHORS   Sakai,H., Sakuragi,J., Shibata,R., Kiyomasu,T., Hayami,M.,
            Ishimoto,A. and Adachi,A.
  TITLE     Genetic characterization of the simian immunodeficiency virus
            isolated from an African mandrill
  JOURNAL   J. Virol. (1991) Submitted
COMMENT     Kindly submitted prior to publication and in a computer readable
            form by A. Adachi. Sequence from pMD-121 (SIV mandrill infectious
            clone).
BASE COUNT      240 a    111 c    168 g    191 t
ORIGIN
        1 gctagcagat actatggagg aaataaagga agaagcaaga aagcacttcc ctctcattat
       61 cctaaatgca gtatcagaat attgtgtgca aaacacaggg agtgaggaag aggcctgtga
      121 gaaatttatt accttaatga atagagccat ttgggtccat ctagctcaag ggtgtgATGg
                                                              tat start ->
      181 aaccttcagg gaaagaagac cacaactgcc tccctcagga tttaggccaa gaggagatag
      241 attataagca actgctagaa gagtattatc agcctttgca agcttgtgag aataaaatgct
      301 ggtgcaagaa atgctgcttt cattgtatgc tttgctttca taagaagggt ttaggaataa
      361 ggtaccATGt ctacaggaaa cgtggaccag gaactaataa gaagatacct ggtggtggtg
      rev start ->
      421 aagaagctat acgaagGTag gtatgaagtg tccaggtctt tttcttatac tatgtttagc
                  5' sj /\
      481 tatagtaggt attataggaa aacaatatgt gacagtcttc tatggagtac cagtatggaa
      541 ggaagctaaa acacatttga tttgtgctac agataattca agtctctggg taaccactaa
      601 ttgcatacct tcattgccag attatgatga ggtagaaatt cctgatataa aggaaaattt
      661 tacaggacct ataagggaaa atcagatagt ttatcaagca tggcatgcta

LOCUS       SIVMND122      697 bp ss-DNA
DEFINITION  Simian immunodeficiency virus isolated from a mandrill, infectious
            clone pMD-121; tat and rev exons (second expressed exons).
ACCESSION   M62884
SEGMENT     2 of 2
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Sakai,H., Sakuragi,J., Shibata,R., Kiyomasu,T., Hayami,M.,
            Ishimoto,A. and Adachi,A.
  TITLE     Genetic characterization of the simian immunodeficiency virus
            isolated from an African mandrill
  JOURNAL   J. Virol. (1991) Submitted
COMMENT     Sequence from pMD-121.
BASE COUNT      223 a    126 c    177 g    171 t
ORIGIN
        1 aagctttgct gtctcaatgg gggtgttcat gggcacaggt gtgtcatact totgtagagt
       61 ggactaatac aagcatcact ccaaattgga catcagaaac ttgg/aggaa tgggagacaa
      121 gaactgatta tctgcaacaa aacattacag aaatgttaaa acaggcatat gatcgagagc
      181 aaagaaacac atatgaatta cagaagttag gagaccttac atcttgggca agttggtttg
      241 actttacttg gtgggttcaa tacttaaaat gggggagttt cttagtgtta ggaattatag
      301 gattaagaat tttgttagcc ttatggaata caataagtag gtttagggAG ggctatcgac
                                                          /\ 3' sj
      361 ctgtgctttt tcaacaggac ttgcagcaga acccatacca caaacgggca gacaacggag
      421 aagaagaaag caacagctTA Gaactaggag agcacaactc cgagaacttg aaggaagaat
                           <- tat end
      481 cctaaaacag atccttgata gaggacctga ccagctttgc cagggagtag ccaatttggc
      541 tttggctgaa aagtctgaaa gcagcaatTG Aatatgggtt cctcgcagtc caagaagcga
                                      <- rev end
      601 tcagaagctt gggttcgcta ctcgtcagct ttgcgggcaat tagttggagg gccgttacca
      661 ccggatgggt acaagcaaat agaatcttca cagggtg
```

II

AMINO ACID SEQUENCES

The amino acid sequence alignments that follow were generated by the PIMA program ("pattern induced multiple alignment"; R. F.Smith and T .F.Smith, *PNAS* 87: 118–122, 1990). The alignments were also subject to some hand-editing performed with MASE. For information concerning both these UNIX tools, contact Dr. Randall Smith, MBCRR, Dana Farber Cancer Institute, Harvard School of Public Health, 44 Binney St., Boston MA 02115, (617)-732-3746.

With a few exceptions, only full-length protein sequences have been included in this compilation. One simian and one HIV1 sequence were not ready for publication at this time, but are reflected in the consensus sequences and patterns. If you have a question about a particular stretch of alignment, contact us for help, as it has been sometimes possible in the past to obtain an author's approval to release sequence data over a limited region.

The reference sequences for the alignments are mixed case consensus sequences in which upper case letters refer to amino acid residues which are conserved 100% and lower case letters represent amino acid residues conserved in at least 50% of the sequences. For type 1 viral sequences, separate consensus sequences have been generated for the North American and African samples.

Within the sequences, "-"s are used to indicate residues conserved with respect to the reference sequence and "."'s represent actual gaps. Blank spaces within the alignment indicate lack of sequence information over that region. Annotation of the ENV amino acid sequences utilizes "*" (or sometimes "#") for conserved cysteine residues and "^" for potential N-linked glycosylation sites.

Following each alignment the consensus sequence for that alignment is shown with the corresponding AACC ("amino acid class covering") PATTERN. The AACC patterns are described in the above-mentioned reference of Smith and Smith and the amino acid class hierarchy used to construct the patterns is shown in the diagram below. Each position in a pattern represents the minimum inclusive amino acid class that covers the range of residues observed or anticipated at that position for the entire set of sequences in the alignment (except that the SIVcpz sequences have not been included with the HIV-1 AACC patterns). Ordinary consensus sequences have the disadvantage of being "averages," thus they reduce the information represented in the variation; however, they are relatively insensitive to sequence errors and defective viral sequences. The AACC patterns, sometimes called "consensus-like" patterns retain more of the information, but they are highly subject to outliers and spurious variants. By presenting both kinds of consensus sequences, we cover the information from both directions.

In the AACC patterns, upper case letters indicate identities, i.e., 100% conservation of a particular amino acid at that position, with exception of the wild-card "X". Gaps are represented by "."'s. Positions at which all the observed or anticipated amino acid residues belong to a particular amino acid class are indicated by "^"'s, with all the members of that class listed vertically below that position. This listing does not imply that all such substitutions have occurred at that site, nor does it preclude the possibility of a non-conserved replacement occurring at that site.

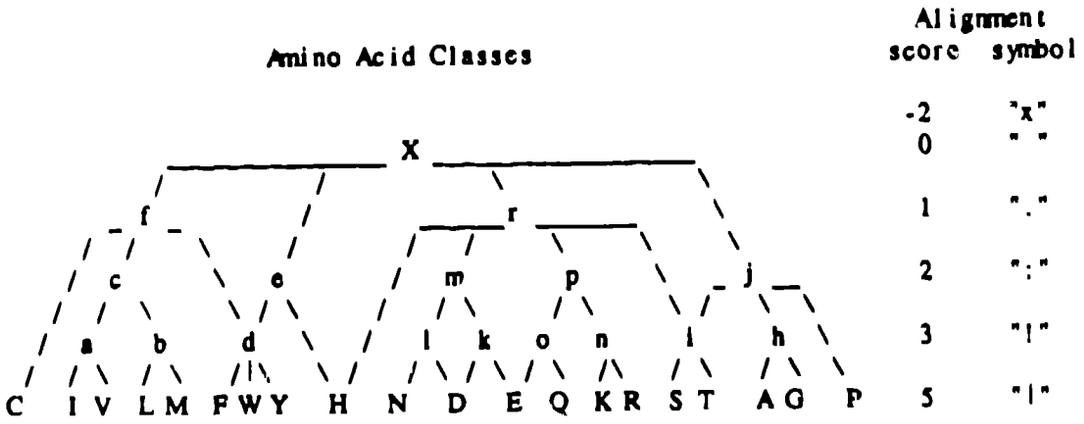
The AACC analysis utility calculates the information content, or the number of amino acid equivalents contained in each pattern: totally variable sites (wild-card X) have 0 information and totally invariant sites (a specific amino acid) have information of 1.0; all other sites, more or less variable, take on values between 0 and 1.0 amino acid equivalents. For this calculation, all 20 amino acids are assumed to have equal informational content and regions separated by gaps are assumed to be statistically independent of one another. An additional calculation derives a length-normalized information content, or information density, from each pattern. A program (subpat)

recently provided by Drs. Randall Smith and Temple Smith, has been used to select maximally information dense subregions from the patterns based upon a user defined threshold. For the most part, this analysis confirms the invariant stretches discovered through ordinary consensus sequences; however, it also identifies variable but conserved regions that might be otherwise overlooked.

The absolute information density thresholds used in this report are indicated above each consensus/pattern alignment. The minimum information density required for each subpattern was chosen in order to select well defined regions with information densities significantly above the average for that pattern. The regions containing maximally information dense subpatterns are indicated by "*"s placed over the appropriate region of the alignment. Thus a catalog of conserved and invariant protein structures becomes possible. Such a catalog of HIV-1 proteins is included.

Because the AACC patterns are acutely sensitive to rare amino acid replacements, and sequencing errors, the catalog of invariant and conserved patterns is highly stringent. The annotation of this catalog is, as the user can tell, in its earliest stage. The annotation of epitopes heavily relied upon notes provided by Dr. James Bradac, Vaccine Branch, AIDS Division, NIAID.

The HIV-1 principal neutralization determinant ("PND" or V3 loop) is not sufficiently conserved to be included in the catalog; accordingly, its analysis and patterns are found in Part III.



PART II. AMINO ACID SEQUENCES AND THEIR VARIATION

INTRODUCTION II-1

GAG

HIV-1 5
HIV-1 Consensus Sequences and Patterns 8
HIV-2/SIV 13
HIV-2/SIV Consensus Sequences and Patterns 16
SIVAGM 18
SIVAGM Consensus Sequences and Patterns 19

POL

HIV-1 21
HIV-1 Consensus Sequences and Patterns 26
HIV-2/SIV 33
HIV-2/SIV Consensus Sequences and Patterns 38
SIVAGM 42
SIVAGM Consensus Sequences and Patterns 44

VIF

HIV-1 48
HIV-1 Consensus Sequences and Patterns 49
HIV-2/SIV 50
HIV-2/SIV Consensus Sequences and Patterns 52
SIVAGM 53
SIVAGM Consensus Sequences and Patterns 54

VPX

HIV-2/SIV 55
HIV-2/SIV Consensus Sequences and Patterns 56
SIVAGM 57
SIVAGM Consensus Sequences and Patterns 58

VPR

HIV-1 59
HIV-1 Consensus Sequences and Patterns 60
HIV-2/SIV 61
HIV-2/SIV Consensus Sequences and Patterns 62

TAT

HIV-1 63
HIV-1 Consensus Sequences and Patterns 64
HIV-2/SIV 65
HIV-2/SIV Consensus Sequences and Patterns 66
SIVAGM 67
SIVAGM Consensus Sequences and Patterns 68

REV

HIV-1 69
HIV-1 Consensus Sequences and Patterns 70
HIV-2/SIV 71
HIV-2/SIV Consensus Sequences and Patterns 72
SIVAGM 73
SIVAGM Consensus Sequences and Patterns 74

VPU

HIV-1 75
HIV-1 Consensus Sequences and Patterns 76

ENV

HIV-1 77
HIV-1 Consensus Sequences and Patterns 84
HIV-2/SIV 90
HIV-2/SIV Consensus Sequences and Patterns 94
SIVAGM 98
SIVAGM Consensus Sequences and Patterns 100

NEF

HIV-1 103
HIV-1 Consensus Sequences and Patterns 105
HIV-2/SIV 107
HIV-2/SIV Consensus Sequences and Patterns 108
SIVAGM 109
SIVAGM Consensus Sequences and Patterns 110

SIVMND PROTEINS 111

```
CONSENSUS1 MGARASVLSGGeLDkWEkIRLRPgGKKkYkLKHiVWASReLERFAvNPgLLEtseGCRQIlgQLQPsLqT   70
HIVLAI     ----------E--R--K-----G---K-K---I-----E-----V--G---TSE-----LG----S-Q-   70
HIVHXB2R   ----------E--R--K-----G---K-K---I-----E-----V--G---TSE-----LG----S-Q-   70
HIVMN      ----------E--R--N-----G---K-K---V-----E-----V--G---TSE-----LG----S-Q-   70
HIVJH3     ----------E--R--K-----G---K-K---I-----E-----V--S---TSE-----LG----S-Q-   70
HIVJRCSF   ----------E--+  X-----G---K-R---I-----E-----V--G---SSE-----LG----S-K-   70
HIVOYI     ----------E--'. -K-----G---K-Q---I-----E-----I--G---TSE-----LG----S-K-   70
HIVSF2     ----------E--K--K-----G---K-K---I-----E-----V--G---TSE-----LG----S-Q-   70
HIVNY5CG   ----------E--K--K-----G---Q-R---I-----E-----V--G---TSE-----LR----S-Q-   70
HIVNL43    ----------E--K--K-----G---Q-K---I-----E-----V--G---TSE-----LG----S-Q-   70
HIVCDC4    ----------E--R--K-----G---Q-R---I-----K-----V--G---TSK-----LG----S-Q-   70
HIVHAN     ----------E--K--K-----G---K-Q---I-----E-----V--G---TSE-----MG----S-Q-   70
HIVRF      ----------K--K--K-----R---R-K---I-----E-----V--S---TAE-----LG----A-Q-   70
CONSENSUS2 MGARASVLsGgKLD?WEkIRLRPGGkKkYrlKHlvWASRELErfalnPGLlEtseGCkqiigGLqpa?qT  68
HIVELI     --------S-G---K--KI------K-K-RL--IV-------RYALN---L-TSE--KQIIG--QPAIQ-   70
HIVZ2Z6    --------S-G---A--KI------K-K-RL--LV-------RFALN---L-TSD--KQIIG--QPAIR-   70
HIVNDK     --------S-G---T--RI------K-K-AL--LI-------RFTLN---L-TSE--KQIIG--QPSIQ-   70
HIVMAL     --------S-G---A--KI------K-K-RL--LV-------RFALN---L-TGE--QQIME--QSTLK-   70
HIVU455    --------S-K---S--KI------N-K-RL--LV-------KFTLN---L-TAE--QQILG--QPALQ-   70
SIVCPZ     --------T-G---R--KV------R-R-MM--LV-------RFACD---M-SKE--TKLLQ--EPALK-   70

CONSENSUS1 GSEElrSLyNtvAtLYCVHQrIevkDTKEALdKiEEEQnkS.kKKAQQ??.....aAAdt???Gnss  126
HIVLAI     ----LR--Y-TV-T------R-EIK------D-I----NK-.K-----.....A--DT...-HSS  126
HIVHXB2R   ----LR--Y-TV-T------R-EIK------D-I----NK-.K-----.....A--DT...-HSN  126
HIVMN      ----LK--Y-TV-T------K-EIK------E-I----NK-.K-----.....A--DTGNR-NSS  129
HIVJH3     ----LK--F-TV-T------R-EVK------E-I----NK-.K-----.....A--DT...-NSS  126
HIVJRCSF   ----LT--Y-TV-T------R-EIK------E-I----TK-.M-----.....A--DT...-NSS  126
HIVOYI     ----IR--Y-TV-T------K-EVK------D-I----NK-.K-----.....T--DT...-NSS  126
HIVSF2     ----LR--Y-TV-T------R-DVK------E-I----NK-.K-----AA.....A--GT...-NSS  128
HIVNY5CG   ----RR--F-TV-V------R-DVK------D-I----NK-.K-----.....A--DT...-NSS  126
HIVNL43    ----LR--Y-TI-V------R-DVK------D-I----NK-.K-----.....A--DT...-NNS  126
HIVCDC4    ----LR--Y-TV-T------R-EVR------D-I----NK-.K-----.....A--DT...-NSS  126
HIVHAN     ----LR--Y-TV-T------K-EVK------D-V----NN-.K-----.....E--DA...-NRN  126
HIVRF      ----LK--Y-AV-T------N-EVR------D-I----NK-.K-----.....A--DT...-NGS  126
CONSENSUS2 GsEelrSLyNTvAtLyCvH?rIdVkDTkeAlekmeeeQnKs?kkktQQ??????????aaadt???gnss 123
HIVELI     -T-ELR--Y--V-T-Y-V-KG-D-K--KE-LEKMEEE-N-S.KKKA--.....AAADT...GNNS  126
HIVZ2Z6    -S-ELR--F--V-T-Y-V-ER-E-K--KE-LEKMEEE-N-SKNKKA--.....AAADA...GNNS  127
HIVNDK     -S-EIR--Y--V-T-Y-V-ER-E-K--KE-VEKMEEE-N-S.KKKT--.....AAAD......SS  123
HIVMAL     -S-EIK--Y--V-T-Y-V-QR-D-K--KE-LDKIEEI-N-S.RQKT--AAAAQQ....AAAAT...KNSS 132
HIVU455    -T-ELR--Y--V-V-Y-V-QR-D-K--KE-LNKIEEM-N-N.KQRT--.....AAANT...G.S.  124
SIVCPZ     -S-GLR--F--L-V-W-I-SD-T-E--QK-L.....E-L-R.HHGE--SKTESNSGSREGGASQGASASA 134

           p17 \/ p24
CONSENSUS1 qVSQNYPIVQNlqGQMVHQaISPRTLNAWVKVvEEKAFSPEVIPMFsALsEGATPQDLNTMLNTVGGHQA  196
HIVLAI     Q----------IQ------A-----------V-------------S--S-------------------  196
HIVHXB2R   Q----------IQ------A-----------V-------------S--S-------------------  196
HIVMN      Q----------IE------A-----------V-------------S--S-------------------  199
HIVJH3     K----------IQ------A-----------V-------------S--S-------------------  196
HIVJRCSF   Q----------LQ------A-----------I-------------S--S-------------------  196
HIVOYI     Q----------LQ------P-----------V-------------S--A-------------------  196
HIVSF2     Q----------LQ------A-----------V-------------S--S-------------------  198
HIVNY5CG   Q----------LQ------A-----------V-------------S--S-------------------  196
HIVNL43    Q----------LQ------A-----------V-------------S--S-------------------  196
HIVCDC4    Q----------LQ------A-----------I-------------A--S-------------------  196
HIVHAN     Q----------LQ------A-----------V-------------S--S-------------------  196
HIVRF      Q----------LQ------A-----------V-------------S--S-------------------  196
CONSENSUS2 qvSqNYPlVQN?CGQmvHQAlSPRTLNAWVKVlEeKAFSPEVIPMFSALSEGAtPQDlNtMLNtVGGHQa 192
HIVELI     QV-Q---I---L---MV---I-----------I-E-----------------T---L-T---T-----A  196
HIVZ2Z6    QV-Q---I---L---MV---I-----------I-E-----------------T---L-T---T-----A  197
HIVNDK     QV-Q---I---L---MV---I-----------I-E-----------------T---L-T---T-----A  193
HIVMAL     SV-Q---I---A---Ml---I-----------I-E-----------------T---L-M---I-----A  202
HIVU455    ..-Q---I---A---PV---L-----------V-D-----------------T---L-M---V-----A  197
SIVCPZ     Gl-G---L---A---MV---I-----------V-E-----------------L---V-T---A-----G  204
```

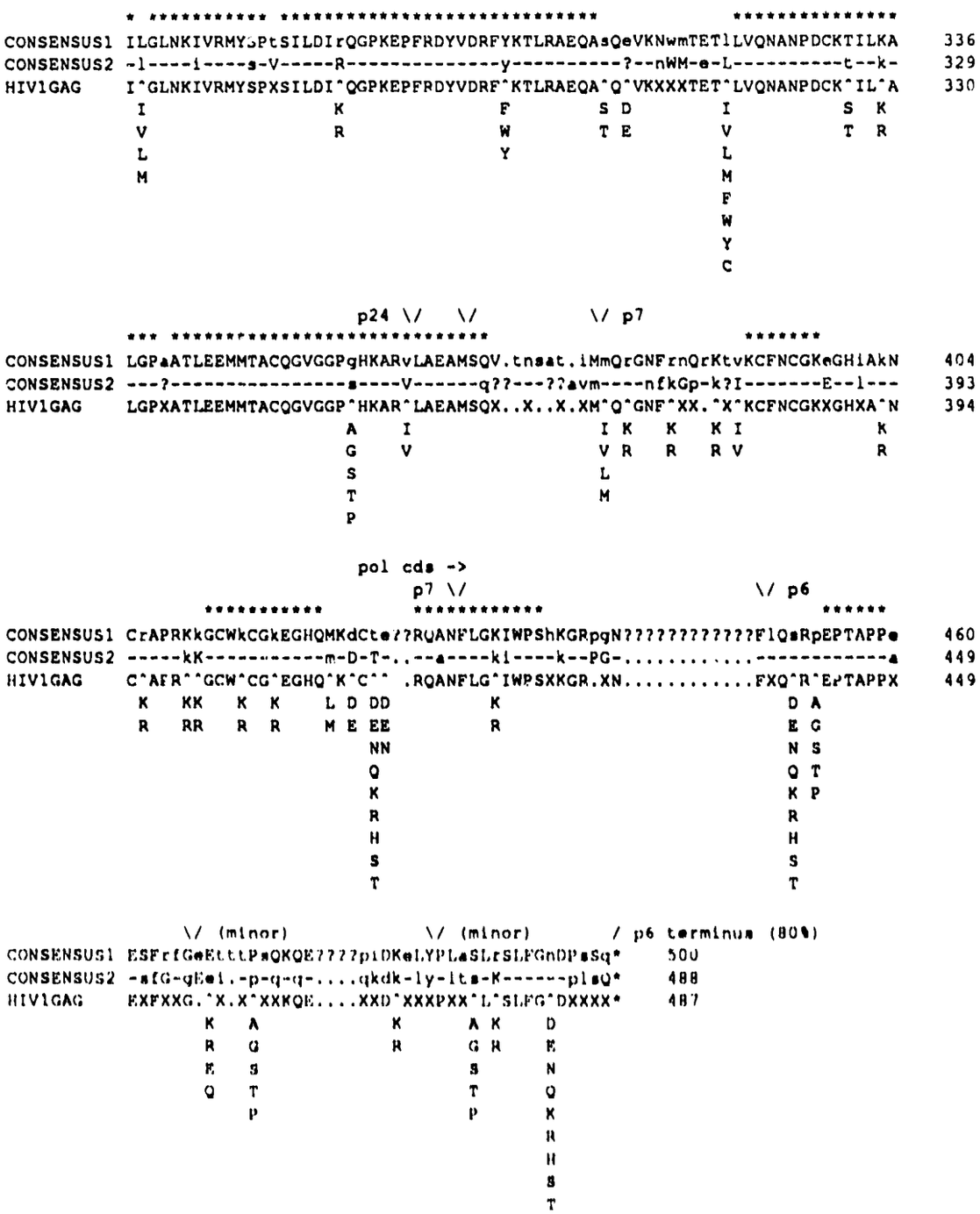
```
CONSENSUS1 AMQMLKETINEEAAEWDRlHPvhAGPIaPGQMREPRGSDIAGTTSTLQEQIGWMTnNPPiPVGEIYKRWI  266
HIVLAI     ------------------V--VH----A--------------------------N---I----------  266
HIVHXB2R   ------------------V--VH----A--------------------------N---I----------  266
HIVMN      ------------------L--VH----T--------------------------N---I----------  269
HIVJH3     ------------------L--AQ----A--------------------------S---I----------  266
HIVJRCSF   ------------------L--VH----A--------------------------N---I----------  266
HIVOYI     ------------------L--VH----A--------------------------N---I----------  266
HIVSF2     ------------------V--VH----A--------------------------N---I----------  268
HIVNY5CG   ------------------L--VH----A--------------------------H---I----------  266
HIVNL43    ------------------L--VH----A--------------------------H---I----------  266
HIVCDC4    ------------------L--VH----A--------------------------N---T----------  266
HIVHAN     ------------------L--VH----A--------------------------N---I----------  266
HIVRF      ------------------L--VH----A--------------------------N---I----------  266
CONSENSUS2 AMQmLKetINeEAAeWDRlHPvHAGPiaPGQmREPRGSDIAGTTSTlQEQI?WmTsNPPIPVG?lYkRWi  260
HIVELI     ---M--ET--E---E---L--V----IA---M--------------L----A-M-S-------EI-K--I  266
HIVZ2Z6    ---M--ET--E---E---L--V----IA---M--------------L----A-M-S-------EI-K--I  267
HIVNDK     ---M--ET--D---E---L--V----VA---M--------------L----A-M-S-------EI-K--I  263
HIVMAL     ---M--DT--E---D---V--V----IP---M--------------L----G-M-S-------DI-K--I  272
HIVU455    ---M--DT--E---E---L--V----IP---M--------------V----G-M-G-------DI-R--I  262
SIVCPZ     ---V--EV--E---E---L--T----IA---L--------------L----G-T-A-------DV-R--V  274

CONSENSUS1 ILGLNKIVRMYSPtSILDIrQGPKEPFRDYVDRFYKTLRAEQAsQeVKNwmTETlLVQNANPDCKTILKA  336
HIVLAI     -------------T-----R----------------------S-E---WM---L----------------  336
HIVHXB2R   -------------T-----R----------------------S-E---WM---L----------------  336
HIVMN      -------------S-----R----------------------S-E---RT---L----------------  339
HIVJH3     -------------S-----R----------------------S-E---WM---L----------------  336
HIVJRCSF   -------------V-----R----------------------T-E---WM---L----------------  336
HIVOYI     -------------T-----R----------------------S-D---WM---L----------------  336
HIVSF2     -------------T-----R----------------------S-D---WM---L----------------  338
HIVNY5CG   -------------T-----R----------------------S-E---WM---L----------------  336
HIVNL43    -------------T-----R----------------------S-E---WM---L----------------  336
HIVCDC4    -------------I-----R----------------------S-E---WM---L----------------  336
HIVHAN     -------------T-----K----------------------T-E---WM---L----------------  336
HIVRF      -------------I-----R----------------------S-D---WM---F----------------  336
CONSENSUS2 IlGLNKlVRMYsPVSILDIRQGPKEPFRDYVDRFyKTLRAEQAsQ?VKnWMTeTLI?QNANPDCKtILkA  329
HIVELI     -V----I----S----------------------Y--------S-D--N---E------------T--K-  336
HIVZ2Z6    -L----I----S----------------------Y--------S-E--G---E------------T--K-  337
HIVNDK     -L----I----S----------------------Y--------S-D--N---E------------T--K-  333
HIVMAL     -L----I----S----------------------F--------T-E--N---E------------T--K-  342
HIVU455    -L----I----S----------------------F--------T-D--N---E------------S--R-  332
SIVCPZ     -L----V----C----------------------Y--------S-E--N---D------------Q--K-  344

                                  p24 \/   \/          \/ p7
CONSENSUS1 LGPaATLEEMMTACQGVGGPgHKARvLAEAMSQV.tnsat.lMmQrGNFrnQrKtvKCFNCGKeGHlAkN  404
HIVLAI     ---A----------------G----V--------.TNSAT.I-M-R---RN-R-IV-------E--I-R-  404
HIVHXB2R   ---A----------------G----V--------.TNSAT.I-M-R---RN-R-IV-------E--T-R-  404
HIVMN      ---A----------------G----V--------.TNSAT.I-M-R---RN-R-II-------E--I-K-  407
HIVJH3     ---A----------------G----V--------.TNSTT.I-M-R---RN-R-II-------E--L-R-  404
HIVJRCSF   ---A----------------G----V--------.TNPAT.I-M-R---RN-R-NV-------E--I-R-  404
HIVOYI     ---A----------------G----V--------..NSVT.V-M-K---KN-R-TV-------E--I-K-  403
HIVSF2     ---A----------------G----V--------.TNPAN.I-M-R---RN-R-TV-------E--I-K-  406
HIVNY5CG   ---A----------------G----V--------.TNPAT.I-I-R---RN-K-TV-------E--I-K-  404
HIVNL43    ---G----------------G----V--------.TNPAT.I-I-K---RN-R-TV-------E--I-K-  404
HIVCDC4    ---A----------------G----V--------.TNSAT.I-M-R---RR-G-TV-------E--I-R-  404
HIVHAN     ---A----------------G----V--------.TGSAA.I-M-R---RN-R-TV-------E--I-R-  404
HIVRF      ---A----------------S----I--------.TNSAT.I-L-K---RD-R-IV-------V--I-K-  404
CONSENSUS2 LGP?ATLEEMMTACQGVGGPsHKARVLAEAMSq??tns??avmmQrGnfkGprk?IKCFNCGKEGHlAkN  393
HIVELI     ---Q----------------S----------QA.TNSVTTAMM-R-NFK-PRKI----------I-K-  405
HIVZ2Z6    ---Q----------------S----------QA.TNSAAAVMM-R-NFK-PRKT----------I-K-  406
HIVNDK     ---Q----------------G----------QV.TGSATAVMM-R-NFK-PRKS----------T-K-  402
HIVMAL     ---G----------------S----------QA.TNSTAAIMM-R-NFK-Q.KR----------L-R-  410
HIVU455    ---G----------------G----------QVQ.Q..TSIMM-R-NFR-P.RR----------L-K-  398
SIVCPZ     ---G----------------S----------MVQNQGRADVFF-K-QGA-PKRK----------L-R-  414
```

```
                                  pol cds ->
                                      p7 \/                          \/ p6
CONSENSUS1 CrAPRKkGCWkCGkEGHQMKdCte??RQANFLGKIWPShKGRpgN????????????FlQsRpEPTAPPe    460
HIVLAI     -R----K---K--K------D-TE..------------Y---PG-FLQSRPEPTAPP-L-S-P------E    472
HIVHXB2R   -R----K---K--K------D-TE..------------Y---PG-............-L-S-P------E    460
HIVMN      -R----R---K--K------D-TE..------------C---.R-............-P-S-T------E    462
HIVJH3     -R----K---K--K------D-NE..------------S---PG-............-L-S-P------E    460
HIVJRCSF   -R----K---K--K------E-TE..------------Y---PG-............-L-S-P------E    460
HIVOYI     -R----K---K--R------D-TE..------------H---PG-............-L-N-P------A    459
HIVSF2     -R----K---R--R------D-TE..------------Y---PG-............-L-S-P------E    462
HIVNY5CG   -R----K---K--K------D-TE..------------H---PG-............-L-S-P------E    460
HIVNL43    -R----K---K--K------D-TE..------------H---PG-............-L-S-P------E    460
HIVCDC4    -K----K---K--R------D-TE..------------H---PG-............-L-S-P------E    460
HIVHAN     -R----K---K--K------D-TE..------------H---PG-............-L-S-P------E    460
HIVRF      -R----K---K--K------D-TNEG------------H---PG-............-L-S-P------E    462
CONSENSUS2 CrAPRkKGCWkCGkEGHQmKDCTe..RQaNFLGkiWPShkGRPGN............FlQsRpEPTAPPa    449
HIVELI     -R---K----R--K----L----E..--A----RI---HK-----.......  ....-L-S-P------A    461
HIVZ2Z6    -R---R----K--K----L----E..--A----KI---HK-----............-L-S-P------A    462
HIVNDK     -R---K----K--R----M----E..--A----KI---HK-----............-L-S-P------A    458
HIVMAL     -R---K----K--K----M----E..--A----KI---HK-----............-L-S-P------A    466
HIVU455    -R---K----K--K----M----E..--A----KI---NK-----............-P-S-P------A    454
SIVCPZ     -K---R----R--Q----M----G..--V----KG---RS-----............-V-N-T------I    470

                \/ (minor)            \/ (minor)       / p6 terminus (80%)
CONSENSUS1 ESFrfGeEtttPsQKQE????piDKeLYPLaSLrSLFGnDPsSq*    500
HIVLAI     ---RS-V-TTT-S----....PI--E----T--R----N--S-Q-    512
HIVHXB2R   ---RS-V-TTT-P----....PI--E----T--R----N--S-Q-    500
HIVMN      ---RF-E-TTT-Y----KKQETI--D----A--K----N--L-Q-    506
HIVJH3     ---RF-E-TTT-S----....PR--E----A--R----N--S-P-    500
HIVJRCSF   ---RF-E-TAT-S----QKQEPI--E----T--R----N--S-Q-    504
HIVOYI     ---GF-E-TTT-P----....PI--G----T--R----N--S-Q-    499
HIVSF2     ---RF-E-KTT-S----....PI--E----T--R----N--S-Q-    502
HIVNY5CG   ---RF-E-TTT-S----....PI--E----A--R----S--S-Q-    500
HIVNL43    ---RF-E-TTT-S----....PI--E----A--R----S--S-Q-    500
HIVCDC4    ---RF-D-TTT-S----....PR--E----A--R----N--S-Q-    500
HIVHAN     ---RF-.-ATA-S----....PI--E----A--K----S--S-Q-    499
HIVRF      ---RF-E-T.T-S----....KI--E----A--K----N--S-Q-    501
CONSENSUS2 EsfGfgEei.tpsqKqE....qkdkelyPltsLKSLFGnDplsQ-    488
HIVELI     -SF-FG-EI.TPSQ-Q-....QKDKELY-LTS------N-PLS--    500
HIVZ2Z6    -SF-FG-EI.TPSQ-Q-....QKDKELY-STA------N-PLL--    501
HIVNDK     -SF-FG-EI.TPSQ-Q-....QKDKELY-LAS------N-PSS--    497
HIVMAL     -SF-FG-EI.KPSQ-Q-....QKDKELY-LAS------N-QLS--    505
HIVU455    -IF-MG-KM.TSPA-Q-....LKDREQT-LVS------N-PLS--    493
SIVCPZ     -SY-YQ-EE.KSQE-K-....G.ESSLY-PTS------S-PSS--    508
```

HIV1 GAG

Information dense stretches (0.93 or greater) are indicated by ****.

```
           **********        *****
CONSENSUS1 MGARASVLSGGeLDkWEkIRLRPgGKKkYkLKHiVWASReLERFAvNPgLLEtseGCRQIlgQLQPsLqT     70
CONSENSUS2 --------s-gK--?---i----G-k---rl--lv----E--rfaln-G-l------kqii---qpa?--     68
HIV1GAG    MGARASVLSGX^LDXWE^IRLRPXG^K^YXLKH^^WASR^LE^^^^NP^LLE^^^GC^QI^XQLQ^^^^T     70
                      K     D       D K     II    K  KFAI  A   SAD  K  I    AAIK
                      R     E       E R     VV    R  RWGV  G   TGE  R  V    GGVR
                      E     N       N E     L     E   YSL  S    SN  E  L    SSLE
                      Q     Q       Q Q     M     Q    TM  T    TQ  Q  M    TTMQ
                            K       K                  P   P    PK          PP
                            R       R                           R
                            H       H                           H
                            S       S                           S
                            T       T                           T

                         *****      *****
CONSENSUS1 GSEElrSLyNtvAtLYCVHQrIevkDTKEALdKiEEEQnkS.kKKAQQ??........aAAdt???Gnss    126
CONSENSUS2 -s-e------T----y-v-?--dV---ke-lekmeee--Ks?-kkt----????????-aa-----g---    123
HIV1GAG    G^EEX^SL^N^^AXLYCVH^XI^^^DTKEA^^K^EEXQ^^^.X^^^QQ..........XAAX......X.    122
            S   D  F AI       K  DIK     ID I    DDD  KKA
            T   E  W GV       R  EVR     VE V    EEE  RRG
                N  Y S        E          LN L    NNN  E S
                Q    T        Q          M  M    QQQ  Q T
                K    P                           KKK    P
                R                                RRR
                H                                HHH
                S                                SSS
                T                                TTT

            p17 \/ p24
             *********          ************************    *********     ******
CONSENSUS1 qVSQNYPIVQNlqGQMVHQaISPRTLNAWVKVvEEKAFSPEVIPMFsALsEGATPQDLNTMLNTVGGHQA    196
CONSENSUS2 -v-q---i---?Q--mv--Ai-----------i-e-----------S--S---t---l-t---t-----a    192
HIV1GAG    ..SQNYPIVQNX^GQX^HQ^^SPRTLNAWVKV^E^KAFSPEVIPMF^AL^EGATPQDLNXMLNXVGGHQA    190
                       E   I  AI           I D           A  A
                       Q   V  GV           V E           G  G
                              SL                         S  S
                              TM                         T  T
                              P                          P  P

           ******************          ****************** ********     **********
CONSENSUS1 AMQMLKETINEEAAEWDRlHPvhAGPIaPGQMREPRGSDIAGTTSTLQEQIGWMTnNPPiPVGEIYKRWI    266
CONSENSUS2 ---m--et--e---e-------H---i----m-------------l----?-m-s---l---?i-k--i    260
HIV1GAG    AMQMLK^TIN^EAA^WDR^HPX^AGP^^PGQMREPRGSDIAGTTST^QLQI^WMTXNPPXPVG^IY^RWI    260
                 D   D   D   I   D   IA                  I    A           D  K
                 E   E   E   V   E   VG                  V    G           E  R
                             L   N    S                  L
                             M   Q    T                  M
                                 K    P
                                 R
                                 H
                                 S
                                 T
```


CONSENSUS1 ILGLNKIVRMYSPtSILDIrQGPKEPFRDYVDRFYKTLRAEQAsQeVKNwmTETlLVQNANPDCKTILKA 336
CONSENSUS2 -l----i----s-V-----R----------y----------?--nWM-e-L----------t--k- 329
HIV1GAG I^GLNKIVRMYSPXSILDI^QGPKEPFRDYVDRF^KTLRAEQA^Q^VKXXXTET^LVQNANPDCK^IL^A 330
p24
p7
CONSENSUS1 LGPaATLEEMMTACQGVGGPgHKARvLAEAMSQV.tnsat.lMmQrGNFrnQrKtvKCFNCGKeGHiAkN 404
CONSENSUS2 ---?----------------s----V------q??---??avm----nfkGp-k?I-------E--l--- 393
HIV1GAG LGPXATLEEMMTACQGVGGP^HKAR^LAEAMSQX..X..X.XM^Q^GNF^XX.^X^KCFNCGKXGHXA^N 394
pol cds ->
p7
p6
CONSENSUS1 CrAPRKkGCWkCGkEGHQMKdCte??RQANFLGKIWPShKGRpgN????????????FlQsRpEPTAPPe 460
CONSENSUS2 -----kK----------m-D-T-..--a----k1----k--PG-..........----------a 449
HIV1GAG C^AFR^^GCW^CG^EGHQ^K^C^^ .RQANFLG^IWPSXKGR.XN.....FXQ^R^EpTAPPX 449
(minor)
(minor)
p6 terminus (80%)
CONSENSUS1 ESFrfGeEtttPsQKQE????p1DKeLYPLaSLrSLFGnDPsSq* 500
CONSENSUS2 -sfG-qEel.-p-q-q-....qkdk-ly-lts-K------plsQ* 488
HIV1GAG EXFXXG.^X.X^XXKQE....XXD^XXXPXX^L^SLFG^DXXXX* 487

HIV1GAG:0 MGARASVLSG

Myristoylation site (Schultz et al., Ann. Rev. Cell Biol. 4: 611-47, 1988).

Together with the following peptide, IRLRP, a 23-residue peptide with a potential amidated C-terminus appears to mimic a gonadoliberin molecule (Glass, J. Theor. Biol. 150:489-496, 1991).

HIV1GAG:18 IRLRP

"IRLR" is part of a conformational neutralization epitope. Papsidaro et al., J. Virol. 63:267-272 (1989).

See above; the LRPG is common to all known gonadoliberin peptides.

HIV1GAG:84 LYCVH

"YCVH" is part of the peptide HGP-30, which can serve as a neutralizing antibody epitope (Sarin et al., Science 232: 1135-37 (1986) and Naylor et al., PNAS 84:2951-55 (1987)), and as a cytotoxic T-cell epitope (Achour et al., PNAS 87: 7045-49 (1990)).

HIV1GAG:95 DTKEA

This fragment is part of the peptide HGP-30, which can serve as a neutralizing antibody epitope (Sarin et al., Science 232: 1135-37 (1986) and Naylor et al., PNAS 84:2951-55 (1987)), and as a cytotoxic T-cell epitope (Achour et al., PNAS 87: 7045-49 (1990)).

This p17 fragment also shows similarity to thymosin alpha one (Sarin et al., Science 232: 1135-37, 1986) and to the human heavy neurofilament subunit (Myers, Korber and Nelson, T-Division, LANL, unpublished).

HIV1GAG:127 SQNYPIVQN

The p17/p24 cleavage site is located within this peptide Henderson et al., J. Virol. 62:2587-95, 1988).

HIV1GAG:146 SPRTLNAWVKVaEkKAFSPEVIPMF

*This fragment is contained within a large conserved pattern of HIV2, that also contains the following two HIV1 patterns, GAG:175 and GAG:189. The HIV-2 pattern is: TLNAWVKLvKeKKFG AEVVpGFQALSEGCtPYDINQMLNCVGdHQAAMQIIReIINeEAA.

HIV1GAG:175 EGATPQDLN

*See HIV1GAG:146

HIV1GAG:189 VGGHQAAMQMLKkTINkEAAkWDR

An insertion in the "EAAE" causes a dominant mutation in p24 which impairs cellular capacity to support HIV replication of wild-type virus. Trono et al., Cell 59:113-120 (1989).

Contains a cytotoxic T-cell epitope. Claverie et al., Eur. J. Immunol. 121:547-51 (1990).

*See HIV1GAG:146

HIV1GAG:223 PGQMREPRGSDIAGTTST
An insertion in the "TST" causes a dominant mutation in p24 which impairs cellular capacity to support HIV replication of wild-type virus. Trono et al., Cell 59:113-120 (1989). "PGQMREPRG" is part of a cytotoxic T-cell epitope. Claverie et al., Eur. J. Immunol. 12:1547-53 (1990).

This fragment contains a region identical to a conserved pattern in HIV2, "GSDIAGTTS".

HIV1GAG:242 QEQIhWMT

HIV1GAG:255 PVGkIYnRWII
Part of this fragment is similar to a conserved pattern in HIV2, "IYRRWIQ?GLQKCVR".

HIV1GAG:267 GLNKIVRMYSP
"GLNKIVRMY" is part of a cytotoxic T-cell epitope. Nixon et al., Nature 336:484-487 (1988).

Part of this fragment is similar to a conserved pattern in HIV2, "IYRRWIQ?GLQKCVR".

HIV1GAG:279 SILDInQGPKEPFRDYVDRFdKTLRAEQA
This fragment contains a region identical to a conserved pattern in HIV2, "QGPKE".

Part of this fragment is similar to a conserved pattern in HIV2, "FQSYVDRFYKSLRAEQTDpAVKNWMTQTLL".
Both T- and B-cell epitopes are reported to be contained within this fragment by Chong et al., FEBS 264: 231-234, 1990.

HIV1GAG:320 LVQNANPDCKiILnALGP
Contains a cytotoxic T-cell epitope. Gotch et al., Internat. Immunol. 2:707-712 (1990).

Part of this fragment is similar to a conserved pattern in HIV2, "NPDCKLVLKGLG".

HIV1GAG:339 ATLEEMMTACQGVGGPjHKARaLAEAMSQ
Contains the p24 C-terminus (Henderson et al., J. Virol. 62: 2587-95, 1988).

Part of this fragment is similar to a conserved pattern in HIV2, "NPTLEEMLT".

HIV1GAG:386 KCFNCGK
Crucial zinc finger motif in the p7 protein required for normal RNA packaging and infectivity (Gorelick et al., J. Virol. 64: 3207-11, 1990).

Part of this fragment is similar to a conserved pattern in HIV2, "WNCGK".

HIV1GAG:407 GCWnCGnFCHQ

Part of this fragment is similar to a conserved pattern in HIV2, "PRRQGCWKCG".

HIV1GAG:425 RQANFLGnIWPS

This segment is in the region where a ribosomal frame shift occurs that allows the gag and pol genes to overlap in different reading frames. The shift occurs at the "L". Jacks et al., Nature 331:280-283 (1988).

"RQANF" is part of a cytotoxic T-cell epitope. Claverie et al., Eur. J. Immunol. 12:1547-53 (1990).

Part of this fragment is similar to a conserved pattern in HIV2, "RQAGFLG".

HIV1GAG:451 EPTAPP

Part of a cytotoxic T-cell epitope -- peptide of Claverie et al., Eur. J. Immunol. 18:1547-1553 (1988).

```
CONSENSUS  MGarnSVL?GKkaDELEkiRLRPgGKKkY?LKH?VWAaNeLDrFGLaESLLesKEGCQKIlsVLaPlVPT     67
HIV2ROD    --ARN---R--KA----RI----G---K-R---I---A-K--R---A----ES-------LT--D-M---     70
HIV2NIHZ   --ARN---R--KA----KI----G---K-K---I---A-E--R---A----ES-------LT--D-L---     70
HIV2ISY    --AKN---R--KA----KI----G---K-R---I---A-E--R---T----ES-------IS--E-L---     70
HIV2ST     --ARN---R--KA----KI----G---K-R---I---A-E--R---A----ES-------LT--D-L---     70
HIV2BEN    --ARN---R--KA----KV----G---K-R---I---A-E--K---A----ES-------LR--D-L---     70
HIV2D194   --ARN---R--KA----KV----N---R-R---V---A-E--R---A----ES-------LK--E-L---     70
HIV2GH1    --ARN---R--KA----KI----S---K-R---I---A-E--K---A----ES-------LT--D-L---     70
HIV2D205   --ARG---S--KT----KV----G---K-M---V---V-E--R---A----ES-------LK--A-L---     70
SIVMM251   --ARN---S--KA----KI----G---K-M---V---A-E--R---A----EN-------LS--A-L---     70
SIVMM142   --ARN---S--KA----KI----G---K-M---V---A-E--R---A----EN-------LS--A-L---     70
SIVMM239   --VRN---S--KA----KI----N---K-M---V---A-E--R---A----EN-------LS--A-L---     70
SIVMNE     --ARN---S--KA----KI----G---K-M---V---A-E--R---A----EN-------LS--A-L---     70
SIVSMMH4   --ARN---S--EA----KV----N---K-M---V---A-E--R---A----DN-------LS--A-L---     70
SIVSMMPBJ  --ARN---S--KA----KI----G---R-Q---I---A-E--R---A----EN-------LS--A-L---     70

                                                                        p16 \/ p28
CONSENSUS  GSENLKSL?NtvCViwCiHAEEKvKdTEeAKqivqRHLvaeTgtaekMP?tsrPTAPpSg?ggNyPVQq?    133
HIV2ROD    --------F-TV--IW-I-----V-D--G--QIVR---VAE-GTAEK--STSR----S-EKGG-Y---H.    139
HIV2NIHZ   --------F-TV--IW-I-----V-D--G--QIVQ---VAE-GTAEK--NTSR----P-GK..-F---Q.    137
HIV2ISY    --------Y-TT--IW-L-----V-D--E--RIVG---VAE-ETAEK--NISR----P-GKGG-F---Q.    139
HIV2ST     --------F-TV--IW-I-----A-D--E--QKVQ---VAE-KTTEK--STSR----P-GNGG-F---Q.    139
HIV2BEN    --------F-TV--IW-L-----V-D--E--KLAQ---VAE-GTAEK--NTSR----P-GKRG-Y---Q.    139
HIV2D194   --------F-TV--IW-L-----V-D--E--KLAQ---VAE-GTAEK--NISR----P-GKGG-F---Q.    139
HIV2GH1    --------F-TV--IW-L-----V-D--E--KLVQ---GAE-GTAEK--STSR----P-GRGR-F---QT    140
HIV2D205   --------F-IV--IF-L-----V-D--E--KIAQ---AAD-...EK--ATNK----P-G..G-Y---Q.    134
SIVMM251   --------Y-TV--IW-I-----V-H--E--QIVQ---VVE-GTAET--KTSR----S-GRGG-Y---Q.    139
SIVMM142   --------Y-TV--IW-I-----V-H--E--QIVQ---VME-GTAET--KTSR----F-GRGG-Y---Q.    139
SIVMM239   --------Y-TV--IW-I-----V-H--E--QIVQ---VVE-GTTET--KTSR----S-GRGG-Y---Q.    139
SIVMNE     --------Y-TV--IW-I-----V-H--E--QIVQ---VVE-GTAET--KTSR----S-GRGG-Y---Q.    139
SIVSMMH4   --------Y-TV--IW-I-----V-H--E--QIVQ---VVE-GTADR--ATSR----P-GRGG-Y---Q.    139
SIVSMMPBJ  --------Y-TV--LW-I-----V-H--E--QIVQ---VVE-GTADK--ATSR----P-GKGG-Y---Q.    139
          (p28 cleavage site determined for sivmac and sivmne; presumed to apply to HIV2)

CONSENSUS  ?qGNY?HlPLSPrTLNAWVKLvEeKKFGAEVVpGFQALSEGCtPYDINQMLNCVGdHQAAMQIIReIINe    201
HIV2ROD    VG---T-I----R--------V-E--------P---------T----------D---------E---E    209
HIV2NIHZ   VA---T-I----G--------V-E--------P---------T----------D---------E---E    207
HIV2ISY    IG---V-L----R--------V-E--------P---------T----------D---------E---E    209
HIV2ST     VA---T-V----R--------V-E--------P---------T----------D---------E---E    209
HIV2BEN    AG---V-V----R--------V-E--------P---------T----------D---------E---E    209
HIV2D194   AG---I-V----R--------V-E--------P---------T----------D---------E---E    209
HIV2GH1    GG---I-V----R--------V-D--------P---------T----------D---------E---D    210
HIV2D205   LA---V-L----R--------V-E--------P---------T----------E---------E---E    204
SIVMM251   IG---V-L----R--------I-E--------P---------T----------D---------D---E    209
SIVMM142   IG---T-L----R--------I-E--------S---------L----------D---------D---E    209
SIVMM239   IG---V-L----R--------I-E--------P---------T----------D---------D---E    209
SIVMNE     VG---T-L----R--------I-E--------P---------T----------E---------E---E    209
SIVSMMH4   VG---V-L----R--------V-E--------P---------T----------E---------E---E    209
SIVSMMPBJ  IG---T-L----R--------I-E--------P---------T----------E---------E---E    209
```

HIV2/SIV GAG

```
CONSENSUS  EAAdWD?qHPlpgPlpaGQLRePrGSDIAGTTStV?EQIqWMyRqqNPvPVGnIYRRWIQ?GLQKCVRmY  268
HIV2ROD    ---E--VQ--IPG-LPA----E-R---------T-E---Q--F-PQ--V---N-------I-------M-  279
HIV2NIHZ   ---D--VA--IPG-LPA----E-R---------T-E---Q--F-PQ--V---N-------I-------M-  277
HIV2ISY    ---D--VQ--IPG-LPA----D-R---------T-E---E--Y-QE--V---N-------I-------M-  279
HIV2ST     ---D--AQ--IPG-LPA----E-R---------T-E---Q--F-PQ--V---S-------I-------M-  279
HIV2BEN    ---D--SQ--IPG-LPA----D-R---------T-D---Q--Y-PQ--V---N-------I-------K-  279
HIV2D194   ---D--AQ--IPG-LPA----D-R---------T-D---Q--Y-QP--V---N-------I-------M-  279
HIV2GH1    ---D--AQ--IPG-LPA----D-R---------T-E---Q--Y-PQ--V---N-------I-------M-  280
HIV2D205   ---D--QQ--SPG-MPA----D-R---------T-E---Q--Y-AQ--V---N-------L-------M-  274
SIVMM251   ---D--LQ--QPA-.QQ----E-S---------S-D---Q--Y-QQ--I---N-------L-------M-  278
SIVMM142   ---D--LQ--QQA-.QQ----E-S---------T-E---Q--Y-QQ--I---N-------L-------M-  278
SIVMM239   ---D--LQ--QPA-.QQ----E-S---------S-D---Q--Y-QQ--I---N-------L-------M-  278
SIVMNE     ---D--LQ--QQA-.QQ----E-S---------T-D---Q--Y-QQ--I---N-------L-------M-  278
SIVSMMH4   ---D--LQ--QPG-LPA----E-R---------T-D---Q--Y-QQ--I---N-------L-------M-  279
SIVSMMPBJ  ---D--LQ--QPG-IPP----E-R---------T-D---Q--Y-QQ--I---N-------L-------M-  279

CONSENSUS  NPTnILDvkQGPKEpFQSYVDRFYKSLRAEQTDpAVKNWMTQTLLiQnaNPDCKLVLKGLGmNPTLEEML  338
HIV2ROD    ---N---IK-----P-----------------P----------V-NA----------M--------  349
HIV2NIHZ   ---N---IN-----P-----------------P----------I-NA----------M--------  347
HIV2ISY    ---N---IK-----S-----------------A----------V-.S----------M--------  348
HIV2ST     ---N---IK-----P-----------------P----------V-NA----------I--------  349
HIV2BEN    ---N---IK-----P-----------------P----------I-NA----------M--------  349
HIV2D194   ---N---VK-----S-----------------P----------I-NA----------M--------  349
HIV2GH1    ---N---VK-----P-----------------P----------I-NA----------M--------  350
HIV2D205   ---N---IK-----P-----------------P----------I-NA----------M--------  344
SIVMM251   ---N---VK-----P-----------------A----------I-NA----------V--------  348
SIVMM142   ---N---VK-----P-----------------P----------I-NA----------T--------  340
SIVMM239   ---N---VK-----P-----------------A----------I-NA----------V--------  348
SIVMNE     ---S---VK-----P-----------------P----------I-NA----------M--------  348
SIVSMMH4   ---N- -VK-----P-----------------P----------I-NA----------M--------  349
SIVSMMPBJ  ---N---VK-----P-----------------P----------I-NA----------I--------  349

                  p28 \/                \/ p8
CONSENSUS  TaCQGvGGPqQKARLMAEALKeal?papiPFAAaQq????rk?ikcWNCGKeGHaArQcraPRRQGCWKC  402
HIV2ROD    -A---V---G----------EVIGPAPI----A-Q....RKAFKC-----E--S-R-CRA---------  415
HIV2NIHZ   -T---V---G----------EVMAPAPI----A-Q....RKTFKC-----E--S-R-WSA---------  413
HIV2ISY    -A---I---G----------EAMRPAPI----A-Q....KRAIKC-----E--S-R-CRA---------  414
HIV2ST     -A---V---G----------EAMAPAPI----A-Q....RRTIKC-----E--S-R-CRA---------  415
HIV2BEN    -A---V---G----------EAMGPSPI----A-Q....RKAIRY-----E--S-R-CRA---------  415
HIV2D194   -A---V---S----------EALTPAPI----A-Q....RKAIRC-----E--S-K-CRA---------  415
HIV2GH1    -A---V---G----------EALTPPPI----A-Q....RKVIRC-----E--S-R-CRA---------  416
HIV2D205   -A---I---G----------EALTPAPI----V-QKAGKRGTVTC-----Q--T-R-CRA---------  414
SIVMM251   -A---V---G----------EALAPVPI----A-KRGP.RKPIKC-----E--S-R-CRA---------  417
SIVMM142   -A---V---G----------EALAPAPI----A-QKGP.RKPIKC-----E--S-R-CRA---------  417
SIVMM239   -A---V---G----------EALAPVPI----A-QRGP.RKPIKC-----E--S-R-CRA---------  417
SIVMNE     -A---V---G----------EALAPGPL----A-QKGP.RKPIKC-----E--S-R-CRT---------  417
SIVSMMH4   -A---V---G----------EALRPDQL----V-QKG.QRKTIKC-----E--S-K-CRA---------  418
SIVSMMPBJ  -A---V---G----------DALTQGPL----V-QKG.QRKIIKC-----E--S-R-CRA---------  418
```

```
                pol cds ->
                          p8 \/p1         p1\/ p6
CONSENSUS   Gk?gHvMakCPeRQAGFLGlGpwGKkpRNFPmaq?pqGl??????????????????tPtAPP?dPavdl     451
HIV2ROD     -KPG-I-TN--D-------L-PW--KP----VAQVPQ-L.................T-T---VD-AVDL     467
HIV2NIHZ    -KSG-V-AN--D-------I-HW--KA----AAQVPQ-L.................T-T---LD-AVDL     465
HIV2ISY     -KSG-I-AN--D-------L-PW--KP----VVPSSQ-L.................T-T---MD-AVDL     466
HIV2ST      -KAG-I-AK--E-------L-PW--KP----VAQIPQ-L.................T-T---ID-VEDL     467
HIV2BEN     -KPG-I-AN--E-------L-PR--KP----VTQAPQ-L.................I-T---AD-AAEL     467
HIV2D194    -KSG-I-AN--E-------M-PR--QP----AAQAPQ-L.. ..............I-T---ID-AVDL     467
HIV2GH1     -KTG-V-AK--E-------M-PW--KP----VAQAPP-L.................I-T---AD-AVDL     468
HIV2D205    -KTG-I-SK--E-------L-PW--KP----MTQVPQ-VTPSAPPMNPAEGMTPRGAT-S---AD-AVEM     484
SIVMM251    -KMD-V-AK--D-------L-PW--KP----MAQVHQ-L.................T-T---ED-AVDL     469
SIVMM142    -KMD-V-AK--N-------L-PW--KP----MAQVHQ-L.................T-T---EE-AVDL     469
SIVMM239    -KMD-V-AK--D-------L-PW--KP----MAQVHQ-L.................M-T---ED-AVDL     469
SIVMNE      -QMG-V-AK--D-------F-PW--KP----MAQMHQ-L.................T-T---ED-AVDL     469
SIVSMMH4    -KTG-V-AK--E-------L-PW--KP----MAQMPQ-L.................I-T---ED-AVDL     470
SIVSMMPBJ   -KAG-V-AK--E-------L-PW--KP----MAQMPQ-L.................T-T---ED-AVDL     470

           \/ (HIV2 only; not minor)          \/ (HIV2 only; not minor)/ p6 terminus
CONSENSUS   L??YMqqG??QRE?q???rerPYKEV??????????????????TEDLLHLNSLFG?DQ*     483
HIV2ROD     -EK--QQ-KR---.Q...RER-----TEDLLHLEQGETPYREPP-----------K---     522
HIV2NIHZ    -EK--QQ-KR---.Q...RER-----TEDLLRFEQAETPCRE.T-----------K---     519
HIV2ISY     -EK--QQ-RK---.Q...RQR-----TEDLLHLEQGETPHRE.T-----------N---     520
HIV2ST      -EK--QQ-KR---.Q...RER-----TEDFLQLEKQETPCRE.T-----------K---     521
HIV2BEN     -ER--QQ-RK---.Q...RER-----TEDLLHLEQRETPHRE.E-----------K---     521
HIV2D194    -EK--QQ-RK---.Q...RER-----TEDLLHLEQGETPHRG.A-----------K---     521
HIV2GH1     -ER--QQ-RE---.Q...RER-----TEDLLHLEQGKAPHRE.A-----------K---     522
HIV2D205    -KS--QM-RQ---....SRER-----.................-----------E---     521
SIVMM251    -KN--QL-KQ---....SREK-----.................-----------G---     506
SIVMM142    -KN--HL-KQ---....SRGK-----.................-----------G---     506
SIVMM239    -KN--QL-KQ---KQRESREK-----.................-----------G---     510
SIVMNE      -KN--QL-KQ---....SKRK-----.................-----------E---     506
SIVSMMH4    -KN--KM-RK---....NRER-----.................-----------E---     507
SIVSMMPBJ   -KN--KM-RR---....NRER-----.................-----------E---     507
```

HIV2/SIV GAG

```
   Information dense stretches (0.93 or greater) are indicated by ****.

                                                                  *******        ***
CONSENSUS  MGarnSVL?GKkaDELEkiRLRPgGKKkY?LKH?VWAaNeLD:FGLaESLLesKEGCQKIlsVLaPlVPT      67
HIV2SIVGAG MGX^XSVL^GK^^DELE^^RLRPXGKK^YXLKH^VWAXN^LD^FGL^ESLL^^KEGCQKI^^VLXP^VPT      70
              K    D  KA    KI        K     I     K  K   A    DD       ID    L
              R    E  RG    RV        R     V     R  R   G    EE       VE    M
                   N  ES                          E      S     N       LN
                   Q  QT                          Q      T     Q       MQ
                   K   P                                 P     K        K
                   R                                             R        R
                   H                                             H        H
                   S                                             S        S
                   T                                             T        T

                                                                        p16 \/ p28
           ********          *****
CONSENSUS  GSENLKSL?NtvCViwCiHAEEKvKdTEeAKqivqRHLvaeTgtaekMP?tarPTAPpSg?ggNyPVQq?     133
HIV2SIVGAG GSENLKSL^NXXCV^^C^HAEEKXK^TEXAK^XXXRHLXX^T...^^MPXX^^PTAPXSX...N^PVQ^.     133
                   F      IF I       D     K        D    DD    DK           F   D
                   W      VW V       E     R        E    EE    ER           W   E
                   Y      LY L       N     E             N     N            Y   N
                          M  M       Q     Q             Q     Q                Q
                                     K                   K     K                K
                                     R                   R     R                R
                                     H                   H     H                H
                                     S                   S     S                S
                                     T                   T     T                T
        (p28 cleavage site determined for sivmac and sivmne; presumed to apply to HIV2)

                     ************************************************************
CONSENSUS  ?gGNY?HlPLSPrTLNAWVKLvEeKKFGAEVVpGFQALSEGCtPYDINQMLNCVGdHQAAMQIIReIINe     201
HIV2SIVGAG X^GNYXH^PLSP\TLNAWVKL^E^KKFGAEVV^GFQALSEGCXPYDINQMLNCVG^HQAAMQIIR^IIN^     203
            A     I            I D       A                        D         D   D
            G     V            V E       G                        E         E   E
                  L                      S
                  M                      T
                                         P

           ***                  *********                    ***************
CONSENSUS  EAAdWD?qHPipgPlpaGQLRePrGSDIAGTTStV?EQIqWMyRqqNPvPVGnIYRRWIQ?GLQKCVRmY     268
HIV2SIVGAG EAA^WDXXHPXX^P.XXGQLR^P^GSDIAGTTS^V^EQI^WM^RXXNP^PVG^IYRRWIQ^GLQKCVRXY     272
              D        A        D D         S D   E  F      I   D       I
              E        G        E E         T E   Q  W      V   E       V
                                  N                  Y          N       L
                                  Q                             Q       M
                                  K                             K
                                  R                             R
                                  H                             H
                                  S                             S
                                  T                             T
```

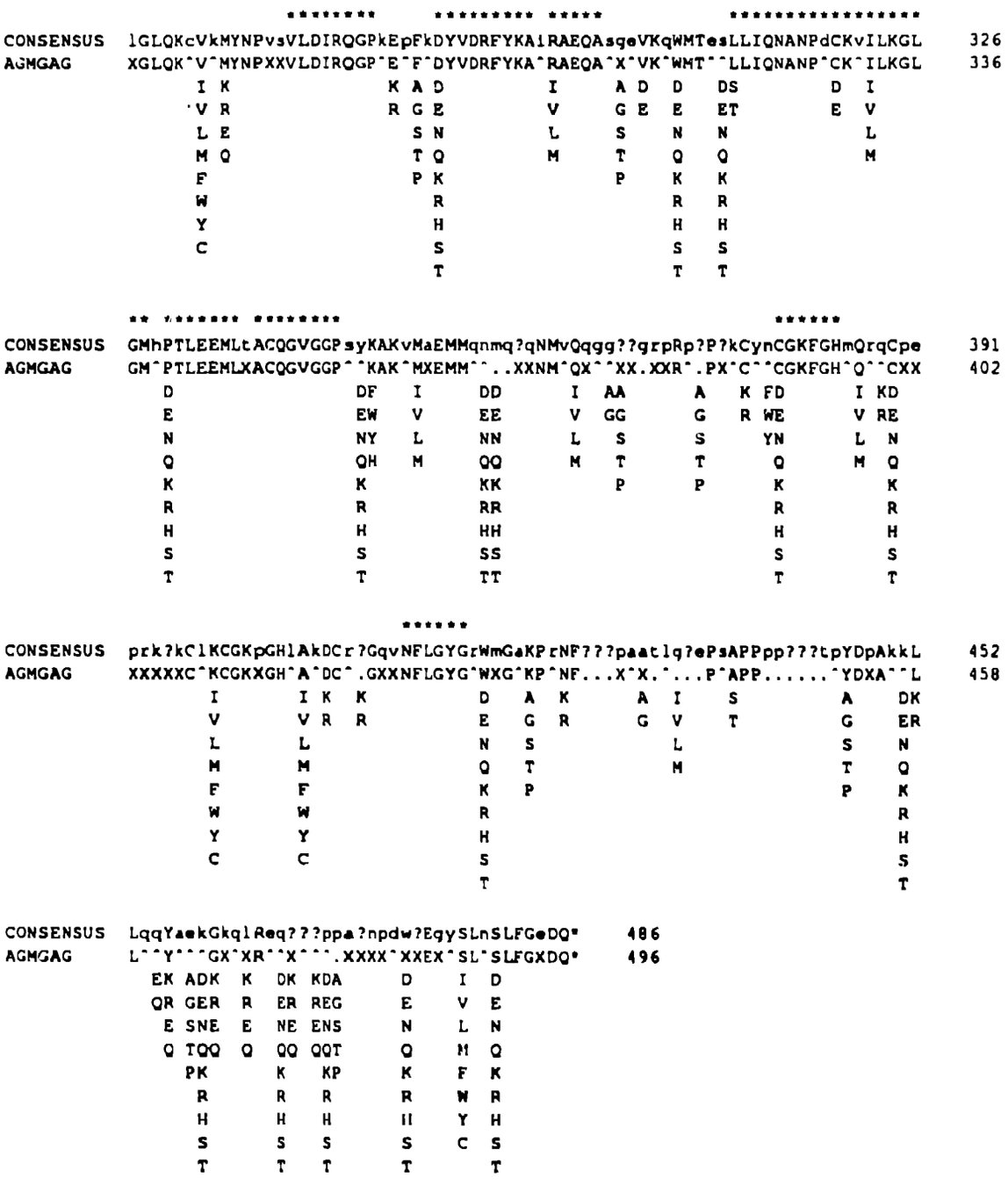
```
                    ***** ******************************    ************ ********
CONSENSUS  NPTnILDvkQGPKEpFQSYVDRFYKSLRAEQTDpAVKNWMTQTLLiQnaNPDCKLVLKGLGmNPTLEEML    338
HIV2SIVGAG NPT^ILD^^QGPKE^FQSYVDRFYKSLRAEQTD^AVKNWMTQTLL^Q.^NPDCKLVLKGLGXNPTLEEML    341
              D   ID     A                  A           I  A
              E   VE     G                  G           V  G
              N    N     S                  S              S
              Q    Q     T                  T              T
              K    K     P                  P              P
              R    R
              H    H
              S    S
              T    T

                      p28 \/                \/ p8
           *          ***********                        *****          *********
CONSENSUS  TaCQGvGGPgQKARLMAEALKeal?papiPFAAaQq????rk?ikcWNCGKeGHsArQcraPRRQGCWKC    402
HIV2SIVGAG T^CQG^GGP^QKARLMAEALK^X^XXXX^PFAAXQ^....^XX^^^WNCGK^GH^A^Q^^^PRRQGCWKC    407
            A   I   A           D I    I      K    K  IDI     E  S K IDA
            G   V   G           E V    V      R    R  VEV     Q  T R VEG
            S       S             L    L      E       LNL            LNS
            T       T             M    M      Q       MQM            MQT
            P       P                                  FKF            FKP
                                                       WRW            WR
                                                       YHY            YH
                                                       CSC            CS
                                                        T              T

              pol cds ->
                       p8 \/p1        p1\/ p6
           *           *******
CONSENSUS  Gk?gHvMakCPeRQAGFLGlGpwGKkpRNFPmaq?pqGl?????????????????tPtAPP?dPavdl    451
HIV2SIVGAG G^XXH^M^^CP^RQAGFLG^GXXGK^^RNFPXXXXXXG^.................XP^APPX^PXX^^    459
            K   I AD  D       I     KA           I                    S    D   DL
            R   V GE  E       V     RG           V                    T    E   EM
            E     SN  N       L     ES           L
            Q     TQ          M     QT           M
                  PK          F      P
                   R          W
                   H          Y
                   S          C
                   T

          \/ (HIV2 only; not minor)          \/ (HIV2 only; not minor)/ p6 terminus
                                 *****                  ************
CONSENSUS  L^?YMqqG??QRE?q???rerPYKEV?????????????????TEDLLHLNSLFG?DQ*    483
HIV2SIVGAG L^^YM^XG^^QRE.....^X^PYKEV.................TEDLLHLNSLFGXDQ*    495
            KD  D  KK        K K
            RE  E  RR        R R
            EN  N   E
            QQ  Q   Q
             K  K
             R  R
             H  H
             S  S
             T  T
```

```
CONSENSUS  MGaatSALnrRqLD?FEhIRLRPnGKKKYQiKHLIWAGKeMeRFGLHEkLLEteEGCkkIIEVL?PLEPT   68
SIVAGMTYO  --AAT---NR-Q--Q--K-----N------I--------E-E------R---TE---KR-----Y-----   70
SIVAGM155  --AAT---NR-Q--E--H-----N------I--------K-D------K---TE---KK-----S-----   70
SIVAGM3    --AAT---NR-Q--K--H-----T------I--------E-E------R---SE---KK-----Y-----   70
SIVAGM692  --SGS---SG-K--Q--H-----N------L--------E-E------K---TE---KK-----L-----   70
SIVAGM677  --GGH---SG-S--T--K-----N------I--------E-E------K---TK---QK-----T-----   70

CONSENSUS  GSEGlKsLfNLvCVl?C?H?e?kVKDTEEAva?vrQ?cHLVdKe??A????t????gqkKn?tg???ppG  118
SIVAGMTYO  ----L-S-F--V--LY-L-KEQK-------VATVR-HC---E-EKS-....TETSSGQK-NDKGIAAPP-  136
SIVAGM155  ----M-S-Y--V--LL-V-QEKK-------LAIVR-CC---D-EKT-V...T.PPGGQQ-NNTGGTATP-  136
SIVAGM3    ----L-S-F--V--LF-V-KDKE-------VAIVR-CC---E-ERN-ERNTTETSSGQK-NDKGVTVPP-  140
SIVAGM692  ----L-S-F--T--IC-I-QEAK-------VIRIK-QC---D-GEN-AK.......GID-..TT.PTPS-  130
SIVAGM677  ----L-A-F--C--IW-I-AEQK-------VVTVK-HY---D-NEK-AK.........K-NETT.APPG-  130

CONSENSUS  ?SqNfPaqqQgNaWvHvPLSPRTLNAWVKavEEKkfGAEiVPMFQALSEGCtpYDiNQMLNVlGDHQGAL  187
SIVAGMTYO  C-Q-F-AQQ-G-A-V-V------------AV---KF---I----------TP--I------L-------  206
SIVAGM155  G-Q-F-AQQ-G-A-V-V------------AV---KF---I----------TP--I------L-------  206
SIVAGM3    G-Q-F-AQQ-G-A-I-V------------AV---KF---I----------TP--I------L-------  210
SIVAGM692  R-Q-Y-AQQ-N-V-V-V------------VI---KF---I----------TP--V------L-------  200
SIVAGM677  E-Q-Y-VVN-N-A-V-Q------------CV---RW---V----------LS--V------I-------  200

CONSENSUS  QIvKEiINEEAAqWD?tHppPAGPLPAGQLRDPrGSDIAGTTStvQEQlEWiytANPRvDVGAiYRrWiI  256
SIVAGMTYO  --V--I------Q--VT-PL-----------R---------SV---L--IYT----V----I--R-I-  276
SIVAGM155  --V--I------Q--VT-PP-----------G---------TV---L--IYT----V----I--R-I-  276
SIVAGM3    --V--I------Q--IA-PP-----------R---------TV---L--IYT----V----I--R-I-  280
SIVAGM692  --V--V------Q--IT-PP-----------R---------TI---L--IYT----I----I--R-V-  270
SIVAGM677  --L--V------E--RT-RP-----------T---------SI---I--TFN----I----Q--K-V-  270

CONSENSUS  lGLQKcVkMYNPvsVLDIRQGPkEpFkDYVDRFYKAiRAEQAsgeVKqWMTesLLIQNANPdCKvILKGL  326
SIVAGMTYO  L----C-K----VS--------K-P-K---------I-----SGE--Q---ES--------D--V-----  346
SIVAGM155  L----C-K----VS--------K-P-K---------I-----SGE--Q---ES--------D--V-----  346
SIVAGM3    L----C-K----VS--------K-A-K---------I-----SGE--Q---ES--------D--V-----  350
SIVAGM692  A----C-R----TG--------R-S-S---------L-----SQD--N---DT--------E--V-----  340
SIVAGM677  L----V-Q----QK--------K-P-Q---------L-----PQD--N---QT--------D--L-----  340

CONSENSUS  GMhPTLEEMLtACQGVGGPsyKAKvMaEMMqnmq?qNMvQqgg??grpRp?P?kCynCGKFGHmQrqCpe  391
SIVAGMTYO  --H-------T--------SY---V-A---QTMQNQ--V-QGGPK.RQ-P.-LR-YN------M-RQ-PE  414
SIVAGM155  --H-------T--------SY---V-A---QNLQSQ--V-QGGGRGRP-P.-PK-YN------M-RQ-PE  415
SIVAGM3    --H-------T--------SY---V-A---QNMQSQ--M-QGGQRGRP-P.-VK-YN------M-RQ-PE  419
SIVAGM692  --H-------T--------QY---L-V---NQMQGV--V-QAGIGGRG-GR-VK-YK------V-KN-TQ  410
SIVAGM677  --N-------I--------QH---L-V---SN..GQ--V-VGPQKKGP-G.-LK-FN------M-RE-KA  407

CONSENSUS  prk?kClKCGKpGHlAkDCr?GqvNFLGYGrWmGaKPrNF???paatlg?ePsAPPpp???tpYDpAkkL  452
SIVAGMTYO  PRKTK-L----L--L-K--R.-QV------R-M-A--R--...PAATLGAE-S---PPSGTTP--P-KK-  480
SIVAGM155  PRKIK-L----P--L-K--R.-QV------R-M-T--R--...PAATLGAE-S---PPNNSTP--P-KK-  481
SIVAGM3    PRKMR-L----P--L-K--R.-QV------R-M-A--R--...PAATLGVE-T---PPS...P--P-KK-  482
SIVAGM692  KGPVV-L----P--F-R--R.-AV------R-M-A--K--LEHRAA.V...-S---PPHNPGA--E-TR-  475
SIVAGM677  PRQIK-F----I--M-K--KN-QA------H-G-A--R--VQYRGDTVGLE-T---...META--P-KK-  474

CONSENSUS  LqqYaekGkqlReq???ppa?npdw?EgySLnSLFGeDQ*    486
SIVAGMTYO  -QQ-AEK-KQL-EQKRNPPAMNPDWT-GY--N----E---    519
SIVAGM155  -QQ-AEK-KQM-NQNRNPPANNPDWN-GY--N----E---    520
SIVAGM3    -QQ-ADK-KQL-EQRKKPPAVNPDWT-GY--N----E---    521
SIVAGM692  -EK-TQE-AQQ-RKVEKS.SQAGREE-DY--K-- -E---    513
SIVAGM677  -QQ-AEK-QRL-EEREQTRKQKEKEV-DV--S----G---    513
```

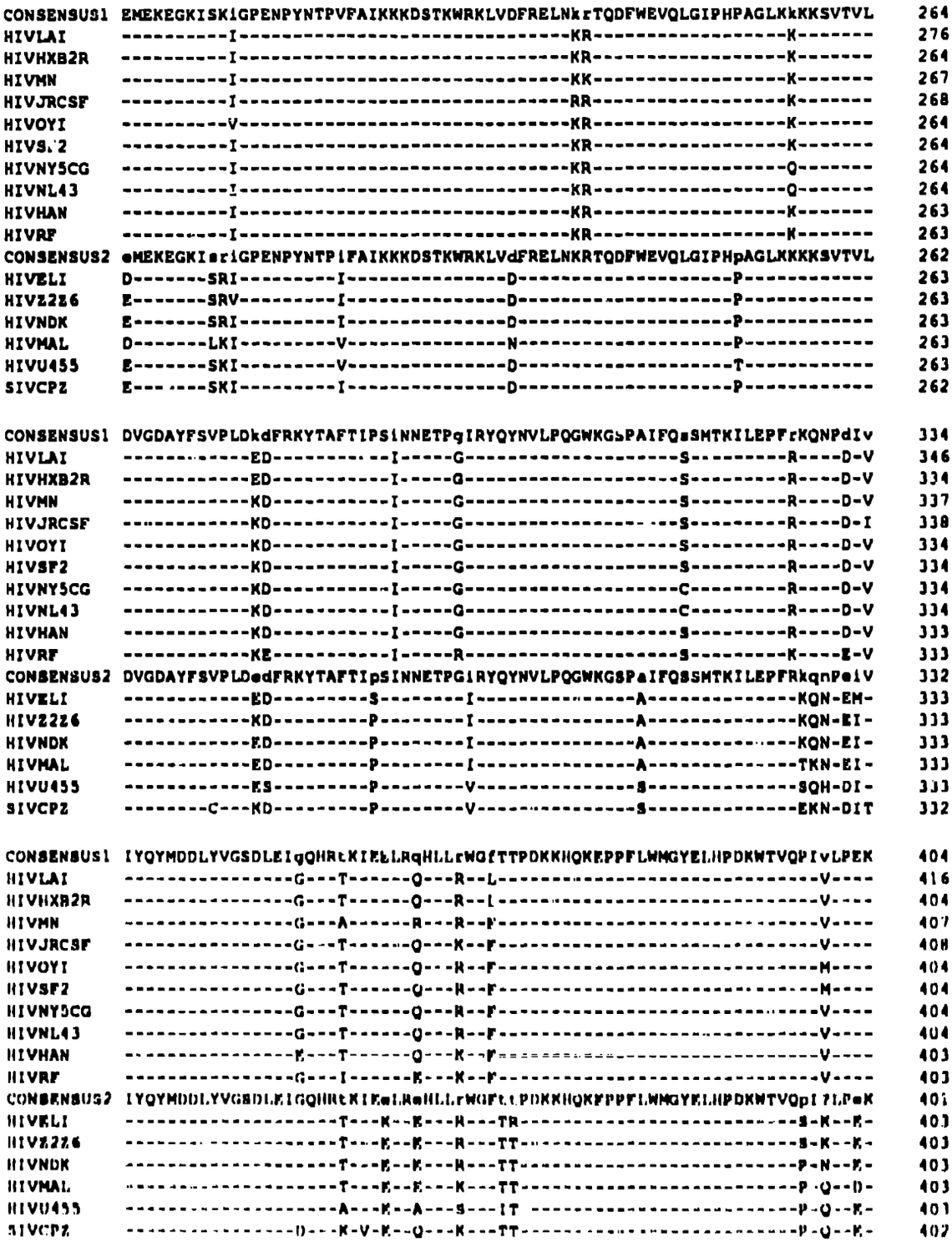
Information dense regions (0.93 or greater) are indicated by ****.

```
                                 ***** ***************   **********       ***** *****
CONSENSUS  MGaatSALnrRqLD?FEhIRLRPnGKKKYQiKHLIWAGKeMeRFGLHEkLLEteEGCkkIIEVL?PLEPT     68
AGMGAG     MG^^^SAL^XR^LD^FE^IRLRP^GKKKYQ^KHLIWAGK^M^RFGLHE^LLE^^EGC^^IIEVLXPLEPT     70
              AAD   D  D  D  D     D      I        K D      K   SK   KK
              GGE   E  E  E  E     E      V        R E      R   TR   RR
              S N   N  N  N  N     N      L        E             E   E
              T Q   Q  Q  Q  Q     Q      M        Q             Q   Q
              P K   K  K  K  K     K
                R   R  R  R  R     R
                H   H  H  H  H     H
                S   S  S  S  S     S
                T   T  T  T  T     T

           ****                    *******
CONSENSUS  GSEGlKsLfNLvCVl?C?H?e?kVKDTEEAva?vrQ?cHLVdKe??A????t????gqkKn?tg???ppG    118
AGMGAG     GSEG^K^L^NLXCV^^C^HX^X^VKDTEEA^XX^^QX^HLV^KX^^A...........^K..^^.^X^^G    126
                L A F     II I  D K       I  IK  I   D  KD            D   DA A AA
                M G W     VV V  E R       V  VR  V   E  RE            E   EG G GG
                  S Y     LL L    E       L      L      EN            N   NS S SS
                  T       MM M    Q       M      M      QQ            Q   QT T TT
                  P        F                      F       K            K   KP P PP
                           W                      W       R            R   R
                           Y                      Y       H            H   H
                           C                      C       S            S   S
                                                          T            T   T

                                 ************       ***************     **************
CONSENSUS  ?SqNfPaqqQgNaWvHvPLSPRTLNAWVKavEEKkfGAEiVPMFQALSEGCtpYDiNQMLNVlGDHQGAL    187
AGMGAG     XS^N^PXX^QXNXW^HXPLSPRTLNAWVKX^EEK^^GAE^VPMFQALSEGCX^YD^NQMLNV^GDHQGAL    196
              K F   D     I               I   KF   I            A  I      I
              R W   E     V               V   RW   V            G  V      V
              E Y   N                          Y                S         L
              Q     Q                                           T         M
                    K                                           P
                    R
                    H
                    S
                    T

           ************        ************* *********           *********
CONSENSUS  QIvKEiINEEAAqWD?tHppPAGPLPAGQLRDPrGSDIAGTTStvQEQlEWiytANPRvDVGAiYRrWiI    256
AGMGAG     QI^KE^INEEAA^WDX^HXXPAGPLPAGQLRDPXGSDIAGTTS^^QEQ^EWX^^ANPR^DVGAXYR^W^I    266
              I  I      E   A                          SI   I   FD    I       K I
              V  V      Q   G                          TV   V   WE    V       R V
              L             S                               L   YN
              M             T                               M    Q
                            P                                   K
                                                                R
                                                                H
                                                                S
                                                                T
```

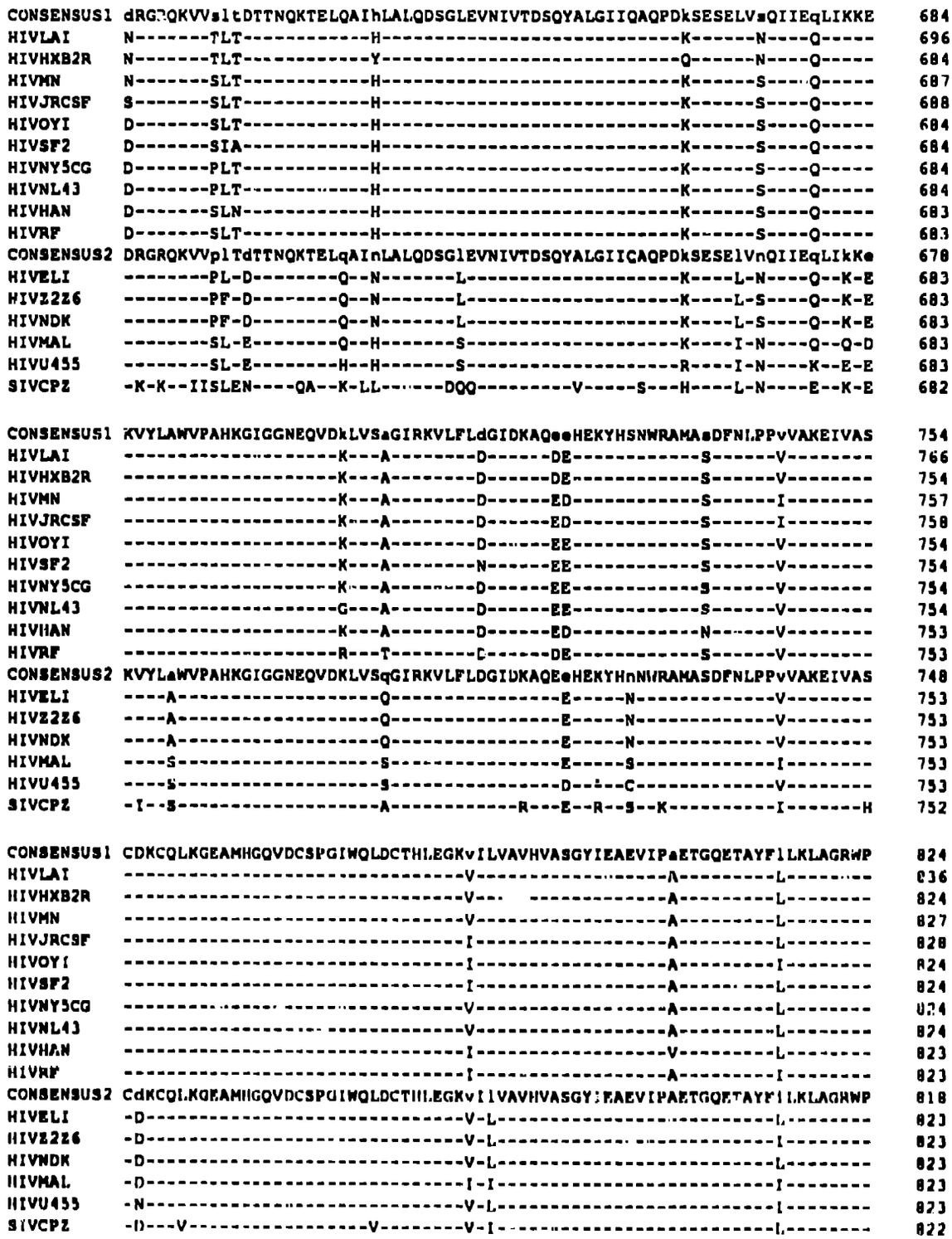
AGMGAG


CONSENSUS lGLQKcVkMYNPvsVLDIRQGPkEpFkDYVDRFYKAlRAEQAsqeVKqWMTesLLIQNANPdCKvILKGL 326
AGMGAG XGLQK^V^MYNPXXVLDIRQGP^E^F^DYVDRFYKA^RAEQA^X^VK^WMT^^LLIQNANP^CK^ILKGL 336
CONSENSUS GMhPTLEEMLtACQGVGGPsyKAKvMaEMMqnmq?qNMvQqgg??grpRp?P?kCynCGKFGHmQrqCpe 391
AGMGAG GM^PTLEEMLXACQGVGGP^^KAK^MXEMM^^..XXNM^QX^^XX.XXR^.PX^C^^CGKFGH^Q^^CXX 402
CONSENSUS prk?kClKCGKpGHlAkDCr?GqvNFLGYGrWmGaKPrNF???paatlq?ePsAPPpp???tpYDpAkkL 452
AGMGAG XXXXXC^KCGKXGH^A^DC^.GXXNFLGYG^WXG^KP^NF...X^X.^...P^APP.....^YDXA^^L 458
CONSENSUS LqqYaekGkqlReq???ppa?npdw?EqySLnSLFGeDQ* 486
AGMGAG L^^Y^^^GX^XR^^X^^^.XXXX^XXEX^SL^SLFGXDQ* 496

```
CONSENSUS1 FFREdLAFpQGkAref???????????sSEQtRANSPTrRELqVWGrdnnSlS????EAGaDRQGtVSf   54
HIVLAI     ----D---L--K-REFSSEQTRANSPTIS---T------R---Q---RDNN-L-....---A----T--F   66
HIVHXB2R   ----D---L--K-REF.............S---T------R---Q---RDNN-P-....---A----T--F   54
HIVMN      ----D---L--K-.EF.............S---N------R---Q---RDNN-L-EAGE---D----P--F   57
HIVJRCSF   ----D---L--K-REF.............P---T------R---Q---RDSN-L-EAGA---A----I--F   58
HIVOYI     ----D---P--K-REF.............S---T------S---R---RDNN-P-....---A----T--F   54
HIVSF2     ----D---L--K-REF.............S---T------R---Q---GENN-L-....---A----T--F   54
HIVNY5CG   ----D---P--K-REF.............S---T------R---Q---RDNN-L-....---A----T--F   54
HIVNL43    ----D---P--K-REF.............S---T------R---Q---RDNN-L-....---A----T--F   54
HIVHAN     ----D---P--E-RKF.............S---T------R---Q---.SNS-L-....---A----T--L   53
HIVRF      ----N---P--K-REL.............S---T------R---Q---RDN.-L-....---E----T--F   53
CONSENSUS2 FFREnLAFpQGkAgEf.............sseQTRANSPTSReLrvwGqdn.pLs...?eTGAERQGtvSF   53
HIVELI     ----N---P--K-G-L.............SPK----------E-RVW-RDN.P-S....K-------TV--   53
HIVZ2Z6    ----D---P--K-G-L.............SSE----------E-RVW-RDN.P-S....E-------TV--   53
HIVNDK     ----D---P--K-G-F.............SSE----------E-RVW-GDN.P-S....E-------TV--   53
HIVMAL     ----N---P--K-R-F.............PSE----------E-RVW-GDK.T-S....E-------IV--   53
HIVU455    ----N---Q--E-R-F.............SSE----------N-.WD-GKD.D-P...CE-------TD--   53
SIVCPZ     ----R---P-RE-RQL.............CAE-N-T-G--D-E-WVP-GR..EEP...GEERGRE-.SI-T   52

           \/ protease  <- gag cds end
CONSENSUS1 sfPQITLWQRPlVTikIGGQLKEALLDTGADDTVleeMnLPgrWKPKMIGGIGGFIKVrQYDQIlieICG   124
HIVLAI     NF--I------L--IK-----------------LEE-S--GR--------------R-----LIE---   136
HIVHXB2R   NF--V------L--IK-----------------LEE-S--GR--------------R-----LIE---   124
HIVMN      SF--I------I--IK-----------------LGE-N--RR--------------R-----TIG---   127
HIVJRCSF   NF--I------L--IK-----------------LED-D--GR--------------R-----PID---   128
HIVOYI     NL--I------I--IK-----------------LEE-N--GR--------------R-----LIE---   124
HIVSF2     NF--I------L--IR-----------------LEE-N--GK--------------R-----PVE---   124
HIVNY5CG   SF--I------L--IK-----------------LEE-N--GR--------------R-----LIE---   124
HIVNL43    SF--I------L--IK-----------------LEE-N--GR--------------G-----LIE---   124
HIVHAN     SL--I------L--IK-----------------VEE-S--GR--------------R-----LVE---   123
HIVRF      SF--I------I--VK-----------------LEE-N--GK--------------R-----LIE---   123
CONSENSUS2 sfPQITLWQRPlVtikIGGQLkEALLDTGADDTVLEeiNLPGKWKPKmIGGIGGFIKVRQYDQIIIEICG   123
HIVELI     NF---------L-AIKI----K--------------EM---------M---------------P-----   123
HIVZ2Z6    NC---------L-TIKI----K--------------EM---------M---------------L-----   123
HIVNDK     SF---------L-TIKI----K--------------EI---------M---------------L-----   123
HIVMAL     SF---------V-TVRV----K--------------EI---------M---------------L-----   123
HIVU455    SF---------L-TVKI----I--------------DI---------I---------------L-----   123
SIVCPZ     NL---------LIPVKVE---C------------I-RIQ-Q-L----M----------K-F-NVH---E-   122

                                  protease \/ p66, p51
CONSENSUS1 HKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKaLvEICT   194
HIVLAI     ---I-------------------I-------------------------------------A-V----   206
HIVHXB2R   ---I-------------------I-------------------------------------A-V----   194
HIVMN      ---I-------------------L-------------------------------------A-I----   197
HIVJRCSF   ---V-------------------I-------------------------------------A-V----   198
HIVOYI     ---I-------------------L-------------------------------------V-I----   194
HIVSF2     ---I-------------------I-------------------------------------A-V----   194
HIVNY5CG   ---I-------------------I-------------------------------------A-V----   194
HIVNL43    ---I-------------------I-------------------------------------A-V----   194
HIVHAN     ---I-------------------I-------------------------------------A-I----   193
HIVRF      ---I-------------------I-------------------------------------A-V----   193
CONSENSUS2 ?KaIGTvLVGPTPVNIIGRNILTQIGCTLNFPISPIETVPVKLKPqMDGPkVKQWPLTEEKIKALTEICt   192
HIVELI     Q-AI--V-------------L-------------------------G----K----------------T   191
HIVZ2Z6    H-AI--V-------------L-------------------------G----K----------------T   191
HIVNDK     Y-AM--V-------------L-------------------------G----K----------------T   191
HIVMAL     K-AI--I-------------M-------------------------G----R----------------K   191
HIVU455    K-TI--V-------------M-------------------------E----K----------------N   191
SIVCPZ     R-VV--V-------------I---L----V----S----------G----K------SA----------Q   192
```



```
CONSENSUS1 DSWTVNDIQKLVGKLNWASQIYaGIKVkqLCKLLRGtKALTEVipLTeEAELELAENREILKEPVHgVYy  474
HIVLAI     ----------------------P----RQ-------T------IP--E----------------G--Y  486
HIVHXB2R   ----------------------P----RQ-------T------IP--E----------------G--Y  474
HIVMN      ----------------------A----KQ-------T------IP--E----------------G--Y  477
HIVJRCSF   ----------------------A----KQ-------T------IP--K----------------G--Y  478
HIVOYI     ----------------------A----KN-------T------IP--E----------------G--Y  474
HIVSF2     ----------------------A----KQ-------T------IP--E----------------E--Y  474
HIVNY5CG   ----------------------A----RQ-------I------VP--E----------------G--Y  474
HIVNL43    ----------------------A----RQ-------T------VP--E----------------G--Y  474
HIVHAN     ----------------------A----KQ-------P------IP--K----------------G--C  473
HIVRF      ----------------------A----KQ-------T------VQ--K----------------G--Y  473
CONSENSUS2 eSWTVNDIQkLVgkLNWASQIYpGIKVkQLCKLLRGtKALTevvpLTeEAELELAENREILKePVHGVYY  471
HIVELI     E--------N--ER--------P----R--------T----EVIP--E-------------E-------  473
HIVZ226    E--------K--GK--------P----R--------T----EVIP--E-------------E-------  473
HIVNDK     E--------K--GK--------A----K--------T----EVVP--E-------------E-------  473
HIVMAL     E--------K--GK--------P----K--------A----DIVP--A-------------E-------  473
HIVU455    D--------K--GK--------A----K--------A----DIVT--E-------------D-------  473
SIVCPZ     EV-------K-IGK--------P---IK-----I--T-K--DVVP--P------------VST-------  472

CONSENSUS1 DPSKDLiaEiQKQGqGQWTYQIyQEPFKNLKTGKYArmrGaHTNDvKQLTEAVQKiatEsIVIWGKtPKF  544
HIVLAI     ------IA-I----Q-------Y-------------RTR-A----V---------ITT-S------T---  556
HIVHXB2R   ------IA-I----Q-------Y-------------RMR-A----V---------ITT-S------T---  544
HIVMN      ------IA-V----Q-------Y-------------RMR-A----V---------IAT-S------T---  547
HIVJRCSF   ------IV-I----Q-------F-------------RTR-A----V---------IAN-S------I---  548
HIVOYI     ------VA-L----Q-------Y-------------RMR-A----V---------ITQ-S------T---  544
HIVSF2     ------VA-I----Q-------Y-------------RMR-A----V---------VST-S------I---  544
HIVNY5CG   ------IA-I----Q-------Y-------------RMK-A----V---------IAT-S------T---  544
HIVNL43    ------IA-I----Q-------Y-------------RMK-A----V---------IAT-S------T---  544
HIVHAN     ------VA-I----E-------Y-------------RMR-A----I---------IAT-G------T---  543
HIVRF      ------IA-I----Q-------Y-------------RMR-A----V---------VAT-S------T---  543
CONSENSUS2 DPSKDLiAEiQKQG?qQWTYQIYQEpfKNLKTGKYAR?rqAHTNDVKQLtEaVQkiatESIVIWGktPKF  539
HIVELI     ------I--I----HG---------PF----------MRG---------A-A--RIST-------RT---  543
HIVZ226    ------I--I----HG---------PF----------MRG---------A-V--KIST-------KT---  543
HIVNDK     ------I--L----DG---------PF----------TRG---------T-A--KIAT-------KT---  543
HIVMAL     ------I--I----QG---------QY----------IKS---------T-A--KIAQ-------KT---  543
HIVU455    ------V--I----QD---------PF----------KRS---------T-V--KVST-------KI---  543
SIVCPZ     --D-E-I--I----NC------F--PH----------QRS-----IR--A-A--KIAT-------KT---  542

                                                                    p51 \/
CONSENSUS1 kLPIQKETWeaWWtaYwqATWIPEWEfVNTPPLVKLWYQLEKePIvGAETFYVDGAAnRETklGKAGYVT  614
HIVLAI     K--------ET--TE-WQ--------F----------------E--V----------S---KL-------  626
HIVHXB2R   K--------ET--TE-WQ--------F----------------E--V----------N---KL-------  614
HIVMN      R--------ET--TE-TX--------V----------------E--V----------N---KK-------  617
HIVJRCSF   K--------ET--TE-WQ--------F----------------E--V----------N---KL-------  618
HIVOYI     K--------EA--TE-WQ--------F----------------D--V----------N---KL-------  614
HIVSF2     K--------EA--ME-WQ--------F----------------E--V----------N---KL-------  614
HIVNY5CG   K--------EA--TE-WQ--------F----------------E--I----------N---KL-------  614
HIVNL43    K--------EA--TE-WQ--------F----------------E--I----------N---KL-------  614
HIVHAN     R--------EA--TE-WQ--------F----------------E--V----------N---KL-------  613
HIVRF      K--------EA--TE-WQ--------F----------------E--I----------N---KL-------  613
CONSENSUS2 rLPIQKETWEtWW?EYWQATWIPEWEFVNTPPLVKLWYQLEkePIiGAETFYVDGAANRETKlGKAGYVT  608
HIVELI     R----------T--A-----------------------------KE--I----------------L-------  613
HIVZ226    R----------T--V-----------------------------KE--I----------------L-------  613
HIVNDK     K----------T--I-----------------------------KE--I----------------L-------  613
HIVMAL     R----------A--T-----------------------------TE--V----------------K-------  613
HIVU455    R----------A--M-----------------------------KD--A----------------L-------  613
SIVCPZ     R--V---T--A--A-----------------I-------  ---T--TE--PTTD-Y----------T-------  612
```

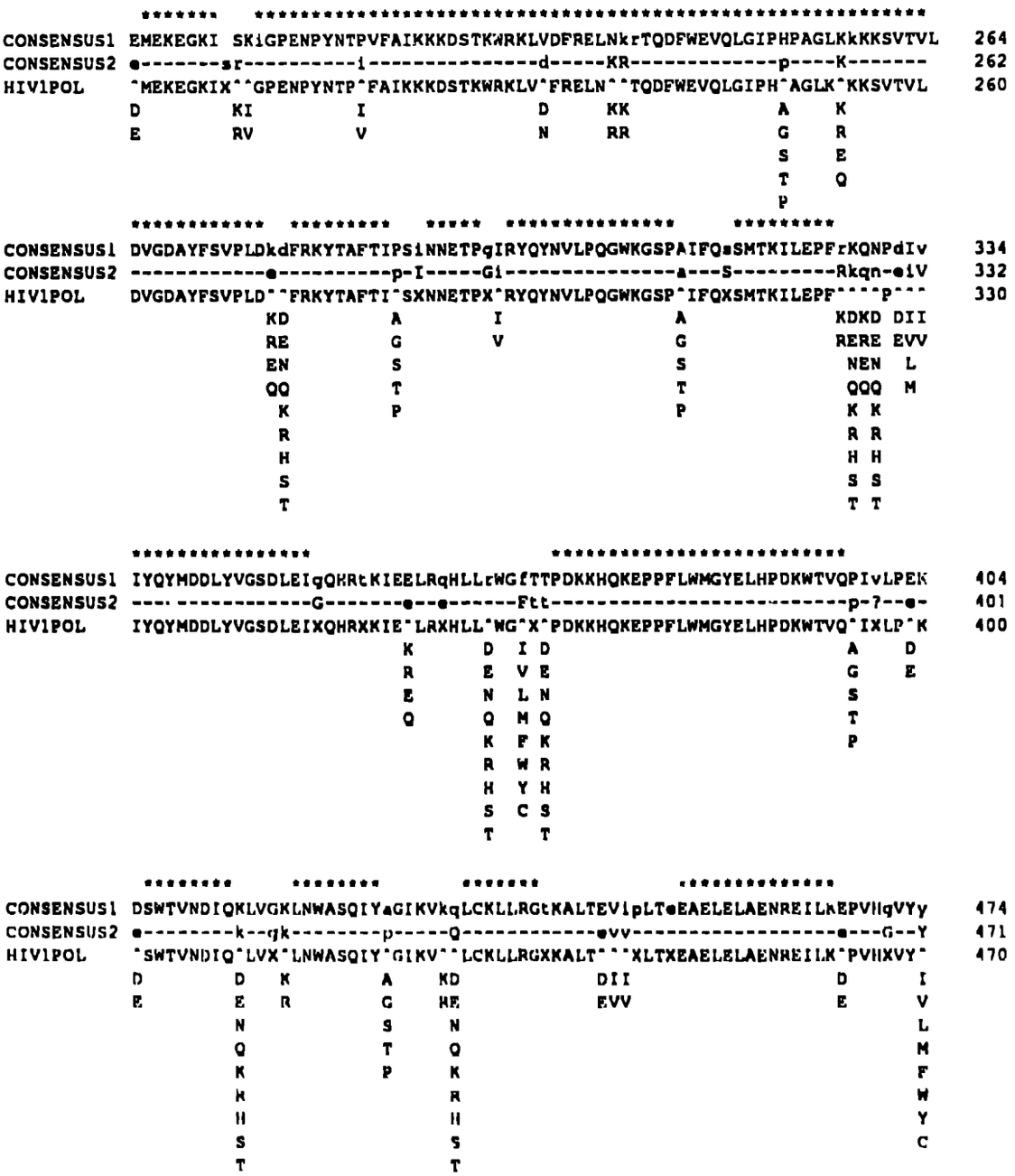

CONSENSUS1 dRGRQKVVslTdTTNQKTELQAIhLALQDSGLEVNIVTDSQYALGIIQAQPDkSESELVsQIIEqLIKKE 684
HIVLAI N-------TLT----------H-------------------------K------N----Q----- 696
HIVHXB2R N-------TLT----------Y-------------------------Q------N----Q----- 684
HIVMN N-------SLT----------H-------------------------K------S----Q----- 687
HIVJRCSF S-------SLT----------H-------------------------K------S----Q----- 688
HIVOYI D-------SLT----------H-------------------------K------S----Q----- 684
HIVSF2 D-------SIA----------H-------------------------K------S----Q----- 684
HIVNY5CG D-------PLT----------H-------------------------K------S----Q----- 684
HIVNL43 D-------PLT----------H-------------------------K------S----Q----- 684
HIVHAN D-------SLN----------H-------------------------K------S----Q----- 683
HIVRF D-------SLT----------H-------------------------K------S----Q----- 683
CONSENSUS2 DRGRQKVVplTdTTNQKTELqAInLALQDSGlEVNIVTDSQYALGIICAQPDkSESElVnQIIEqLIkKe 678
HIVELI --------PL-D--------Q--N-------L--------------------K----L-N----Q--K-E 683
HIVZ2Z6 --------PF-D--------Q--N-------L--------------------K----L-S----Q--K-E 683
HIVNDK --------PF-D--------Q--N-------L--------------------K----L-S----Q--K-E 683
HIVMAL --------SL-E--------Q--H-------S--------------------K----I-N----Q--Q-D 683
HIVU455 --------SL-E--------H--H-------S--------------------R----I-N----K--E-E 683
SIVCPZ -K-K--IISLEN----QA--K-LL--------DQQ---------V-----S---H----L-N----E--K-E 682
CONSENSUS1 KVYLAWVPAHKGIGGNEQVDkLVSaGIRKVLFLdGIDKAQeeHEKYHSNWRAMAsDFNLPPvVAKEIVAS 754
HIVLAI ----------K---A--------D------DE----------S------V-------- 766
HIVHXB2R ----------K---A--------D------DE----------S------V-------- 754
HIVMN ----------K---A--------D------ED----------S------I-------- 757
HIVJRCSF ----------K---A--------D------ED----------S------I-------- 758
HIVOYI ----------K---A--------D------EE----------S------V-------- 754
HIVSF2 ----------K---A--------N------EE----------S------V-------- 754
HIVNY5CG ----------K---A--------D------EE----------S------V-------- 754
HIVNL43 ----------G---A--------D------EE----------S------V-------- 754
HIVHAN ----------K---A--------D------ED----------N------V-------- 753
HIVRF ----------R---T--------D------DE----------S------V-------- 753
CONSENSUS2 KVYLaWVPAHKGIGGNEQVDKLVSqGIRKVLFLDGIDKAQEeHEKYHnNWRAMASDFNLPPvVAKEIVAS 748
HIVELI ----A----------Q----------E-----N----------V-------- 753
HIVZ2Z6 ----A----------Q----------E-----N----------V-------- 753
HIVNDK ----A----------Q----------E-----N----------V-------- 753
HIVMAL ----S----------S----------E-----S----------I-------- 753
HIVU455 ----S----------S----------D--I--C----------V-------- 753
SIVCPZ -I--S----------A----------R---E--R--S--K----------I-------H 752
CONSENSUS1 CDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKvILVAVHVASGYIEAEVIPaETGQETAYFlLKLAGRWP 824
HIVLAI ----------V----------A---------L--------- 836
HIVHXB2R ----------V----------A---------L--------- 824
HIVMN ----------V----------A---------L--------- 827
HIVJRCSF ----------I----------A---------L--------- 828
HIVOYI ----------I----------A---------I--------- 824
HIVSF2 ----------I----------A---------L--------- 824
HIVNY5CG ----------V----------A---------L--------- 824
HIVNL43 ----------V----------A---------L--------- 824
HIVHAN ----------I----------V---------L--------- 823
HIVRF ----------I----------A---------I--------- 823
CONSENSUS2 CdKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKvILVAVHVASGYIEAEVIPAETGQETAYFlLKLAGRWP 818
HIVELI -D----------V-L----------L--------- 823
HIVZ2Z6 -D----------V-L----------I--------- 823
HIVNDK -D----------V-L----------L--------- 823
HIVMAL -D----------I-I----------I--------- 823
HIVU455 -N----------V-L----------I--------- 823
SIVCPZ -D---V----------V---------V-I----------L--------- 822

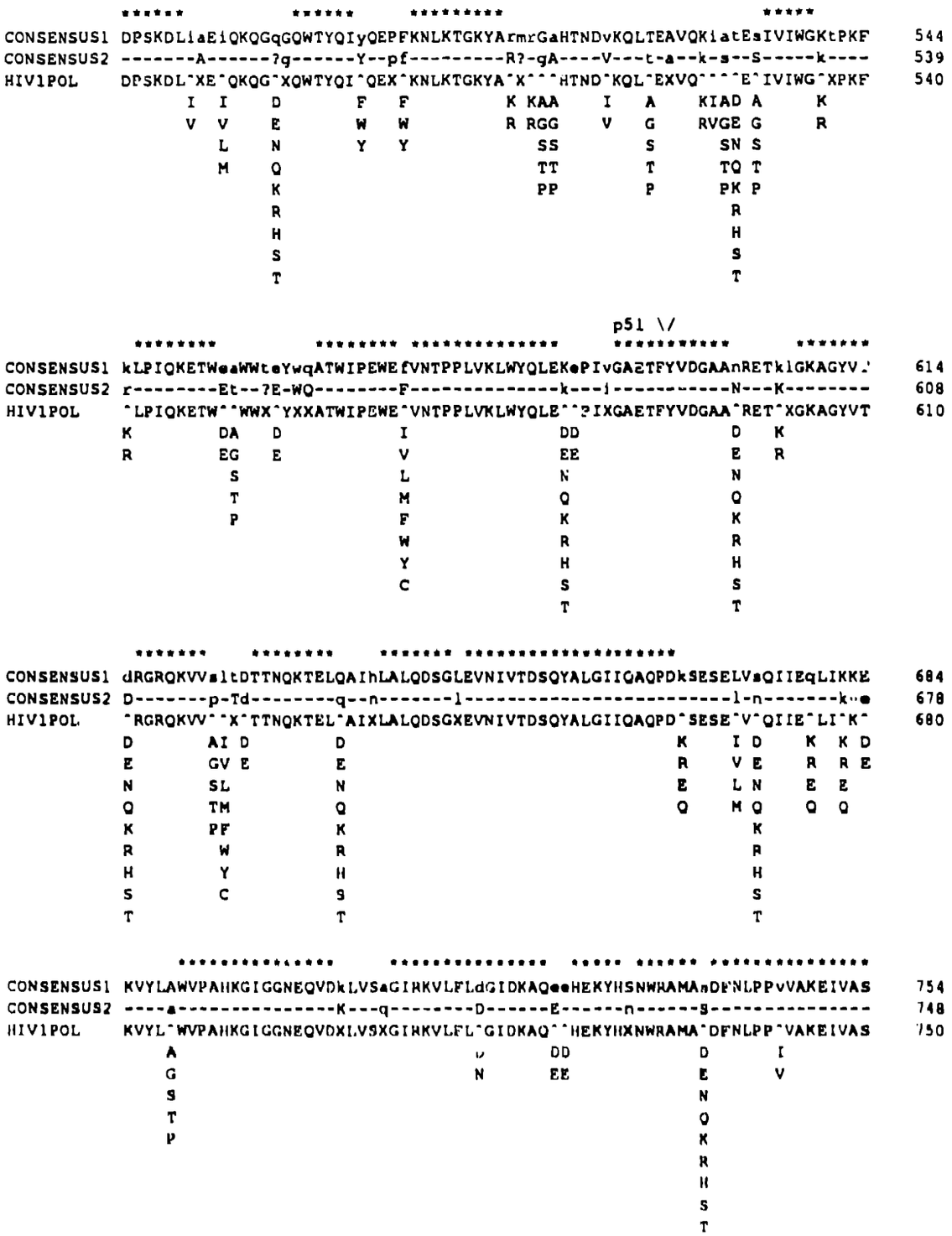
```
CONSENSUS1 VktiHTDNGsNFtatTVkAACWWaGIKQEFGIPYNPQSQGVvESMNkeLKkIIGQVRDQAEHLKtAVQMA  894
HIVLAI     -KTI-----S--TST--K-----A----------------V----KE--K----------T-----  906
HIVHXB2R   -KTI-----S--TGA--R-----A----------------V----KE--K----------T-----  894
HIVMN      -KTI-----P--TST--K-----T----------------I----KE--K----------R-----  897
HIVJRCSF   -TTI-----S--TST--K-----A----------------V----KE--K----------T-----  898
HIVOYI     -KTI-----S--TST--K-----A----------------V----NE--K----------T-----  894
HIVSF2     -KTI-----S--TST--K-----A----------------V----NE--K----------T-----  894
HIVNY5CG   -KTV-----S--TST--K-----A----------------I----KE--K----------T-----  894
HIVNL43    -KTV-----S--TST--K-----A----------------I----KE--K----------T-----  894
HIVHAN     -KTV-----P--TST--K-----A----------------V----KE--K----------T-----  893
HIVRF      -KVI-----S--TST--K-----A----------------V----KQ--Q----------T-----  893
CONSENSUS2 VKvvHTDNGSNFTSAaVKAaCWWAgIkQEFGIPYNPQSQGVVESMNKELKKIIGQVRdQAEHLKTAVQMA  888
HIVELI     --VV----------A---A----G-K-----------------------------D----------  893
HIVZ2Z6    --IV----------A---A----G-K-----------------------------D----------  893
HIVNDK     --VV----------T---A----G-K-----------------------------D----------  893
HIVMAL     --VV----------A---A----N-K-----------------------------E----------  893
HIVU455    --VI----------A---V----N-Q-----------------------------E----------  893
SIVCPZ     --TI-----P-----A---A----D-K----------------L----------D----------  892

CONSENSUS1 VFIHNFKRKGGIGGYSAGERIvdIIATDIQtKELQKQITKIQNFRVYYRDsrdPlWKGpAKLLWKGEGAV  964
HIVLAI     ---------------------VD-------T------------------SRD-L---P----------  976
HIVHXB2R   ---------------------VD-------T------------------SRN-L---P----------  964
HIVMN      ---------------------VG-------T------------------SRD-L---P----------  967
HIVJRCSF   ---------------------ID-------T------------------NRD-I---P----------  968
HIVOYI     ---------------------VD-------T------------------SRE-L---P----------  964
HIVSF2     ---------------------VD-------T------------------NKD-L---P----------  964
HIVNY5CG   ---------------------VD-------I------------------SRD-V---P----------  964
HIVNL43    ---------------------VD-------T------------------SRD-V---P----------  964
HIVHAN     ---------------------VD-------T------------------SRE-F---P----------  963
HIVRF      ---------------------VD-------T------------------SRD-L---H----------  963
CONSENSUS2 VFIHNFKRkgGIGGYSAGERIIDiIATDIQTkELQKQI?KIQNFRVYYRDsRDPIWKGPAKLLWKGEGAV  957
HIVELI     --------RR-------------I-------K------I----------S------------------  963
HIVZ2Z6    --------KG-------------I-------K------T----------S------------------  963
HIVNDK     --------KG-------------I-------R------I----------S------------------  963
HIVMAL     --------KG-------------M-------K------T----------N------------------  963
HIVU455    --------KG-------------I-------K------S----------S------------------  963
SIVCPZ     --------KG-----T-------I-------S------L-V-K-------S---------T---------  962

                        vif cds ->
CONSENSUS1 VIQDNsDIKVVPRRKaKiIRDYGKQmagDDCVASRQDED*   1003
HIVLAI     -----S---------A-I-------MAG------------   1015
HIVHXB2R   -----S---------A-I-------MAG------------   1003
HIVMN      -----N---------A-V-------TAG------------   1006
HIVJRCSF   -----S---------V-I-------MAG------------   1007
HIVOYI     -----S---------A-I-------MAG------------   1003
HIVSF2     -----S---------A-I-------MAG------------   1003
HIVNY5CG   -----S---------A-I-------MAG------------   1003
HIVNL43    -----S---------A-I-------MAG------------   1003
HIVHAN     -----S---------A-I-------MGS------------   1002
HIVRF      -----S---------A-I-------MAG------------   1002
CONSENSUS2 VIQDnSDIKVVPRRKvKIIRDYGKQMAGDDCvAsrQDED*    996
HIVELI     ----K----------V---------------V-SR-----   1002
HIVZ2Z6    ----N----------V---------------V-SR-----   1002
HIVNDK     ----N----------V---------------V-SR-----   1002
HIVMAL     ----N----------A---------------V-GG-----   1002
HIVU455    ----N----------A---------------M-GR-----   1002
SIVCPZ     ----QGGL-------A---------------V-SR-N---   1001
```

V1 POL

Information dense regions (0.95 or greater) are indicated by ****.

```
                                                  ******
CONSENSUS1 FFREdLAFpQGkAref????????????sSEQtRANSPTrRELqVWGrdnnSlS????EAGaDRQGtVSf     54
CONSENSUS2 ----n--------gE-............-se-T------S-e-rvw-g--.pLs...-eT-AE----v-F     53
HIV1POL    FFRE^LAFXQG^A.^^............^^^Q^RANSPT^R^L.^XG.^^.XX^....^^GX^RQGXXS^     50
               D      K  KI            AAK D      D D  I   DD   A    KA  D      I
               N      R  RV            GGR E      E E  V   EE   G    RG  E      V
                      E  EL            SSE N      N N  L   NN   S    ES         L
                      Q  QM            TTQ Q      Q    M   QQ   T    QT         M
                          F            PP  K      K    F   KK   P     P         F
                          W                R      R    W   RR                   W
                          Y                H      H    Y   HH                   Y
                          C                S      S    C   SS                   C
                                           T      T        TT

            \/ protease  <- gag cds end
                ******          ************              ********** *****
CONSENSUS1 sfPQiTLWQRPlVTikIGGQLKEALLDTGADDTVleeMnLPgrWKPKMIGGIGGFIKVrQYDQIlieICG    124
CONSENSUS2 ----I--------t--i----k------------LE-iN--GK----m----------R------IE---    123
HIV1POL    ^^PQ^TLWQRP^V^^^^GGQLXEALLDTGADDTV^X^^^LPX^WKPK^IGGIGGFIKVXQYDQIX^XICG    120
           DI  I      I AIKI                 I DID   K    I                 I
           EV  V      V GVRV                 V EVE   R    V                 V
           NL         L S                    L  LN        L
           QM         M T                    M  MQ        M
           KF           P                        K
           RW                                    R
           HY                                    H
           SC                                    S
           T                                     T

                                 protease \/ p66, p51
                  ************************************ ****************
CONSENSUS1 HKAiGTVLVGPTPVNIIGRNLLTQiGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKaLvEICT    194
CONSENSUS2 ?-a---v-------------l---I-------------------g----k------------A-T---t    192
HIV1POL    XK^^GT^LVGPTPVNIIGRN^LTQ^GCTLNFPISPIETVPVKLKPXMDGP^VKQWPLTEEKIKXLXEIC^    190
             AI  I             L   I                          K                 D
             GV  V             M   V                          R                 E
             SL                    L                                            N
             TM                    M                                            Q
             P                                                                  K
                                                                                R
                                                                                H
                                                                                S
                                                                                T
```


CONSENSUS1 EMEKEGKI SKiGPENPYNTPVFAIKKKDSTKWRKLVDFRELNkrTQDFWEVQLGIPHPAGLKkKKSVTVL 264
CONSENSUS2 e-------sr----------i--------------d-----KR------------p----K------- 262
HIV1POL ^MEKEGKIX^^GPENPYNTP^FAIKKKDSTKWRKLV^FRELN^^TQDFWEVQLGIPH^AGLK^KKSVTVL 260
CONSENSUS1 DVGDAYFSVPLDkdFRKYTAFTIPSiNNETPqIRYQYNVLPQGWKGSPAIFQsSMTKILEPFrKQNPdIv 334
CONSENSUS2 ------------e----------p-I-----Gi--------------a---S---------Rkqn-eiV 332
HIV1POL DVGDAYFSVPLD^^FRKYTAFTI^SXNNETPX^RYQYNVLPQGWKGSP^IFQXSMTKILEPF^^^^P^^^ 330
CONSENSUS1 IYQYMDDLYVGSDLEIqQHRtKIEELRqHLLrWGfTTPDKKHQKEPPFLWMGYELHPDKWTVQPIvLPEK 404
CONSENSUS2 ---- ------------G-------e--e------Ftt----------p-?--e- 401
HIV1POL IYQYMDDLYVGSDLEIXQHRXKIE^LRXHLL^WG^X^PDKKHQKEPPFLWMGYELHPDKWTVQ^IXLP^K 400
CONSENSUS1 DSWTVNDIQKLVGKLNWASQIYaGIKVkqLCKLLRGtKALTEVipLTeEAELELAENREILkEPVHgVYy 474
CONSENSUS2 e--------k--qk--------p-----Q------------evv----------e---G--Y 471
HIV1POL ^SWTVNDIQ^LVX^LNWASQIY^GIKV^^LCKLLRGXKALT^^^XLTXEAELELAENREILK^PVHXVY^ 470


CONSENSUS1 DPSKDLiaEiQKQGqGQWTYQIyQEPFKNLKTGKYArmrGaHTNDvKQLTEAVQKiatEsIVIWGKtPKF 544
CONSENSUS2 -------A------?g------Y--pf---------R?-gA----V---t-a--k-s--S-----k---- 539
HIV1POL DPSKDL^XE^QKQG^XQWTYQI^QEX^KNLKTGKYA^X^^^HTND^KQL^EXVQ^^^^E^IVIWG^XPKF 540
CONSENSUS1 kLPIQKETWeaWWteYwqATWIPEWEfVNTPPLVKLWYQLEKePIvGAzTFYVDGAAnRETklGKAGYV. 614
CONSENSUS2 r--------Et--?E-WQ--------F-------------k---j----------N---K-------- 608
HIV1POL ^LPIQKETW^^WWX^YXXATWIPEWE^VNTPPLVKLWYQLE^^?IXGAETFYVDGAA^RET^XGKAGYVT 610
CONSENSUS1 dRGRQKVVsltDTTNQKTELQAIhLALQDSGLEVNIVTDSQYALGIIQAQPDksESELVsQIIEqLIKKE 684
CONSENSUS2 D-------p-Td--------q--n-------l-------------------------l-n-------k-e 678
HIV1POL ^RGRQKVV^^X^TTNQKTEL^AIXLALQDSGXEVNIVTDSQYALGIIQAQPD^SESE^V^QIIE^LI^K^ 680
CONSENSUS1 KVYLAWVPAHKGIGGNEQVDkLVSaGIRKVLFLdGIDKAQeeHEKYHSNWRAMAnDFNLPPvVAKEIVAS 754
CONSENSUS2 ----a--------------K---q--------D------E------n------s-------------- 748
HIV1POL KVYL^WVPAHKGIGGNEQVDXLVSXGIRKVLFL^GIDKAQ^^HEKYHXNWRAMA^DFNLPP^VAKEIVAS 750

```
          ************************************************** ******************
CONSENSUS1 CDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKvILVAVHVASGYIEAEVIPaETGQETAYFlLKLAGRWP    824
CONSENSUS2 -d------------------------------l----------------A---------i--------    818
HIV1POL    C^KCQLKGEAMHGQVDCSPGIWQLDCTHLEGK^I^VAVHVASGYIEAEVIPXETGQETAYF^LKLAGRWP    820
            D   .                          I I                          I
            N                              V V                          V
                                             L                          L
                                             M                          M
          *   *****                 ********************     ************* ******
CONSENSUS1 VktiHTDNGsNFtstTVkAACWWaGIKQEFGIPYNPQSQGVvESMNkeLKkIIGQVRDQAEHLKtAVQMA    894
CONSENSUS2 -Kvv-----S--TSAa-K-a---Ag-k--------------V----KE--K------d------T-----    888
HIV1POL    V^X^HTDNG^NFX^^^V^AXCWW^XI^QEFGIPYNPQSQGV^ESMN^^LK^IIGQVR^QAEHLK^AVQMA    890
            D I      A    AAA K      A  K               I     DE  K       D       D
            E V      G    GGG R      G  R               V     EQ  R       E       E
            N        S    SSS        S  E                     N   E               N
            Q        T    TTT        T  Q                     Q   Q               Q
            K        P    PPP        P                        K                   K
            R                                                 R                   R
            H                                                 H                   H
            S                                                 S                   S
            T                                                 T                   T

          ********  ***********   ******  ****** ***********          ***********
CONSENSUS1 VFIHNFKRKGGIGGYSAGERIvdIIATDIQtKELQKQITKIQNFRVYYRDsrdPlWKGpAKLLWKGEGAV    964
CONSENSUS2 --------kg-----------IDi------Tk------?-----------RD-I---P-----------    957
HIV1POL    VFIHNFKR^XGIGGYSAGERI^X^IATDIQX^ELQKQIXKIQNFRVYYRD^^^P^WKGXAKLLWKGEGAV    960
                   K            I I       K                   DKD I
                   R            V V       R                   ERE V
                                  L                           N N L
                                  M                           Q   M
                                                              K   F
                                                              R   W
                                                              H   Y
                                                              S   C
                                                              T

                        vif cds ->
          ****  *********   *******
CONSENSUS1 VIQDNsDIKVVPRRKaKiIRDYGKQmagDDCVASRQDED*   1003
CONSENSUS2 ----nS----------v-I-------MAG---v-sr----*    996
HIV1POL    VIQD^^DIKVVPRRKXK^IRDYGKQX^^DDC^A^XQDED*    999
              DD           I        AA   I A
              EE           V        GG   V G
              NN                     S   L S
              QQ                     T   M T
              KK                     P     F
              RR
              HH
              SS
              TT
```

HIV1 POL

There are many regions in conserved patterns of HIV-1 Pol and HIV-2 Pol with a high degree of similarity; because of the extensive homology between the two viral polymerases we did not specify each of these regions here.

HIV1POL:22 RANSPT

HIV1POL:61 TLWQRP
Part of a cytotoxic T-cell epitope. Hosmalin et al., PNAS 87:2344-2348 (1990).

Immediately C-terminal to the protease NH2-terminus.

HIV1POL:77 EALLDTGADDTV
The "DTG" is a conserved feature of retroviral aspartic proteases, with D being a catalytic residue. Weber et al. Science 243:928-30 (1989).

HIV1POL:103 IGGIGGFIKV

HIV1POL:115 QYDQI

HIV1POL:133 LVGPTPVNIIGRNbLTQcGCTLNFPISPIETVPVKLKP
Part of a cytotoxic T-cell epitope -- peptide 40 of Walker et al., PNAS 86:9514-18 (1989).

Mutational analyses -- Louis et al., BBRC: 164, 30-38, 1989; Loeb et al., Nature 340: 397-400, 1989.

The protease/p66,51 cleavage site is between F and P in TLNFPI (Lightfoote et al., J. Virol. 60: 771-5, 1986 and di Marzo Veronese et al., Science 231: 1289-91, 1986).

HIV1POL:171 MDGPnVKQWPL
Part of a cytotoxic T-cell epitope -- peptide 40 of Walker et al., PNAS 86:9514-18 (1989).

HIV1POL:197 MEKEGKI

HIV1POL:207 GPENPYNTPaFAIKKKDSTKWRKLVlFRELNnnTQDFWEVQLGIPHjAGLKpKKSVTV
LDVGDAYFSVPLD
*Sites that can contribute to AZT resistance when mutated. Larder and Kemp, Science 246:1155-58 (1989).

HIV1POL:280 FRKYTAFTI
HIV1POL:292 NNETP

HIV1POL:299 RYQYNVLPQGWKGSP

HIV1POL:319 SMTKILEPF
Part of a cytotoxic T-cell epitope -- peptide 49 of Walker et al., PNAS 86:9514-18 (1989).

HIV1POL:336 IYQYMDDLYVGSDLEI
Part of a cytotoxic T-cell epitope -- peptide 50 of Walker et al., PNAS 86:9514-18 (1989).

HIV1POL:371 PDKKHQKEPPFLWMGYELHPDKWTVQ
*Site that can contribute to AZT resistance when mutated. Larder and Kemp, Science 246:1155-58 (1989).

HIV1POL:407 SWTVNDIQ

HIV1POL:420 LNWASQIY

HIV1POL:435 LCKLLRG

HIV1POL:454 EAELELAENREILK
Part of a cytotoxic T-cell epitope -- peptide 57 of Walker et al., PNAS 86:9514-18 (1989).

HIV1POL:476 DPSKDL

HIV1POL:492 QWTYQI
Part of a cytotoxic T-cell epitope -- peptide 59 of Walker et al., PNAS 86:9514-18 (1989).

HIV1POL:503 KNLKTGKYA
Part of a cytotoxic T-cell epitope -- peptide 59 of Walker et al., PNAS 86:9514-19 (1989).

HIV1POL:536 IVIWG

HIV1POL:547 LPIQKETW

HIV1POL:564 ATWIPEWE

HIV1POL:573 VNTPPLVKLWYQLE

HIV1POL:592 GAETFYVDGAA

HIV1POL:609 GKAGYVT

HIV1POL:617 RGRQKVV

HIV1POL:628 TTNQKTEL

HIV1POL:640 LALQDSG

HIV1POL:648 EVNIVTDSQYALGIIQAQPD

HIV1POL:658 LGIIQ

HIV1POL:691 WVPAHKGIGGNEQVD

HIV1POL:711 GIRKVLFLDGIDKAQ

HIV1POL:728 HEKYH

HIV1POL:734 NWRAMA

HIV1POL:741 DFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVA
VHVASGYIEAEVIP

HIV1POL:808 ETGQETAYFLLKLAGRWPV

HIV1POL:830 HTDNG

HIV1POL:851 QEFGIPYNPQSQGVVESMN

HIV1 POL

HIV1POL:877 IIGQVRkQAEHLK

HIV1POL:891 AVQMAVFIHNFKR

HIV1POL:906 GIGGYSAGERI

HIV1POL:920 IATDIQ

HIV1POL:928 ELQKQI

HIV1POL:935 KIQNFRVYYRD
Part of a helper T-cell epitope -- peptide pol3 of Schrier et al., J. Immunol. 142:1166-76 (1989).

HIV1POL:955 AKLLWKGEGAVVIQD

HIV1POL:972 DIKVVPRRK

HIV1POL:984 IRDYGKQ

```
CONSENSUS  TggFFR?wpmGKeapQlPhqps?sGa??????????????????dtnc?p?gsS?Gs??elhaag?k???   41
HIV2ROD    -GR---TGPL--EAP-L-RGPSSA-A.....................DTNSTPSGS-S-STGEIYAAREK...   49
HIV2NIHZ   -GW---DWPL--EGP-L-RGPSPA-A.....................NTNSTPIGS-S-PTGEIYAARKK...   49
HIV2ISY    -GW---AWTM--EAP-L-RGPKFA-A.....................NTNSTPNGS-S-PTGEVHAAREK...   49
HIV2ST     -GG---VGPM--EAP-F-CGPNPA-A.....................DTNSTPDRP-R-PTREVHAAREK...   49
HIV2BEN    -GR---VGPT--EAS-L-RDPSPS-A.....................DTNSTSGRS-S-TVGEIYAAREK...   49
HIV2D194   -GR---DGPT--AAP-L-RGPSSS-A.....................DTNSTPNRS-S-PVGEIYAAREK...   49
HIV2GH1    -GR---DGSM--EAP-L-RGPSSS-A.....................DTNSTPSRS-S-SIGKIYAAGER...   49
HIV2D205   -GG---VRTL--EAS-L-HDPSAS-SDTICTPDEPSRGHDTSGGDTICAPCRS-S-DAEKLHADGET...   67
SIVMM251   -GG---PWSM--EAP-F-HGSSAS-A.....................DANCSPRGP-C-SAKELHAVGQA...   49
SIVMM142   -GG---PWPL--EAP-F-HGSSAS-A.....................DANCSPRRT-C-SAKELHALGQA...   49
SIVMM239   -GG---PWSM--EAP-F-HGSSAS-A.....................DANCSPRGP-C-SAKELHAVGQAAER   52
SIVMNE     -GR---LWPM--EAP-F-HGPNAS-A.....................DANCSPRGP-C-SAKKLHAVGQT...   49
SIVSMMH4   -GG---AWPM--EAP-F-HGPDAS-A.....................DTNCSPRGS-C-STEELHEDGQK...   49
SIVSMMPBJ  -SG---AWPM--EAP-F-HGPDAS-A.....................DTNCSPRGS-C-STEELHEDGQK...   49

                                                          <- gag cds
CONSENSUS  ?aer??retlQgg??????????????????drGfAaPQFSLWrRPvVtA?iEgQpvEVLLDTgadDSIV   89
HIV2ROD    .TERAERETI-GSDRGLTAPRAGGDTIQGATNR-L-A------K--V-T-YI-G-PV------GAD----  118
HIV2NIHZ   .AKGAERETV-GSDRGLTAFRAGRDTMQGD.DR-L-A------K--V-T-HI-G-PV------RAN----  117
HIV2ISY    .TERAETKTI-RSDRGLAASRARRDTTQRD.DR-L-A------K--V-T-YI-D-PV------GAD----  117
HIV2ST     .AERAEREAI-RSDRGLPAARETRDTMQRD.DR-L-A------K--V-T-HV-G-PV------GAD----  117
HIV2BEN    .AEGAEGETI-RGDGGLAAPRAERDTSQRG.DR-L-A------K--V-T-YI-D-PV------GAD----  117
HIV2D194   .AERAEGETI-GGDGGLTAPRAGRDAPQRG.DR-L-T------K--V-T-FI-D-PV------GAD----  117
HIV2GH1    .AEGAEGETI-RGDGRLTAPRAGKSTSQRG.DR-L-A------K--V-T-YI-V-PV------GAD----  117
HIV2D205   .TEREPRETL-GG.................DR-F-A------R--V-K-CI-G-SV------GVD----  113
SIVMM251   .AERKQREAL-GG.................DR-F-A------R--V-T-HI-G-PV------GAD----  100
SIVMM142   .AERKQREAL-GG.................DR-F-A------R--V-T-HI-G-PV------GAD----  100
SIVMM239   KAERKQREAL-GG.................DR-F-A------R--V-T-HI-G-PV------GAD----  104
SIVMNE     .AERKQKEAL-GG.................DR-F-A------R--V-T-HI-G-PA------GAD----  100
SIVSMMH4   .AEGEQRETL-GG.................DR-F-A------R--V-T-YI-E-PV------GAD----  100
SIVSMMPBJ  .AEGEQRETL-GG.................NG-F-A------R--I-T-YI-E-PV------GAD----  100

CONSENSUS  aGIeLG?nYtPKlVGGIGGFINTKEyKnveIeVlgKrvkaTIMTGDTPINIFGRN?LtalgMaLNlI?AK  157
HIV2ROD    A-IE--NN-S--I-----------Y-NVE-E-LN-KVRA---------------I-TALG-S--L-V--  188
HIV2NIHZ   A-IE--SN-S--I-----------Y-NVE-E-LG-RVKA---------------V-TALG-S--L-V--  187
HIV2ISY    A-IE--SN-S--I-----------Y-DVE-R-LN-KVRA---------------I-TALG-S--L-V--  187
HIV2ST     A-VE--SN-S--I-----------Y-NVE-R-LN-RVRA---------------I-TALG-S--L-V--  187
HIV2BEN    A-IE--DN-T--I-----------Y-NVE-K-LN-RVRA---------------I-TALG-S--L-V--  187
HIV2D194   A-IE--DN-T--I-----------Y-NVE-K-LN-RVRA---------------I-ATLG-S--L-V--  187
HIV2GH1    A-IQ--DN-V--I-----------I-NIE-K-LN-RVRA---------------I-TALG-S--L-I--  187
HIV2D205   A-IE--SN-T--I-----------Y-DVE-E-VG-RVRA---------------I-NTLG-T--F-V--  188
SIVMM251   T-IE--PH-T--I-----------Y-NVK-E-LG-RIKG---------------L-TALG-S--L-I--  170
SIVMM142   T-IE--PH-T--I-----------Y-NVE-E-LG-RIKG---------------L-TALG-S--L-I--  170
SIVMM239   T-IE--PH-T--I-----------Y-NVE-E-LG-RIKG---------------L-TALG-S--F-I--  174
SIVMNE     A-IE--PH-T--I-----------Y-NVK-E-LG-RIKG---------------L-TALE-S--F-I--  170
SIVSMMH4   A-IE--PN-T--I-----------Y-DVK-K-LG-VIKG---------------L-TAMG-S--L-I--  170
SIVSMMPBJ  A-IE--PN-T--I-----------Y-DVK-K-LG-VIKG---------------L-TAMG-S--L-I--  170
```

HIV2/SIV POL

```
CONSENSUS  vePiKvtLKPGKdGPklrQWPLskEKIeALrEICEKMEkeGQLEEAPPTNPYNTPTFAIKKKDKNKWRmL  227
HIV2ROD    VE-I-IM-----D--KLR----TK---E--K-------KE----------------------------M-  258
HIV2NIHZ   IE-I-IM-----D--RLK----TK---E--K-------KE----------------------------M-  257
HIV2ISY    IE-V-VT-----D--KQR----TR---E--R-------RE----------------------------M-  257
HIV2ST     IE-I-IM-----D--KLR----TK---E--K-------RE----------------------------M-  257
HIV2BEN    IE-I-VT-----D--RLK----TK---E--K-------KE----------------------------M-  257
HIV2D194   LD-I-VT-----D--RLK----TK---E--K-------RE----------------------------M-  257
HIV2GH1    IE-I-VT-----D--RLR----TK---E--R-------KE----------------------------M-  257
HIV2D205   VE-V-VE-----D--KIR----SR---L--K-------KE----------------------------M-  258
SIVMM251   VE-V-VT-----V--KLK----SK---V--R-------KD----------------------------M-  240
SIVMM142   VE-V-SP-----D--KLK----SK---V--R-------KD----------------------------M-  240
SIVMM239   VE-V-VA-----D--KLK----SK---V--R-------KD----------------------------M-  244
SIVMNE     VE-V-VT-----D--KLR----SK---V--R-------KD----------------------------I-  240
SIVSMMH4   VE-I-VT-----E--KLR----SK---I--R-------KD----------------------------M-  240
SIVSMMPBJ  VE-I-VT-----D--KLR----SK---I--R-------KD----------------------------M-  240

CONSENSUS  IDFRELNkVTQDFTEvqlgiPhpaglAkkkRItvlDvGDAYFSIPLde?FRqYTAFTLPsvNNaEPgKRY  296
HIV2ROD    -------K-------IQLGI-HPAGL-KKR--TVL-V---------HED--P-------SV--A--G---  328
HIV2NIHZ   -------K-------IQLGI-HPAGL-KKR--TVL-V---------HED--Q-------SV--A--G---  327
HIV2ISY    -------K-------VQLGI-HPAGL-KKR--TVL-V---------YED--Q-------SV--A--G---  327
HIV2ST     -------K-------IQLGI-HPAGL-KKK--TVL-V---------HED--Q-------SI--A--G---  327
HIV2BEN    -------K-------IQLGI-HPAGL-KKK--SIL-V---------HED--Q-------AV--M--G---  327
HIV2D194   -------R-------IQLGI-HPAGL-KKK--TVL-V---------HED--Q-------SV--A--E---  327
HIV2GH1    -------R-------IQLGI-HPAGL-KKK--TVL-V---------HED--Q-------SV--A--G---  327
HIV2D205   -------K-------VNWVF-.TRQV-EKR--TVI-V---------DPN--Q-------SV--A--G---  327
SIVMM251   -------R-------VQLGI-HPAGL-KRK--TVL-I---------DEE--Q-------SV--A--G---  310
SIVMM142   -------R-------VQLGI-HPAGL-KRK--TVL-I---------DEE--Q-------SV--A--G---  310
SIVMM239   -------R-------VQLGI-HPAGL-KRK--TVL-I---------DEE--Q-------SV--A--G---  314
SIVMNE     -------K-------VQLGI-HPAGL-KRK--TVL-V---------DEE--Q-------SV--A--G---  310
SIVSMMH4   -------K-------VQLGI-HPAGL-KRR--TVL-V---------DEE--Q-------SV--A--G---  310
SIVSMMPBJ  -------K-------VQLGI-HPAGL-KRR--TVL-V---------DEE--Q-------SV--A--G---  310

CONSENSUS  iYKVlPQGWKGSpaIfQytMR?vLePFRkANpDvilIQYMDDILiASDRtdLEHDrVVlQlKELLNglGF  365
HIV2ROD    I---L--------PA-F-HT--QV-E---K--K-VIII-------I----TD----R--L-L-----GL--  398
HIV2NIHZ   I---L--------PA-F-YT--QI-E---K--E-VIII-------I----TD----K--L-L-----GL--  397
HIV2ISY    I---L--------PA-F-YT--QV-E---K--P-VIIV-------I----TD----K--L-L-----GL--  397
HIV2ST     I---S--------PA-F-YT--QV-E---K--P-IILI-------I----TD----R--L-L-----GL--  397
HIV2BEN    I---L--------PA-F-YT--QV-E---K--P-VILI-------I----TG----K--L-L-----GL--  397
HIV2D194   V---L--------PA-F-FM--QI-E---K--P-VILI-------I----TG----K--L-L-----GL--  397
HIV2GH1    I---L--------PA-F-HT--QV-E---K--P-VILI-------I----TG----K--L-L-----GL--  397
HIV2D205   I---L--------QS-C-YS--KV-D---K--S-VIII-------I----SD----R--S-L-----DM--  397
SIVMM251   I---L--------PA-F-YT--HV-E---K--P-VTLV-------I----TD----R--L-L-----SI--  380
SIVMM142   I---L--------PA-F-YT--HV-E---K--P-VTLV-------I----TD----R--L-L-----SI--  380
SIVMM239   I---L--------PA-F-YT--HV-E---K--P-VTLV-------I----TD----R--L-S-----SI--  384
SIVMNE     I---L--------PA-F-HT--NV-E---K--P-VTLV-------V----TD----R--L-L-----SI--  380
SIVSMMH4   I---L--------PA-F-YT--NV-E---K--P-VTLI-------I----TD----R--L-L-----GI--  380
SIVSMMPBJ  I---L--------PA-F-HT--NV-E---K--P-VTLI-------I----TD----R--L-L-----SI--  380
```

```
CONSENSUS  StPeEKFQKDPFfqWMGyELWPtKWKLQkiqLPqke?WTVNdIQKLVGVLNWAAQiypGIKTkhlCrLIr  434
HIV2ROD    -T-D--------YH---Y----T-----KIQ--QKEI----D-------------LYP----KHL-R--R  468
HIV2NIHZ   -T-D--------YR---Y----T-----KIQ--QKEV----D-------------IYP----KHL-R--R  467
HIV2ISY    -T-D--------YQ---Y----T-----KIQ--QKEV----D-------------IYP----KHL-K--R  467
HIV2ST     -T-D--------YQ---Y----T-----RIQ--QKEV----D-------------IYP----RNL-R--R  467
HIV2BEN    -T-D--------FQ---C----T-----KLQ--QKDI----D-------------IYS----KHL-R--R  467
HIV2D194   -T-D--------FQ---Y----T-----KIQ--QKEI----D-------------IYP----KHL-K--R  467
HIV2GH1    -T-D--------LQ---Y----T-----KLQ--QKEI----D-------------IYP----KHL-R--K  467
HIV2D205   -T-E--------FK---Y----K-----KIQ--EKEV----A-------------LFP----RHI-K--R  467
SIVMM251   -T-E--------FQ---Y----T-----KIE--QRET----D-------------IYP----KHL-R--R  450
SIVMM142   -S-E--------FQ---Y----T-----KIE--QRET----D-------------IYP----KHL-R--R  450
SIVMM239   -T-E--------FQ---Y----T-----KIE--QRET----D-------------IYP----KHL-R--R  454
SIVMNE     -T-E--------FQ---Y----T-----KIE--QKET----D-------------IYP----KHL-R--R  450
SIVSMMH4   -T-E--------FQ---Y----T-----KIE--QRET----D-------------IYP----KHL-R--R  450
SIVSMMPBJ  -T-E--------FQ---Y----T-----KIE--QRET----D-------------IYP----KHL-R--R  450

CONSENSUS  GKMtlTEEVQWTElAEAEleENkIILsQeQEG?YYqE?kplEATv?K?qdnQW?YKiHQe?kiLKvGK?a  497
HIV2ROD    ---TL--------L----LE--R---S-E---H--Q-EKEL---VQ-DQEN--T--I--EEKI--V--YA  538
HIV2NIHZ   ---TL--------L----LE--R---S-K---H--Q-EKKL---VQ-DQDN--T--V--GEKI--G--IC  537
HIV2ISY    ---TP--------L----LE--K---S-E---H--Q-EKEL---VQ-DQDN--T--V--GEKI--V--YA  537
HIV2ST     ---TL--------L----LE--K---S-E---C--Q-EKEL---VQ-DQDN--T--I--GGKI--V--YA  537
HIV2BEN    ---TL--------L----LE--K---S-E---Y--Q-EKEL---IQ-S(GH--T--I--EEKI--V--YA  537
HIV2D194   ---TL--------L----LE--K---S-E---S--Q-EEEL---VI-SQDN--A--I--GERV--V--YA  537
HIV2GH1    ---TL--------L----LE--K---S-E---Y--Q-EKEL---IQ-N(DN--T--I--EEKI--V--YA  537
HIV2D205   ---TL--------L----LQ--K---E-E---S--K-RVPL---VQ-NLAN--T--I--GNKV--V--YA  537
SIVMM251   ---TL--------M----YE--K---S-E---C--Q-GKPL---VI-SQDN--S--I--EDKI--V--FA  520
SIVMM142   ---TL--------M----YE--K---S-E---C--Q-GKPL---VI-SQDN--S--I--EDKI--V--FA  520
SIVMM239   ---TL--------M----YE--K---S-E---C--Q-GKPL---VI-SQDN--S--I--EDKI--V--FA  524
SIVMNE     ---TL--------M----YE--K---S-E---C--Q-GKPL---VI-NQDN--S--I--EDKI--V--FA  520
SIVSMMH4   ---TL--------M----YE--K---S-E---C--Q-GKPI---VI-SQDN--S--I--EDKV--V--FA  520
SIVSMMPBJ  ---TL--------M----YE--K---S-E---C--Q-GKPL---VI-SQDN--S--I--EDKI--V--FA  520

CONSENSUS  KlKnthtnqvRLLAhVvQKIGKEAlVIWG?iPkFHLPVEre?WeQWW?dYWQVTWiPeWDFvSTPPLvRL  564
HIV2ROD    -V-NTHTNGI----Q-V-------L----RI-K------REI-E---DN------I-D---V-----V--  608
HIV2NIHZ   -D-KYPYQRV----Q-V-------L----RI-K------RDT-E---DN------I-D---V-----V--  607
HIV2ISY    -I-NTHTNGV----Q-V-------L----RI-K------RET-E---DN------I-D---V-----V--  607
HIV2ST     -V-NTHTNGV----Q-V-------L----RI-K------RDT-E---DN------I-D---I-----V--  607
HIV2BEN    -I-NTHTNGV----Q-V-------L----RI-K------RET-E---DN------I-E---V-----V--  607
HIV2D194   -I-NTHTNGV----Q-V-------L----RV-K------RDT-E---DN------V-E---V-----V--  607
HIV2GH1    -I-NTHTNGV----Q-V-------L----RI-K------RET-E---DN------I-E---V-----V--  607
HIV2D205   -V-NTHTNGV----H-V-------L----EI-V------RET-D---TD------I-E---V-----I--  607
SIVMM251   -I-NTHTNGV----H-I-------I----QV-K------RDV-E---TD------I-E---I-----V--  590
SIVMM142   -I-NTHTNGV----H-I-------I----QV-K------KDV-E---TD------I-E---I-----V--  590
SIVMM239   -I-NTHTNGV----H-I-------I----QV-K------KDV-E---TD------I-K---I-----V--  594
SIVMNE     -I-NTHTNGV----H-I-------I----QV-K------KDV-E---TD------ I-K---I-----V--  590
SIVSMMH4   -V-NTHTNGV----H-V-------L----EV-K------REI-E---TD------I-D---V-----V--  590
SIVSMMPHJ  -I-NTHTNGV----H-V-------I----QV-R------REI-K---TD------I-E---V-----V--  590
```

HIV2/SIV POL

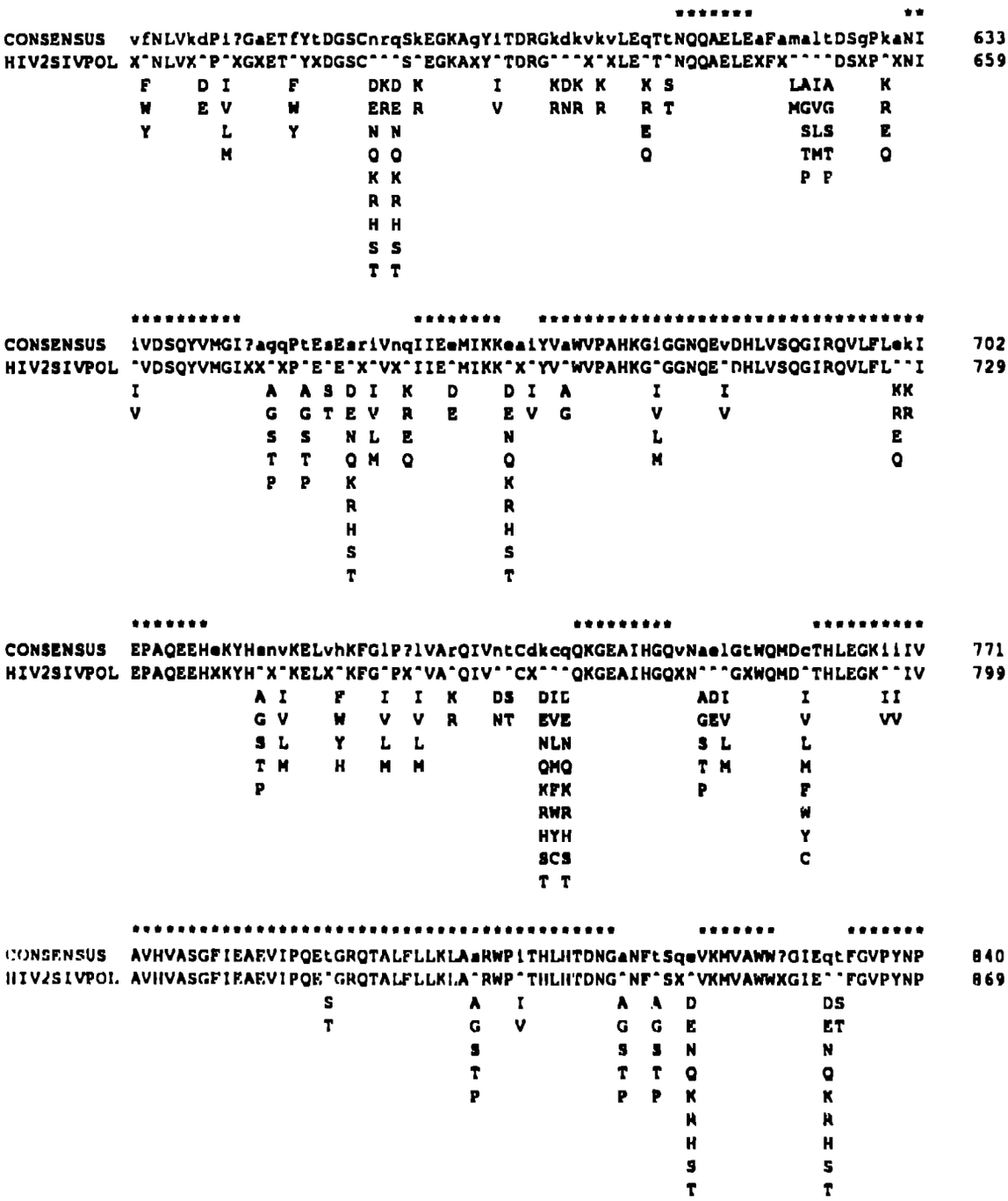
```
CONSENSUS  vfNLVkdPi?GaETfYtDGSCnrqSkEGKAgYiTDRGkdkvkvLEqTtNQQAELEaFamaltDSqPkaNI   633
HIV2ROD    AF---GD-IP-A--F-T----NRQ-K----G-V----KDKVKK--Q-T-------A-AMALT--G-KV--   678
HIV2NIHZ   AF---GE-VP-A--F-T----NRQ-K----G-I----RDRVKV--Q-T-------A-AMALT--G-KA--   677
HIV2ISY    AF---KD-IP-A--F-T----NRQ-K----G-I----KDKVRI--Q-T-------A-AMAVT--G-KV--   677
HIV2ST     VF---KD-IL-A--F-T----NKQ-R----G-I----RDKVRL--Q-T-------A-AMAVT--G-KA--   677
HIV2BEN    TF---GD-IP-A--F-T----NRQ-K----G-V----KDKVKV--Q-T-------V-RMALA--G-KV--   677
HIV2D194   TF---GD-IP-T--F-T----NRQ-K----G-V----RDRVRV--Q-S-------A-AMALA--G-KV--   677
HIV2GH1    TF---GD-IP-A--F-T----NRQ-K----R-V----RDKVRV--R-T-------A-AMTLT--G-KV--   677
HIV2D205   AY---KD-LE-R--Y-T----NRT-K----G-V----KDKVKV--Q-T-------A-ALALT--E-QV--   677
SIVMM251   VF---KD-IE-E--Y-T----NKQ-K----G-I----KDKVKV--Q-T-------A-LMALT--G-KT--   660
SIVMM142   VF---KD-IE-E--Y-V----SKQ-K----G-I----KDKVKV--Q-T-------A-LMALT--G-KA--   660
SIVMM239   VF---KD-IE-E--Y-T----NKQ-K----G-I----KDKVKV--Q-T-------A-LMALT--G-KA--   664
SIVMNE     VF---KD-IK-E--Y-V----NKQ-K----G-I----KDKVKV--Q-T-------A-LMALA--G-KA--   660
SIVSMMH4   VF---KE-IQ-A--F-V----NRQ-R----G-V----RDKAKL--Q-T-------A-YLALA--G-KA--   660
SIVSMMPBJ  VF---KE-IQ-A--F-V----NRQ-R----G-V----RDKAKL--Q-T-------A-YLALA--G-KA--   660

CONSENSUS  iVDSQYVMGI?aqqPtEsEsriVnqIIEeMIKKeaiYVaWVPAHKGiGGNQEvDHLVSQGIRQVLFLekI   702
HIV2ROD    I---------SASQ-T-S-SKI-NQ---E----EAI--A-------I-----V--------------EK-   740
HIV2NIHZ   I---------VAGQ-T-S-NRI-NQ---E----EAI--A-------I-----V--------------EK-   747
HIV2ISY    V---------VTGQ-A-S-SRI-NK---E----EAI--A-------I-----I--------------ER-   747
HIV2ST     I---------VAGQ-T-S-SKI-NQ---E----EAI--A-------I-----V--------------EK-   747
HIV2BEN    I---------VAGQ-T-S-NRI-NQ---E----EAV--A-------I-----V--------------EK-   747
HIV2D194   I---------VAGQ-T-S-NRI-NQ---D----EAV--A-------I-----V--------------EK-   747
HIV2GH1    I---------VVGQ-T-S-SRI-NQ---D----EAV--A-------I-----V--------------ER-   747
HIV2D205   I---------IAAQ-T-T-SPI-AK---E----EAV--G-------L-----V--------------EK-   747
SIVMM251   I---------ITGC-T-S-SRL-NQ---E----SEI--A-------I-----I--------------EK-   730
SIVMM142   I---------ITGC-T-S-SRL-NQ---E----TEI--A-------I-----I--------------EK-   730
SIVMM239   I---------ITGC-T-S-SRL-NQ---E----SEI--A-------I-----I--------------EK-   734
SIVMNE     I---------ITGC-T-S-SRL-NQ---E----TEI--A-------I-----I--------------EK-   730
SIVSMMH4   I---------IAGQ-T-S-SRL-NQ---E----EAI--A-------I-----V--------------KK-   730
SIVSMMPBJ  I---------VAGQ-T-S-SRL-NQ---E----EAI--A-------I-----V--------------EK-   730

CONSENSUS  EPAQEEHeKYHsnvKELvhKFGlP?lVArQIVntCdkcqQKGEAIHGQvNaelGtWQMDcTHLEGKiiIV   771
HIV2ROD    -------E---SNV---SH---I-NL--R---NS-AQCQ---------V-AEL-T----C------II--   818
HIV2NIHZ   -------E---SNI---SH---I-KL--R---NT-AHVQ---------V-AEL-T----C------VI--   817
HIV2ISY    -------G---SNV---AH---L-NL--R---NT-AQCQ---------V-AEL-T----C------II--   817
HIV2ST     -------E---SNV---SH---L-KL--R---NT-TQCQ---------V-AEL-T----C------II--   817
HIV2BEN    -------E---SII---TH---I-LL--R---NS-AQCQ---------V-AEI-V----Y------II--   817
HIV2D194   -------E---SNI---TH---I-QL--R---NT-AQCQ---------V-AEI-V----C------II--   817
HIV2GH1    -------E---SNM---TH---I-QL--R---NT-AQCQ---------V-AEI-V----C------II--   817
HIV2D205   -------E---GNV---VH---I-QL--K---NS-DKCQ---------V-ADL-T----C------II--   817
SIVMM251   -------D---SNV---VF---L-RI--R---DT-DKCH---------V-SDL-T----C------IV--   800
SIVMM142   -------S---SNI---VF---L-RL--K---DT-DKCH---------V-SDL-T----C------IV--   800
SIVMM239   -------D---SNV---VF---L-RI--R---DT-DKCH---------A-SDL-T----C------II--   804
SIVMNE     -------D---SNV---VF---L-RL--K---DT-DKCH---------V-SDL-T----C------II--   800
SIVSMMH4   -------E---SNV---VF---L-RL--K---DT-DKCH---------V-AEL-T----C------II--   800
SIVSMMPBJ  -------E---SNV---VF---L-RL--K---DT-DKCH---------V-AEL-T----C------II--   800
```

```
CONSENSUS  AVHVASGFIEAEVIPQEtGRQTALFLLKLAsRWPiTHLHTDNGaNFtSqeVKMVAWW?GIEqtFGVPYNP   840
HIV2ROD    ----------------S-----------S---I--------A--T-QE-------I---QS-------   888
HIV2NIHZ   ----------------S-----------S---I--------A--T-QE-------V---QT-------   887
HIV2ISY    ----------------S-----------S---I--------A--T-QE-------V---QS-------   887
HIV2ST     ----------------S-----------S---I--------A--T-QE-------I---QS-------   887
HIV2BEN    ----------------S-----------S---I--------P--T-QE-------V---QS-------   887
HIV2D194   ----------------S-----------S---I--------P--T-QE-------I---QS-------   887
HIV2GH1    ----------------S-----------S---I--------S--T-QE-------I---QS-------   887
HIV2D205   ----------------T-----------S---I--------A--T-PS-------V---QT-------   887
SIVMM251   ----------------T-----------G---I--------A--A-QE-------A---HT-------   870
SIVMM142   ----------------T-----------S---I--------A--A-QE-------A---HT-------   870
SIVMM239   ----------------T-----------G---I--------A--A-QE-------A---HT-------   874
SIVMNE     ----------------T-----------S---I--------A--A-QE-------A---HT-------   870
SIVSMMH4   ----------------T-----------G---I--------A--T-QE-------A---QT-------   870
SIVSMMPBJ  ----------------T-----------S---I--------A--T-QE-------A---QT-------   870

CONSENSUS  QSQGVVEAMNHHLKnQIdRiReQAnt?ETiVLMAvHCMNFKrRGGiGDMTPaERliNMiTtEQEIQF?Q?   907
HIV2ROD    --------------N--S-I-E--NTI--I----I------R---I-----S--LI--I-T------L-A   958
HIV2NIHZ   --------------N--D-I-E--NTV--I----V------R---I-----A--II--I-T------L-A   957
HIV2ISY    --------------N--E-I-E--NTM--I----V------R---I-----V--LV--I-T------L-A   957
HIV2ST     --------------N--S-I-E--NTV--I----V------R---I-----A--LI--V-A------L-A   957
HIV2BEN    --------------N--S-I-E--NTI--I----V------R---I-----A--LI--I-T------L-R   957
HIV2D194   --------------N--S-I-E--NTI--I----V------R---I-----A--LI--I-T------L-R   957
HIV2GH1    --------------N--S-I-E--NTI--I----V------R---I-----A--LI--I-T------L-R   957
HIV2D205   --------------N--D-L-D--VSI--V----T------R---I-----A--LV--I-T------F-A   957
SIVMM251   --------------N--D-I-E--NSV--I----V------R---I-----A--LI--I-T------Q-S   940
SIVMM142   --------------N--D-I-E--NSV--I----V------R---I-----A--LI--I-T------Q-S   940
SIVMM239   --------------N--D-I-E--NSV--I----V------R---I-----A--LI--I-T------Q-S   944
SIVMNE     --------------N--D-I-E--NSM--I----V------R---I-----A--LL--I-T------Q-S   940
SIVSMMH4   --------------T--D-I-E--NSI--I----V------R---I-----A--LV--I-T------Q-S   940
SIVSMMPBJ  --------------T--D-I-E--NSI--I----V------R---I-----A--LV--I-T------Q-S   940

                                                              vif cds ->
CONSENSUS  KNakfknFrVYyREGRdQLWkGPGELLWKGeGAVivKVGtdIKvvPRRKAKIIkdYGqgke?DasshlEd   976
HIV2ROD    --SKLKD-R--F----D---K---------E---LV---TD--II--------RD--GRQEM-SGSHL-G  1028
HIV2NIHZ   --SKLKN-R--F----D---K---------D---IV---TE--VV--------KD--GRQEM-SGSHL-G  1027
HIV2ISY    --SKLKN-R--F----N---Q---------D---IV---TD--VI--------RD--PRQEM-SGSHL-G  1027
HIV2ST     --SKLQN-R--F----D---K---------D---IV---AD--II--------KD--GRQEM-SGSNL-G  1027
HIV2BEN    --SNFKN-Q--Y----D---K---------E---IV---TD--VV--------RD--GRQEL-SSPHL-G  1027
HIV2D194   --SNFKK-Q--Y----D---K---------D---IV---AD--VV--------RD--GRQEL-SSSHL-G  1027
HIV2GH1    --SNFKN-Q--Y----D---K---------D---IV---AD--VI--------RD--GRQEL-SS.HL-G  1026
HIV2D205   --LKFQN-Q--Y----D---K---------E---II---TE--VV--------RH--GGKGL-CSADM-D  1027
SIVMM251   --SKFKN-R--Y----D---K---------E---IL---TD--VV--------KD--GGKEV-SSSHM-D  1010
SIVMM142   --SKFKN-R--Y----D---K---------E---IL---TD--VV--------KD--GGKEM-SSSHM-D  1010
SIVMM239   --SKFKN-R--Y----D---K---------E---IL---TD--VV--------KD--GGKEV-SSSHM-D  1014
SIVMNE     --SKFKN-R--Y----D---K---------E---VL---TD--VV--------KD--GGKEV-SSSHM-D  1010
SIVSMMH4   --SKFKN-R--Y----D---K---------E---IL---TE--VV--------KD--GGKEL-SGSHL-D  1010
SIVSMMPBJ  --SKFKN-R--Y----D---K---------E---IL---TE--VV--------KD--GGKEL-SGSHL-D  1010

CONSENSUS  tr?eaqEvA*???    985
HIV2ROD    AR.EDG-M--      1036
HIV2NIHZ   AR.EDG-M--      1035
HIV2ISY    AR.EDG-M--      1035
HIV2ST     AR.EDG-V--      1035
HIV2BEN    AR.EDG-M-CPCQ >  1122
HIV2D194   AR.EDG-V--      1035
HIV2GH1    ARREDG-V--      1035
HIV2D205   TR.QAR-M-QSD*   1038
SIVMM251   TG.EAR-V--      1018
SIVMM142   TG.EAR-V--      1018
SIVMM239   TG.EAR-V--      1022
SIVMNE     TG.EAR-V        1018
SIVSMMH4   TG.EAR-V--      1018
SIVSMMPBJ  TG.EAR-V        1018
```

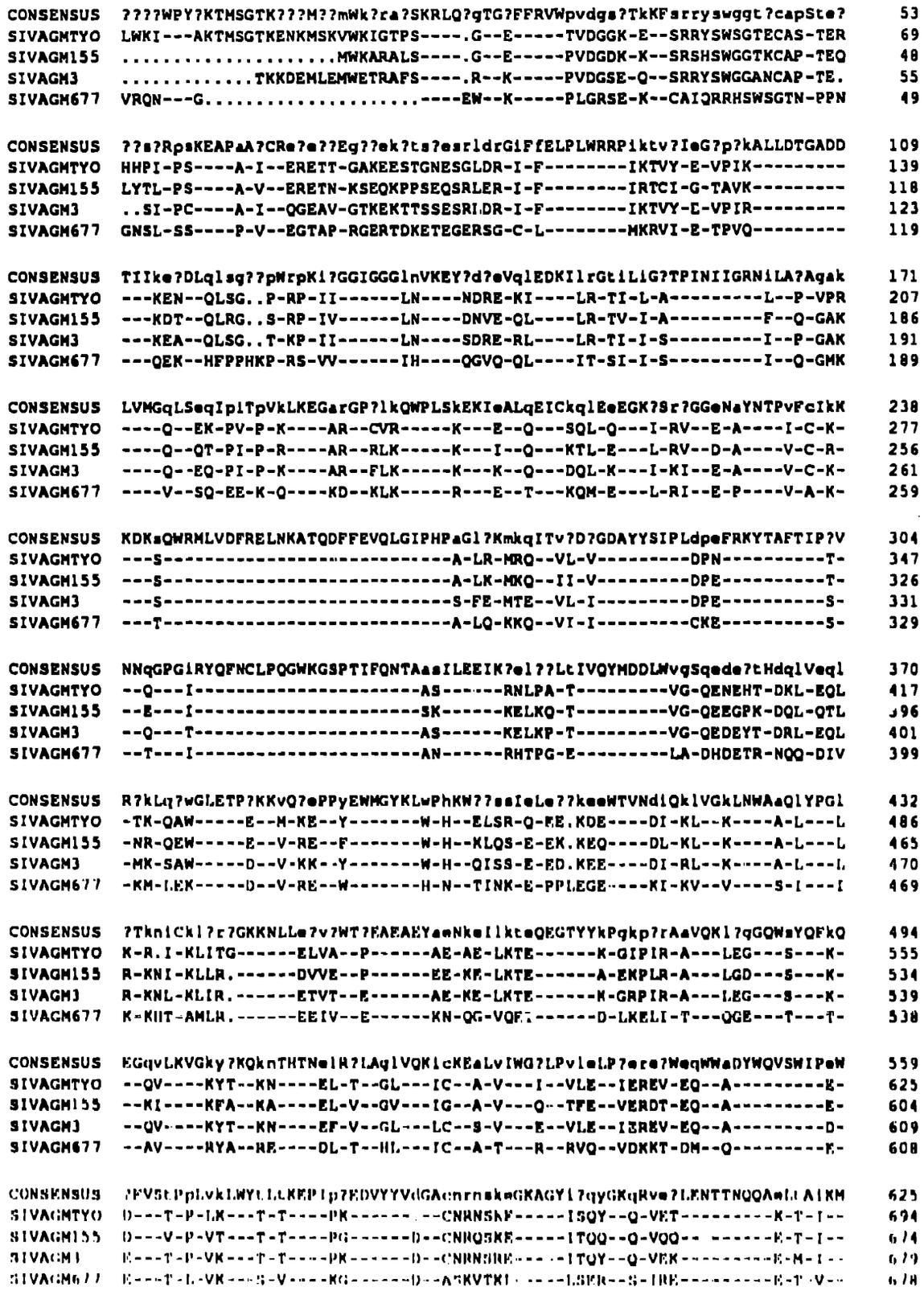
Information dense stretches (0.9 or greater) are indicated by ****.

```
CONSENSUS  TggFFR?wpmGKeapQlPhgps?sGa?????????????????dtnc?p?gsS?Gs??elhaag?k???     41
HIV2SIVPOL T^XFFRXX^XGKX^^Q^PXX^^X^G^..................^^XX^^XX^SXGXXX^^^XXX^X...     49
            A      A'   AA I   AD A A                  DA  AA  A      KIF   K
            G      G    GG V   GE G G                  NG  GG  G      RVW   R
            S      S     S L   SN S S                   S  SS  S      ELY   E
            T      T     T M   TQ T T                   T  TT  T      QMH   Q
            P      P     P F   PK P P                   P  PP  P
                           W    R
                           Y    H
                           C    S
                                T

                                                        <- gag cds
                                                  *********          ******
CONSENSUS  ?aer??retlQgg?????????????????drGfAaPQFSLWrRPvVtA?iEgQpvEVLLDTgadDSIV     89
HIV2SIVPOL .^^XXXX^^^QX^..................^XG^A^PQFSLW^RP^V^AX^EXQ^XEVLLDTXX^DSIV    100
            AK    DAI  A                  D  I A      K  I D  I   A         D
            GR    EGV  G                  N  V G      R  V E  V   G         N
            SE    NSL  S                     L S           N      S
            TQ    QTM  T                     M T           Q      T
            P     KP   P                     F P           K      P
                  R                          W             R
                  H                          Y             H
                  S                          C             S
                  T                                        T

                     ***************              ***************
CONSENSUS  aGieLG?nYtPKiVGGIGGFINTKEyKnveIeVlgKrvkaTIMTGDTPINIFGRN?LtalgMsLNlPvAK    157
HIV2SIVPOL ^G^^LGX^YXPK^VGGIGGFINTKE^K^^^I^V^XKX^^^TIMTGDTPINIFGRN^LX^^XM^LN^P^AK    170
           A IE   D    I            I DID K I   IKA               I  AL  S  I I
           G VQ   E    V            V NVE R V   VRG               V  GM  T  V V
           S      N                 L   N E L                     L  S      L
           T      Q                 M   Q Q                       M  T      M
           P      K                 F   K                            P      F
                  R                 W   R                                   W
                  H                 Y   H                                   Y
                  S                 C   S                                   C
                  T                     T

                  *****      ********* ****************************************
CONSENSUS  veP!KvtLKPGKdGPklrQWPLskEKIeALrEICEKMEkeGQLEEAPPTNPYNTPTFAIKKKDKNKWRmL    227
HIV2SIVPOL ^^P^KXXLKPGKXGP^X^QWPL^^EKIXAL^EICEKME^^GQLEEAPPTNPYNTPTFAIKKKDKNKWR^L    240
           ID I           K K    SK      K       KD                            I
           VE V           R R    TR      R       RE                            V
           L                                                                   L
           M                                                                   M
           ***************                      *********      *******        ***
CONSENSUS  IDFRELNkVTQDFTEvqlqiPhpaglAkkkRItvlDvGDAYFSIPLde?FRqYTAFTLPsvNNaEPqKRY    296
HIV2SIVPOL IDFRELN^VTQDFTE^^^X^P.^XX^A^^^RI^^^D^GDAYFSIPLXXXFRXYTAFTLP^^NNXEPXKRY    309
                  K       IDI I  A  I KKK  SII I                      AI
                  R       VEV V  G  V RRR  TVV V                      GV
                           NL L  S  L E      L                         S
                           QM M  T  M Q      M                         T
                           KF F  P                                     P
                           RW W
                           HY Y
                           SC C
                           T
```

```
           **** *******                          ************  *******   *****  **
CONSENSUS  iYKVlPQGWKGSpaIfQytMR?vLePFRkANpDviliQYMDDILiASDRtdLEHDrVVlQlKELLNglGF    365
HIV2SIVPOL ^YKVXPQGWKGSX^I^Q^XMR^^L^PFR^ANXD^X^^QYMDDIL^ASDR^XLEHD^VVXQXKELLNX^GF    379
           I            A I F   DI D   K    I II       I    S     K           I
           V            G V W   EV E   R    V VV       V    T     R           V
                        S L Y   N             L                               L
                        T M H   Q             M                               M
                        P F     K
                          W     R
                          Y     H
                          C     S
                                T

           ************           *****              *************
CONSENSUS  StPeEKFQKDPPfqWMGyELWPtKWKLQkiqLPqke?WTVNdIQKLVGVLNWAAQiypGIKTkhlCrLIr    434
HIV2SIVPOL S^P^EKFQKDPP^^WMG^ELWP^KWKLQ^^^LP^^^XWTVNXIQKLVGVLNWAAQ^^^GIKT^^^C^LI^    449
            S D        ID   I    D     KIE  EKD                   IFA    KDI K  K
            T E        VE   V    E     RVQ  QRE                   VWG    REV R  R
                       LN   L    N      L                         LYS     NL
                       MQ   M    Q      M                         M T     QM
                       FK   F    K                                  P     K
                       WR   W    R                                        R
                       YH   Y    H                                        H
                       CS   C    S                                        S
                        T        T                                        T

                ************
CONSENSUS  GKMtlTEEVQWTElAEAEleENkIILsQeQEG?YYqE?kplEATv?K?qdnQW?YKiHQe?kiLKvGK?a    497
HIV2SIVPOL GKM^XTEEVQWTE^AEAE^XEN^IIL^Q^QEGXYY^EXXX^EAT^XK^XX^QW^YK^HQXX^^LKXGK^X    519
              A         L    I   K   D K      K    I   I  D  D  A  I    KI     I
              G         M    V   R   E R      R    V   V  E  E  G  V    RV     V
              S              L       N E      E    L      N  N  S              L
              T              M       Q Q      Q    M      Q  Q  T              M
              P              F       K                     K  K  P              F
                             W       R                     R  R                 W
                             Y       H                     H  H                 Y
                             C       S                     S  S                 C
                                     T                     T  T

                            ************    ******          ******   ************
CONSENSUS  KiKnthtngvRLLAhVvQKIGKEAlVIWG?iPkFHLPVEre?WeQWW?dYWQVTWiPeWDFvSTPPLvRL    564
HIV2SIVPOL KXK^XXX^X^RLLA^V^QKIGKEA^VIWG^^PXFHLPVE^^XW^QWWX^YWQVTW^P^WDF^STPPL^RL    589
              D   D I    D I       I    XI        KD  D    D      I D   I     I
              E   E V    E V       V    RV        RE  E    N      V E   V     V
              N   N      N         L    E                         N
              Q   Q      Q         M    Q                         Q
              K   K      K                                        K
              R   R      R                                        R
              H   H      H                                        H
              S   S      S                                        S
              T   T      T                                        T
```

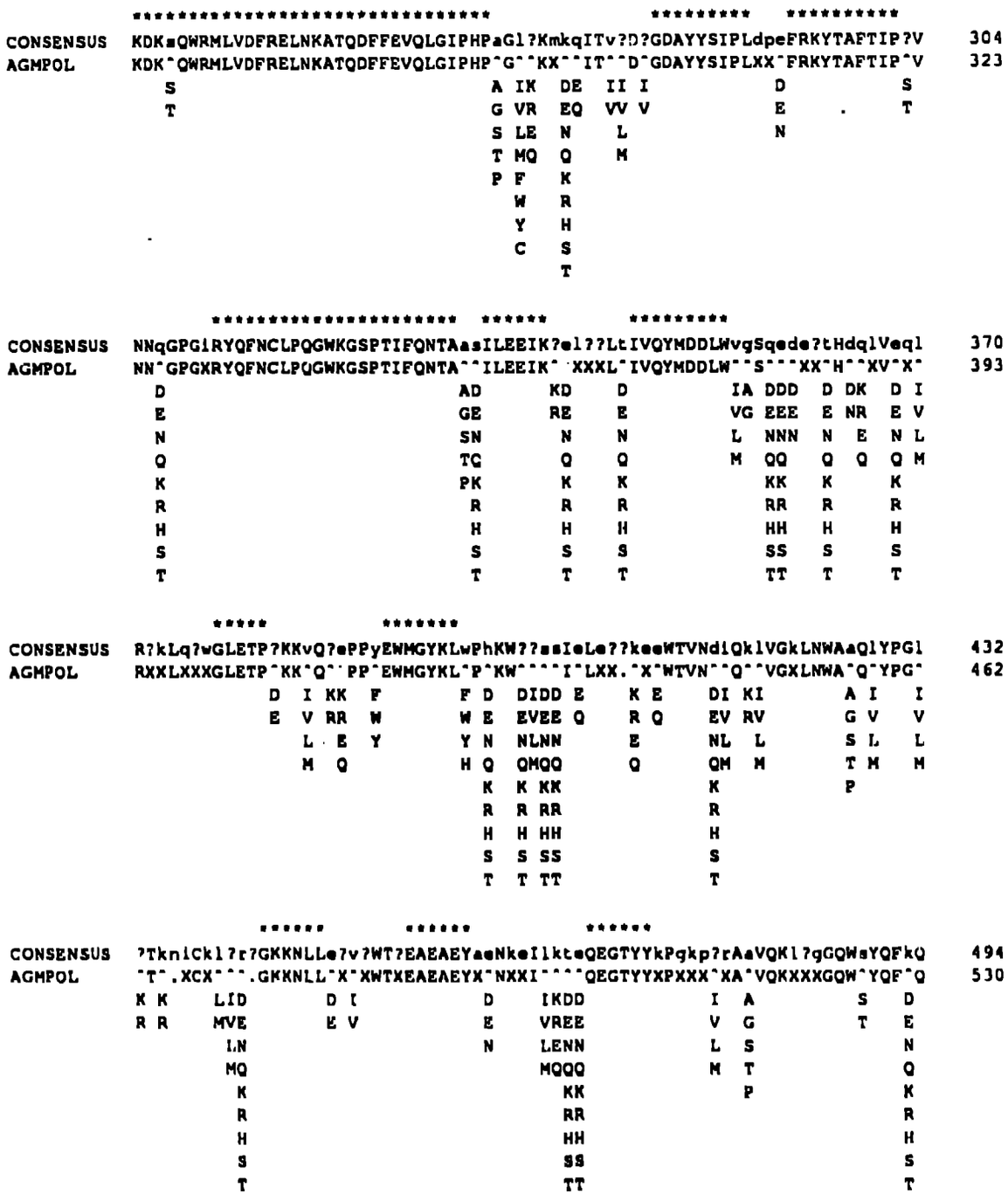

CONSENSUS vfNLVkdPi?GaETfYtDGSCnrqSkEGKAgYiTDRGkdkvkvLEqTtNQQAELEaFamaltDSgPkaNI 633
HIV2SIVPOL X^NLVX^P^XGXET^YXDGSC^^^S^EGKAXY^TDRG^^^X^XLE^T^NQQAELEXFX^^^^DSXP^XNI 659
CONSENSUS iVDSQYVMGI?aqqPtEaEariVnqIIEeMIKKeaiYVaWVPAHKGiGGNQEvDHLVSQGIRQVLFLekI 702
HIV2SIVPOL ^VDSQYVMGIXX^XP^E^E^X^VX^IIE^MIKK^X^YV^WVPAHKG^GGNQE^DHLVSQGIRQVLFL^^I 729
CONSENSUS EPAQEEHeKYHenvKELvhKFGlP?lVArQIVntCdkcqQKGEAIHGQvNaelGtWQMDcTHLEGKiiIV 771
HIV2SIVPOL EPAQEEHXKYH^X^KELX^KFG^PX^VA^QIV^^CX^^^QKGEAIHGQXN^^^GXWQMD^THLEGK^^IV 799
CONSENSUS AVHVASGFIEAEVIPQEtGRQTALFLLKLAaRWPiTHLHTDNGaNFtSqeVKMVAWW?GIEqtFGVPYNP 840
HIV2SIVPOL AVHVASGFIEAEVIPQE^GRQTALFLLKLA^RWP^THLHTDNG^NF^SX^VKMVAWWXGIE^^FGVPYNP 869

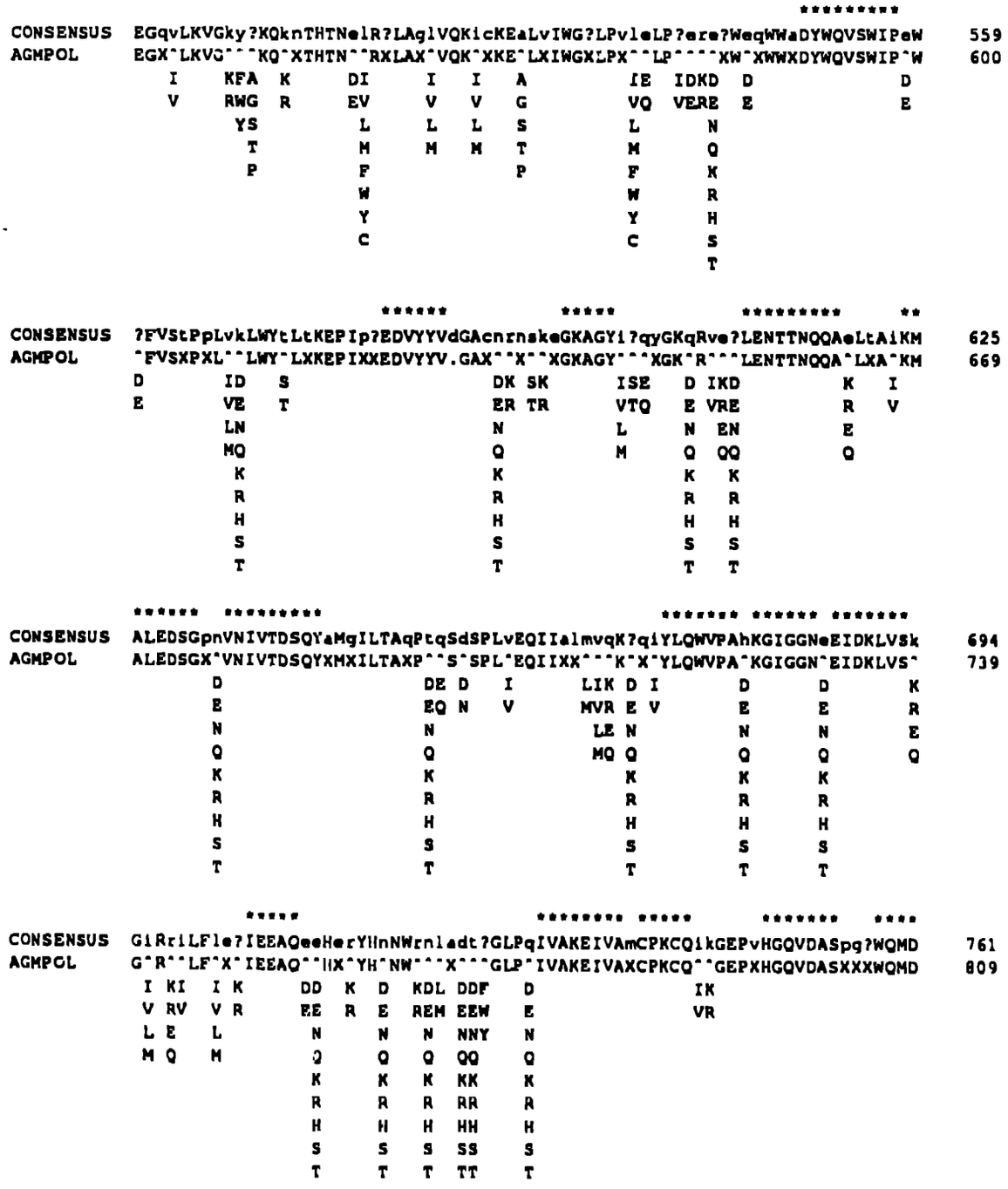
```
           **************              ******* ****************          ******
CONSENSUS  QSQGVVEAMNHHLKnQIdRiReQAnt?ETiVLMAvHCMNFKrRGGiGDMTPaERliNMiTtEQEIQF?Q?    907
HIV2SIVPOL QSQGVVEAMNHHLK^QI^R^R^QAX^^ET^VLMAXHCMNFK^RGG^GDMTPXER^^NM^T^EQEIQFXQX    939
                         D  D I D   SI  I           K   I        II  I A
                         E  E V E   TV  V           R   V        VV  V G
                         N  N L      L                  L        LL    S
                         Q  Q M      M                  M        MM    T
                         K  K                                         P
                         R  R
                         H  H
                         S  S
                         T  T

                                                        vif cds ->
                       **********************          ********
CONSENSUS  KNskfknFrVYyREGRdQLWkGPGELLWKGeGAVivKVGtdIKvvPRRKAKIIkdYGqqke?DssshlEd    976
HIV2SIVPOL KNX^^^^F^VY^REGR^QLW^GPGELLWKG^GAV^^KVG^^IK^^PRRKAKII^^YG^X^X^DX^.^^EX   1008
              DIKD K  F    D   K         D   II   AD  II        KD  A K I  A DL
              EVRE R  W    N   R         E   VV   GE  VV        RE  G R V  G EM
              NLEN E  Y        E             LL   S              N  S E L  S N
              QMQQ Q           Q             MM   T              Q  T Q M  T Q
              KF K                                P              K  P      P K
              RW R                                               R           R
              HY H                                               H           H
              SC S                                               S           S
              T  T                                               T           T

CONSENSUS  tr?eagEvA???Q*    985
HIV2SIVPOL ^X.^XXE^A*       1016
           A  E   I
           G  Q   V
           S      L
           T      M
           P
```



```
CONSENSUS  ALEDSGpnVNIVTDSQYaMgILTAqPtqSdSPLvEQIIalmvqK?qiYLQWVPAhKGIGGNeEIDKLVSk   694
SIVAGMTYO  ------PN---------A-G----Q-TQ-D---V----ALMIQ-QQI-------H------E-------K   764
SIVAGM155  ------PK---------A-G----Q-TQ-D---V----AQMVQ-EAI-------H------E-------K   744
SIVAGM3    ------PN---------A-G----Q-TQ-D---I----ALMVQ-HQI-------D------E-------Q   749
SIVAGM677  ------EN---------V-N----C-QE-N---V----QALMK-RQV-------H------T-------K   748

CONSENSUS  GiRriLFle?IEEAQeeHerYHnNWrnladt?GLPqIVAKEIVAmCPKCQikGEPvHGQVDASpg?WQMD   761
SIVAGMTYO  -I-RV--LEK-----EK-ER--N--KNLADTY---Q--------M-----IK---V-------PGT----   834
SIVAGM155  -V-RI--IGR-----EE-DR--S--RNLADTF---Q--------M-----VK---I-------PGV----   814
SIVAGM3    -M-KI--LEK-----EE-ER--N--RNLADTY---Q--------M-----IK---V-------PGV----   819
SIVAGM677  -I-QI--LDR-----DD-AK--N--RSMVQEF---N--------A-----IR---K-------IET----   818

CONSENSUS  CTHlEqKv?IVAVHVASGFIEAEVIPRETGKETAkFLLKilsRWPit?LHTDNGPNFTSQeVAAmCWWGk   829
SIVAGMTYO  ---L-K-VV-------------------------K----ILS---ITQ-----------E---I----K   904
SIVAGM155  ---I-G-IV-------------------------K----IIG---ITH-----------E---M----K   884
SIVAGM3    ---L-G-VI-------------------------K----ILS---ITQ-----------E---M----K   889
SIVAGM677  ---L-G-VI-------------------------H----LLA---VKH-----------N---V----N   888

CONSENSUS  iEHTTG?PYNPQSQGSiESMNkQLKEIIgkIRDDCqytEtAVlMAchiHNFKRKGGIGGltsAERLiNmi   898
SIVAGMTYO  I-----I---------I----K------GK-----QYT-A--L--CIL----------QTS----I-II   974
SIVAGM155  V-----V---------I----K------GK-----QYT-T--L--CHI----------LTA----I-MI   954
SIVAGM3    I-----V---------I----K------GK-----QYT-T--L--CHI----------LTP----I-MI   959
SIVAGM677  I-----I---------V----R------SQ-----ERL-T--Q--THI----------ISS----V-ML   958

CONSENSUS  TTQLE???LQtkIQKILNFrVYYREGRDPVWKGPa?LIWKGEGAVV?KeGe?lKvvPRRKAKIIKDYepr   962
SIVAGMTYO  -----IQH--TK-------R-------------AQ----------L-D-SDL-VV----------EPK  1044
SIVAGM155  -----INT--TK-------R-------------AR----------L-E-EEL-VV----------EPR  1024
SIVAGM3    -----LQH--TK-------R-------------GQ----------I-G-VEL-EY----------EPR  1029
SIVAGM677  -----LNT--NQ-------K-------------AR----------I-E-EDI-VV----------GER  1028

CONSENSUS  k?mgdEg?lEGa?g?dn?M?GDS?          978
SIVAGMTYO  QRVGN-GDV--TRGSDN*               1061
SIVAGM155  KTLGD-THL--AGGSDHQ-A---*         1047
SIVAGM3    KRMGD-SNL--AGGADN*               1046
SIVAGM677  KTMDS-GSM--VREANKQ-E---DLQDQE*   1057
```

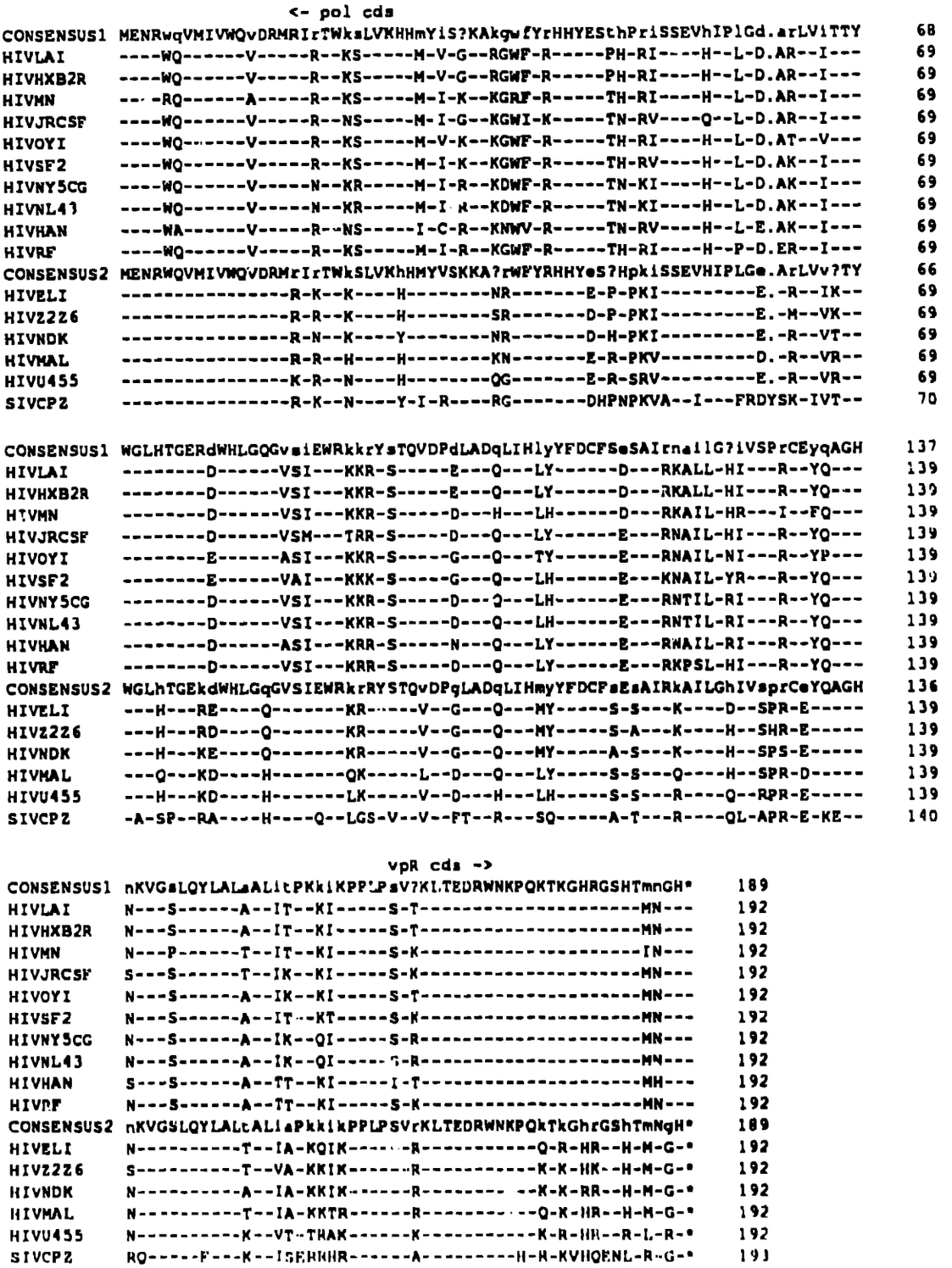
```
   Information dense regions (0.95 or greater) are indicated by ****.

                                          *****     *****
CONSENSUS  ????WPY?KTMSGTK???M??mWk?ra?SKRLQ?gTG?FFRVWpvdgs?TkKFsrryswggt?capSte?      53
AGMPOL     ....................^W^XX^XSKRLQ.XTG^FFRVW^^XXX^T^KFXXXX^XX^X^X^XS^X.      47
                           I K  A          K      AI    K K       D  A D A   A
                           V R  G          R      GV    R R       E  G E G   G
                           L E  S          E      SL    E E       N  S N S   S
                           M Q  T          Q      TH    Q Q       Q  T Q T   T
                                P                 P               K  P K P   P
                                                                  R     R
                                                                  H     H
                                                                  S     S
                                                                  T     T

                                                       ********           *********
CONSENSUS  ??s?RpsKEAPaA?CRe?e??Eg??ek?ts?esrldrGiFfELPLWRRPiktv?IeG?p?kALLDTGADD     109
AGMPOL     ..^^R^XKEAP^A^CR^X^^XEX^^^^XXX^^^XX^XG^F^ELPLWRRP^^^^^IXGX^^^ALLDTGADD     115
             AI A      A I   E DA    AKKD    DEA  D  I I          IKDII     AIK
             GV G      G V   Q EG    GRRE    EQG  E  V V          VREVV     GVR
             SL S      S       NS    SEEN    N S  N  L L          L NLL     S E
             TM T      T       QT    TQQQ    Q T  Q  M M          M QMM     T Q
             P  P      P       KP    P  K    K P  K  F F            KFF     P
                               R        R     R    R  W W            RWW
                               H        H     H    H  Y Y            HYY
                               S        S     S    S  C C            SCC
                               T        T     T    T                 T

           ***                      ******                        *********
CONSENSUS  TIIke?DLqlsg??pWrpKi?GGIGGGlnVKEY?d?eVqlEDKIlrGtiLiG?TPINIIGRNiLA?Aqak     171
AGMPOL     TII^^XDL^^X^..^W^^K^^GGIGGG^^VKEY^XX^V^^EDKI^^G^^L^G^TPINIIGRN^LAXAXX^     183
             KD    DI A   A KA II       ID     D  E KI     ID SI I A           I       K
             RE    EV G   G RG VV       VE     E  Q RV     VE TV V G           V       R
             E     NL S   S  S          LN     N     EL    LN    L S           L
             Q     QM T   T  T          MQ     Q     QM    MQ    M T           M
                   KF P   P  P            K     K           K      P           F
                   RW                     R     R           R                  W
                   HY                     H     H           H                  Y
                   SC                     S     S           S                  C
                   T                      T     T           T

                                     *****                                    *
CONSENSUS  LVMGqLSeqIpiTpVkLKEGarGP?lkQWPLSkEKIeALqEICkqlEeEGK?Sr?GGeNaYNTPvFcIkK     238
AGMPOL     LVMGXLS^^IXXTXV^LKEGX^GPX^^QWPLS^EKIXAL^EIC^^^E^EGK^S^^GG^N^YNTP^FXI^K     253
                DD       K      D   IK      K       D   DDL K   I KI  D A     I   K
                EE       R      E   VR      R       E   EEM R   V RV  E G     V   R
                NN       E      N   L               N   NN  E   L         S
                QQ       Q      Q   M               Q   QQ  Q   M         T
                KK              K                   K   KK              P
                RR              R                   R   RR
                HH              H                   H   HH
                SS              S                   S   SS
                TT              T                   T   TT
```


CONSENSUS KDKaQWRMLVDFRELNKATQDFFEVQLGIPHPaGl?KmkqITv?D?GDAYYSIPLdpeFRKYTAFTIP?V 304
AGMPOL KDK^QWRMLVDFRELNKATQDFFEVQLGIPHP^G^^KX^^IT^^D^GDAYYSIPLXX^FRKYTAFTIP^V 323
CONSENSUS NNqGPGiRYQFNCLPQGWKGSPTIFQNTAasILEEIK?el??LtIVQYMDDLWvgSqede?tHdqlVeql 370
AGMPOL NN^GPGXRYQFNCLPQGWKGSPTIFQNTA^^ILEEIK^^XXXL^IVQYMDDLW^^S^^^XX^H^^XV^X^ 393
CONSENSUS R?kLq?wGLETP?KKvQ?aPPyEWMGYKLwPhKW??asIaLe??keeWTVNdiQklVGkLNWAaQlYPGl 432
AGMPOL RXXLXXXGLETP^KK^Q^^PP^EWMGYKL^P^KW^^^^I^LXX.^X^WTVN^^Q^^VGXLNWA^Q^YPG^ 462
CONSENSUS ?TknlCkl?r?GKKNLLe?v?WT?EAEAEYaaNkeIlkteQEGTYYkPgkp?rAaVQKl?gGQWsYQFkQ 494
AGMPOL ^T^.XCX^^^.GKKNLL^X^XWTXEAEAEYX^NXXI^^^^QEGTYYXPXXX^XA^VQKXXXGQW^YQF^Q 530


CONSENSUS EGqvLKVGky?KQknTHTNelR?LAglVQKicKEaLvIWG?LPvleLP?ere?WeqWWaDYWQVSWIPeW 559
AGMPOL EGX^LKVG^^^KQ^XTHTN^^RXLAX^VQK^XKE^LXIWGXLPX^^LP^^^^XW^XWWXDYWQVSWIP^W 600
CONSENSUS ?FVStPpLvkLWYtLtKEPIp?EDVYYVdGAcnrnskeGKAGYi?qyGKqRve?LENTTNQQAeLtAiKM 625
AGMPOL ^FVSXPXL^^LWY^LXKEPIXXEDVYYV.GAX^^X^^XGKAGY^^^XGK^R^^^LENTTNQQA^LXA^KM 669
CONSENSUS ALEDSGpnVNIVTDSQYaMgILTAqPtqSdSPLvEQIIalmvqK?qiYLQWVPAhKGIGGNeEIDKLVSk 694
AGMPOL ALEDSGX^VNIVTDSQYXMXILTAXP^^S^SPL^EQIIXX^^^K^X^YLQWVPA^KGIGGN^EIDKLVS^ 739
CONSENSUS GiRriLFle?IEEAQeeHerYHnNWrnladt?GLPqIVAKEIVAmCPKCQikGEPvHGQVDASpg?WQMD 761
AGMPOL G^R^^LF^X^IEEAQ^^HX^YH^NW^^^X^^^GLP^IVAKEIVAXCPKCQ^^GEPXHGQVDASXXXWQMD 809

```
           ***      *************************              ************
CONSENSUS  CTHlEgKv?IVAVHVASGFIEAEVIPRETGKETAkFLLKilsRWPit?LHTDNGPNFTSQeVAAmCWWGk    829
AGMPOL     CTH^EXK^^IVAVHVASGFIEAEVIPRETGKETA^FLLK^^^RWP^^^LHTDNGPNFTSQ^VAA^CWWG^    879
              I   II                              D    IIA   IDD             D   I    D
              V   VV                              E    VVG   VEE             E   V    E
              L                                   N    LLS    NN             N   L    N
              M                                   Q    MMT    QQ                 M    Q
                                                  K      P    KK                      K
                                                  R           RR                      R
                                                  H           HH                      H
                                                  S           SS                      S
                                                  T           TT                      T

            ***************************  *****            ************
CONSENSUS  iEHTTG?PYNPQSQGSiESMNkQLKEIIgkIRDDCqytEtAVlMAchiHNFKRKGGIGGltsAERLiNmi    898
AGMPOL     ^EHTTG^PYNPQSQGS^ESMN^QLKEII^^IRDDC^XXE^AVXMAXX^HNFKRKGGIGGX^^AERL^N^^    949
           I     I         I    K      AK     E   A       I             SA     I II
           V     V         V    R      GR     Q   G       V             TG     V VV
                                       SE         S       L              S       LL
                                       TQ         T       M              T       MM
                                       P          P                      P
           *****       **********************  **********         ***********
CONSENSUS  TTQLE???LQtkIQKILNFrVYYREGRDPVWKGPa?LIWKGEGAVV?KeGe?lKvvPRRKAKIIKDYepr    962
AGMPOL     TTQLE^^^LQ^^IQKILNF^VYYREGRDPVWKGP^^LIWKGEGAVV^KXGX^^KX^PRRKAKIIKD^XX^   1019
                IDD  DK       K              AK          I    DI  I              K
                VEE  ER       R              GR          V    EV  V              R
                LNN  NE                      E           L     L  L
                MQQ  QQ                      Q           M     M  M
                 KK  K                                          F
                 RR  R                                          W
                 HH  H                                          Y
                 SS  S                                          C
                 TT  T

CONSENSUS  k?mgdEg?lEGa?g?dn?M?GDS?    978
AGMPOL     ^^^X^E^^^EGXXX^^^*         1036
           KDI D ADI      ADD
           REV E GEV      GNE
           ENL N SNL      S N
           QQM Q TQM      T Q
            K  K PK       P K
            R  R  R         R
            H  H  H         H
            S  S  S         S
            T  T  T         T
```

HIV1 VIF


<- pol cds
CONSENSUS1 MENRwqVMIVWQvDRMRIrTWksLVKHHmYiS?KAkgwfYrHHYESthPriSSEVhIPlGd.arLViTTY 68
HIVLAI ----WQ------V-----R--KS-----M-V-G--RGWF-R-----PH-RI----H--L-D.AR--I--- 69
HIVHXB2R ----WQ------V-----R--KS-----M-V-G--RGWF-R-----PH-RI----H--L-D.AR--I--- 69
HIVMN --·-RQ------A-----R--KS-----M-I-K--KGRF-R-----TH-RI----H--L-D.AR--I--- 69
HIVJRCSF ----WQ------V-----R--NS-----M-I-G--KGWI-K-----TN-RV----Q--L-D.AR--I--- 69
HIVOYI ----WQ------V-----R--KS-----M-V-K--KGWF-R-----TH-RI----H--L-D.AT--V--- 69
HIVSF2 ----WQ------V-----R--KS-----M-I-K--KGWF-R-----TH-RV----H--L-D.AK--I--- 69
HIVNY5CG ----WQ------V-----N--KR-----M-I-R--KDWF-R-----TN-KI----H--L-D.AK--I--- 69
HIVNL43 ----WQ------V-----N--KR-----M-I·R--KDWF-R-----TN-KI----H--L-D.AK--I--- 69
HIVHAN ----WA------V-----R--NS-----I-C-R--KNWV-R-----TN-RV----H--L-E.AK--I--- 69
HIVRF ----WQ------V-----R--KS-----M-I-R--KGWF-R-----TH-RI----H--P-D.ER--I--- 69
CONSENSUS2 MENRWQVMIVWQVDRMrIrTWkSLVKhHMYVSKKA?rWFYRHHYeS?HpkiSSEVHIPLGe.ArLVv?TY 66
HIVELI ----------R-K--K----H--------NR-------E-P-PKI---------E.-R--IK-- 69
HIVZ2Z6 ----------R-R--K----H--------SR-------D-P-PKI---------E.-M--VK-- 69
HIVNDK ----------R-N--K----Y--------NR-------D-H-PKI---------E.-R--VT-- 69
HIVMAL ----------R-R--H----H--------KN-------E-R-PKV---------D.-R--VR-- 69
HIVU455 ----------K-R--N----H--------QG-------E-R-SRV---------E.-R--VR-- 69
SIVCPZ ----------R-K--N----Y-I-R----RG-------DHPNPKVA--I---FRDYSK-IVT-- 70
CONSENSUS1 WGLHTGERdWHLGQGvsiEWRkkrYsTQVDPdLADqLIHlyYFDCFSeSAIrnalIG?iVSPrCEyqAGH 137
HIVLAI --------D------VSI---KKR-S-----E---Q---LY------D---RKALL-HI---R--YQ--- 139
HIVHXB2R --------D------VSI---KKR-S-----E---Q---LY------D---RKALL-HI---R--YQ--- 139
HIVMN --------D------VSI---KKR-S-----D---H---LH------D---RKAIL-HR---I--FQ--- 139
HIVJRCSF --------D------VSM---TRR-S-----D---Q---LY------E---RNAIL-HI---R--YQ--- 139
HIVOYI --------E------ASI---KKR-S-----G---Q---TY------E---RNAIL-NI---R--YP--- 139
HIVSF2 --------E------VAI---KKK-S-----G---Q---LH------E---KNAIL-YR---R--YQ--- 139
HIVNY5CG --------D------VSI---KKR-S-----D---Q---LH------E---RNTIL-RI---R--YQ--- 139
HIVNL43 --------D------VSI---KKR-S-----D---Q---LH------E---RNTIL-RI---R--YQ--- 139
HIVHAN --------D------ASI---KRR-S-----N---Q---LY------E---RNAIL-RI---R--YQ--- 139
HIVRF --------D------VSI---KRR-S-----D---Q---LY------E---RKPSL-HI---R--YQ--- 139
CONSENSUS2 WGLhTGEkdWHLGqGVSIEWRkrRYSTQvDPqLADqLIHmyYFDCFsEsAIRkAILGhIVsprCeYQAGH 136
HIVELI ---H---RE----Q-------KR-----V--G---Q---MY-----S-S---K----D--SPR-E----- 139
HIVZ2Z6 ---H---RD----Q-------KR-----V--G---Q---MY-----S-A---K----H--SHR-E----- 139
HIVNDK ---H---KE----Q-------KR-----V--G---Q---MY-----A-S---K----H--SPS-E----- 139
HIVMAL ---Q---KD----H-------QK-----L--D---Q---LY-----S-S---Q----H--SPR-D----- 139
HIVU455 ---H---KD----H-------LK-----V--D---H---LH-----S-S---R----Q--RPR-E----- 139
SIVCPZ -A-SP--RA----H----Q--LGS-V--V--FT--R---SQ-----A-T---R----QL-APR-E-KE-- 140
vpR cds ->
CONSENSUS1 nKVGsLQYLALsALitPKkiKPPLPsV?KLTEDRWNKPQKTKGHRGSHTmnGH* 189
HIVLAI N---S------A--IT--KI-----S-T----------MN--- 192
HIVHXB2R N---S------A--IT--KI-----S-T----------MN--- 192
HIVMN N---P------T--IT--KI-----S-K----------IN--- 192
HIVJRCSF S---S------T--IK--KI-----S-K----------MN--- 192
HIVOYI N---S------A--IK--KI-----S-T----------MN--- 192
HIVSF2 N---S------A--IT--KT-----S-K----------MN--- 192
HIVNY5CG N---S------A--IK--QI-----S-R----------MN--- 192
HIVNL43 N---S------A--IK--QI-----S-R----------MN--- 192
HIVHAN S---S------A--TT--KI-----I-T----------MH--- 192
HIVRF N---S------A--TT--KI-----S-K----------MN--- 192
CONSENSUS2 nKVGSLQYLALtALiaPkkikPPLPSVrKLTEDRWNKPQkTkGhrGShTmNgH* 189
HIVELI N----------T--IA-KQIK------R----------Q-R-HR--H-M-G-* 192
HIVZ2Z6 S----------T--VA-KKIK------R----------K-K-HK--H-M-G-* 192
HIVNDK N----------A--IA-KKIK------R---------K-K-RR--H-M-G-* 192
HIVMAL N----------T--IA-KKTR------R----------Q-K-HR--H-M-G-* 192
HIVU455 N----------K--VT-THAK------K----------K-R-HR--R-L-R-* 192
SIVCPZ RQ-----F---K--ISERHHR------A---------H-R-KVHQENL-R-G-* 193

Information dense stretches (0.8 or greater) are indicated by ****.

```
                          <- pol cds
                 ******                          *******                      **
CONSENSUS1 MENRwqVMIVWQvDRMRIrTWksLVKHHmYiS?KAkgwfYrHHYESthPriSSEVhIPlGd.arLViTTY     68
CONSENSUS2 ----WQ------V---r-----S---h-M-V-K--?rWF-R---e-?Hpk-----H--L-e.A---v?--     66
HIV1VIF    MENRXXVMIVWQXDRM^I^TW^^LVK^H^Y^SXKA^XXXY^HHY^SX^^^^SSEV^IPXG^.XXLV^^TY     69
                          K D  DD   F I I    D    K   D  DAKI    D    D     ID
                          R E  EE   W V V    E    R   E  EGRV    E    E     VE
                            N  NN   Y L L    N           NS      N           N
                            Q  QQ   H M M    Q           QT      Q           Q
                            K  KK       F    K           KP      K           K
                            R  RR       W    R           R       R           R
                            H  HH       Y    H           H       H           H
                            S  SS       C    S           S       S           S
                            T  TT            T           T       T           T

           *************                          *****
CONSENSUS1 WGLHTGERdWHLGQGvsiEWRkkrYsTQVDPdLADqLIHlyYFDCFSeSAIrnailG?iVSPrCEyqAGH    137
CONSENSUS2 ---h---k-----q-VSI----rR-S--v--g-------m------sEs--RkAIL-hI-sp--eYQ---    136
HIV1VIF    WGL^TGE^^WHLG^GX^^EWRX^XY^TQ^DPXLAD^LIHX^YFDCF^^^AI^^^X^GXXV^XXC^^XAGH    139
              D   KD    D  AI    K  D  I      D    F     ADA  KDA I    D   DF
              E   RE    E  GV    R  E  V      E    W     GEG  REG V    E   EW
              N         N  SL       N  L      N    Y     S S   NS L    N    Y
              Q         Q  TM       Q  M      Q    H     T T   QT M    Q
              K         K  P        K         K          P P   KP      K
              R         R           R         R                R       R
              H         H           H         H                H       H
              S         S           S         S                S       S
              T         T           T         T                T       T

                                   vpR cds ->
            **********                 ***************
CONSENSUS1 nKVGsLQYLALaALitPKkiKPPLPsV?KLTEDRWNKPQKTKGHRGSHTmnGH*    189
CONSENSUS2 ----S------t---a-k--k----S-r-----------k-k-hr--h--Ng-*    189
HIV1VIF    ^KVG^LQYLALXALXXP^^X^PPLPXV^KLTEDRWNKPQ^T^G^^GS^T^^XH*    192
           D   A            DK K      D           K K DK  D ID
           E   G            ER R      E           R R ER  E VE
           N   S            NE        N           E   N   N LN
           Q   T            QQ        Q           Q   Q   Q MQ
           K   P            K         K               K   K  K
           R                R         R               R   R  R
           H                H         H               H   H  H
           S                S         S               S   S  S
           T                T         T               T   T  T
```

HIV1VIF:6 VMIVWQ

HIV1VIF:39 YnHHYkS

HIV1VIF:68 TYWGLrTGEnkWHLG

HIV1VIF:111 YFDCF

HIV1VIF:141 KVGjLQYLAL

This fragment is similar to a conserved pattern in HIV2, "VPSLQyLAL" (and in all known lentiviral vif or Q orf proteins).

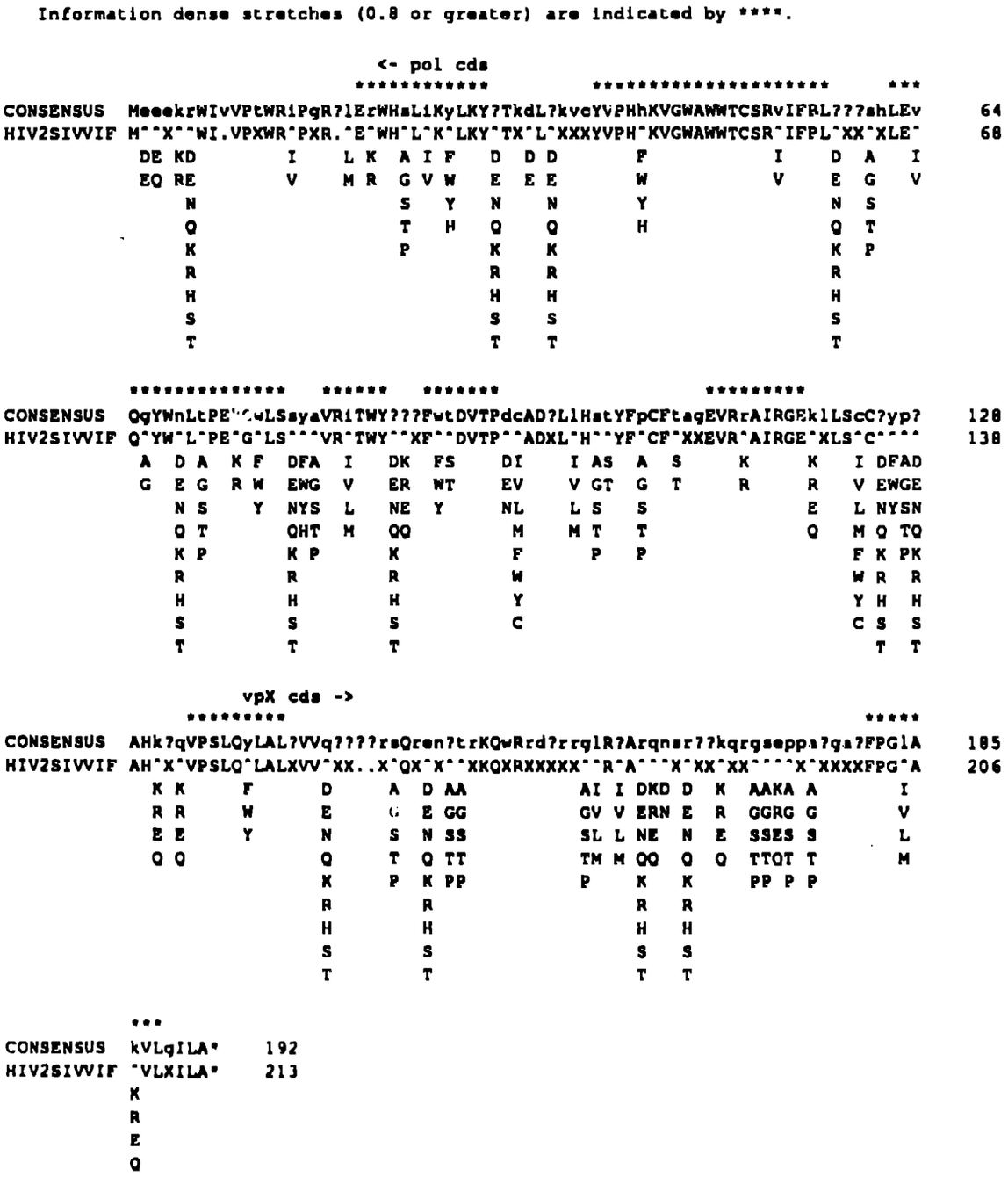
HIV1VIF:168 KLTEDRWNKPQpTnG

HIV2/SIV VIF

```
                                <- pol cds
CONSENSUS  MeeekrWIvVPtWRiPqR?lErWHsLiKyLKY?TkdL?kvcYVPHhKVGWAWWTCSRvIFPL???shLEv   64
HIV2ROD    -EEDKR--V--T--V-G-.M-K--S-V-Y---K-KD-EKVC----H----------V----KGNSH--I   69
HIV2NIHZ   -EEGKR--V--I--V-G-.M-R--S-V-Y---R-KD-EKVC----H----------V----KENSH--I   69
HIV2ISY    -DQGKR--A--T--V-G-.M-K--S-I-Y---R-KD-EQVR----H----------V----KGNSH--I   69
HIV2ST     -EEGKR--A--T--V-G-.M-R--S-I-Y---R-GD-EKVC----H----------V----KGESH--I   69
HIV2BEN    -EEDRN--V--T--V-G-.M-K--A-V-Y---R-KD-EEVR----H----------V----QGKSH--I   69
HIV2D194   -EEGKN--V--T--V-G-.M-R--S-V-H---R-KD-EEVR----H----------V----EGESH--I   69
HIV2GH1    -EEGKN--.--T--V-G-RM-R--S-V-Y---R-RD-EEVR----H----------V----KGESH--I   69
HIV2D205   -EEEKD--V--T--I-G-.L-R--S-I-Y---R-GE-QQVS----H----------I----NKGAW--V   69
SIVMM251   -EEEKR--A--T--I-E-.L-R--S-I-Y---K-KD-QKVC----F----------V----QEGSH--V   69
SIVMM142   -EEEKR--V--T--I-E-.L-R--S-I-Y---K-KD-QKAC----H----------V----QEGSH--V   69
SIVMM239   -EEEKR--A--T--I-E-.L-R--S-I-Y---K-KD-QKVC----F----------V----QEGSH--V   69
SIVMNE     -EEEKR--A--T--I-E-.L-R--S-I-Y---K-KD-GKVC----H----------V----QEKSQ--V   69
SIVSMMH4   -EEEKN--V--T--I-E-.L-R--S-I-H---N-KD-QMAC----H----------V----RDETH--V   69
SIVSMMPBJ  -EEEKN--V--T--I-G-.L-K--S-I-H---N-KD-QKAC----H----------V----KDEAH--V   69

CONSENSUS  QqYWnLtPEkGwLSsyaVRiTWY???FwtDVTPdcAD?LlHstYFpCFtagEVRrAIRGEklLScC?yp?  128
HIV2ROD    -A--N-T--K-W--SYS--I---TEK-WT----DC--V-I-ST--P--TAG---R-----KL--C-NYPR  139
HIV2NIHZ   -A--N-T--K-W--SHS--I---TEK-WT----DC--T-I-ST--S--TAG---R-----KL--C-KYPR  139
HIV2ISY    -A--N-T--K-W--SYS--M---SEG-WT----DC--T-I-ST--S--TAG---R-----KS--C-NYPQ  139
HIV2ST     -A--N-T--K-W--SYS--L---TEK-WT----DC--S-I-ST--S--TAG---R-----KL--C-NYPQ  139
HIV2BEN    -A--N-T--K-W--SHA--L---TEK-WT----DC--I-I-ST--S--TAG---R-----KL--C-NYPQ  139
HIV2D194   -A--N-T--K-W--SHS--L---TEK-WT----DC--S-I-ST--S--TAG---R-----KL--C-NYPQ  139
HIV2GH1    -A--N-T--K-W--SHS--I---TER-WT----DY--I-I-ST--S--TAG---R-----KL--C-NYPQ  139
HIV2D205   -G--N-T--R-F--SYA--L---ERN-YT----DV--Q-L-GS--S--SAN---R-----KI--Y-NYPS  139
SIVMM251   -G--H-T--R-W--TYA--I---SRN-WT----DY--I-L-ST--P--TAG---R-----QL--C-KFPR  139
SIVMM142   -G--N-T--R-W--TYA--I---SKD-WT----EY--I-L-ST--P--TAG---R-----RL--C-RFPR  139
SIVMM239   -G--H-T--K-W--TYA--I---SKN-WT----NY--I-L-ST--P--TAG---R-----QL--C-RFPR  139
SIVMNE     -G--N-T--R-W--THA--I---SRN-WT----DC--I-L-ST--P--TAG---R-----QL--C-RFPR  139
SIVSMMH4   -G--N-A--K-W--TYA--I---SRN-WT----DY--T-L-ST--P--SEG---K-----KL--C-KFPK  139
SIVSMMPBJ  -G--N-T--K-W--TYA--I---SRN-WT----DY--T-L-GT--P--SEG---R-----KL--C-KFPK  139

                   vpX cds ->
CONSENSUS  AHk?qVPSLQyLAL?VVq????rsQren?trKQwRrd?rrglR?Arqnsr??kqrqseppa?ga?FPGlA  185
HIV2POD    --RAQ-----F---V--QQND.RP-RDSTTR--R-RDYRRGL-L-KQDSRSHKQRSSESPTPRTY---V-  208
HIV2NIHZ   --RSQ-----F---V--QQND.RS-GNSATR--R-GDYRRGL-M-RQDSRGYKQRGSESPPTRAH---L-  208
HIV2ISY    --KSK-----F---V--QQND.KP-RDNTTR--W-RNYRRGL-L-RQDGRSHKQRGSEPPAQGAY---V-  208
HIV2ST     --KYQ-----I---V--QQNG.RP-RDNTTR--W-RHYRRGL-V-RQDGRSHKQRGSEPPAPRAY---V-  208
HIV2BEN    --KAQ-----Y---V--QQND.RP-RKGTAR--W-RDHWRGL-V-REDHRSLKQGGSEPSAPRAH---V-  208
HIV2D194   --KAQ-----Y---V--QQNG.RP-RKGAAR--W-RDHWRGL-V-RQDYRSLKQGGSEPSAPRAH---V-  208
HIV2GH1    --KVQ-----Y---V--QQND.RP-RKGTAR--W-RDHWRGL-V-RQDYRSLKQRGSEPSAPRAH---V-  208
HIV2D205   --EGQ-----F---R--QEGKNGS-GESATR--R-RNSRRSI-L-RKNNNRAQQGSGQPFAPRTY---L-  209
SIVMM251   --RYQ-----Y---K--.SDV.RS-GENPTW--W-RDNRRGL-M-KQNSRGDKQRGSKPPTKGAD---L-  207
SIVMM142   --KHQ-----Y---R--.SHV.RS-GENPTW--W-RDNRRSL-V-KQNSRGDKQRGGKPPTEGAN---L-  207
SIVMM239   --KYQ-----Y---K--.SDV.RS-GENPTW--W-RDNRRGL-M-KQNSRGDKQRGGKPPTKGAN---L-  207
SIVMNE     --KTQ-----Y---R--.SYV.RS-RENPTW--W-RDNRRSL-M-KQNSRGDKQRGSKPPTKGVD---L-  207
SIVSMMH4   --KNQ-----Y---T--.SHV.RS-GEDPTW--W-RNNRKGL-M-KQNSRRNKQGSSKSPAEGAN---L-  207
SIVSMMPBJ  --KNQ-----Y---T--.SHV.RS-GENPTW--W-RNNRRGL-L-RQNSRRNKQGSSESFAEGTN---L-  207
```

CONSENSUS	kVLgILA*	192
HIV2ROD	E--E----	215
HIV2NIHZ	E--E----	215
HIV2ISY	K--E----	215
HIV2ST	K--E----	215
HIV2BEN	K--E----	215
HIV2D194	K--G----	215
HIV2GH1	K--E----	215
HIV2D205	E--G----	216
SIVMM251	K--G----	214
SIVMM142	K--G----	214
SIVMM239	K--G----	214
SIVMNE	K--G----	214
SIVSMMH4	K--G----	214
SIVSMMPBJ	K--G----	214

HIV2/SIV VIF


Information dense stretches (0.8 or greater) are indicated by ****.
<- pol cds
vpX cds ->
CONSENSUS
HIV2SIVVIF

```
CONSENSUS  MnqeKeWvmRvtWkVpeelItkWqGIVrYwmr?rnLkWeYrmHyQitWawyTmSryvIPlgkhgeIhVd?   68
SIVAGMTYO  -NPN-E-VM-VT-K-PGDL-TK-Q---R-WMRQRN-K-N-YM-Y-IT-AWY-M-RYV--IGKHGE-C-DL   70
SIVAGM155  -SQE-H-VM-LT-K-QEEV-TK-Q---R-WMNKRN-K-E-KM-Y-IT-AWY-M-RYV--LPGSGE-H-DI   70
SIVAGM3    -NQE-E-VM-VT-K-PEEL-TK-Q---R-WMRTRK-D-K-RM-Y-IT-AWY-M-RYE--LGQHGS-H-DL   70
SIVAGM677  -ERE-Q-IV-VV-R-SERQ-SR-R---T-KIRNKQ-P-E-RH-W-VQ-QFW-Y-QFI--LSKDDY-E-NI   70

CONSENSUS  YwhLtPeqGWLStyavGiqYvsn???kYrTElDPaTADsmIH?hYF?CFt?RAIqqAlrG?rftfCqFpe  131
SIVAGMTYO  -WH-T-EQ----TYAV-IQ-VSNLESK-R--L--A---SI--GH--N--KE---QQ-LR-HRFVF-Q-PE  140
SIVAGM155  -WH-A-KQ----TYAV-IQ-VSLVNDK-R--L--N---SM--CH--T--TD--QQ-LR-NRFIF-Q-PG  140
SIVAGM3    -WH-T-EK----TYAE-IQ-LSNRDPW-R--L--A---SL--TH--T--TE--RK-LL-QRFTF-Q-PE  140
SIVAGM677  -HN-T-ER----SHGV-LS-YHQKG..-K--V--G---RM--LY--N--TD---QQ-IR-EKYTW-T-KE  138

CONSENSUS  GHK?tGQVpSLQyLALlAhqNGlRkRS?RgkTrrtrNlGs?QGAvGrMAkR?arrsq??sq?AfW?rtpv  193
SIVAGMTYO  ---ST---P---Y---L-HQ--L-E--K-GK-RRSR-L-SK---V-Q--K-YVTRSQPGGEA-F-ERTPV  210
SIVAGM155  ---LT---P---Y---L-HQ--L-K--Q-GE-RRTR-L-SQ---V-R--Q-YGRRNQQRSQT-F-PRTPI  210
SIVAGM3    ---KT---P---Y---L-HQ--L-Q--Q-SK-GGTR-M-FE---V-R--K-HARRYQSGSQD-F-ARAPV  210
SIVAGM677  ---..---Q---L---V-YT--I-K--K-TF-RMAG-L-SR---M-R--T-HAQGSKRRSQK-L-NEHAN  206

CONSENSUS  PSMELLsqG?r?t??sH?gKGl?       207
SIVAGMTYO  ------SG-RRKTWYS-DG--LQIL*    235
SIVAGM155  ------SG-RGETGKT-SG--I*       232
SIVAGM3    ------SG-GRKESHS-AR--L*       232
SIVAGM677  ------CR-GKET*                219
```

Information dense stretches (.7 or greater) are indicated by ****.

```
                                 ********
CONSENSUS  MnqeKeWvmRvtWkVpeelItkWqGIVrYwmr?rnLkWeYrmHyQitWawyTmSryvIPlgkhgeIhVd?     68
AGMVIF     M^X^K^W^^R^XW^VXX^XI^^W^GIV^YX^^^^^LXW^YXXH^Q^^WX^^T^S^^XIP^^X^XXIXV^^     70
            D D D II I  K   D  SK K   D  IDDKD   D    F ID  FF I KF   IA D     DI
            E E E VV V  R   E  TR R   E  VEERE   E    W VE  WW V RW   VG E     NV
            N N N  L L      N     E   N  LNN N   N    Y  N  YY L EY   LS N      L
            Q   Q  M M      Q     Q   Q  MQQ Q   Q       Q     M Q    MT Q      M
            K   K           K         K   KK K   K       K     F       P K
            R   R           R         R   RR R   R       R     W         R
            H   H           H         H   HH H   H       H     Y         H
            S   S           S         S   SS S   S       S     C         S
            T   T           T         T   TT T   T       T               T

              **********              ****************************
CONSENSUS  YwhLtPeqGWLStyavGiqYvsr???kYrTElDPaTADsmIH?hYF?CFt?RAIqqAlrG?rftfCqFpe    131
AGMVIF     Y^^L^P^^GWLS^^^XG^^Y^^XXX..Y^TE^DPXTAD^^IHX^YF^CF^^RAI^^A^XG^^^X^C^FXX     38
            FD A KK    SFA  ID ID      K  I      DI   F  D  DD   KK I  DKF F D
            WE G RR    TWG  VE VE      R  V      EV   W  E  EE   RR V  ERW W E
            YN S EE     Y   LN LN         L      NL   Y  N  N    EE L  N Y Y N
            HQ T QQ     H   MQ MQ         M      QM   H  Q  Q    QQ M  Q     Q
             K P             K FK               K       K  K          K     K
             R               R WR               R       R  R          R     R
             H               H YH               H       H  H          H     H
             S               S CS               S       S  S          S     S
             T               T  T               T       T  T          T     T

                 ************************           ********
CONSENSUS  GHK?tGQVpSLQyLALlAhqNGlRkRS?RgkTrrtrNlGs?QGAvGrMAkR?arrsq??sq?AfW?rtpv    193
AGMVIF     GHK..GQVXSLQ^LAL^A^^NG^R^RS^R^XTXX^XN^GX^QGA^G^MA^R^X^XX^XX^^XA^WX^X^X    206
                        I   I FD  I K  K A    A  L  K   I K  D F D  K  AE  I  K A
                        V   V WE  V R  R G    G  M  R   V R  E W E  R  GQ  V  R G
                        L   L YN  L E  E S    S     E   L E  N Y N  E  S   L  E S
                        M   M HQ  M Q  Q T    T     Q   M Q  Q H Q  Q  T   M  Q T
                        F      K         P    P              K   K     P   F    P
                        W      R                             R   R         W
                        Y      H                             H   H         Y
                        C      S                             S   S         C
                               T                             T   T

            ******
CONSENSUS  PSMELLsqG?r?t??sH?qKGl*        207
AGMVIF     PSMELLXXGXX^^*                 219
                       KD
                       RE
                       EN
                       QQ
                        K
                        R
                        H
                        S
                        T
```

```
                                                                  <- vif cds
CONSENSUS  M?dPREriPPGNSgEEtigeAFeWL?RTve?iNReAVNHlPRELIFQVWqRSW?YWhdeqGMs?SYtKYR   65
HIV2ROD    -TD---TV-----G--TIGE--A--N--VEAI--E----L---------Q---R--HDEQ--SE--T---   70
HIV2NIHZ   -TD---TV-----G--TIEE--A--D--VEAI--E----H---------Q---R--HDEQ--ST--T---   70
HIV2ISY    -TN---TI-----G--TIEE--D--D--VEAI--E----L---------Q---R--HDEQ--SR--T---   70
HIV2ST     -AG---TI-----G--TIGE--E--D--VEAI--E----L---------Q---R--HDEQ--SI--T---   70
HIV2BEN    -TD---RV-----G--TIGE--E--E--IEAL--E----L---------Q---R--HDEQ--SA--T---   70
HIV2D194   -AD---RV-----G--TIGE--E--D--IEAL--E----L---------Q---A--HDEQ--ST--T---   70
HIV2GH1    -TD---RV-----G--TIGE--E--D--IEAL--E----L---------Q---R--HDDQ--SP--T---   70
HIV2D205   -.D---RV-----D--TVGE--A--E--ITEL--V----L---------Q---A--REEQ--SI--T---   69
SIVMM251   -SD---RI-----G--TIGE--E--N--VEEI--E----L---------Q---E--HDEQ--SQ--V---   70
SIVMM142   -SD---RI-----G--TIGE--E--N--VEEI--E----L---------Q---E--HDEQ--SQ--T---   70
SIVMM239   -SD---RI-----G--TIGE--E--N--VEEI--E----L---------Q---E--HDEQ--SP--V---   70
SIVMNE     -SD---RI-----G--TIGE--E--N--VEEI--E----L---------Q---E--HDEQ--SQ--V---   70
SIVSMMH4   -SD---RI-----G--TIGK--E--N--VEEI--A----L---------R---E--HDEM--SE--T---   70
SIVSMMPBJ  -SD---RI-----G--TIGE--D--D--VEEI--A----L---------R---E--HDEM--SV--T---   70

CONSENSUS  YLclIQkA?fmH?kkgcrClgeghG?gGw?rpGPPPPPPPGLa*    104
HIV2ROD    --CII-K-VYM-VRKGCT-LGRGH-PG-W.RP----------V-    112
HIV2NIHZ   --CLI-M-MYM-AKRDGT-LGGGM-QK-.GDQ----------V-    112
HIV2ISY    --CLM-K-VFM-FKKGCT-RGEGH-PG-W.RS----------V-    112
HIV2ST     --CLM-K-MFI-SKRGCT-LGGGH-PG-W.RS----------V-    112
HIV2BEN    --CLM-K-IFT-FKRGCT-WGEDM-RE-LEDQ----------V-    113
HIV2D194   --CIM-K-VYI-FKKGCT-LGRGH-PG-W.RP----------V-    112
HIV2GH1    --CLM-K-VFI-FKRGCT-LGGGH-PG-W.RS----------V-    112
HIV2D205   --LLM-K-MFV-YTKGCR-LQEGH-PG-W.RS----------A-    111
SIVMM251   --CLM-K-LFM-CKKGCR-LGEGH-AG-W.RP----------A-    112
SIVMM142   --CLI-K-LFM-CKKGCR-LGEGH-AG-W.RP----------A-    112
SIVMM239   --CLI-K-LFM-CKKGCR-LGEGH-AG-W.RP----------A-    112
SIVMNE     --CLI-K-LFM-CKKGCR-LGEGH-AG-W.RP----------A-    112
SIVSMMH4   --CLI-K-LFV-CKKGCR-LGEEH-AG-W.RT----------A-    112
SIVSMMPBJ  --CLI-K-MFM-CKKGCR-LGGEH-AG-W.RP----------A-    112
```

HIV2/SIV VPX

Information dense stretches (0.7 or greater) are indicated by ****.

```
                                                                  <- vif cds
              **********                        ******************              ***
CONSENSUS  M?dPREriPPGNSgEEtigeAFeWL?RTve?iNReAVNHlPRELIFQVWqRSW?YWhdeqGMs?SYtKYR     65
HIV2SIVVPX M.XPRE^^PPGNSXEE^^X^AFXWL^RT^^X^NRXAVNHXPRELIFQVW^RSWXYW^^^XGM^XSYXKYR     69
                DI        AI K     D  ID I                 K      DDD   A
                EV        GV R     E  VE V                 R      EEE   G
                N         S  E     N   N L                 E      N     S
                Q         T  Q     Q   Q M                 Q      Q     T
                K         P        K   K                           K     P
                R                  R   R                           R
                H                  H   H                           H
                S                  S   S                           S
                T                  T   T                           T

           **                              **********
CONSENSUS  YLcliQkA?fmH?kkgcrClgeghG?gGw?rpGPPPPPPPGLa*    104
HIV2SIVVPX YL^^^QXA^^XHX^^XX^CXXXXXGXXG..^XGPPPPPPPGLX*    110
             III   IF   DK  D            D
             VVV   VW   ER  E            E
             LLL   LY   N   N            N
             MMM   M    Q   Q            Q
             F          K   K            K
             W          R   R            R
             Y          H   H            H
             C          S   S            S
                        T   T            T
```

```
CONSENSUS  MASGRDPREarPGelEIWDLsREPWDeWLrDMled?Nq..A?lHFGrelLfqVWNYCqEEGeRhgtPm?Er     67
SIVAGMTYO  ---------AR--EV-----S-----K--R--LQDL-Q--RL---REL-FQ-----Q---E-HGT-MM-R     70
SIVAGM155  ---------ER--GL-----S-----E--R--VEEI-N--KL---REL-YQ-----Q---E-QGR-IA-R     70
SIVAGM3    ---------AR--EL-----S-----E--R--LEDI-Q--KM---REL-FQ-----Q---E-NRT-ML-R     70
SIVAGM677  ---------PL--WL-----D-----E--Q--LRDL-E--RR---MNM-IR-----V---R-HNT-WN-I     70

CONSENSUS  aYKYYrlVQKalFVHFRCGCRRRqPFePYEERR?GQGGGragrvPPGLd*    115
SIVAGMTYO  A----RL---AL----------Q--E------D-----RANRV----E-    119
SIVAGM155  A----RL---AL----------Q--E------N-----RPGRV----D-    119
SIVAGM3    A----KL---AL----------Q--E------D-----RAGRV----D-    119
SIVAGM677  G----RI---SM----------G--S------N-----.APPP----A-    118
```

Information dense stretches (0.9 or greater) are indicated by "*".

```
          *********      ***** *****                          *****
CONSENSUS MASGRDPREarPGelEIWDLsREPWDeWLrDMled?NqEA?lHFGrelLfqVWNYCqEEGeRhgtPm?Er      67
AGMVPX    MASGRDPREXXPGX^EIWDL^REPWD^WL^DM^^^^N^EA^XHFGX^^L^^VWNYCXEEG^R^X^F^XEX      70
                      I          K  K  IKDI D  K      DL IK          K D D I
                      V     E    R  R  VREV E  R      EM VR          R E E V
                      L     N    E  E  LE L N         N  LE          E N N L
                      M     Q    Q  Q  MQ M Q             MQ          Q Q Q M
                            K                K                        K K F
                            R                R                        R R W
                            H                H                        H H Y
                            S                S                        S S C
                            T                T                        T T

                     ***********    ************
CONSENSUS aYKYYrlVQKalFVHFRCGCRRRqPFePYEERR?GQGGGragrvPPGLd*     115
AGMVPX    ^YKYY^^VQK^^FVHFRCGCRRRXPF^PYEERR^GQGGG.^XXXPPGLX*     118
          A     KI   AL             D       D       A
          G     RV   GM             E       N       G
                 L   S              N               S
                 M   T              Q               T
                     P              K               P
                                    R
                                    H
                                    S
                                    T
```

```
                          <- vif cds
CONSENSUS1 MEQaPEDQGPQRePyNeWtLELLeeLKsEAVRHFPRiWLh?LgQhIYETYGDTWaGVEAiIRiLQQ.LFI   69
HIVLAI     ---A--------E-H-E-T----EE--N--------I--HG-G-H---------A----I--I-------   70
HIVHXB2R   ---A--------E-H-E-T----EE--N--------I--HG-G-H---------A----I--I-------   70
HIVMN      ---A--------E-Y-Q-A----EE--N--------I--HG-G-H---------A----I--I-------   70
HIVSC      ---A--------E-Y-E-T----RK--S--------I--HG-G-H---------A----I--I-------   70
HIVJRCSF   ---A--------E-Y-E-T----EE--N--------I--HS-G-Y---------A----I--I-------   70
HIVOYI     ---A--------E-Y-E-T----EE--S--------I--HS-E-Y---------E----I--I-------   70
HIVSF2     ---A--------E-Y-E-T----EE--R--------P--HS-G-Y---------A----I--I-------   70
HIVNY5CG   ---A--------E-Y-E-T----EE--S--------I--HN-G-H---------A----I--I-------   70
HIVNL43    ---A--------E-Y-E-T----EE--S--------I--HN-G-H---------A----I--I-------   70
HIVHAN     ---A--------E-Y-A-T----EE--S--------I--HS-G-Y---------A----I--I-------   70
HIVRF      ---A--------E-Y-E-T----EE--S--------L--HS-G-H---------A----I--I-------   70
CONSENSUS2 MEQAPeDQGPQREPynEwtLELLEELKsEaVRHFPRiWLHsLGQhIYeTYGDTW?GVEAiIRiLQQLLFI   69
HIVELI     -----A--------YN-WA--------S-A------I---S---H--E------V----I--I-------   70
HIVZ2Z6    -----E--------YN-WT--------S-A------I---S---Y--E------A----L--I-------   70
HIVNDK     -----E--------YN-WT--------S-A------I---S---H--E------A----I--I-------   70
HIVMAL     -----A--------HN-WT--------Q-A------I---S---H--E------E----I--S-------   70
HIVU455    -----E--------YA-XA--------H-D------Q---G---H--N------E----I--I-------   70
SIVCPZ     -----E----P---YQ-WA--T-----N-A------P---Q---F--D------V----I--I--H----   70

                       tat cds ->
CONSENSUS1 HFrigcrhSRIGitrQRRa?RNGaSRS*     95
HIVLAI     --RIGCRH----VTQ---A.---A----     96
HIVHXB2R   --QNWVST                         78
HIVMN      --RIGCRH----IIR---A.---A----     96
HIVSC      --RIGCRH----IIQ---S.---S----     96
HIVJRCSF   --RIGCRH----ITR---A.---A----     96
HIVOYI     --RIGCQH----ITR---A.---A----     96
HIVSF2     --RIGCQH----IIQ---AR---A----     97
HIVNY5CG   --RIGCQH----IIR---A.---A----     96
HIVNL43    --RIGCRH----VTR---A.---A----     96
HIVHAN     --RIGCRH----ITP---A.---A----     96
HIVRF      --RIGCQH----ITR---A.---A----     96
CONSENSUS2 HFRIGCQHSRIqItrqRRa.RNGsSRS*     95
HIVELI     ----------G-IRQ--A.---S----     96
HIVZ2Z6    ----------G-TRQ--A.---S----     96
HIVNDK     ----------S-TRQ--A.---S----     96
HIVMAL     ----------G-TRQ--A.---S----     96
HIVU455    ----------G-IPG--G.---A----     96
SIVCPZ     ---L-------G-LPQ--R.S--SN---     96
```

HIV1 VPR

Information dense stretches (0.8 or greater) are indicated by ****.

```
                              <- vif cds
                  ********                ******         *************************
CONSENSUS1 MEQaPEDQGPQRePyNeWtLELLeeLKsEAVRHFPRiWLh?LgQhIYETYGDTWaGVEAiIRiLQQLLFI     69
CONSENSUS2 ---A-e------E--nEw-----EE----a---------Hs-G----e------?---------------     69
HIV1VPR    MEQ^PXDQGPQR^P^XXX^LELL^^LK^EXVRHFPRXWLXXLXQ^IY^TYGDTWXGVEA^IRXLQQLLFI     70
              A        E F   A    KK  D                F  D           I
              G        Q W   G    RR  E                W  E           V
              S          Y   S    EE  N                Y  N           L
              T          H   T    QQ  Q                H              M
              P              P        K
                                      R
                                      H
                                      S
                                      T

                       tat cds ->
           ***********
CONSENSUS1 HFrigcrhSRIGitrQRRa?RNGaSRS*     95
CONSENSUS2 --RIGCQH---gI--q---.---s---*     95
HIV1VPR    HFRIGC^HSRI^^XXXRR^.RNG^SRS*     96
                 K    AI     A    A
                 R    GV     G    G
                 E    S      S    S
                 Q    T      T    T
                      P      P    P
HIV1VPR:6       DQGPQRoP
        See Tristem et al., Nature 347:341-1, 1990.

HIV1VPR:30      VRHFPR
        This fragment is similar to a conserved pattern in HIV2, "EALkHFDP".
        See Tristem et al., Nature 347: 341-2, 1990.

HIV1VPR:45      IYmTYGDTWXGVEAcIRXLQQLLFIHFRIGCpHSRI
        This fragment contains a region similar to a conserved pattern in HIV2,
        "LQRAL".

HIV1VPR:90      RNGjSRS
```

```
        <- vpX cds                                              tat cds ->
CONSENSUS  Mtea??e?PPede?PqRepwdeWv?e?L?E1keEALkHFDPrLLtaLGnyIy?RHGDTleGAgeLIrILQ    62
HIV2ROD    -AEAPTEL--VDGT-L-EPGDE-IIEI-R-IKE---K----R--IA--KY-YT-----LE--RE--KV--    70
HIV2NIH2   -TEAPTEL--EDRT-P-EPGDA-VIEI-R-IEE---R----R--TA--RY-YT-----LE--RE--RI--    70
HIV2ISY    -TEAPAEF--EDGT-P-EPGDE-VIEI-R-IKE---K----R--TA--YY-YT-----LE--RE--RV--    70
HIV2ST     -TEAPTES--EDRT-P-EPGDE-VIET-R-IKX---K----R--IT--NY-YA-----LE--RG--RI--    70
HIV2BEN    -TEAPTEF--EDGT-R-DLGSD-VIET-R-IKE---R----R--IA--YY-HN-----LE--RE--KT--    70
HIV2D194   -TEAPTEF--EDGT-R-ELGST-VIET-K-IKE---K----C--IA--NY-YN-----LE--RE--RV--    70
HIV2GH1    -TEAPTEF--EDGT-R-ELGGD-VIRI-G-IKE---K----R--IA--NY-HS-----PE--RE--RI--    70
HIV2D205   -AEAAPEI--ENEN-Q-EPWEE-IGEI-E-IKQ---K----R--TA--NF-YS-----LA--GE--KI--    70
SIVMM251   -.E...ER--ENEG-Q-EPWDE-VVEV-E-LKE---K----R--TA--NH-YN-----LE--GE--RI--    66
SIVMM142   -.E...ER--ENEG-Q-EPWDE-VVEV-K-LKE---K----R--TA--NH-YN-----LE--GE--RI--    66
SIVMM239   -.E...ER--ENEG-Q-EPWDE-VVEV-E-LKE---K----R--TA--NH-YN-----LE--GE--RI--    66
SIVMNE     -.E...ER--EDEG-Q-EPWDE-VVEV-E-LKE---K----R--TA--NH-YN-----LE--GE--KI--    66
SIVSMMH4   -A....ER--EDEA-Q-EPWDE-VVEV-E-LKE---K----R--TA--NY-YD-----LE--GE--RI--    66
SIVSMMPBJ  -T....ER--EDEA-Q-EPWDE-VVEV-E-IKE---N----R--TA--NY-YD-----LE--GE--RI--    66

CONSENSUS  RALf?HfRaGC?hSRigq??ggnpls?ipp?r?m?*     89
HIV2ROD    ---FT-F-A--GH--IGQTRGGNPLSAIPTPRNMQ-    105
HIV2NIH2   ---FA-F-A--GH--IGQTRGGNPLSAIPTPRGMH-    105
HIV2ISY    ---FT-F-A--GH--IGQPRGRNPLSAIPTPRNMQ-    105
HIV2ST     ---LL-F-A--GR--IGQPRGRNPLSAIPTPRGMR-    105
HIV2BEN    ---FV-F-A--NR--IG*                       87
HIV2D194   ---FV-I-A--DR--KGQTRRRAPCPAAPTPRGMH-    105
HIV2GH1    ---FV-L-A--NR--ISQTRRRTPFPAAPTPRGMY-    105
HIV2D205   ---FL-F-A--QH--IGQSGGGNPLSTIPPP*        101
SIVMM251   ---FM-F-G--NH--IGQPGGGNPLSTIPPS*         97
SIVMM142   ---FI-F-S--SH--IGQPGGGNPLSTIPPSRSML-    101
SIVMM239   ---FM-F-G--IH--IGQPGGGNPLSAIPPSRSML-    101
SIVMNE     ---FM-F-G--TH--IGQSGGGNPLSTIPPSRRML-    101
SIVSMMH4   ---FI-F-S--AH--IGQSRGGNPLSTIPPSRGVL-    101
SIVSMMPBJ  ---FI-F-G--RH--IGQSGGGNPLSTIPPSRGVL-    101
```

HIV2/SIV VPR

Information dense stretches (0.7 or greater) are indicated by ****.

```
        <- vpX cds                                        tat cds ->
                                                ********          *****          **
CONSENSUS  Mtea??e?PPede?PqRepwdeWv?e?L?EikeEALkHFDPrLLtaLGnyIy?RHGDTleGAgeLIriLQ     62
HIV2SIVVPR M.....^XPPX^XXPXR^XXXXW^X^XLXE^^XEAL^HFDPXLLX^LGX^I^XRHGDTXXGAXXLI^XLQ     65
                D    D     D     I K    IK    D        A   F F              K
                E    N     E     V R    VR    E        G   W W              R
                N                  E    LE    N        S   Y Y
                Q                  Q    MQ    Q        T   H H
                K                            K        P
                R                            R
                H                            H
                S                            S
                T                            T

           ***
CONSENSUS  RALf?HfRaGC?hSRigq??ggnpls?ipp?r?m?*     89
HIV2SIVVPR RAL^XH^R^GCX^SRX^*                       82
              I  I A   D   A
              V  V G   E   G
              L  L S   N   S
              M  M T   Q   T
              F  F P   K   P
              W  W     R
              Y  Y     H
              C  C     S
                       T
```

```
                                                    \/3'sj \/3'sj
                                                  rev cds ->
CONSENSUS1 MePVDPrLEPWKHPGSQPktactnCYCKkCCfHCqvCFttKgLgISYGRKKRRQ.RrrapqdSqtHQvSL     69
HIVLAI     -E----R-----------KTACTT----K--F--QV--TT-A-G----------.-RRPPQG-QT--V--     69
HIVHXB2R   -E----R-----------KTACTN----K--F--QV--IT-A-G----------.-RRAHQN-QT--A--     69
HIVNL43    -E----R-----------KTACTN----K--F--QV--MT-A-G----------.-RRAHQN-QT--A--     69
HIVMN      -E----R-----------KTACTT----K--F--QV--TK-A-G----------.-RRAPED-QT--V--     69
HIVSC      -D----R-----------KAACTS----K--F--QV--TT-G-G----------.-RRAPQD-QT--V--     69
HIVBAL1    -E----K-----------KTACTT----K--F--QV--IT-G-G----------.-RRAPQD-ET--V--     69
HIVJRCSF   -E----3-----------KTACTN----K--L--QV--TT-G-G----------.-RRPPQD-QT--V--     69
HIVOYI     -E----R-----------KTASNN----R--L--QV--TK-G-G----------.-RRAPQD-KT--V--     69
HIVSF2     -E----N-----------RTACNN----K--F--YA--TR-G-G----------.-RRAPQD-QT--A--     69
HIVNY5CG   -E----R-----------KTACTN----R--F--QV--TT-A-G----------.-RRAPQD-ET--V--     69
HIVSF162   -E----R-----------KTACTN----K--F--QV--IT-G-G----------.-RRAPPD-EV--V--     69
HIVCDC4    -E----R-----------KTACTN----K--F--QV--TK-A-G----------.-RRAHQD-QN--A--     69
HIVSF33    -E----N-----------RTACTN----K--F--QV--IT-G-G----------.-QRAPDS-QN--D--     69
HIVHAN     -E----R-----------RTPCTN----K--L--QV--TR-G-S----------.-RRAPQD-QT--V--     69
HIVRF      -E----R-----------KTACNN----K--Y--QV--LT-G-G----------.-RGPPQG-QT--V--     69
CONSENSUS2 MdPVDPNlEpWnHPGSQPrTaCnkChCKkCCyHCqvCFitKGLGISYGrKKRrq.RR?PpQG?qaHQ?pi     66
HIVELI     -D-----L-P-N------R-P-NK-H--K--Y--PV--LN--------R---RQ.--G-P--GQA--VPI     69
HIVZ2Z6    -D-----I-P-N------K-A-NR-H--K--Y--QV--IT--------R---RQ.--R-S--GQT--DPI     69
HIVNDK     -D-----L-S-N------R-A-NK-H--K--Y--QV--IT--------R---RQ.--K-P--DQA--VPI     69
HIVMAL     -D-----L-P-N------R-P-NK-Y--K--Y--QM--IT--------R---RQ.--R-P--NQA--DPL     69
HIVU455    -E-----L-P-K------T-A-SN-Y--V--W--QL--LK--------K---KP.--G-P--SKD--TLI     69
SIVCPZ     -D-I--DL-P-K------R-V-NN-Y--A--Y--IY--TK--------R---TTR--TAPA-SKNN-DSI     70

         exon\/exon
CONSENSUS1 sKQP?sqprGD.PtqPkesKkkVErEtEtdP?d?      99
HIVLAI     S---TSQPR--.-TG-KE*    86
HIVHXB2R   S---TSQPR--.-TG-KE*                     86
HIVNL43    S---TSQSR--.-TG-KE*                     86
HIVMN      P---APQFR--.-TG-KES-KK--R-T-TH-VD*     101
HIVSC      P---ASQAR--.-TG-KES-KK--R-T-TD-VD*     101
HIVBAL1    S---TSHPR--.-TG-KES-KK--R-T-TD-VH*     101
HIVJRCSF   P---SSQQR--.-TG-KES-KK--R-T-TD-DN*     101
HIVOYI     S---ASQPR--.-TG-KES-KK--R-T-TD-ED*     101
HIVSF2     S---ASQSR--.-TG-TES-KK--R-T-TD-FD*     101
HIVNY5CG   S---ASQPR--.-TG-KES-KK--R-T-TD-DH*     101
HIVSF162   P---ASQPQ--.-TG-KES-KK--R-T-TD-VH*     101
HIVCDC4    S---SSQTR--.-TG-KEP-KE--R-A-TD-LD*     101
HIVSF33    S---SSQPR--.-TG-KES-KE--R-T-TD-LD*     101
HIVHAN     P---SSQQR--.-DS-KKS-KK--R-T-AD-FD*     101
HIVRF      S---TSQPR--.-TG-KES-EK--R-T-TD-AVQ*    102
CONSENSUS2 PkQPssQ?rgd?ptGpke?KKKVES?A?TD??A*      92
HIVELI     -K--SS-PRGD.PT-PKEQ------E-E--P*        99
HIVZ2Z6    -K--SS-PRGD.PT-PKE*                     86
HIVNDK     -E--SS-SRGD.PT-PKK*                     86
HIVMAL     -E--SS-HRGDHPT-PKE*                     87
HIVU455    -K--LP-SQRV.SA-QEES------K-K--RF-*     101
SIVCPZ     -K--LS-SRGN.KE-SEKST-E-A-KTEA-Q*       100
```

HIV1 TAT

Information dense stretches (0.8 or greater) are indicated by ****.

```
                                                  \/3'sj \/3'sj
                                                rev cds ->
          ******      ******                     ************
CONSENSUS1 MePVDPrLEPWKHPGSQPktactnCYCKkCCfHCqvCFttKgLgISYGRKKRRQ.RrrapqdSqtHQvSL     69
CONSENSUS2 -d----Nl-p-n------rT-Cnk-h-----y------i--G-G----r---rq.-R?P-QG?-a--?pi     66
HIV1TAT    M^PVDP^^E^W^HPGSQP^^^X^^C^CKXCC^HCXXCFX^K^L^ISYG^KKR^X.R^X^XXXX^XHQXX^     69
            D    DI A D      DAA DD F     I       D A A    K   K   K A    K     I
            E    EV G E      EGG EE W     V       E G G    R   R   R G    R     V
                 NL S N      NSS NN Y     L       N   S            E S    E     L
                 QM T Q      QTT QQ H     M       Q   T            Q T    Q     M
                 K  P K      KPP KK       F       K   P              P
                 R    R      R   RR       W       R
                 H    H      H   HH       Y       H
                 S    S      S   SS       C       S
                 T    T      T   TT               T

        exon\/exon
CONSENSUS1 sKQP?sqprGD.PtgPkesKkkVErEtEtdP?d?*     99
CONSENSUS2 Pk--s-Q?-gd?p-Gp--?-KK--S?A?TD?-A*      92
HIV1TAT    ^^QPX^^X^XX.^X^X^^XK^^VEX^^^^^X*        99
           AK   AD K   A A DK  KK   KAKAD
           GR   GE R   G G ER  RR   RGRGE
           SE   SN E   S S NE  EE   ESESN
           TQ   TQ Q   T T QQ  QQ   QTQTQ
           P    PK     P P K         P PK
                 R         R            R
                 H         H            H
                 S         S            S
                 T         T            T
HIV1TAT:0       MkPVDP
```

Acidic residues in the "k" and "D" (second and fifth) positions are required for full Tat activity. Rappaport et al., The New Biologist 1:101-110 (1989).

MAb -- Brake et al., J. Virol. 64:962-5, 1990.

```
HIV1TAT:12      HPGSQP

HIV1TAT:40      KhLjISYGnKKR
```

The "GRKKR" has been implicated in nuclear localization. Ruben et al., J. Virol. 63:1-8 (1989) and Hauber et al., J. Virol. 63:1181-87 (1989). Part of region III as defined by Green and Lowenstein, Cell:55, 1179-88, 1988.

MAb -- Brake et al., J. Virol. 64:962-5, 1990.

This fragment contains a region similar to a conserved pattern in HIV2, "KGLGI".

```
                                                        <- vpR cds
CONSENSUS  MEtpLkeqEsSLessnepsSctsE?d??tqelanqgeEiLsQLyrPLeaC?N?CyCKkCcyhCQ?CF1kK    64
HIV2ROD    --TP-KAP-S--KSCNEPF-RTS-QDVATQELARQGE-I-S--YR--ET-N-S-Y--R-CYH--M--LN-    70
HIV2NIHZ   --TP-KAP-S--ESCNEPS-RTS-QDVATQELARQGE-I-S--YR--EA-T-S-Y--K-CYD--L--LQ-    70
HIV2ISY    --TP-KAP-S--GSYNEPS-RTS-QDVATQELANQGE-I-S--YR--ET-N-K-F--G-CFH--L--LN-    70
HIV2ST     --TP-KAP-G--GSYNEPS-CTS-QDAAAQGLVSPGD-I-Y--YQ--EA-D-K-Y--K-CYH--M--LN-    70
HIV2BEN    --TP-KAP-S--KPYNEPS-CTS-RDVTAQELAKQGE-L-A--HR--EP-T-K-Y--R-SFH--L--SK-    70
HIV2D194   --TP-KEP-S--ESYNEPS-CTS-RDVTAQERAKQGE-L-A--HR--EA-T-S-Y--Q-SYH--L--LK-    70
HIV2GH1    --TH-KAP-S--ESYNEPS-CTS-QGVTAQELAKQGE-L-S--HR--EA-T-S-Y--Q-SFH--L--LK-    70
HIV2D205   --IP-QEQ-S--KSSNEPS-STS-PVVNTQGLDNQGE-I-S--YR--KA-D-T-Y--K-CYH--L--LK-    70
SIVMM251   --TP-REQ-N--ESSNERS-CIL-ADATTPESANLGE-I-S--YR--EA-Y-T-Y--K-CYH--F--LK-    70
SIVMMP11   --TP-REQ-N--ESSNERS-CIL-ADATTPESANLGE-I-S--YR--EA-Y-T-Y--K-CYH--F--LK-    70
SIVMM142   --TP-REQ-N--ESSNERS-YIS-AAAAIPESANLGE-I-S--YR--EA-Y-T-Y--K-CYH--F--LK-    70
SIVMM239   --TP-REQ-N--ESSNERS-CIS-ADASTPESANLGE-I-S--YR--EA-Y-T-Y--K-CYH--F--LK-    70
SIVMNE     --TP-REQ-N--KSSNGRS-CTS-AAAPTLESANLEE-I-S--YR--EE-Y-T-Y--K-CYH--F--LK-    70
SIVSMMH4   --TP-KEQ-N--ESCREHS-SIS-VDVPTPESANLEE-I-S--YR--EA-Y-K-Y--R-CYH--H--LK-    70
SIVSMMPBJ  --TP-KEQ-S--ESSREHS-SIS-VDADTPESASLEE-I-S--YR--EA-Y-K-Y--R-CYH--H--LK-    70

        rev cds ->                  exon\/exon
CONSENSUS  GLGIcYer??rRRRtpkktKa??ssasd?ksistRtr?sQp?K?qKktve??v?t??glGr*    112
HIV2ROD    ----C-ERKGR---TPKKT-THPSPTPD.KSIST-TGDS-PT-KQ-KTVEATVETDTGP-R-    130
HIV2NIHZ   ----W-DRKGR---TPKKT-AHPSSASD.KSIST-TRNS-PE-KQ-KTLEATVETDLGL-R-    130
HIV2ISY    ----C-DRKGR---SPKKT-AHSSPASD.KSIST-TGNS-TE KQ-KTLGTTLEADPGL-R-    130
HIV2ST     ----W-ERKGR---TPKKT-AHSSSASD.KSIST-TGNS-PE-KQ-KTLETALETIGGP-R-    130
HIV2BEN    ----S-ERKGR---TPRKT-TPSPSAPD.KSIST-TGDS-PT-EQ-KTSEATVVTTCGL-Q-    130
HIV2D194   ----W-ARQGR---TPRKT-THPPPASD.KSIST-TGDS-PT-KQ-KTPETTVVSACGL-H-    130
HIV2GH1    ----W-ARKSR---TPRKT-THSSSASD.KSIST-TGDS-PT-EQ-KTTETTMVTTCSL-R-    130
HIV2D205   ----C-DRS..---SAKRA-TTAPSAP* (exon not sequenced)                   95
SIVMM251   ----C-EQSRK---TPKKA-ANTSSASN.KLIPN-TRHC-PE-AK-ETVEKAVATAPGL-R-    130
SIVMMP11   ----C-EQSRK---TPKKA-ANTSSASN.KLIPN-TRHC-PE-AK-ETVEKAVATAPGL-R-    130
SIVMM142   ----S-EKSHR---TPKKA-ANTSSASN.EPIPN-IRLC-PK-AK-ETVEAAVATAPGL-R-    130
SIVMM239   ----C-EQSRK---TPKKA-ANTSSASN.KPISN-TRHC-PE-AK-ETVEKAVATAPGL-R-    130
SIVMNE     ----C-EQSRR---TPKKA-ANTSSASN.KPISD-TRHC-PK-EQ-ETVEAAVATAPGL-R-    130
SIVSMMH4   ----C-EQ.H.---TPKKT-ANPFPASN.KSLST-ARNR-PK-EK-ETVETEVAADLGL-R-    128
SIVSMMPBJ  ----C-EQQ..---TPKKT-ANTSSASD.KSLSR-ARNC-PK-EK-ETVEAEVATDLGL-R-    128
```

```
   Information dense stretches (0.7 or greater) are indicated by ****.

                                                                  <- vpR cds
                                                                                  *
CONSENSUS  MEtpLkeqEsSLessnepsSctsE?d??tqelanqgeEiLsQLyrPLeaC?N?CyCKkCcyhCQ?CFlkK      64
HIV2SIVTAT MEXXL^XXEXSLX^X^XXXSXXXEXXXXXXXXX^XX^E^LXQL^^PL^XCXN^C^CKXCX^^CQXCFX^K      70
                K       A D                 D  D I    FK  K    D F     FD      D
                R       G E                 E  E V    WR  R    E W     WE      E
                E       S N                 N    L    YE  E    N Y     YN      N
                Q       T Q                 Q    M    HQ  Q    Q        Q      Q
                        P K                 K                  K        K      K
                          R                 R                  R        R      R
                          H                 H                  H        H      H
                          S                 S                  S        S      S
                          T                 T                  T        T      T

        rev cds ->                 exon\/exon
           ****
CONSENSUS  GLGIcYer??rRRRtpkktKa??ssasd?ksistRtr?sQp?K?qKktve??v?t??glGr*    112
HIV2SIVTAT GLGIXYX^...RRRXX^^^K^X^X^^^^.^X^^^RXXXXQ^^KX^K^^XXXX^X^XX^XG^*    127
                 K        KKA A A AAAD K IAD      AD  K KD    I A  A  D
                 R        RRG G G GGGN R VGE      GE  R RE    V G  G  E
                 E          S S S SSS  E LSN      SN  E EN    L S  S  N
                 Q          T T T TTT  Q MTQ      TQ  Q QQ    M T  T  Q
                            P P P PPP     PK      PK     K      P  P  K
                                          R       R     R            R
                                          H       H     H            H
                                          S       S     S            S
                                          T       T     T            T
```

```
CONSENSUS  MDKgEeeq?vshQDLieqyrkPLqtCtNKC?CKkCCYHCqlCFLqKGLGvtYHapRtRRKkiasl?r?P?     65
SIVAGMTYO  ---G-AEQIVSH---SEDYQK--QT-K---F--K-----QL---Q----VT--AP-T---KIRSLNLA-L     70
SIVAGM155  ---G-EDQDVSH---IKQYRK--ET-T---F--K-----QF---R----IT--AF-T---KIASADRI-V     70
SIVAGM3    ---G-DEQGAYH---IEQLKA--KR-T---Y--C-----QL---Q----VT--AP-I---KIAPLDRF-E     70
SIVAGM677  ---.-EEPHPLL---.H..R.--QP-T---Y--K-----EL---Q----VR--VS-K---TSTQDNQD-I     65

CONSENSUS  qqQSIStr?gRd?QTTqe?qekVEt?A?an????RKNL????????GA??*      96
SIVAGMTYO  QH----TKWG-DG---PTSQEK--TT-GSN*                         100
SIVAGM155  PQ----IR.G-DS---QESQKK--EQ-KANLRIS----GDETRGPV--GN-     119
SIVAGM3    QK----TR.G-DS---QKGQEK--TS-RTAPSLG----AQQSGRAT--SD-     119
SIVAGM677  RQ----T.VQ-NG---EEGKTE--KA-AAN*                          94
```

Information dense stretches (.7 or greater) are indicated by ****.

```
                                          **************************
CONSENSUS  MDKgEeeq?vshQDLieqyrkPLqtCtNKC?CKkCCYHCqlCFLqKGLGvtYHapRtRRKkiasl?r?P?      65
AGMTAT     MDK.EX^XXXXXQDL.^..^.PL^XC^NKC^CKXCCYHC^^CFL^KGLG^^YHXXRXRRK^XXXX^XXPX      65
                D         D  K   K  D   F        EI   K    ID         D    D
                E         E  R   R  E   W        QV   R    VE         E    N
                          N  E   E  N   Y         L   E     N         N
                          Q  Q   Q  Q             M   Q     Q         Q
                          K         K             F         K         K
                          R         R             W         R         R
                          H         H             Y         H         H
                          S         S             C         S         S
                          T         T                       T         T

CONSENSUS  qqQSIStr?gRd?QTTqe?qekVEt?A?an????RKNL????????GA??*      96
AGMTAT     X^QSISX..XR^^QTTX^^^^^VE^XAX^X*                        93
            D         DA    DAKDK  D   A
            E         NG    EGRER  E   G
            N          S    NSENE  N   S
            Q          T    QTQQQ  Q   T
            K          P    KP K   K   P
            R               R  R   R
            H               H  H   H
            S               S  S   S
            T               T  T   T
```

```
            \/ 3' sj              exon\/exon
CONSENSUS1 MAGrsgdsde.ellktVRlIKfLYQSNPpPspeG.TrqARRNRRRRWReRQRqIrsiSewilstyLGRsa    68
HIVLAI     ---RSGDSDE.DLLKA--L--F------P-NPE-.-RQ----------E---Q-HSI-ERILSTY---SA    68
HIVHXB2R   ---RSGDSDE.ELIRT--L--L------P-NPE-.-RQ----------E---Q-HSI-ERILGTY---SA    68
HIVNL43    ---RSGDSDE.ELIRT--L--L------P-NPE-.-RQ----------E---Q-HSI-ERILSTY---SA    68
HIVMN      ---RSGDSDE.ELLKT--L--F------P-SSE-.-RQ----------E---H-RSI-AWILSNY---PA    68
HIVBRVA              ELLKT--V--F------P-RPE-.-RQ----------A---Q-QSL-GWILSTF---SA    58
HIVSC      ---RSGDSEE.ELLKT--L--F------P-KPE-.-RQ----------E---Q-RSI-GWILSNY---LA    68
HIVJH3        KKRRQRR.RAAQA--L--F------P-SPE-.-RQ----------Q---Q-RSI-GWIISNY---PA    65
HIVBAL1    ---RSGDSDE.ELLKT--L--F------P-ILE-.-RQ----------E---Q-RSI-ERILGAY---SA    68
HIVJRCSF   ---RSGDSDE.DLLKT--L--F------P-SNE-.-RQ----------E---Q-RTI-ERILSTY---PA    68
HIVOYI     ---RSGDSDE.ELLKT--L--F------P-NPE-.-RQ----------E---Q-RKI-GWILSTY---SA    68
HIVSF2     ---RSGDSDE.ELLRT--L--L------P-SPE-.-RQ----------E---Q-RSI-GWILSTY---SA    68
HIVNY5CG   ---RSGDSDE.ELLKT--L--F------P-SPE-.-RQ----------E---Q-RTI-EWILSTH---SA    68
HIVSF162   ---RSGDSDE.ELLQT--F--F------L-SPK-.-RQ----------E---Q-QSI-AWIISTH---ST    68
HIVCDC4    ---RSGDSDE.ELIKT--I--H------P-KPE-.-RQ----------E---R-HSI-AWLLSTC---SA    68
HIVSF33    ---RSGGSDR.ELLTA--I--I------H-SPE-.-RQ----------E---Q-HSI-ERILSTV---SS    68
HIVHAN     ---RSGDSDE.ELLKT--L--F------P-SNE-.-PT----------E---Q-RSI-ERILSTF---PA    68
HIVRF      ---RRGDSDE.DLLKA--L--S------P-SPE-.-RQ----------E---Q-RRC-EWILDTY---SV    68
CONSENSUS2 MAgRSGdsDE.dLLkAvRlIK?LYQSNPpPspeG?tRQARrNRRRRWRARQRQI?sigERIlstyLGRpe    66
HIVELI     --G---DS--.D--K-V-L--F------P-SPE-.T----R------------REIA---LGTY---PA    68
HIVZ2Z6    --G---DR--.D--K-V-L--I------P-SPE-.T----R------------HSIG---LSTY---SE    68
HIVNDK     --G---DS--.N--K-I-L--F------P-SPE-.T----R------------HSIG---ICTF---PE    68
HIVMAL     --G---DS--.D--R-I-L--I------P-NTE-TT----R------------NSIG---LSTY---PE    69
HIVU455    --R---NP--.D--K-V-I--L------C-NPR-.S----K------------DSLS---LSDC---PA    68
SIVCPZ     --G--EPQ-DAR--Q-VKI--I------Y-SPE-.T-K--R----------K--SEISG-VLATY---PP    69

CONSENSUS1 ePVpLQLPPleRLtLDcseDCGtSGtq......?GVGsPqiLvESP?vLesGtKE*    115
HIVLAI     E--P-----LE--T--CNE---T--TQ.......---S-QI-V---TV-ES-T---    116
HIVHXB2R   E--P-----LE--T--CNE---T--TQ.......---S-QI-V---TV-ES-T---    116
HIVNL43    E--P-----LE--T--CNE---T--TQ.......---S-QI-V---TV-ES-T---    116
HIVMN      E--P-----.Q--T--CSE---T--TQ.......---S-QI-V---TV-ES-T---    115
HIVBRVA    E--P-----LE--T--CNE---T--TQ.......---S-QI-V---TV-ES-T---    116
HIVSC      E--P-----LE--T--CNE---T--TQ.......---S-QI-V---TV-ES-T---    116
HIVJH3     E--P-----LE--T--CNE---T--TQ.......---S-QI-V---TI-ES-T---    113
HIVBAL1    E--P-----LE--T--CDE---T--TS......R---S-QI-V---TV-ES-T---    117
HIVJRCSF   E--P-----LE--T--CNE---T--TQ.......---N-EI-V---TV-ES-T---    116
HIVOYI     E--P-----LE--N--CSE---T--TQ.......---S-EI-V---AV-EP-T---    116
HIVSF2     E--P-----LE--T--CSE---N--AQ.......---S-QI-V---AV-DS-T---    116
HIVNY5CG   E--P-----LE--T--CSE---T--TQ.......---S-QI-L---AV-GS-T---    116
HIVSF162   E--P-----LE--N--CSE---T--TQ.......---S-QV-G---AV-DS-T---    116
HIVCDC4    E--P-----LE--T--CSE---T--TQ.......---S-QI-V---AV-ES-T---    116
HIVSF33    E--P-----LD--T--CSE---T--TQ.......---S-QI-V---AV-ES-T---    116
HIVHAN     E--P-----LE--T--CSE---N--TQ.......---S-QV-V---AV-EP-T---    116
HIVRF      D--Q-----LE--T--SSE---T--TQ.......---S-QV-V---AV-ES-A---    116
CONSENSUS2 EPVPLQLPPlERLnLnCsEdCgtSGTQ??????GVGsPQIsVEspavLeSGTee?C*    115
HIVELI     ---------L---N-N-S-D-RT----.......---H---S--SPTV-E---EEQ--    118
HIVZ2Z6    ---------L---N-N-S-D-GA----.......---S---S--SPTV-E---EEQ--    118
HIVNDK     ---------L---N-N-S-D-GT----.......---S---P--PPAV-E---EE*     116
HIVMAL     ---------L---T-N-N-D-GT----.......---S---S--SPAI-G---EE*     117
HIVU455    ---------I---R-D-S-S-GT----QPQGTET---G---S--SSAV-G---KN*     123
SIVCPZ     K-GD-E--ELDK-S-Q-V-TTQDV--SNTSQPQTAT-ETVPAGGNYSI-GK-AKN*     124
```

HIV1 REV

```
   Information dense stretches (0.8 or greater) are indicated by ****.

           \/ 3' sj            exon\/exon
                                ******          **********
CONSENSUS1 MAGrsgdsde.ellktVRlIKfLYQSNPpPspeG.TrqARRNRRRRWReRQRqIrsiSewilstyLGRsa      68
CONSENSUS2 --gRSG--DE.dLL-Av----?------------?tRQ--r--------A---Q-?--gERI-------pe      66
HIV1REV    ...........^XX^^^R^IKXLYQSNPXPXX^G.^X^AR^NRRRRWRXRQR^I^X^^XX^^XXXLGRXX      58
                     D  DAI I              K  S D  K           D D IA  II
                     E  EGV V              R  T E  R           E E VG  VV
                     N  NS  L              E     N             N N LS  LL
                     Q  QT  M              Q     Q             Q Q MT  MM
                     K  KP  F                    K             K K FP
                     R  R   W                    R             R R W
                     H  H   Y                    H             H H Y
                     S  S   C                    S             S S C
                     T  T                        T             T T

                 *****
CONSENSUS1 ePVpLQLPPleRLtLDcseDCGtSGtq......?GVGsPqiLvESP?vLesGtKE*       115
CONSENSUS2 E--P------E--n-nC-Ed-g---TQ??????------QIsV-spa---S-Tee?C*     115
HIV1REV    ^PVXLQLPP.^RL^L^X^^^CXXSG^^.......GVGXP^^XXE^^^^LX^G^^^*       105
           D          D  D D DKD     AD            EI    AAAI  A AKD
           E          E  E N ERE     GE            QV    GGGV  G GRE
                      N  N   NEN     SN                  SSS   S SEN
                      Q  Q   QQQ     TQ                  TTT   T TQ
                      K  K   K K     PK                  PPP   P P
                      R  R   R R      R
                      H  H   H H      H
                      S  S   S S      S
                      T  T   T T      T
HIV1REV:12      *LYQSNP
        An insertion proximal to the "L" causes an over 100-fold
        decrease in structural gene expression. Trono et al., Cell
        59:113-120 (1989). Mutations in the "Y" and "SN" also
        cause a dramatic loss of function. Malim et al., Cell
        58:205-214 (1989). See also Venkatesh et al., Virology 176:
        39-47, 1990.

HIV1REV:28       ARnNRRRRWR
        The "RRRRW" is implicated in the localization of rev in the
        nucleolus and is required for function. Perkins et al., J.
        Acquired Immune Deficiency Syndrome 2:256-263 (1989); Malim
        et al., Cell 58:205-214 (1989); Venkatesh et al., Virology 176:
        39-47, 1990.

        This fragment contains a region similar to a conserved pattern
        in HIV2, "GtAnQRRnrrRR".

HIV1REV:64       LQLFP*
        *Mutations proximal to the "P" cause a trans-dominant
        repression of Rev function. Malim et al., Cell 58:205-214 (1989).
```

```
                                exon\/exon
CONSENSUS  Ms??a?eEelrrrLRLi?lLHQt?npYP?GpGtAnQRRnrrRRwrqrW?qllALAdriysfPDPPtdtPL   63
HIV2ROD    -NERADE-GLQRK---IRL---T.NP--Q-P-T-S---NRR--WKQR-RQIL---DSIYTF----ADS--   69
HIV2NIHZ   -TERADE-GLQRK---IRL---T.NP--Q-P-T-S---NRR--WKQR-RQIL---DSIYTF----ADS--   69
HIV2ISY    -TERADE-GVRRK---IRL---T.NP--Q-L-T-R---NRR--WERR-KQIL---DRIYTF----ADP--   69
HIV2ST     -NERAEE-ELRRK---IRL---T.NP--Q-P-T-S---NRR--WKQR-RQLV---DKIYTF----ADS--   69
HIV2BEN    -SERADE-GLQGK---LRL---T.NP--Q-P-T-S---NRR--RRRQ-LRLV---NKLCAV----TDS--   69
HIV2D194   -RDRADE-GLQEK---IRL---I.NP--H-Q-T-S---NRR--RRRQ-FRLV---TKLHTI----TDS--   69
HIV2GH1    -HEKADG-ELQER---IRL---T.NP--H-P-T-S---NRR--QRRR-LRLV---DKLYTF----TDS--   69
SIVMM251   -SSHERE-ELRKR---IHL---T.NS--T-P-T-N---QRR--WRRR-QQLL---DRIYSF----TDT--   69
SIVMMP11   -SSHERE-ELRKR---IHL---T.NS--T-P-T-N---QRR--WRRR-QQLL---DRIYSF----TDT--   69
SIVMM142   -RSHTGE-ELRRR---IHL---T.NP--T-S-S-N---QKR--WRQR-QQLL---DRIYSF----TDT--   69
SIVMM239   -SNHERE-ELRKR---IHL---T.NP--T-P-T-N---QRK--WRRR-QQLL---DRIYSF----TDT--   69
SIVMNE     -SSHAEE-ELRRR---IHL---T.NP--T-P-T-N---NRR--WRQR-QQLL---DRIYSF----TNT--   69
SIVSMMH4   -SST..E-ELRRR---IHF---T.NP--Q-P-T-N---RRR--WRQR-QQIL---DRIYSF----VDT--   67
SIVSMMPBJ  -SSN..E-ELRRR---IHL---T.NP--D-P-T-N---RRR--WRQR-QQIL---DRIYSF----VDT--   67

CONSENSUS  dlaiQqLQ?L?I??lPdPPtn?pe?l????????s?q?lAEt*    89
HIV2ROD    DQTI-H--G-T-QEL-D--THLPES..........QRL--T-   100
HIV2NIHZ   DRAI-H--G-T-QDL-D--TNLPES....PESTNSNQRL--A-   107
HIV2ISY    DQTI-Q--G-T-QTL-D--TT..............QRL--TQ >  96
HIV2ST     EQTI-H--G-T-QEL-D--TNLPES....SESIDSSQRL--I-   107
HIV2BEN    DRAI-H--R-T-QEL-D--TDLPES........NSNQGL--T-   103
HIV2D194   DRAI-Q--G-T-QEL-D--TDLPES........NSNQGL--T-   103
HIV2GH1    DRAI-D--R-T-HEL-D--TDLPES........NSNQGL--T-   103
SIVMM251   DLAI-Q--N-A-ESI-D--TNTPEALCDPTKGSRSPQD*      107
SIVMMP11   DLAI-Q--N-A-ESI-D--TNTPEALCDPTKDSRSPQD*      107
SIVMM142   DLAI-Q--N-A-ESI-D--TNIPEALCDPTENSRSPQA*      107
SIVMM239   DLAI-Q--N-A-ESI-D--TNTPEALCDPTEDSRSPQD*      107
SIVMNE     DLVI-Q--N-A-ESI-D--TNIPEILHDPTENPRSPQD*      107
SIVSMMH4   DLAI-Q--G-A-EEL-D--TSAPEPLNDVAK...SP*        100
SIVSMMPBJ  DLAI-Q--R-A-EEL-N--ASAPEPLKDAAE...SP*        100
```

HIV2/SIV REV

```
   Information dense stretches (0.7 or greater) are indicated by ****.

                                exon\/exon
                                           ************
CONSENSUS  Ms??a?eEelrrrLRLi?lLHQt?npYP?GpGtAnQRRnrrRRwrqrW?qllALAdriysfPDPPtdtPL      63
HIV2SIVREV M^^^..XEX^^X^LRL^^^LHQX.^^YP^GXG^A^QRR^^^RRX^^^WX^^^ALA^^^X^^PDPPX^^PL      67
            DDD      IK K   IDI      DA  D    S D    DKK    KKK  KII    DDI AI      DA
            EEE      VR R   VEV      NG  E    T E    ERR    RRR  RVV    EEV GV      EG
            NNN      LE     LNL       S  N      N    N      EEE  ELL    NNL SL      NS
            QQQ      MQ     MQM       T  Q      Q    Q      QQQ  QMM    QQM TM      QT
            KKK              KF       P  K      K    K                  KK  PF      KP
            RRR              RW          R      R    R                  RR   W      R
            HHH              HY          H      H    H                  HH   Y      H
            SSS              SC          S      S    S                  SS   C      S
            TTT              T           T      T    T                  TT          T

CONSENSUS  dlalQqLQ?L?I??lPdPPtn?pe?l?????????s?q?lAEt*      89
HIV2SIVREV ^XX^Q^LQXL^I^^^P^PP^^^                          88
           D  I D    A DDI D  AD
           E  V E    G EEV N  GE
              N    S NNL    SN
              Q    T QQM    TQ
              K    P KK     PK
              R      RR      R
              H      HH      H
              S      SS      S
              T      TT      T
```

```
CONSENSUS  MpLGpEErrllrlIafLy?sNPYPpv?eGTARQRRRARrRWrqqQ?QiralaeRi??tr?eeQlVqaidq     64
SIVAGMTYO  -P--P--RRFVRL-WL-YST----PSGE----------R--RQQ-D-IRVLVE-.L....QE-.-YAVDR     64
SIVAGM155  -P--P--RRLLRL-AF-YRS----SV.E----------R--KNR-K-IYALAE-IWGTRQED-L-QAIDQ     69
SIVAGM3    -P--S--RRLLRL-AF-NKN----PV.E----------R--RQA-E-LRALAE-IWHSRVEE-L-QAIDQ     69
SIVAGM677  -S--K--KQALKI-KT-YGS----QFS.----------Q--RKQ-Q-IDKIAG-VLNTFEDQ-L-AQLQE     69

CONSENSUS  Lvl??qhL?iQqLPDPP?sa*                                                     81
SIVAGMTYO  -ADEAQH-AI-Q-----HSA-                                                     84
SIVAGM155  -VLDTQH-VT-Q-----SQA-                                                     89
SIVAGM3    -VLDQQH-AI-Q-----SSS-                                                     90
SIVAGM677  -QLENKD-VL-H-----HIHQDSSGIPAVWAPATPRGSNRACSSSGEGCEGS LGQTGCYCPIRLSGSHQQ > 147
```

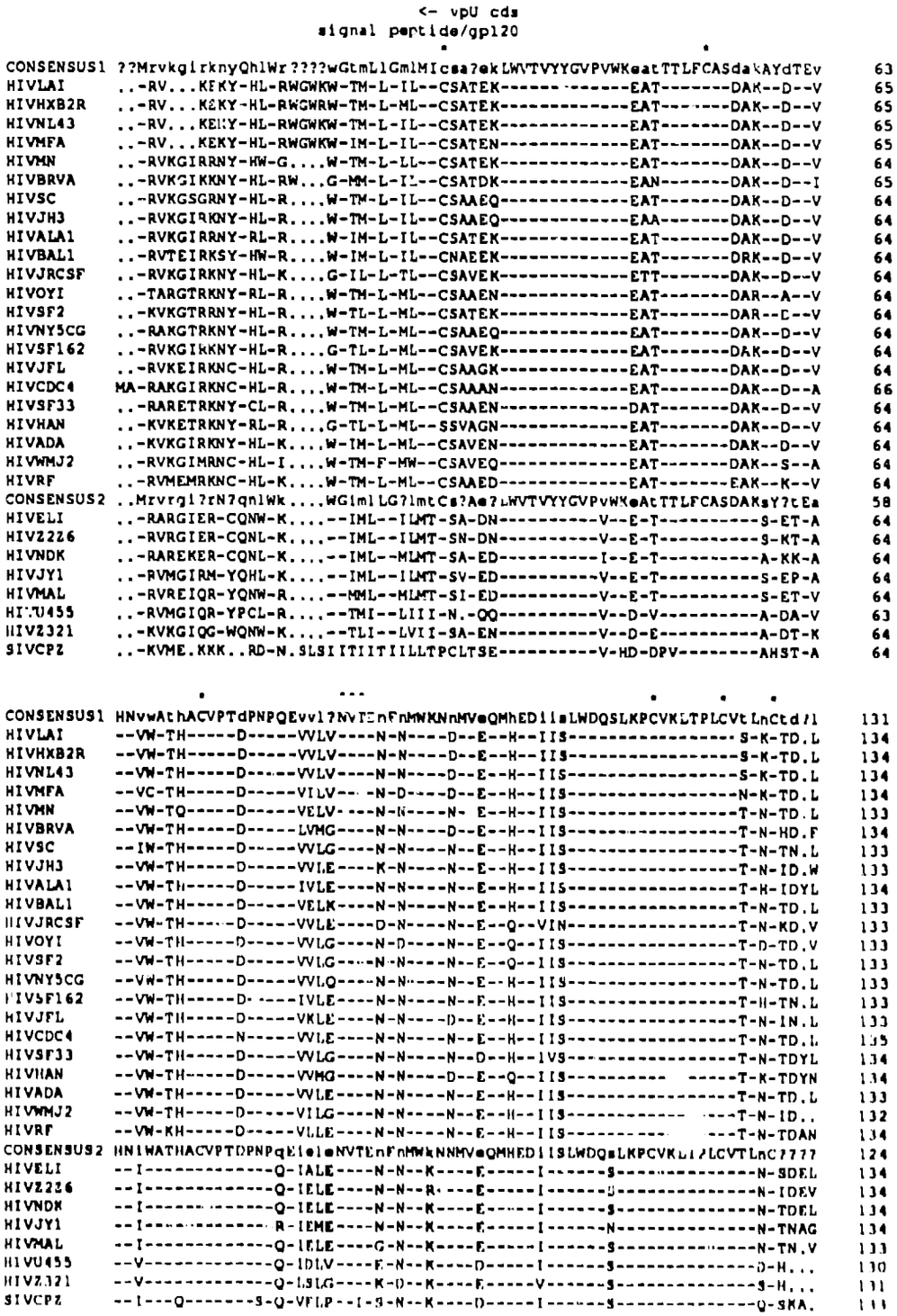
AGMREV

```
   Information dens- stretche: (.7 or greater) are indicated by ****.

                                          **************
CONSENSUS  MpLGpEErrllrlIafLy?sNPYPpv?eGTARQRRRARrRWrqqQ?QiralaeRi??tr?eeQlVqaidq     64
AGMREV     M^LGXEE^^X^^^IXXLXX^NPYPXX..GTARQRRRAR^RW^^XQ^Q^XX^XXR.^....^^Q VYX^^^     62
            A      KK IKI       D                   K  KD  D I  I     I     DD      IDK
            G      RR VRV       E                   R  RE  E V  V     V     EE      VER
            S       E L L       N                   E   N  N L  L     L     NN      LNE
            T       Q M M       Q                   Q   Q  Q M  M     M     QQ      MQQ
            P                   K                       K  K          F     KK       K
                                R                       R  R          W     RR       R
                                H                       H  H          Y     HH       H
                                S                       S  S          C     SS       S
                                T                       T  T                TT       T

                      *****
CONSENSUS  Lvl??qhL?iQqLPDPP?sa*     81
AGMREV     LXX^X^^LXXQ^LPDPP^XX*     82
              D KD     D      D
              E RE     E      E
                EN     N      N
                QQ     Q      Q
                 K     K      K
                 R     R      R
                 H     H      H
                 S     S      S
                 T     T      T
```

```
                                                                  env cds ->
CONSENSUS1 Mqplqi?aIvaLVVaaIiaIvVWs.Iv?ieyRkilrqr...kidRLIDri?ERAEDSGNESeGdqeE.lS     62
HIVLAI     -QPIQI.A-AA---AI-IA-V--S.-VIIEY-KILRQR...KID----RLI----------E-...-.I-     61
HIVNL43    -QPIIV.A-VA---AI-IA-V--S.-VIIEY-KILRQR...KID----RLI----------E-...-.V-     61
HIVMN      -QPLVIAA-VA---AG-IA-V- S.-VFIEY-KIRRQR...KID----RIS----------E-DQE-.L-     64
HIVBRVA    -QPLQISA-VA---AA-IA-V--S.-ALLEY-KLLRQR...KID----RIR----------E-DQE-.L-     65
HIVSC      -QPLQILS-VA---AA-IA-V--S.-VFILI ENIKTK...KNRQVN  NKRKSR------E-IRK-.L-     62
HIVJH3     -QSL IAA-VA---VA-IA-V--S.-VFIEY-KILRQR...KID----RIR----------E-DQE-.L-     64
HIVBAL1;2  -QALQISA-VG---AA-IA-V--T.-VFIEY-KILRQR...KID----RIT----------D-DQE-.L-     65
HIVSF2     -QSLQILA-VS---VA-IA-V--T.-VLIEY-KILRQRK. T.D --E EKKQ.KTVAMKVK-TRRN.YQ     62
HIVJRCSF   -QPLQILA-VA---AG-IA-I--S.-VLIEY-KILRQR...KID----KIR----------E-DQE-.L-     65
HIVNY5CG   -LSF..  ..........NSC-DH.SVHRIQESI.KTKK.SRQDN.  NKRKSRRQWQ -RR-PGRI.I-     48
HIVSF162   -QPLQILA-VA---AA-IA-V--T.-VYIEY-KILRQR...KID----RIT----------E-DQE-.L-     65
HIVHAN     -LSLQIAA-VA---AA-IV-V--T.-VVIEY-KLLRQR...KIN----RLT----------E-DQE-.L-     65
HIVRF      -QSLEILA-VA---AA-LA-V--T.-VGIEI-KTLRQKKK T.G --E EK-Q.KTVAMRVMEMRKN.CQ     63
CONSENSUS2 mQPlvIiAIaALVVaiIiAIVVWT.IVfIEyRrIkrQR...kIDcliDRIrERAEDSGNESEGdrEe.LS     65
HIVELI     M--LG-I--A----AI-L------.--F--Y-R-KK--...R--CLL---T-----------DR-K.--     65
HIVZ2Z6    M--SQ-I--A----AA-I------.--F--Y-R-KR--...K--CII---R-----------DR-E.--     65
HIVNDK     M--LV-I--A----AI-I------.--Y--Y-R-KR--...K--CLI---R-----------ER-E.--     65
HIVMAL     I--LV-L--V----TL-I------.--F--I-K-RR--...K--RLI---R-----------DT-E.--     65
SIVCPZ     MTLLVGL.VLI--GLIAWN-CI-GY-IKWGY-RYKRH-LETE-ERLNLILR----------N-EE-ER-E     69

CONSENSUS1 AL????vem?GH?A.PwdidDl?     77
HIVLAI     --VEMGVEM.--H-.-WDID-L*     81
HIVNL43    --VEMGVEM.--H-.-WDID-L*     81
HIVMN      --....VGM.--D-.-WVIN-L?     81
HIVBRVA    --....VEMG--D-.-WDID-L*     82
HIVSC      --....VEM.--D-.-GDID-L?     79
HIVJH3     --....VEM.--D-.-WDID-L*     80
HIVBAL1;2  --....VEM.--H-.-WDIN-L*     81
HIVSF2     H-WRWGTLLL-MLMI?            78
HIVJRCSF   --....VER.--L-.-WDIN-L*     81
HIVNY5CG   TCGDGARCSL-C  S?            62
HIVSF162   --....VER.--L-.-WDVD-L*     81
HIVHAN     --....VER.--L-.-WDVN-]*     81
HIVRF      H-WKWGTMLL-MLMI?            79
CONSENSUS2 KL....VEM.GHhA.PWD?DDL*     80
HIVELI     --....---.--H-.---I---*     81
HIVZ2Z6    --....---.--H-.---I---*     81
HIVNDK     --....---.--H-.---V---*     81
HIVMAL     --....---.--D-.---V---*     81
SIVCPZ     Q-IH.NYNHNN-F-N-M.F.--*     88
```

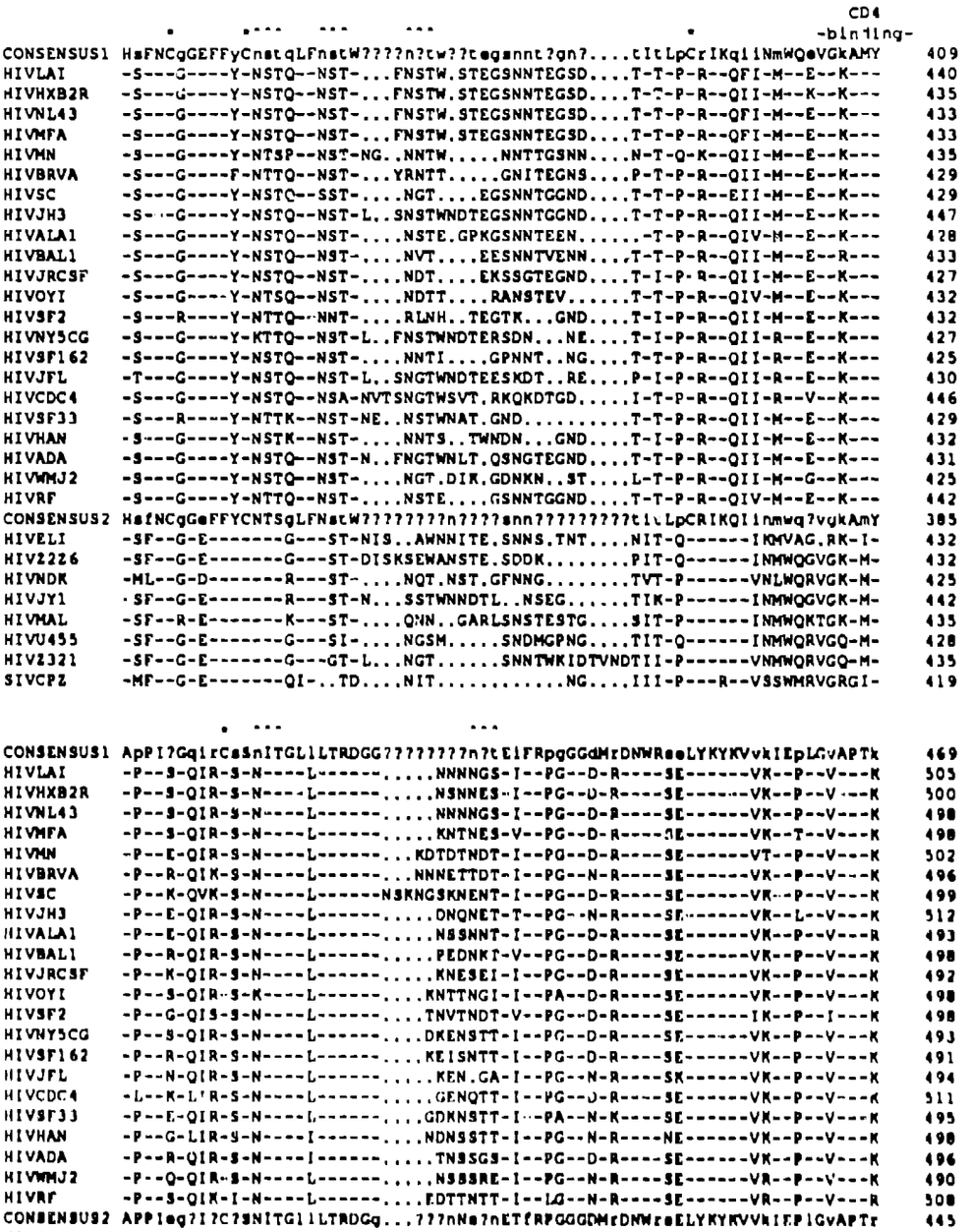
HIV1 VPU

```
   Information dense stretches (0.8 or greater) are indicated by ****.

                                                                  env cds ->
                                                                   ************
CONSENSUS1 Mqplqi?aIvaLVVaaIiaIvVWs.Iv?ieyRkilrqr...kidRLIDri?ERAEDSGNESeGdqeE.lS      62
CONSENSUS2 mQP-vIiA-aA----i--A-V--T.-VfIE--rIk-QR...-IDcli-RIr----------E--rEe.L-      65
HIV1VPU    ^X^X.^.^IX^LVVXXI^XI^V.^.IXX^E^R^XX^Q^...^^^X^^D^^XERAEDSGNES^G...^.^S      59
           I A  I A  A      I  I  S    I I K  K K   KID II KI           D    K I
           V G  V G  G      V  V  T    V V R  R R   RVN VV RV           E    R V
           L S    S  S      L         L L          E   LL  L                E L
           M T    T  T      M         M M          Q   MM  M                Q M
             P    P  P                  F
                                         W
                                         Y
                                         C

CONSENSUS1 AL????vem?GH?A.PwdidDl*     77
CONSENSUS2 K-....VEM.--h-.-WD?D-L*     80
HIV1VPU    XL....^XX.GHXA.PXX^^DX*     75
                 I           ID
                 V           VN
                 L
                 M




HIV1VPU:49        ERAEDSGNESkG
          See Strebel et al., Science 241: 1221-4, 1988.
```

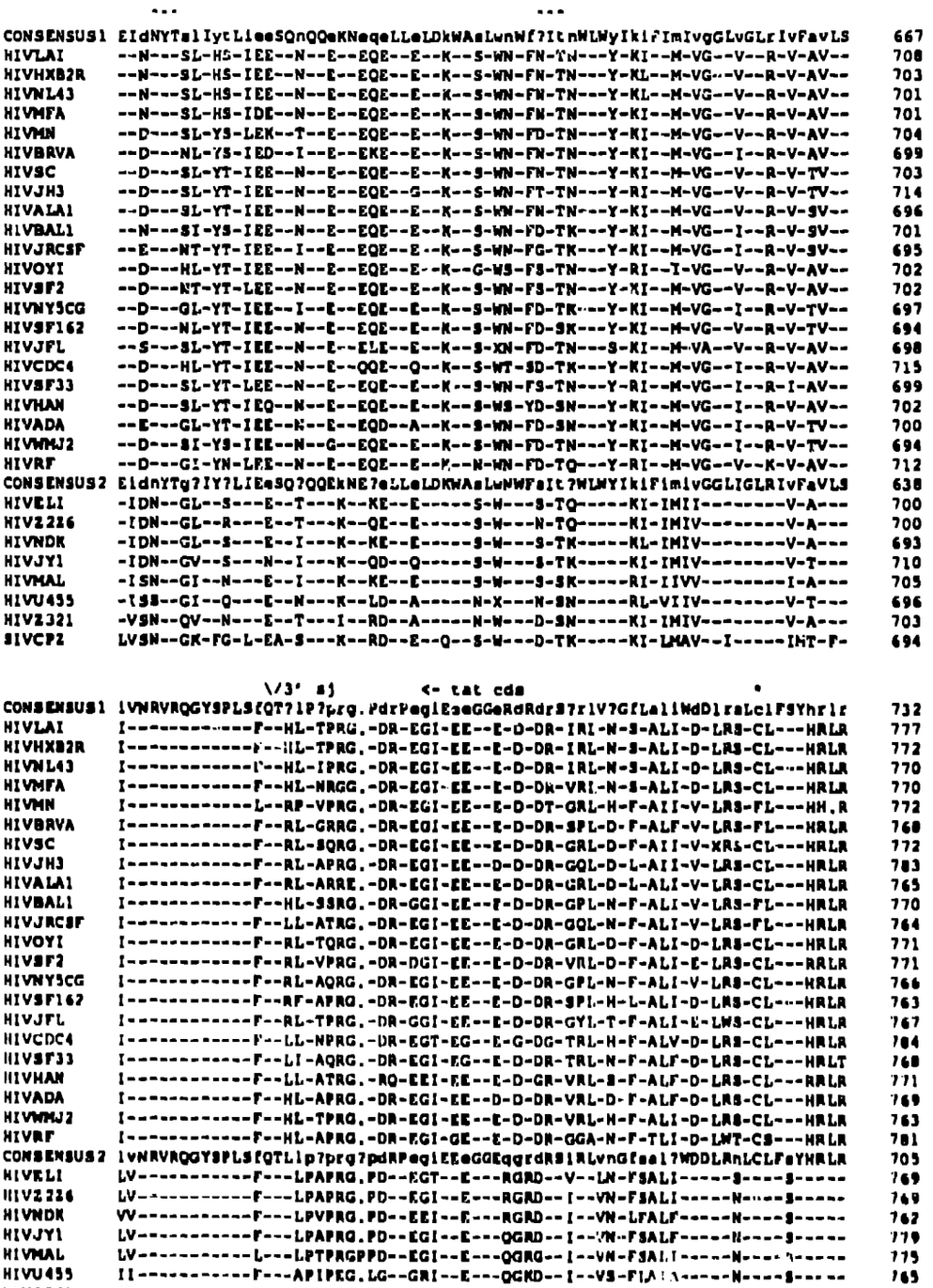

<- vpU cds
signal peptide/gp120

```
                                                *** ***
CONSENSUS1 ??n?T????????nt?s#?????????me?gEikNCSFnittsirdKvqkeyAlfyklDvvpI??dn????   174
HIVLAI     G.NA-NTNSS...NTNSSS...GEMMMEKG-IK----NISTSIRG-VQKEY-FFYKL-IIP-..DN....   191
HIVHXB2R   K.ND-.........NTNSSS...GRMIMEKG-IK----NISTSIRG-VQKEY-FFYKL-IIP-..DN....   186
HIVNL43    K.ND-.........NTNSSS...GRMIMEKG-IK----NISTSIRD-VQKEY-FFYKL-IVP-..DN....   186
HIVMFA     K.ND-.........NTNSSN...GRMIMEKG-IK----NISTSIRN-VQKEY-FFYKL-IRP-..DN....   186
HIVMN      R.NT-NTNNS.TANNNSNS..EGT..IKGG-MK----NITTSIRD-MQKEY-LLYKL-IVS-..DN....   191
HIVBRVA    ..NA-.........NATSNS...GK.MMEGG-MK----NITTSIRD-MQKEY-LFYKL-IVP-..DND...   185
HIVSC      R.ND-STNAT...NTTSSN...RG.KMEGG-MT----NITTSIRS-VQKEY-LFYKL-VVP-..DN....   189
HIVJH3     G.ND-SPNAT...NTTSSG...GE.KMEKG-MK----NITTSIRD-VQKEH-LFYKH-VVP-..NNSTKD   193
HIVALA1    ..NA-.........NTTSSN...GT.KMERG-IK----NITTSIRD-MQKES-LFHKL-VVP-..DN....   184
HIVBAL1    R.NA-NGNDT...NTTSSS...RG.MVGGG-MK----NITTNIRG-VQKEY-LFYKL-IAP-..DN....   189
HIVJRCSF   ..NA-.........NTTSSS...EG.MMERG-IK----NITKSIRD-VQKEY-LFYKL-VVP-..DN....   183
HIVOYI     ..NT-SSSLRNATNTTSSS...WE.TMEKG-LK----NTTTSIRD-MQEQY-LFYKL-VLP-..DK....   191
HIVSF2     G.KA-.........NTNSSN...WKEEI.KG-IK----NITTSIRD-IQKEN-LFRNL-VVP-..DNAST.   187
HIVNY5CG   T.NA-.Y....A.NGSSE.........ERG-IR----NVTTIIRN-IQKEY-LFYRL-IVP-..DK....   181
HIVSF162   K.NA-.........NTKSSN...WKE.MDRG-IK----KVTTSIRN-MQKEY-LFYKL-VVP-..DN....   184
HIVJFL     K.NV-.....NI.NNS......WGK.MEEG-IK----NTTTSIKN-MQREY-LFYKL-VVP-..DK....   183
HIVCDC4    NTNN-T.......NTTELSIIVVWEQRGKG-MR----NITTSIRD-VQREY-LFYKL-VEP-.DDN....   193
HIVSF33    G.NA-.........NTNNSS....GGTVEKE-IK----NITTGIRD-VQKAY-YFYKL-VVP-DDDN....   187
HIVHAN     ..NA-..........NSS....WGR.MEKG-IQ----KVTTNIRD-VQKES-LFYKT-VVP-..DNNKT.   184
HIVADA     R.NV-.....NI.NNSSE.........GMRG-IK----NITTSIRD-VKKDY-LFYRL-VVP-..DN....   182
HIVWMJ2    K.NI-DW.....KNTT.........IIGGG-VK----NITTSRRD-VHKEY-LFYKL-VVP-KGDN....   183
HIVRF      L.NG-.........NVTSSS...GGTMMENG-IK----QVTTSRRD-TQKKY-LFYKL-VVP-EKGNISPK   192
CONSENSUS2 ??N?t??????gnn????????????#ek?#mKNCSfN?ttv?rDK??qvyalFYrLD1VpI.?dn???.   162
HIVELI     RN-GT.M....GNNVTT.........EEKG.M----F-VTTVLK--KQQVYAL--R--I-P-..DNDSS.   186
HIVZ2Z6    ME-VT.M...KNNNVT..........EEIR.M----F-ITTVVR--TKQVHAL--R--I-P-..DNDNS.   186
HIVNDK     R.-.S.....KGN.........GKVEEEEKR..----F-V....R--REQVYAL--K--I-P-..DNNNR.   180
HIVJY1     G.-KT....TNGNNTTNQ....EEQMMEKGEM----F-ITTVIS--KKQVHAL--R--V-P-..DDDNS.   192
HIVMAL     ..-GTAV...NGTNAG.SNRTNAELKMEIGEV----F-ITPVGS--.RQEYAT--N--L-Q-..DDS...   191
HIVU455    ..-IT.INNTN.NNTNIT....DG...VREEM----F-MTTELR--KQKVYSL--R--I-Q-...NKT..   184
HIVZ321    ..-IT.I...KDNNTNVD.......TEMKEEI----Y-MTTELR--QRKIYSL--R--I-P-.GGNS...   184
SIVCPZ     ..-FS..QAKNLTNQTSS.........PPLEM----F-VTTELR--KKQVYSL--VE-V-NL..GNE...   185

        peptide T region      *---  *        *        * *** ***  * ---  *
CONSENSUS1 ???????n?t???yrLi?CntSvITQaCPKvsFePIPiHYCapagFAilkCndKkFnGtGpCtNVSTVQC   232
HIVLAI     ......D.TT..SYT-TS-NT-V---A---VS-E---I---APAG--ILK-NN-T-N-T-P-T-------   252
HIVHXB2R   ......D.TT..SYK-TS-NT-V---A---VS-E---I---APAG--ILK-NN-T-N-T-P-T-------   247
HIVNL43    .........T..SYR-IS-NT-V---A---VS-E---I---APAG--ILK-NN-T-N-T-P-T-------   245
HIVMFA     .........T..TYR-IS-NT-V---A---VS-E---I---APAG--ILK-ND-T-N-T-P-T-------   245
HIVMN      ......D.ST..SYR-IS-NT-V---A---IS-E---I---APAG--ILK-ND-K-S-K-S-K-------   252
HIVBRVA    ......KTNT..RYR-IS-NT-V---A---VT-E---I---APAG--ILK-NN-K-N-T-P-T-------   247
HIVSC      .........T..SYT-IN-NT-V---A---VS-E---I---A.RW--ILN-NN-K-N-T-P-T-------   247
HIVJH3     NIKN..DNST..RYR-IS-NT-V---A---IS-E---I---APAG--IIK-ND-K-N-T-P-T-------   259
HIVALA1    .......NNT..SYI-IS-NT-V---A---VS-Q---I---APAG--ILK-KD-K-N-K-T-K-------   245
HIVBAL1    ......NSNN..RYR-IS-NT-V---A---VS-E---I---APAG--ILK-KD-K-N-K-P-T-------   251
HIVJRCSF   ......KNNT..KYR-IS-NT-V---A---VS-E---I---APAG--ILK-NN-T-N-K-Q-K-------   245
HIVOYI     .......NDT..KFR-IH-NT-T---A---IS-E---M---TPAG--ILK-ND-K-N-T-P-T-------   252
HIVSF2     .....TTNYT .NYR-IH-NR-V---A---VS-E---I---TPAG--ILK-NN-T-N-K-P-T-------   250
HIVNY5CG   ......DNTS...YT-IN-DT-V---A---VS-E---I---APAG--ILK-ND-K-N-T-P-T-------   242
HIVSF162   ......DNTS...YK-IN-NT-V---A---VS-E---I---APAG--ILK-ND-K-N-S-P-T-------   245
HIVJFL     ......DNTS...YT-IN-NT-T---A---VS-E---I---TPAG--ILK-ND-K-N-K-P-K-------   244
HIVCDC4    ...KNTTNNT..KYR-IN-NT-V---A---VS-E---I---TPTG--LLK-ND-K-N-T-P-T-------   258
HIVSF33    ......TNTS...YR-IH-NS-V---T---VS-E---I---APAG--ILK-NN-K-S-K-Q-T-------   248
HIVHAN     ...SNRDNTT..SYM-IH-NR-V---A---VS-E---I---APAG--ILK-ND-K-N-K-P-K-------   249
HIVADA     ......DNTS...YR-IN-NT-T---A---VS-E---I---TPAG--ILK-KD-K-N-T-P-K-------   243
HIVWMJ2    .......NSS..RYR-IN-NT-V---A---VS-E---I---APAG--ILK-ND-K-N-T-P-T-------   244
HIVRF      NNTS..NNTSYGNYT-IH-NS-V---A---VS-E---I---TPAG--ILK-ND-K-N-T-P-K-------   260
CONSENSUS2 ??????tnst..nYRLINCnTSaITQACPKvsFePIPIHyCAPAGFAILKCrDkkFnGtgpC?nVSTVQC   223
HIVELI     ......TNST..N------N--A-------VS-E-----Y----------R-KK-N-TGP-TN------   248
HIVZ2Z6    ......TNST..N------N--A-------VS-E-----Y----------R-KR-N-TGP-TN------   248
HIVNDK     ......TNST..N------D--T-------IS-E-----F----------R-KK-N-TGP-SN------   242
HIVJY1     ANTSN.TNYT..N------N--A-------VT-E-----Y----------K-KK-N-TGP-KR------   259
HIVMAL     ......DN.S..S------N--V-------VT-D-----Y----------N-KK-N-TEI-RN------   252
HIVU455    ......DNN...S------N--T-------VS-E-----Y----------R-PE-N-KGP-RN------   245
HIVZ321    ...SNGD.SS..K------N--A-------VS-E-----Y----------R-EE-E-KGP-RN------   248
SIVCPZ     .......NN...T--I---N-TA-------TS-E-----Y--  ---------N-KD-S-KGK-TN----H-   245
```

```
                                                                  <- V3 neutralization loop
                          ---             ---             ---    *.*     ---
CONSENSUS1 THGIrPvVSTqLLLNGSLAEeevviRS?NFtdNaKtIIVqLneaveINCTRp?nnntrk...si?i??GP    294
HIVLAI     ----R-V---Q---------EEVVI--A--TD-A-T-I-Q-NQSVE-----P.NNNTRK...SIRIQR--    318
HIVHXB2R   ----R-V---Q---------EEVVI--V--TD-A-T-I-Q-NTSVE-----P.NNNTRK...RIRIQR--    313
HIVNL43    ----R-V---Q---------EDVVI--A--TD-A-T-I-Q-NTSVE-----P.NNNTRK...SIRIQR--    311
HIVMFA     ----R-V---Q---------EEGVI--A--TD-A-T-I-Q-NTSVE-----P.NNNTRK...SIRIQR--    311
HIVMN      ----R-V---Q---------EEVVI--E--TD-A-T-I-H-NESVQ-----P.NYNKRK...RIHI..--    316
HIVBRVA    ----R-V---Q---------EEVVI--E--TN-V-T-I-Q-NESVE-----P.NNNTRK...RITM..--    311
HIVSC      ----R-V---H---------EEVVL--E--TD-A-T-I-Q-KEAVE-----PNNNTTR....SIHI..--    311
HIVJH3     ----K-V---Q---------EEVVI--E--TD-A-T-I-Q-KEPVV-----P.SKTTRR...RIHI..--    323
HIVALA1    ----R-V---Q---------EEVVI--D--TD-A-T-I-Q-NESVE-----P.NIYRKG...RIHI..--    309
HIVBAL1    ----R-V---Q---------EEVVI--A--AD-A-V-I-Q-NESVE-----P.NNNTRK...SIHI..--    315
HIVJRCSF   ----R-V---Q---------EKVVI--D--TD-A-T-I-Q-NESVK-----P.SNNTRK...SIHI..--    309
HIVOYI     ----K-V---Q---------EEVII--S--TN-A-I-I-Q-NKSVE-----P.NNNTRN...RISI..--    316
HIVSF2     ----R-I---Q---------EEVVI--D--TN-A-T-I-Q-NESVA-----P.NNNTRK...SIYI..--    314
HIVNY5CG   ----K-V---Q---------GEVVI--E--TN-V-T-I-Q-NESVE-----P.NNNTRK...GIAI..--    306
HIVSF162   ----R-V---Q---------EGVVI--E--TD-A-T-I-Q-KESVE-----P.NNNTRK...SITI..--    309
HIVJFL     ----R-V---H---------GEVVI--E--MD-A-T-I-Q-NKSIA-----P.NNNTRK...SITL..--    308
HIVCDC4    ----R-V---Q---------EEVVI--E--TN-A-T-I-Q-NVSVE-----P.NNHTRK...RVTL..--    322
HIVSF33    ----K-V---Q---------EEVVI--D--TN-A-T-L-Q-NVSVE-----P.NNNRRR...RITS..--    312
HIVHAN     ----R-V---Q---------KEVVI--D--TD-T-T-I-H-NESVE-----P.NNNTRK...GIHI..--    313
HIVADA     ----R-V---Q---------EEVVI--S--TD-A-N-I-Q-KESVE-----P.NNNTRK...SIHI..--    307
HIVWMJ2    ----R-V---Q---------EEIVI--E--TD-A-T-I-H-NESVE-----P.YNNVRR...SLSI..--    308
HIVRF      ----R-V---Q---------EEVVI--E--TD-V-T-I-Q-NASVQ-----P.NNNTRK...SITK..--    324
CONSENSUS2 THGIrPVVSTQLLLNGSLAEeEiiIRSENiTnNaK?IIvqLneav?InJtRP?y?ntRq???rtsi..Gl    284
HIVELI     ----R---------------E-VI-----L-N-A-N--AH-NESVK-T-A--.YQNT-Q...RTPI..-L    312
HIVZ226    ----R---------------E-II-----L-N-A-I--VQ-NESVA-N-T--.YRNI-Q...RTSI..-L    312
HIVNDK     ----R---------------E-II-----L-N-V-T--VQ-NASIV-N-T--.YKYT-Q...RTSI..-L    306
HIVJY1     ----R---------------E-II-----L-N-V-T--VH-NESVE-N-T--DNKIT-Q...STPI..-L    324
HIVMAL     ----K---------------E-IM-----L-D-T-N--VQ-NETVT-N-T--.GNNT-R...GIHF..-P    316
HIVU455    ----K---------------R-IR-----F-N-A-T--VQ-VNPVK-N-S--.Y.NT-KNIRRYSI..-S    311
HIVZ321    ----R---------------G-VR-----F-D-A-I--VQ-VKPVN-T-M--.NNNT-K...SISI..-P    312
SIVCPZ     ----K---T----.------GNITV-V--KSK-TDVW-VQ-VEAVSLN-H--.GNNT-G...EVQI..-P    309

   V3 neutralization loop ->
                          *---    ---             ---
CONSENSUS1 Grafyt...tg?IigdiRqAhCNisrakWnnTLkqiv?kLre??QFg?NRT...iv.FnqSSGGDPEIVm    352
HIVLAI     -RAFVT...IGK-.GNM-Q-H--ISRAK-NA--KQIASK-RE..--GN---...II.-KQ---------T    378
HIVHXB2R   -RAFVT...IGK-.GNM-Q-H--ISRAK-NN--KQIASK-RE..--GN---...II.-KQ---------T    373
HIVNL43    -RAFVT...IGK-.GNM-Q-H--ISRAK-NA--KQIASK-RE..--GN---...II.-KQ---------T    371
HIVMFA     -RAFVT...IGK-.GNM-Q-H--ISRAK-MS--KQIASK-RE..--GN---...VI.-KQ---------T    371
HIVMN      -RAFYT...TKN-IGTI-Q-H--ISRAK-ND--RQIVSK-KE..--.K---...IV.-NQ---------M    376
HIVBRVA    -RVYYT...TGQ-IGDI-R-H--LSRSK-EN--KQIVTK-RV..--.K---...IV.-NR---------M    371
HIVSC      -RAFYA...TGD-IGDI-Q-H--ISRAK-NN--KQIVIK-RD..--.E---...II.-NR---------M    371
HIVJH3     -RAFYT...TKQ-AGDL-Q-H--INRAR-NA--KQIVGK-RK..--.V---...IV.-NR---------M    383
HIVALA1    -RAFHT...TRQ-IENI-Q-H--ISRAK-ED--KQIVEK-RE..--.K---...IV.-NQ---------T    369
HIVBAL1    -RAFYT...TGE-IGDI-Q-H--LSRAK-ND--NKIVIK-RE..--G.---...IV.-KH---------T    375
HIVJRCSF   -RAFYT...TGE-IGDI-Q-H--ISRAQ-NN--KQIVEK-RE..--.N---...IV.-TH---------M    369
HIVOYI     -RAFHT...TKQ-IGDI-Q-H--LSRAT-EK--KQIATK-RK..--.R---...IA.-DR---------M    376
HIVSF2     -RAFHT...TGR-IGDI-K-H--ISRAQ-NN--EQIVKK-RE..--GN---...IV.-NQ---------M    375
HIVNY5CG   -RTLYA...RKK-IGDI-Q-H--LSRAK-ND--KQIVTK-KE..--.R---...IV.-NQ---------M    366
HIVSF162   -RAIYA...TGD-IGDI-Q-H--ISGEK-NN--KQIVTK-QA..--G.---...IV.-KQ---------M    369
HIVJFL     -RAFYT...TGD-IGDI-Q-H--LSRVQ-NN--KQIVIK-RE..--G.---...IV.-KQ---------M    368
HIVCDC4    -RVWYT...TGE-LGNI-Q-H--ISRAQ-NN--QQIATT-RE..--G.---...IA.-NQ---------M    382
HIVSF33    -KVLYT...TGE-IGDI-K-Y--ISRAK-NK--EQVATK-RE..--G.---...IV.-KQ---------M    372
HIVHAN     -RAVYT...TGR-VQDI-L-H--ISRAR-NK--NQIFRK-REIR--.F---...IV.-NR---------M    375
HIVADA     -RAFYT...TGE-IGDI-Q-H--ISRTK-NN--NQIATK-KE..--GN---...IV.-NQ---------M    368
HIVWMJ2    -RAF.R...TRE-IGII-Q-H--ISRAK-NN--KQIVEK-RE..--.K---...IV.-NH---------T    367
HIVRF      -RVIYA...TGQ-IGDI-K-H--LSRAQ-NN--KQVVTK-RE..--.D---...IV.-TS---------L    384
CONSENSUS2 qqalyt???t??i?qdIrqAhCnisr?eWnktlqqVA?kLg?..?ii?nkT?.?ii?FkpssGGDpEITt    337
HIVELI     GQSLYT...TR.SRSI-GQ-H-NISRAQ-SKTLQQ--RK-GT...LL.NK-...-IK-KPSS---P---T    371
HIVZ226    GQALYT...TR.TRSI-GQ-Y-NISKNE-NKTLQQ--IK-GN...LL.NK-T..-I.-KPSS---P---T    371
HIVNDK     RQSLYTITGKKKKTGY-GQ-H-KISRAE-NKALQQ--TK-GN...LL.NK-T..-T.-KPSS---P---S    369
HIVJY1     GQALYT...TR.IKGD-RQ-Y-NISAAA-NKTLQQ--RK-GD...LL.NQ-T..-I.-KPPA---P---T    383
HIVMAL     GQALYT...TG.IVGD-RR-Y-TINETE-DKTLQQ--VK-GS...LL.NK-K..-I.-NSSS---P---I    375
HIVU455    GQAFYV...TGKIIGD-RQ-H-NVSRRD-NRTIQQ--EQ-KK..KF.NNK-...-I.-ASSS---I---T    371
HIVZ321    GRAFFA...TGDIIGD-RQ-H-NVSRTE-NDTLSQ--AQ-RK..HFVNTS-..D-I.-ANSS---V---T    374
SIVCPZ     GMTFYN...IENVVGDTRS-Y-KINGTT-NRTVEF-KK- AT....SSNP-AAN-T.INRAS---P-V-H    371
```

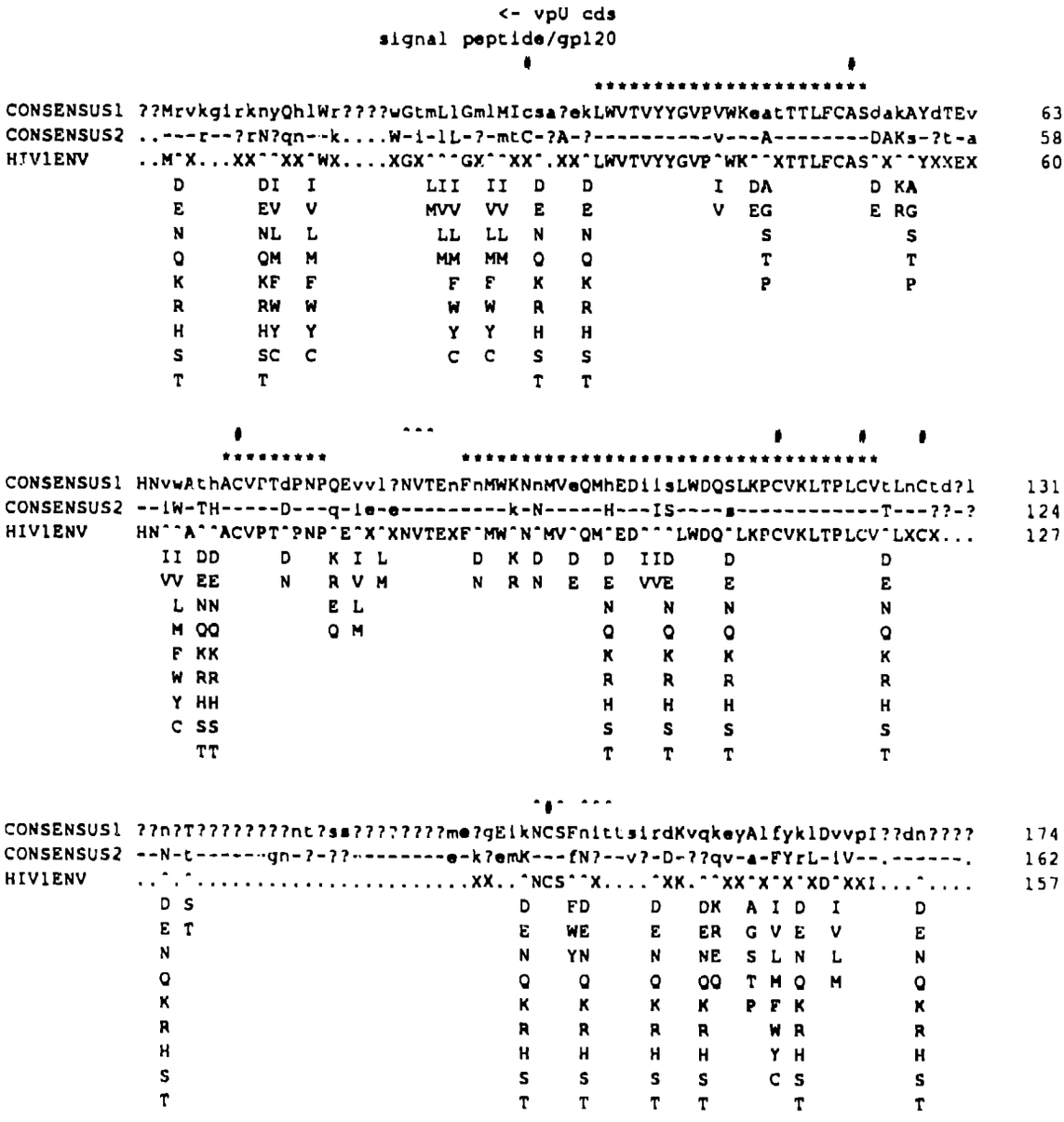

CD4
-binding-
CONSENSUS1 HsFNCgGEFFyCnstqLFnstW????n?tw??tegsnnt?gn?....tItLpCrIKqiiNmWQeVGkAMY 409
HIVLAI -S---G----Y-NSTQ--NST-....FNSTW.STEGSNNTEGSD....T-T-P-R--QFI-M--E--K--- 440
HIVHXB2R -S---G----Y-NSTQ--NST-....FNSTW.STEGSNNTEGSD....T-T-P-R--QII-M--K--K--- 435
HIVNL43 -S---G----Y-NSTQ--NST-....FNSTW.STEGSNNTEGSD....T-T-P-R--QFI-M--E--K--- 433
HIVMFA -S---G----Y-NSTQ--NST-....FNSTW.STEGSNNTEGSD....T-T-P-R--QFI-M--E--K--- 433
HIVMN -S---G----Y-NTSP--NST-NG..NNTW.....NNTTGSNN....N-T-Q-K--QII-M--E--K--- 435
HIVBRVA -S---G----F-NTTQ--NST-....YRNTT.....GNITEGNS....P-T-P-R--QII-M--E--K--- 429
HIVSC -S---G----Y-NSTQ--SST-....NGT....EGSNNTGGND....T-T-P-R--EII-M--E--K--- 429
HIVJH3 -S---G----Y-NSTQ--NST-L..SNSTWNDTEGSNNTGGND....T-T-P-R--QII-M--E--K--- 447
HIVALA1 -S---G----Y-NSTQ--NST-....NSTE.GPKGSNNTEEN.....-T-P-R--QIV-M--E--K--- 428
HIVBAL1 -S---G----Y-NSTQ--NST-....NVT....EESNNTVENN....T-T-P-R--QII-M--E--R--- 433
HIVJRCSF -S---G----Y-NSTQ--NST-....NDT....EKSSGTEGND....T-I-P-R--QII-M--E--K--- 427
HIVOYI -S---G----Y-NTSQ--NST-....NDTT....RANSTEV.....T-T-P-R--QIV-M--E--K--- 432
HIVSF2 -S---R----Y-NTTQ--NNT-....RLNH..TEGTK...GND....T-I-P-R--QII-M--E--K--- 432
HIVNY5CG -S---G----Y-KTTQ--NST-L..FNSTWNDTERSDN...NE....T-I-P-R--QII-R--E--K--- 427
HIVSF162 -S---G----Y-NSTQ--NST-....NNTI....GPNNT..NG....T-T-P-R--QII-R--E--K--- 425
HIVJFL -T---G----Y-NSTQ--NST-L..SNGTWNDTEESKDT..RE....P-I-P-R--QII-R--E--K--- 430
HIVCDC4 -S---G----Y-NSTQ--NSA-NVTSNGTWSVT.RKQKDTGD.....I-T-P-R--QII-R--V--K--- 446
HIVSF33 -S---R----Y-NTTK--NST-NE..NSTWNAT.GND.....T-T-P-R--QII-M--E--K--- 429
HIVHAN -S---G----Y-NSTK--NST-....NNTS..TWNDN...GND....T-I-P-R--QII-M--E--K--- 432
HIVADA -S---G----Y-NSTQ--NST-N..FNGTWNLT.QSNGTEGND....T-T-P-R--QII-M--E--K--- 431
HIVWMJ2 -S---G----Y-NSTQ--NST-....NGT.DIK.GDNKN..ST....L-T-P-R--QII-M--G--K--- 425
HIVRF -S---G----Y-NTTQ--NST-....NSTE....GSNNTGGND....T-T-P-R--QIV-M--E--K--- 442
CONSENSUS2 HsfNCgGeFFYCNTSqLFNstW????????n????snn?????????tiLpCRIKQIinmwq?vgkAmY 385
HIVELI -SF--G-E-------G---ST-NIS..AWNNITE.SNNS.TNT....NIT-Q------IKMVAG.RK-I- 432
HIVZ226 -SF--G-E-------G---ST-DISKSEWANSTE.SDDK.....PIT-Q------INMWQGVGK-M- 432
HIVNDR -ML--G-D-------R---ST-....NQT.NST.GFNNG.....TVT-P------VNLWQRVGK-M- 425
HIVJY1 -SF--G-E-------R---ST-N...SSTWNNDTL..NSEG.....TIK-P------INMWQGVGK-M- 442
HIVMAL -SF--R-E-------K---ST-....QNN..GARLSNSTESTG....SIT-P------INMWQKTGK-M- 435
HIVU455 -SF--G-E-------G---SI-....NGSM.....SNDMGPNG....TIT-Q------INMWQRVGQ-M- 428
HIVZ321 -SF--G-E-------G---GT-L...NGT.....SNNTWKIDTVNDTII-P------VNMWQRVGQ-M- 435
SIVCPZ -MF--G-E-------QI-..TD....NIT.....NG....III-P---R--VSSWMRVGRGI- 419
CONSENSUS1 ApPI?GqirCaSnITGLlLTRDGG????????n?tEIFRpgGGdMrDNWReeLYKYKVvkIEpLGvAPTk 469
HIVLAI -P--S-QIR-S-N----L------.....NNNNGS-I--PG--D-R----SE------VK--P--V---K 505
HIVHXB2R -P--S-QIR-S-N----L------.....NSNNES-I--PG--D-R----SE------VK--P--V---K 500
HIVNL43 -P--S-QIR-S-N----L------.....NNNNGS-I--PG--D-R----SE------VK--P--V---K 498
HIVMFA -P--S-QIR-S-N----L------.....KNTNES-V--PG--D-R----SE------VK--T--V---K 498
HIVMN -P--E-QIR-S-N----L------...KDTDTNDT-I--PG--D-R----SE------VT--P--V---K 502
HIVBRVA -P--R-QIK-S-N----L------...NNNETTDT-I--PG--N-R----SE------VK--P--V---K 496
HIVSC -P--K-QVK-S-N----L------NSKNGSKNENT-I--PG--D-R----SE------VK--P--V---K 499
HIVJH3 -P--E-QIR-S-N----L------.....DNQNET-T--PG--N-R----SE------VK--L--V---K 512
HIVALA1 -P--E-QIR-S-N----L------.....NSSNNT-I--PG--D-R----SE------VK--P--V---R 493
HIVBAL1 -P--R-QIR-S-N----L------.....PEDNKT-V--PG--D-R----SE------VK--P--V---K 498
HIVJRCSF -P--K-QIR-S-N----L------.....KNESEI-I--PG--D-R----SE------VK--P--V---K 492
HIVOYI -P--S-QIR-S-K----L------....KNTTNGI-I--PA--D-R----SE------VK--P--V---K 498
HIVSF2 -P--G-QIS-S-N----L------....TNVTNDT-V--PG--D-R----SE------IK--P--I---K 498
HIVNY5CG -P--S-QIR-S-N----L------....DKENSTT-I--PG--D-R----SE------VK--P--V---K 493
HIVSF162 -P--R-QIR-S-N----L------....KEISNTT-I--PG--D-R----SE------VK--P--V---K 491
HIVJFL -P--N-QIR-S-N----L------.....KEN.GA-I--PG--N-R----SK------VK--P--V---K 494
HIVCDC4 -L--K-LIR-S-N----L------.....GENQTT-I--PG--D-R----SE------VK--P--V---K 511
HIVSF33 -P--E-QIR-S-N----L------....GDKNSTT-I--PA--N-K----SE------VK--P--V---K 495
HIVHAN -P--G-LIR-S-N----I------....NDNSSTT-I--PG--N-R----NE------VK--P--V---K 498
HIVADA -P--R-QIR-S-N----I------.....TNSSGS-I--PG--D-R----SE------VK--P--V---K 496
HIVWMJ2 -P--Q-QIR-S-N----L------....NSSSRE-I--PG--N-R----SE------VR--P--V---K 490
HIVRF -P--S-QIK-I-N----L------....EDTTNTT-I--LG--N-R----SE------VR--P--V---R 508
CONSENSUS2 APPIeg?I?C?SNITGlLlLTRDGg...???nNs?nETfRPGGGDMrDNWreELYKYKVVkIEPlGvAPTr 445
HIVELI ----ERN-L-S-----LI ----G.....IN-STN--F-------R---RS--------Q---L-V---R 497
HIVZ226 ----EGQ-N-S-----LL-----G.....TN-SSN--F-------R---RS--------K---L-V---R 497
HIVNDR ----EGL-K-S-----LL-----G.....AN-SSH--I-------R---RS--------K---I-V---K 490
HIVJY1 ----EGL-K-T-----LL-----G.....VN-STN--F-------K---RN--------R---L-I---R 507
HIVMAL ----AGV-N-L-----LI-----G...NSSD-SDN--L-------R---IS--------R---L-V---K 502
HIVU455 ----QGV-R-E-----LL-----G.....TN-TKN--F-------R---KS--------K---L-V---R 493
HIVZ321 ----KGV-K-V-----LL-----V.....GN-TSN--F-------R---RS--------K---L-V---K 500
SIVCPZ ----RGN-T-N-----LL--S-TP....VTN-SGNL-F--T--N-K-I-RS--------R---LSV---K 485

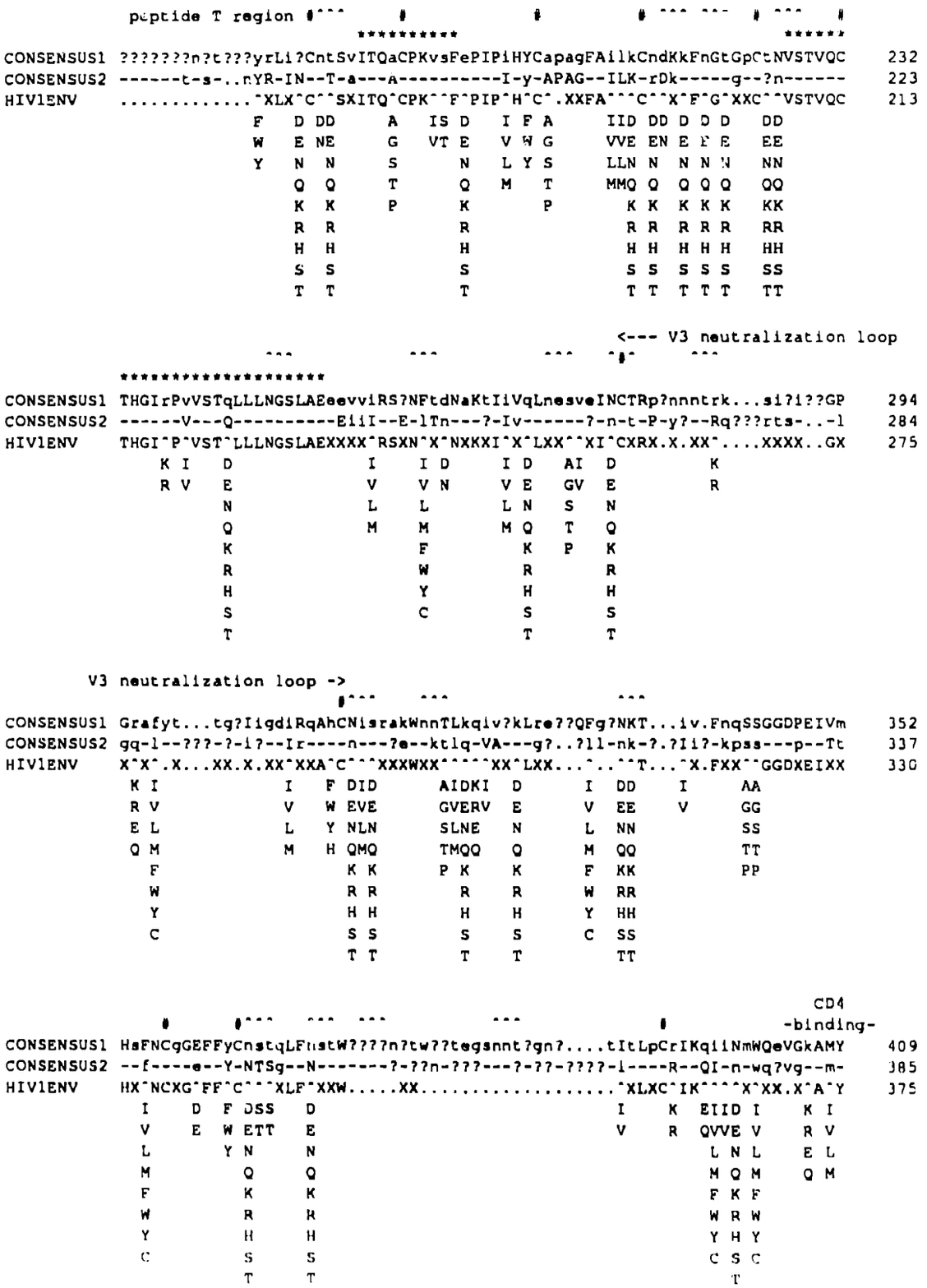
```
                    gp120/gp41
CONSENSUS1 AkRRVV....QReKRAv.q?lGA?FLGFLGAAGSTMGa?amtLTVQaRqLlSqIVQQQNNLLrAIeaQQH  531
HIVLAI     -K----....--E---V.G.I--L-------------ARSMT----A-Q-L-G---------R--EA---  569
HIVHXB2R   -K----....--E---V.G.I--L-------------AASMT----A-Q-L-G---------R--EA---  564
HIVNL43    -K----....--E---V.G.I--L-------------CTSMT----A-Q-L-D---------R--EA---  562
HIVMFA     -K----....--E---V.G.I--L-------------AASMT----A-Q-L-G---------R--EA---  562
HIVMN      -K----....--E---..A.I--L-------------AASVT----A-L-L-G---------R--EA---  565
HIVBRVA    -K----....--E---V.G.L--L-------------AASLT----A-L-L-G---------M--EA---  560
HIVSC      -K----....--E---V.GTI--M-------------ATSMT----A-L-L-G---------R--EA---  564
HIVJH3     -K----....--E---V.G.I--V-------------A.SMT----A-L-L-G---------R--EG---  575
HIVALA1    -K----....--E---V.G.L--L-------------ARSMT----A-Q-L-G---------R--EA---  557
HIVBAL1    -K----....--E---V.G.I--V-------------AAAMT----A-L-L-G---------R--EA---  562
HIVJRCSF   -K----....--E---V.G.I--L-------------ARSMT----A-Q-L-G---------R--EA---  556
HIVOYI     -R----....--E---V.GML--M-------------ARSMT----A-Q-L-G---------R--EA---  563
HIVSF2     -K----....--E---V.GIV--M-------------AVSLT----A-Q-L-G---------R--EA---  563
HIVNY5CG   -K----....--E---V.GAL--L-------------AASMA----T-Q-M-G---------K--EA---  558
HIVSF162   -K----....--E---V..TL--M-------------ARSLT----A-Q-L-G---------R--EA---  555
HIVJFL     -K----....--K---V.GTI--M-------------AASMT----A-L-L-G---------R--EA---  559
HIVCDC4    -K----....--E---V.GML--M-------------ATSMA----A-Q-L-G---------R--KA---  576
HIVSF33    -K----....--E---V.GVI--M-------------AASIT----A-K-L-G---------R--EA---  560
HIVHAN     -K----....--E---V.GML--M-------------ARSLT----A-Q-L-G---------R--EA---  563
HIVADA     -K----....--E---V.GTI--M-------------AASIT----A-L-L-G---------R--EA---  561
HIVWMJ2    -K----....--E---V.GTI--M-------------AGSLT----A-Q-L-G---------R--DA---  555
HIVRF      -K----....--E---V.GTI--M-------------AGSIT----A-H-L-G---------R--EA---  573
CONSENSUS2 AkRRVV....eREKRAi.G.lGA?FLGFLGAAGSTMGAaS?tLTvQARqLlSGIVqQQnNLLRAIEAQQH  507
HIVELI     -K----....E-----I.-.L--M-------------R-VT--V---Q-M----Q--N----------  561
HIVZ2Z6    -K----....E-----I.-.L--M-------------R-LT--V---Q-L----Q--N----------  561
HIVNDK     -R----....E-----I.-.L--V-------------A-VT--V---Q-M----H--N----------  554
HIVJY1     -K----....E-----I.-.L--V-------------V-VA--G---Q-L----Q--N----------  571
HIVMAL     -K----....E-----I.-.L--M-------------A-LT--V---Q-L----Q--N----------  566
HIVU455    -K----....E-----V.-.L--I-------------A-IT--V---Q-L----Q--S----------  557
HIVZ321    -K----....A-----I.-.M--F-------------A-IT--V---R-L----Q--N----------  564
SIVCPZ     -R-HT-ARQKD-Q---AI-.L--L-------------AAVT--V---Q-L----Q--N---K-------  554

                                                .   .     ...   ...
CONSENSUS1 lLqLTVWGIKQLQARvLAvERYLkDQQLlgiWGCSGKLICTTaVPWNaSW.SNKsl??iwnnmTWmeWer  598
HIVLAI     L-Q----------I--V----K----LGI----------A----A--.---SLEQIWNNM--ME-DR  638
HIVHXB2R   L-Q----------I--V----K----LGI----------A----A--.---SLEQIWNHT--ME-DR  633
HIVNL43    L-Q----------I--V----K----LGI----------A----A--.---SLEQIWNNM--ME-DR  631
HIVMFA     L-Q----------I--V----K----LGI----------A----A--.-  SLEQFWNNM--ME-DR  631
HIVMN      M-Q----------V--V----K----LGF----------T----A--.---SLDDIWNNM--MQ-ER  634
HIVBRVA    M-E----------V--V----K----LGI----------A----A--.---SLSDIWDNM--ME-ER  629
HIVSC      L-Q----------V--V----R----LGI----------T----T--.---SLDKIWGNM--ME-ER  633
HIVJH3     L-Q----------I--V----K----LGI----------A----A--.---SLEEIWDNM--ME-ER  644
HIVALA1    L-Q----------V--V----R----LEI----------T----A--.---SLNQIWDNM--ME-ER  626
HIVBAL1    L-Q----------V--V----R----LGI----------A----A--.---SLNKIWDNM--IE-DR  631
HIVJRCSF   M-Q----------V--V----K----MGI----------A----T--.---SLDSIWNNM--ME-EK  625
HIVOYI     L-Q----------V--V----K----LGI----------T----A--.---SLNEIWDNM--MQ-ER  632
HIVSF2     L-Q----------V--V----R----LGI----------A----A--.---SLEDIWDNM--MQ-ER  632
HIVNY5CG   L-Q----------V--V----K----LRI----------T----A--.---SLDKIWDNM--ME-ER  627
HIVSF162   L-Q----------V--V----K----LGI----------A----A--.---SLDQIWNNM--ME-ER  624
HIVJFL     L-Q----------V--V----Q----LGI----------T----A--.---SLDEIXNNM--MQ-ER  628
HIVCDC4    L-Q----------I--V----K----LGF----------A----A--.---TLDQIWNNM--ME-DK  645
HIVSF33    L-Q----------V--V----R----LGI----------T----T--.---SLDKIWNNM--ME-ER  629
HIVHAN     L-Q----------V--V----R----LGI----------T----A--.---TLDQIWNNM--ME-ER  632
HIVADA     L-Q----------V--L----R----LGI----------A----A--.---TLDMIWDNM--ME-ER  630
HIVWMJ2    L-Q----------V--V----R----LGI----------T----A--.---SMNQIWDNL--MS-ER  624
HIVRF      L-Q----------V--V----R----LGI----------T----A--.---SLNMIWNNM--MQ-ER  642
CONSENSUS2 lLqLTVWGIKQLQARvLAVErYLkDQqLLGIWGCSGKkhICtT?VPWNSSW.SNrSl?dIW?nMTWmeWer  573
HIVELI     L-Q----------I----R--K--Q---l-----KH--T-N-------.--R-LNE--QN---ME-ER  630
HIVZ2Z6    L-Q----------I----R--K--Q---l-----KL--T-T-------.--R-LND--QN---ME-ER  630
HIVNDK     L-Q----------V----R--K--Q---I-----RH--T-N-------.--R-LDE--QN---ME-ER  623
HIVJY1     M-Q----------V----S--K--Q---I-----KH--T-T-------.--R-LEE--NN---IE-ER  640
HIVMAL     L-Q----------V----R--Q--R---M-----KH--T-T-------.--R-LDD--NN---MQ-ER  635
HIVU455    L-K----------V----R--Q--Q---I-----KL--T-T-------.--R-QKD--NN---LQ-ER  626
HIVZ321    L-K----------I----R--K--Q---I-----KI--P-N-------.--R-QSD--DK---LE-DK  633
SIVCPZ     L-Q-SI--V------L----R--Q--QI--L-----KAV-Y-T----N--PGSN-TDD--GNL--QQ-DK  624
```

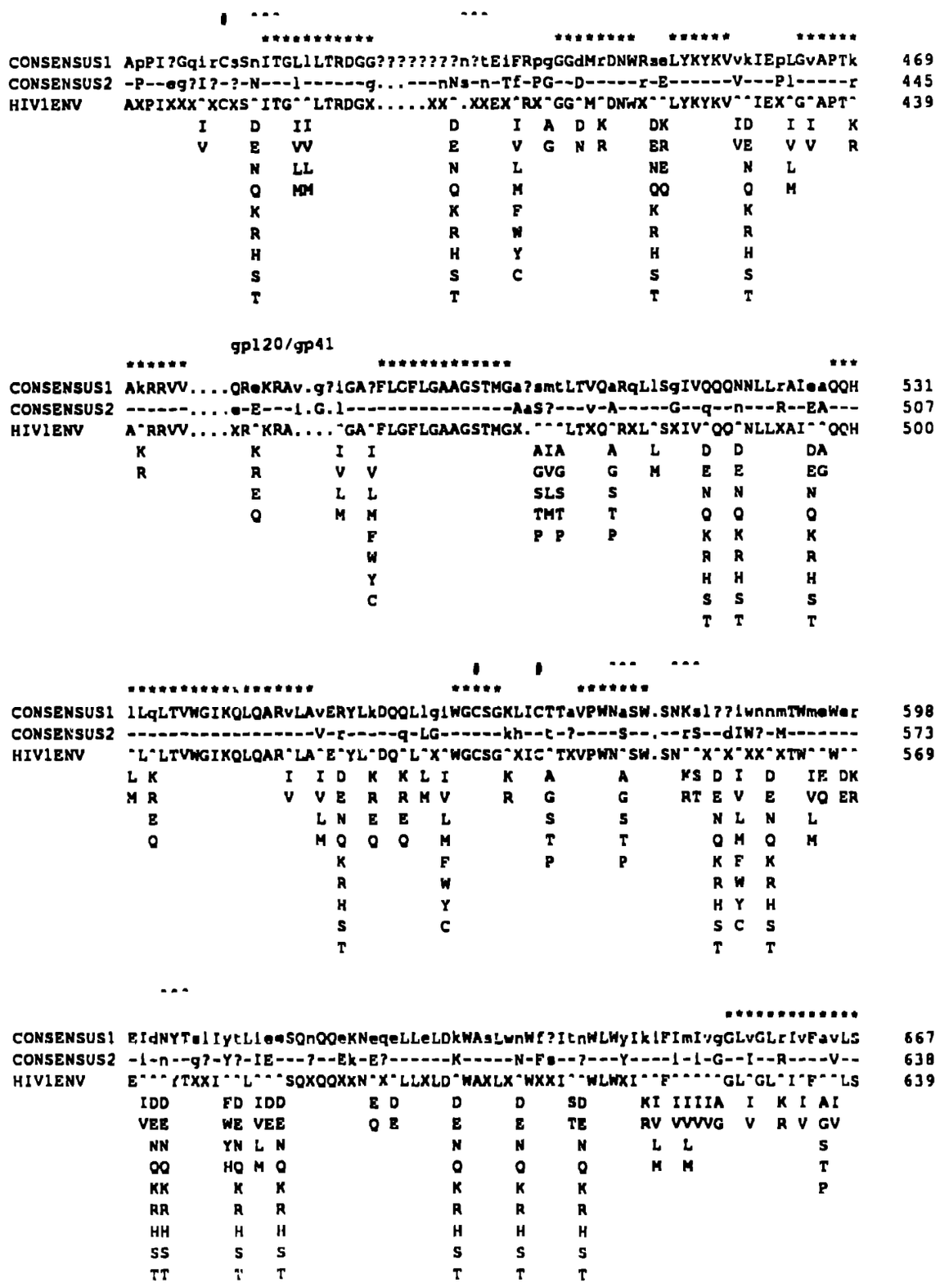


```
                                                    <- rev cds
CONSENSUS1 DLlLlv?rlVElLG......?rrGWEaLKYwwnLLqYWsqELknSavsllnatAlAVaEgTDRvlevvQr   794
HIVLAI     --L-IVTRI--L--.......RR---A---WWN--Q--SQ--KN-AVSLLNAT-I--A-G---VIEVV-G   840
HIVHXB2R   --L-IVTRI--L--.......RR---A---WWN--Q--SQ--KN-AVSLLNAT-I--A-G---VIEVV-G   835
HIVNL43    --L-IVTRI--L--.......RR---A---WWN--Q--SQ--KN-AVNLLNAT-I--A-G---VIEVL-A   833
HIVMFA     --L-IVTRI--L--.......RR---A---WWN--Q--SQ--KN-AVSLLNAT-I--A-G---VIEVV-G   833
HIVMN      --L-IAARI--L--.......RR---V---WWN--Q--SQ--KS-AVSLLNAT-I--A-G---VIEVL-R   835
HIVBRVA    --L-IVTRI--L--.......RR---V---WWN--Q--SQ--KN-AVSLLNAT-I--G-R---AIEVV-P   831
HIVSC      --L-IVTRI--L--.......RR---A---WWN--Q--SQ--RN-AVSLVNAT-I--A-G---VIELL-R   835
HIVJH3     --L-IVTRI--L--.......RR---A---WWN--Q--SQ--KN-AVSLFNAI-I--A-G---VLKIL-R   846
HIVALA1    --L-IVTRI--L--.......RR---V---WWN--Q--SQ--KN-AVSLLNAT-I--A-G---VIEVL-R   828
HIVBAL1    --L-IVMRI--L--......LAG---V---WWN--Q--SQ--KN-AVSLLNAT-V--A-G---VIEVL-R   834
HIVJRCSF   --L-TVTRI--L--.......RR---I---WWN--Q--SQ--KN-AVSLLNAT-I--A-G---IIEVV-R   827
HIVOYI     --I-IVARI--L--.......RR---V---WWN--Q--SQ--KN-VISLLNAT-I--A-G---VIEIV-R   834
HIVSF2     --L-IAART--I--.......HR---A---WWS--Q--IQ--KN-AVSWLNAT-I--T-G---VIEVA-R   834
HIVNY5CG   --L-IVARI--L--.......RR---A---CWN--Q--GQ--KN-AISLLNAT-V--A-G---VIEVA-R   829
HIVSF162   --I-IAARI--L--.......RR---A---WGN--Q--IQ--KN-AVSLFDAI-I--A-G---IIEVA-R   826
HIVJFL     --L-IVTRI--L--.......RR---A---WWN--Q--SQ--KN-AVSLLNAT-I--A-G---IIEVV-R   830
HIVCDC4    --L-IVARI--L--.......RR---V---WWN--Q--SQ--KN-AVSLVNVT-I--A-G---VIEVV-R   847
HIVSF33    --L-IVARI--L--.......RR---V---WWN--L--SQ--KN-AVSLLNAT-I--A-G---VIEVV-R   831
HIVHAN     --L-IVARI--T--.......RR---V---WWN--Q--SQ--KN-AVSLFNTI-I--A-G---VIEVV-R   834
HIVADA     --L-IVARI--L--.......RR---V---WWN--Q--SQ--RN-AVSLLNAT-I--A-G---VIEVV-R   832
HIVWMJ2    --L-IVKRI--L--.......RR---A---WWN--Q--SK--KN-AVGLLNAI-I--A-G---VIEVV-R   826
HIVRF      --L-IVVRI--L--.......RR---A---WWN--Q--SQ--KN-AVSLLNTT-I--A-G---IIEVA-R   844
CONSENSUS2 DliLIaaRlVElLq???????RrGWealKYLwnLlqYW?qELknSalsLldt?AlAVAe?tDRvIEl?qr   764
HIVELI     -LI--AV-I--L-G.......-R--DIL---WN-LQ--SQ--RN-ASS-FDAI-I---EGT--V--IIQR   832
HIVZ226    -LI--AA-I--L-G.......-R--EAL---WN-LQ--SR--KN-ASS-LDTI-I---EGT--V--IVRR   832
HIVNDK     -SI--AA-I--L-G.......-R--EAL---WN-LQ--SQ--RN-ASS-LDTI-I---ERT--V--VVQR   825
HIVJY1     -LI--AT-I--L-G.......-R--EAI---WS-LQ--TQ--KN-FIS-LNAT-I---EGT--I--LIRR   842
HIVMAL     -LL--AT-I--L-G.......-R--EAL---WN-LQ--GQ--KN-AIS-LNTT-I---ECT--V--IGQR   838
HIVU455    -FA--VA-A--L-GRSSLKGL-L--EGL---WN-LL--GR--KI-AIT-LDAV-V---GWI--V--IGQT   835
HIVZ321    -CA--AA-I--T-.......I-R--ETL---GN-VI--GQ--KN-AIN-LDTV-I---DWT--V--VVQR   835
SIVCPZ     SLACNVW-QLKT-GHLILHSL-LLRERLCL-GGIIQ--GK--KI-AIS-LDAT-I---EGT--I--AFQV   833

CONSENSUS1 a?RallhIprRlRQGlERaLl*    814
HIVLAI     AC-AIRH-PR-I---L--I-L-    861
HIVHXB2R   AC-AIRH-PR-I---L--I-L-    856
HIVNL43    AY-AIRH-PR-I---L--I-L-    854
HIVMFA     AY-AIRH-PR-I---L--I-L-    854
HIVMN      AG-AILH-PT-I---L--A-L-    856
HIVBRVA    AF-AILH-PR-I---L--A-Q-    852
HIVSC      AF-AILH-PT-I---L--A-Q-    856
HIVJH3     AF-AILH-PT-I---L--A-L-    867
HIVALA1    AF-AILH-PR-I---F--A-L-    849
HIVBAL1    AV-AILH-PR-I---L--A-L-    855
HIVJRCSF   VY-AILH-PT-I---L--A-L-    948
HIVOYI     AY-AFLN-PR-I---L--A-L-    855
HIVSF2     AY-AILH-RR-I---L--L-L-    855
HIVNY5CG   IC-GILH-PR-I---L--L-L-    850
HIVSF162   IG-AFLH-PR-I---F--A-L-    847
HIVJFL     IG-AILH-PV-I---L--A-L-    851
HIVCDC4    IY-AFLH-PR-I---F--A-L-    868
HIVSF33    VG-AILH-PT-I---F--A-L-    852
HIVHAN     AC-AILH-PR-V---L--A-L-    855
HIVADA     IY-AILH-PT-I---L--L-L-    853
HIVWMJ2    IC-AIIH-PR-I---L--A-L-    847
HIVRF      IL-AFLH-PR-I---L--A-L-    865
CONSENSUS2 a?RA?LnlPrRlRQGlERaLL*    783
HIVELI     AC--V-NI-R-I---L--S---    853
HIVZ226    AC--V-HI-T-I---L--L---    853
HIVNDK     AC--I-NV-R-I---L--L---    846
HIVJY1     AF--V-HI-R-V---L--A---    863
HIVMAL     FG--I-HI-R-I---F--A---    859
HIVU455    IQ--I-NI-R-I---L--A---    856
HIVZ321    AQ--F-NI-R-I---L--A---    856
SIVCPZ     TL-IIRNI-R-I---L--A---    854
```

Information dense stretches (0.8 or greater) are indicated by ****. Cysteine residues are indicated by # and potential N-glycosylation sites by ^^^. The SIVCPZ env sequence has been included in the ordinary CONSENSUS2 (African sequences) but not in the HIV1ENV Pattern.




peptide T region
CONSENSUS1 ???????n?t???yrLi?CntSvITQaCPKvsFePIPiHYCapagFAilkCndKkFnGtGpCtNVSTVQC 232
CONSENSUS2 ------t-s-..nYR-IN--T-a---A----------I-y-APAG--ILK-rDk-----g--?n------ 223
HIV1ENV .....^XLX^C^^SXITQ^CPK^^F^PIP^H^C^.XXFA^^^C^^X^F^G^XXC^^VSTVQC 213
<--- V3 neutralization loop
CONSENSUS1 THGIrPvVSTqLLLNGSLAEeevviRS?NFtdNaKtIiVqLnesveINCTRp?nnntrk...si?i??GP 294
CONSENSUS2 ------V---Q----------EiiI--E-lTn---?-Iv------?-n-t-P-y?--Rq???rts-..-l 284
HIV1ENV THGI^P^VST^LLLNGSLAEXXXX^RSXN^X^NXKXI^X^LXX^^XI^CXRX.X.XX^....XXXX..GX 275
V3 neutralization loop ->
CONSENSUS1 Grafyt...tg?IigdiRqAhCNisrakWnnTLkqiv?kLre??QFg?NKT...iv.FnqSSGGDPEIVm 352
CONSENSUS2 gq-l--???-?-i?--Ir----n---?e--ktlq-VA---g?..?ll-nk-?.?Ii?-kpss---p--Tt 337
HIV1ENV X^X^.X...XX.X.XX^XXA^C^^^XXXWXX^^^^^XX^LXX...^..^^T...^X.FXX^^GGDXEIXX 330
CD4
-binding-
CONSENSUS1 HsFNCgGEFFyCnstqLFnstW????n?tw??tegsnnt?gn?....tItLpCrIKqiiNmWQeVGkAMY 409
CONSENSUS2 --f----e--Y-NTSg--N-------?-??n-???---?-??-????-i----R--QI-n-wq?vg--m- 385
HIV1ENV HX^NCXG^FF^C^^^XLF^XXW.....XX.....^XLXC^IK^^^^X^XX.X^A^Y 375


gp120/gp41
CONSENSUS1 ApPI?GqirCsSnITGLlLTRDGG???????n?tEiFRpgGGdMrDNWRseLYKYKVvkIEpLGvAPTk 469
CONSENSUS2 -P--eg?I?-?-N---l------g...---nNs-n-Tf-PG--D-----r-E------V---Pl-----r 445
HIV1ENV AXPIXXX^XCXS^ITG^^LTRDGX.....XX^.XXEX^RX^GG^M^DNWX^^LYKYKV^^IEX^G^APT^ 439
CONSENSUS1 AkRRVV....QReKRAv.g?iGA?FLGFLGAAGSTMGa?smtLTVQaRqLlSgIVQQQNNLLrAIeaQQH 531
CONSENSUS2 ------....e-E---i.G.l----------AaS?---v-A-----G--q--n---R--EA--- 507
HIV1ENV A^RRVV....XR^KRA....^GA^FLGFLGAAGSTMGX.^^^LTXQ^RXL^SXIV^QQ^NLLXAI^^QQH 500
CONSENSUS1 lLqLTVWGIKQLQARvLAvERYLkDQQLlgiWGCSGKLICTTaVPWNaSW.SNKsl??iwnnmIWmeWer 598
CONSENSUS2 ----------V-r-----q-LG------kh--t-?----S--.--rS--dIW?-M------- 573
HIV1ENV ^L^LTVWGIKQLQAR^LA^E^YL^DQ^L^X^WGCSG^XIC^TXVPWN^SW.SN^^X^X^XX^XTW^^W^^ 569
CONSENSUS1 EIdNYTslIytLIeeSQnQQeKNeqeLLeLDkWAsLwnWf?ItnWLWyIkIFImIvgGLvGLrIvFavLS 667
CONSENSUS2 -i-n--g?-Y?-IE---?--Ek-E?------K-----N-Fs--?---Y----i-i-G--I--R----V-- 638
HIV1ENV E^^^fTXXI^^L^^^SQXQQXXN^X^LLXLD^WAXLX^WXXI^^WLWXI^^F^^^^GL^GL^I^F^^LS 639

```
                    \/3' sj       <- tat cds                        #
          *************
CONSENSUS1 iVNRVRQGYSPLSfQT?lP?prg.PdrPegiEeeGGeRdRdrS?rlV?GfLaliWdDlrsLclFSYhrlr    732
CONSENSUS2 lv--------------L-p----?p-R-----E---Eqgr-R-iRLvn--s--?-D-LRn-CL-s-HRLR    705
HIV1ENV    ^^NRVRQGYSPLS^QTXXXXXXX.XX^PXXXEXXGGX^X^XXSXXX^^GXX^^XWXDXX^L^XFXY^^.^    707
           II           I            K          K K      ID   AI      D I    DD D
           VV           V            R          R R      VE   GV      E V    EE E
           L            L            E          E        LN   SL      N L    NN N
           M            M            Q          Q        MQ   TM      Q M    QQ Q
                        F                                 K   P       K F    KK K
                        W                                 R           R W    RR R
                        Y                                 H           H Y    HH H
                        C                                 S           S C    SS S
                                                          T           T      TT T

                                                   <- rev cds
CONSENSUS1 DLlLiv?  VElLG......?rrGWEaLKYwwnLLqYWsqELknSavsllnatAiAVaEgTDRvievvQr    794
CONSENSUS2 -li-IaaR-----g??????-R---e-l--L---l---?-------i-L-dt?----Ae?t---IEi?q-    764
HIV1ENV    DXXLXXX^XVEXL........XXGW^X^KY^X^L^XYWX^EL^XSXXX^^^XXA^AV^XXXDRX^^^X^X    769
                  K                 D I  I D I    K  K     IID   I  A      IKI K
                  R                 E V  V E V    R  R     VVN   V  G      VRV R
                                      L  L N L    E        LL       S      LEL E
                                      M  M Q M    Q        MM       T      MQM Q
                                         F K               FF       P
                                         W R               WW
                                         Y H               YY
                                         C S               CC
                                           T

CONSENSUS1 a?RailhIprRiRQGlERaLl*    814
CONSENSUS2 ---A?LniP-----------L*    783
HIV1ENV    XXR^^X^^XXR^RQG^ERXLX*    790
              AI DI   I   I
              GV EV   V   V
               L N        L
               M Q        M
               F K        F
               W R        W
               Y H        Y
               C S        C
                 T
```

HIV1ENV:32 LWVTVYYGVPaWiKkjXTTLFCAS

At N-terminus of gp120. This fragment and fragment 91 below have been studied as potential CD4 binding sites (Syu et al., PNAS:87, 3695-9, 1990; See also Maddon et al., Cell 47:333-348, 1986; Cordonnier et al., J. Virol. 63:4464-8, 1989)

This fragment is similar to a conserved pattern in HIV2, "VTVFYGvPaWrNAtiPLfCa".

HIV1ENV:71 ACVPTlPNP

HIV1ENV:91 FlMWnNlMVkQMrEDaarLWDQrLKPCVKLTPLCV*

The peptide "HEDIISLWDQSLK" can serve as both a helper T-cell epitope and a cytotoxic T-cell epitope.

Helper T-cell: Cease et al., PNAS 84:4249-4253 (1987) and Berzofsky et al., Nature 334:706-08 (1988).

Cytotoxic T-cell: Clerici et al., J. Immunol. 146: 2214-19 (1991).

"The carboxyl end borders a site that is functionally linked to an N-linked glycosylation site that is shown below, as well as a site in the V3 loop. Willey et al., J. Virol. 62:139, 1988 and J. Virol. 63:3595-3600, 1989. See also Cordinnier et al., J. Virol. 63:4464-8, 1989).

A small insertion at "EQM" inhibits gp120/gp41 association. Kowalski et al., Science 237:1351-55 (1987).

This fragment is similar to a conserved pattern in HIV2, "QAakDVWrLFETSIKPCVKLiPLC".

HIV1ENV:179 ITQjCPKaiF
Just C-terminal to the "peptide T" region. Potential N-glycosylation site immediately upstream absent in noncytopathic variant (Stevenson et al., J. Virol. 64:3792-3803, 1990).

*

HIVENV:220 VSTVQCTHGInPaVSTrLLLNGSLAE
"N-linked glycosylation site described in HIV1ENV:91, above. The peptide "CTHGIRPVVSTQLLLNGSLAE" can serve as a neutralizing antibody epitope. Ho et al., Science 239:1021-23 (1988). However, see Ronco, AIDS Res. Hum. Retroviruses 7:1-2 (1991).

Mutational analysis -- Willey et al., J. Virol. 63:3595-3600, 1989.

HIV1ENV:416 ITGccLTRDG
Immediately C-terminal to the CD4 binding site defined by Lasky et al., Cell 50: 975-85, 1987.

HIV1ENV:440 GG1MnDNW

A small insertion at "GGD" inhibits gp120/CD4 interactions. Kowalski et al., Science 237:1351-55 (1987).

HIV1ENV:451 LYKYKV
Part of a neutralizing antibody epitope. Charbit et al., AIDS 4:545-51 (1990).

See also Cordinnier et al., J. Virol. 63:4464-8, 1989.

HIV1ENV:463 GaAPTnAnRRVV
Part of a neutralizing antibody epitope. Charbit et al., AIDS 4 545-551 (1990), Chanh et al., EMBO J. 5:30655-71 (1986) and Ho et al., J. Virol. 61:2024-28 (1987).

HIV1ENV:487 FLGFLGAAGSTMG
The entire fragment is embedded in the fusogenic domain of gp41. Freed et al., PNAS 87:4650-54 (1990).
The fragment is embedded in a region of gp41 that is important for oligomeric assembly. Earl et al., PNAS 87:648-52 (1990), and Doms et al., J. Virol. 64:3537-40 (1990).

"FLGFLGA" is part of a neutralizing antibody epitope. Chanh et al., EMBO J. 5:3065-71 (1986) and Ho et al. J. Virol. 61:2024-28 (1987).

A small insertion at "TMG" inhibits gp120/gp41 association. Kowalski et al., Science 237:1351-55 (1987).

This fragment is similar to a conserved pattern in HIV2, "fVLGFLGFLjTAGjAM".

HIV1ENV:530 QQHbLpLTVWGIKQLQARaLA

A mutation in the last "A" can result in a neutralization escape mutant. Reitz et al., Cell 54:57-63 (1988). That "A" is the first residue of the so-called "AVERY" peptide that separates this fragment from the following fragment.

The fragment is embedded in a region of gp41 that is important for oligomeric assembly. Earl et al., PNAS 87:648-52 (1990), and Doms et al., J. Virol. 64:3537-40 (1990).

This fragment is similar to a conserved pattern in HIV2, "QQEbLRLTVWGjKNLQjRViA".

HIV1ENV:564 WGCSG

Part of a 12 amino acid neutralizing antibody epitope. Schrier et al., J. Virol. 62:2531-36 (1988). Also Gnann et al., J. Virol. 61:2639-41, 1987 and Science 237:1346-9, 1987.

The fragment is embedded in a region of gp41 that is important for oligomeric assembly. Earl et al., PNAS 87:648-52 (1990), and Doms et al., J. Virol. 64:3537-40 (1990).

Antibody to this fragment is cross-reactive with CNS tissue, namely an astrocyte protein of 43 kD still to be identified (Yamada et al., J. Virol. 65:1370-6, 1991).

HIV1ENV:576 VPWNjSW

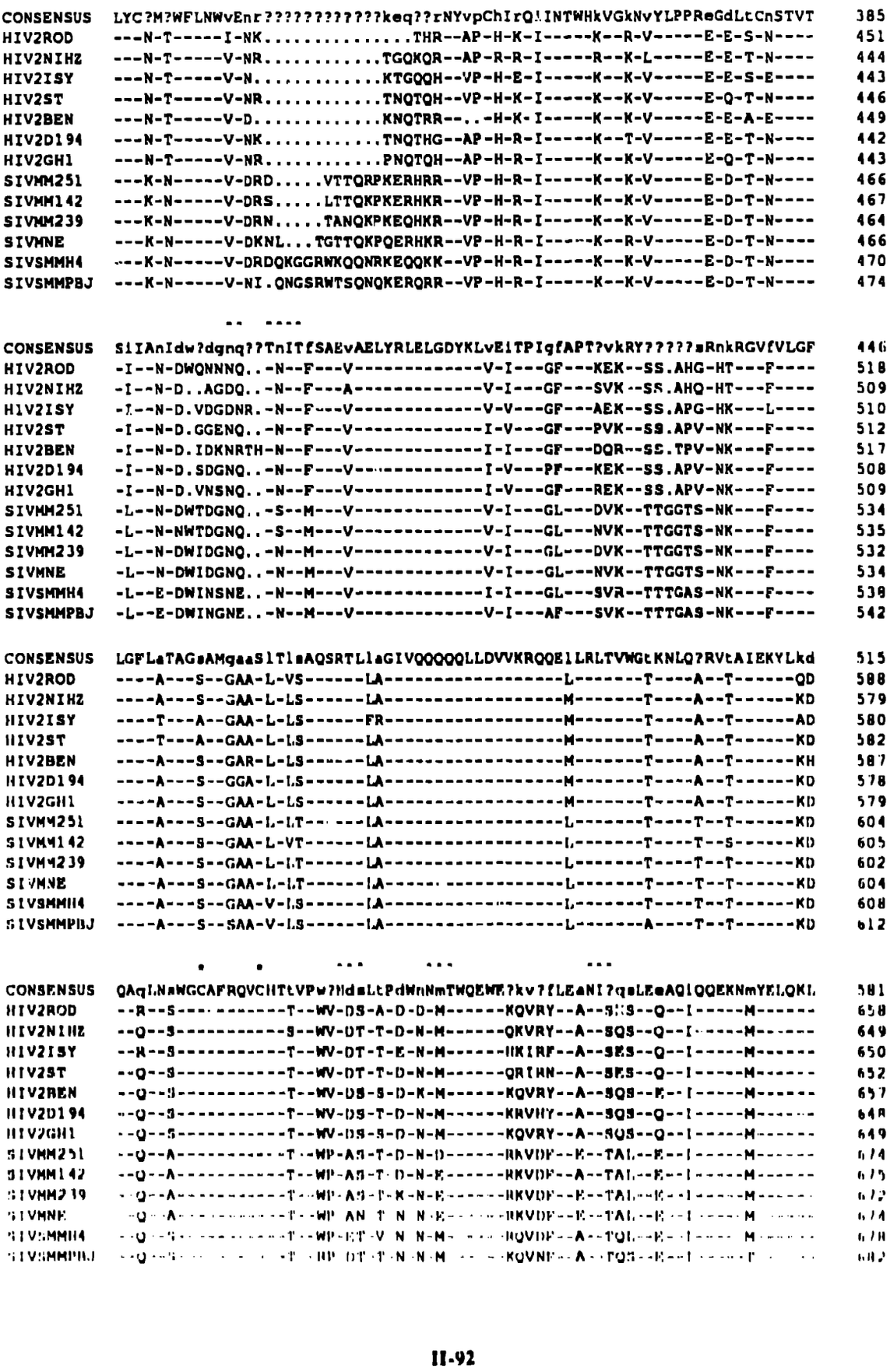
HIV1ENV:660 GLaGLnIaFjaLScaNRVRQGYSPLS

Mutational study -- Helseth et al., J.Virol.64:6314-8, 1990.

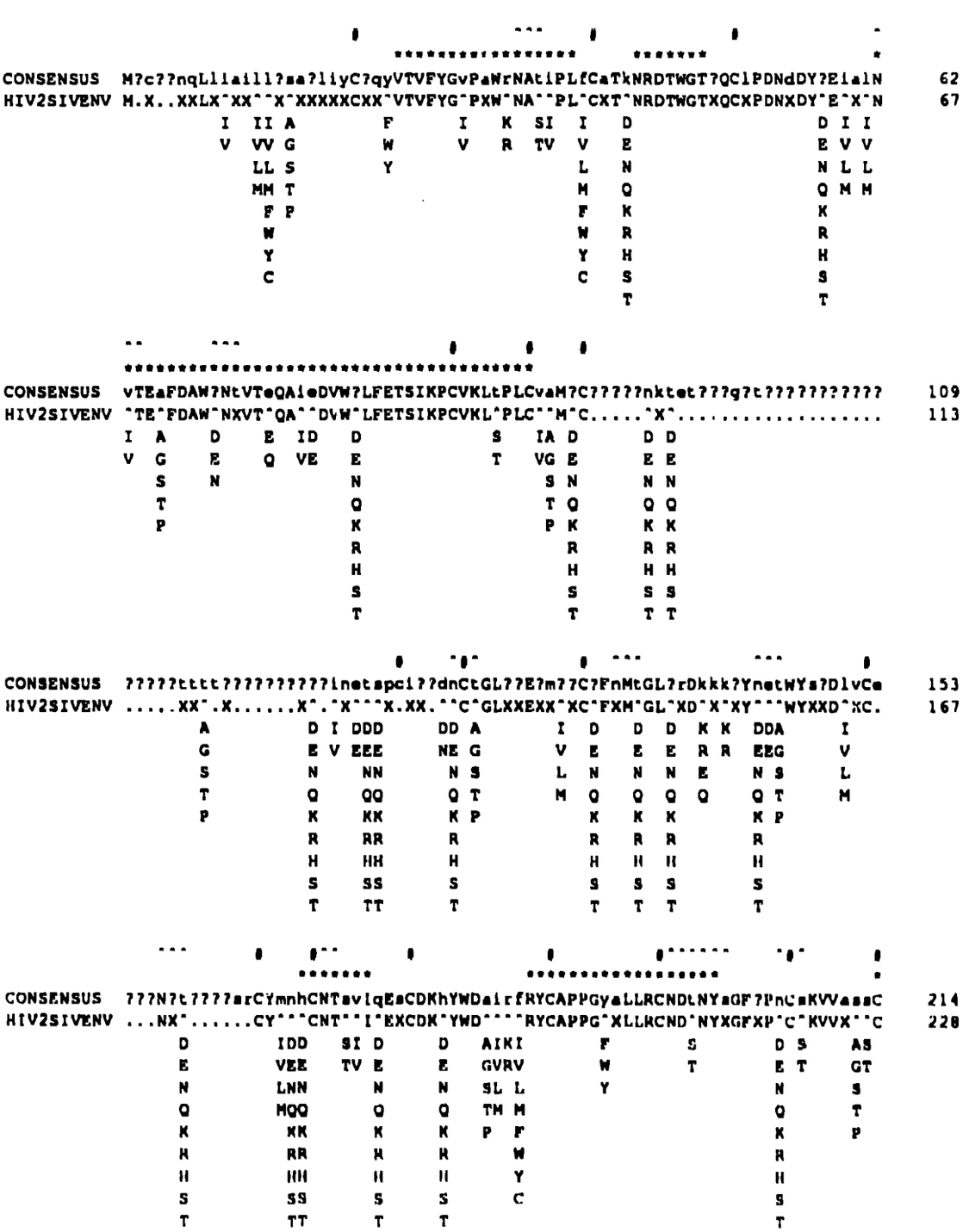
This fragment is similar to a conserved pattern in HIV2, "NRDTWGT".

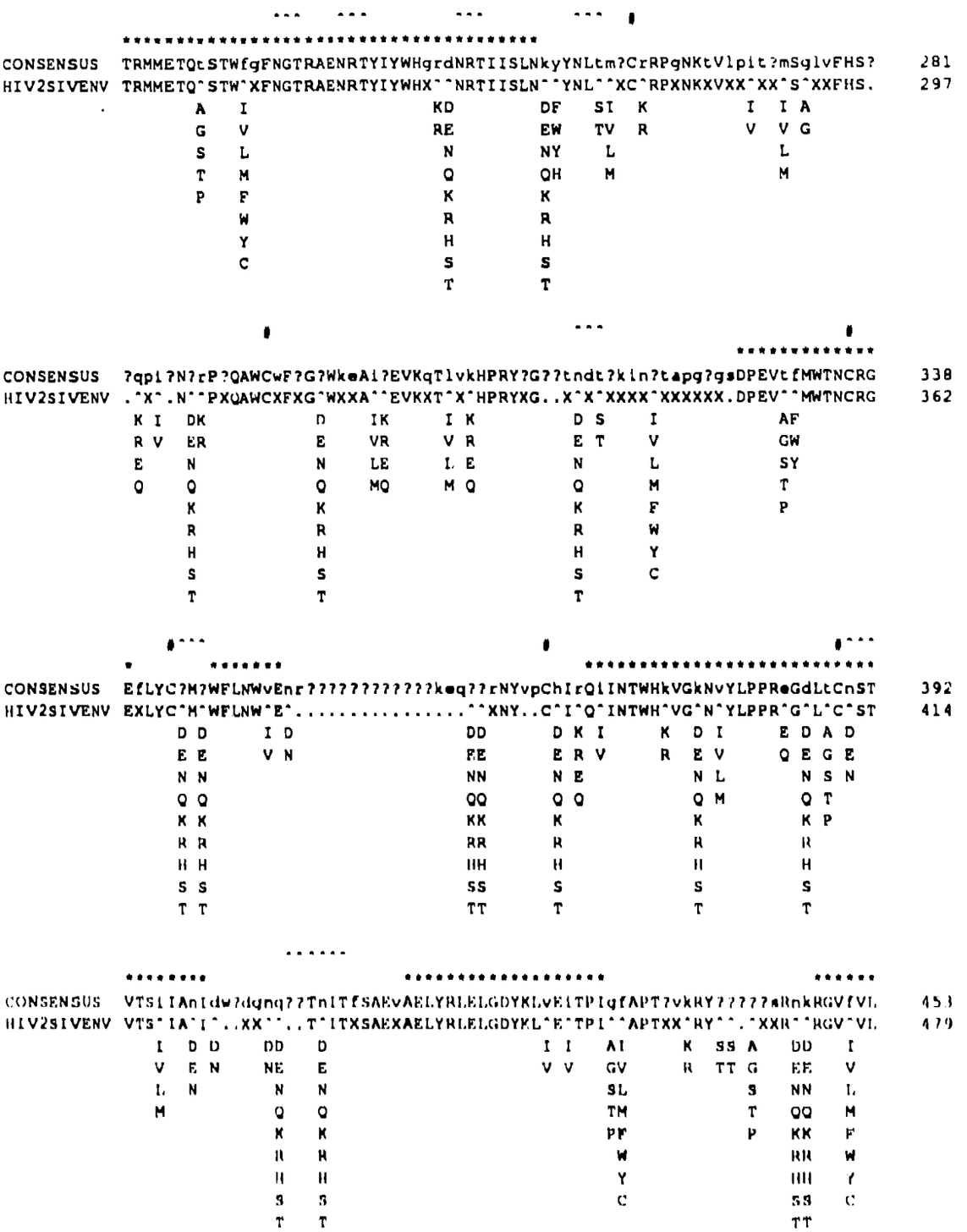
```
                                 *              ***    *           *           .

CONSENSUS  M?c??nqlliaill?aa?liyc?qyvtvfygvpawrnatiplfcatknrdtwgt?qclpdnddyse?alN   62
HIV2ROD    -.M..NQLLIAILLASACLVYCTQYVTVFYGVPTWKNATIPLFCATRNRDTWGTIQCLPDNDDYQEITL-   67
HIV2NIHZ   -KGSKNQLLIAIVLASAYLIHCKQFVTVFYGIPAWRNASIPLFCATKNRDTWGTIQCLPDNDDYQEITL-   70
HIV2ISY    -.SGKIQLLVAFLLTSACLIYCTKYVTVFYGVPVWKNASIPLFCATKNRDTWGTIQCLPDNDDYQEIPL-   69
HIV2ST     -.CGRNQLFVASLLASACLIYCVQYVTVFYGVPVWRNASIPLFCATKNRDTWGTIQCLPDNDDYQEIAL-   69
HIV2BEN    -EPGRNQLFVVILLTSACLVYCSQYVTVFYGIPAWKNASIPLFCATKNRDTWGTIQCLPDNDDYQEIIL-   70
HIV2D194   -EPGRNQLLVAILLTSACLIYCKQYVTVFYGIPAWRNASIPLFCATKNRDTWGTIQCLPDNDDYQEITL-   70
HIV2GH1    -.CGKSLLCVASLLASAYLVYCTQYVTVFYGVPVWRNASIPLFCATKNRDTWGTIQCKPDNDDYQEITL-   69
SIVMM251   -GCLGNQLLIAILLLSVYGIYCTQYVTVFYGVPAWRNATIPLFCATKNRDTWGTTQCLPDNGDYSELAL-   70
SIVMM142   -GCLGNQLLIAILLLSVYGIYCTQYVTVFYGVPAWRNATIPLFCATKNRDTWGTTQCLPDNDDYSELAL-   70
SIVMM239   -GCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFCATKNRDTWGTTQCLPDNGDYSEVAL-   70
SIVMNE     -GCLGNQLLIAILFLSAYGIYCIQYVTVFYGVPAWRNATIPLFCVTRNRDTWGTTQCLPDNDDYSELAL-   70
SIVSMMH4   -GCLGNQLLIALLLVSVLEICCVQYVTVFYGVPAWKNATIPLFCATKNRDTWGTTQCLPDNDDYSELAI-   70
SIVSMMPBJ  -GCLGNQLLIALLLLSASGIYCVQYVTVFYGIPAWRNATVPLFCATKNRDTWGTTQCLPDNGDYSELAI-   70
HIV2D205   -LLSSMAYPHGGTQQFPSFVQPQTETPGELYSVSQTMVTTLRSGTQRLLMHGIIQHNRQMMCGDSLKPP-   70

           **      ***                  *      *    *

CONSENSUS  vteafdaw?ntVteqaiedvw?lfetsikpcvkLtplc?am?c?????n??????????????????t?  103
HIV2ROD    VTEAFDAWNNT-TEQAIEDVWHLFETSIKPCVK-TPLCVAMKC.....SST........ESST.GNNTTS  123
HIV2NIHZ   VTEAFDAWNNT-TEQAVEDVWNLFETSIKPCVK-TPLCVAMNCTR...NMT.....................  118
HIV2ISY    VTEAFDAWDNI-TEQAVEDVWNLFETSIKPCVK-TPLCVTMNC.....NAS.....................  115
HIV2ST     VTEAFDAWNNT-TEQAVEDVWSLFETSIKPCVK-TPLCVAMRC.....NST..............TAKNT  120
HIV2BEN    VTEAFDAWNNT-TEQAVEDVWHLFETSIKPCVK-TPLCVAMNCSRVQGNTT..............TPNPR  126
HIV2D194   VTEAFDAWDNT-TEQAIEDVWRLFETSIKPCVK-TPLCVAMNC.....NIT.....................  116
HIV2GH1    VTEAFDAWDNT-TEQAVEDVWSLFETSIKPCVK-TPLCVAMSC.....NST.................TN  117
SIVMM251   VTESFDAWENT-TEQAIEDVWQLFETSIKPCVK-SPLCITMRC.....NKSETDRWGLTKSST.......  128
SIVMM142   VTESFDAWENT-TEQAIEDVWQLFETSIKPCVK-SPLCITMRC.....NKSETDKWGLTKSST.....TT  130
SIVMM239   VTESF AWNNT-TEQAIEDVWQLFETSIKPCVK-SPLCITMRC.....NKSETDRWGLTKSIT.....TT  130
SIVMNE     ITESFDAWENT-TEQAIEDVWHLFETSIKPCVK-TPLCITMKC.....NKSETDKWGLTKSST.....TT  130
SIVSMMH4   VTEAFDAWDNT-TEQAIEDVWNLFETSIKPCVK-TPLCIAMRC.....NKTETDRWGLTGNAG.....TT  130
SIVSMMPBJ  VTEAFDAWDNT-TEQAIEDVWNLFETSIKPCVK-TPLCITMRC.....NKSETDRWGLTGTPAPTTTQTT  135
HIV2D205   HVSNPHCVWQT-VKPKQTQGMPVVLPPLSLLPP-VGKRLTKQTHAKMTAAQDEKRKCNVILVRDEEMSNN  140

                                  *    **.          *  ***           ***        *

CONSENSUS  ????ttt???????????inetspci???nctgl??e?m??c?fnmtgl?rdkkk?ynetwys?d?vce?  144
HIV2ROD    KSTSTTTTTPTDQ..EQEISEDTPCARADNCSGLGEEETINCQFNMTGLERDKKKQYNETWYSKDVVCET  191
HIV2NIHZ   ....TWTGRTDTQ.NITIINDTSH.ARADNCTGLKEEEMIDCQFSMTGLERDKRKQYTEAWYSKDVVCD.  181
HIV2ISY    TESAVATTSPS...GPDMINDTDPCIQLNNCSGLREEDMVECQFNMTGLELDKKKQYSETWYSKDVVCES  182
HIV2ST     TSTPTTTTTA....NTT.IGENSSCIRTDNCTGLGEEEMVDCQFNMTGLERDKKKLYNETWYSKDVVCES  185
HIV2BEN    TSSSTTS.RPPTS.AASIINETSNC!ENNTCAGLGYEEMMQCEFNMKGLEQDKKRRYKDTWYLEDVVCD.  193
HIV2D194   .SGTTATPSPP...NITIIDENSTCIGDNNCTGLGKEEVVECEFNMTGLEQDKKRKYNDAWYSRDVVCDK  182
HIV2GH1    NTTTTGSTT.....GMSEINETSPSYS.DNCTGLGKEEIVNCQFYMTGLERDKKKQYNETWYSKDVVCES  181
SIVMM251   TITTAAPTSAPVSEKIDMVNETSSCIAQNNCTGLEQEQMISCKFTMTGLKRDKTKEYNETWYSTDLVCEQ  190
SIVMM142   ASTTTTTTAKSVE.TRDIVNETSPCVVHDNCTGLEQEPMISCKFNMTGLKRDKKKEYNETWYSADLVCEQ  199
SIVMM239   ASTTSTTASA....KVDMVNETSSCIAQDNCTGLEQEQMISCKFNMTGLKRDKKKEYNETWYSADLVCEQ  196
SIVMNE     A.PTAIPTKAE...AIKVVNENSPCINHDNCTGLEQEPMISCKFNMTGLKRDKRREYNETWYSADLVCEQ  196
SIVSMMH4   TTAITTTATPSV..AENVINESNPCIKNNSCAGLEQEPMIGCKFNMTGLNRDKKKEYNETWYSHDLICEQ  198
SIVSMMPBJ  TTQASTTPTSPI..TAKVVNDSDPCIKINNCTGLEQEPMVSCKFNMTGLKRDKKREYNETWYSRDLVCEQ  203
HIV2D205   IKTPGTQKISVIIPGSIPADAIEPATQQLSKSHVTNIIGTAGLGIVLPHGFFYDVMIPTIQASCPTAVHR  210
```

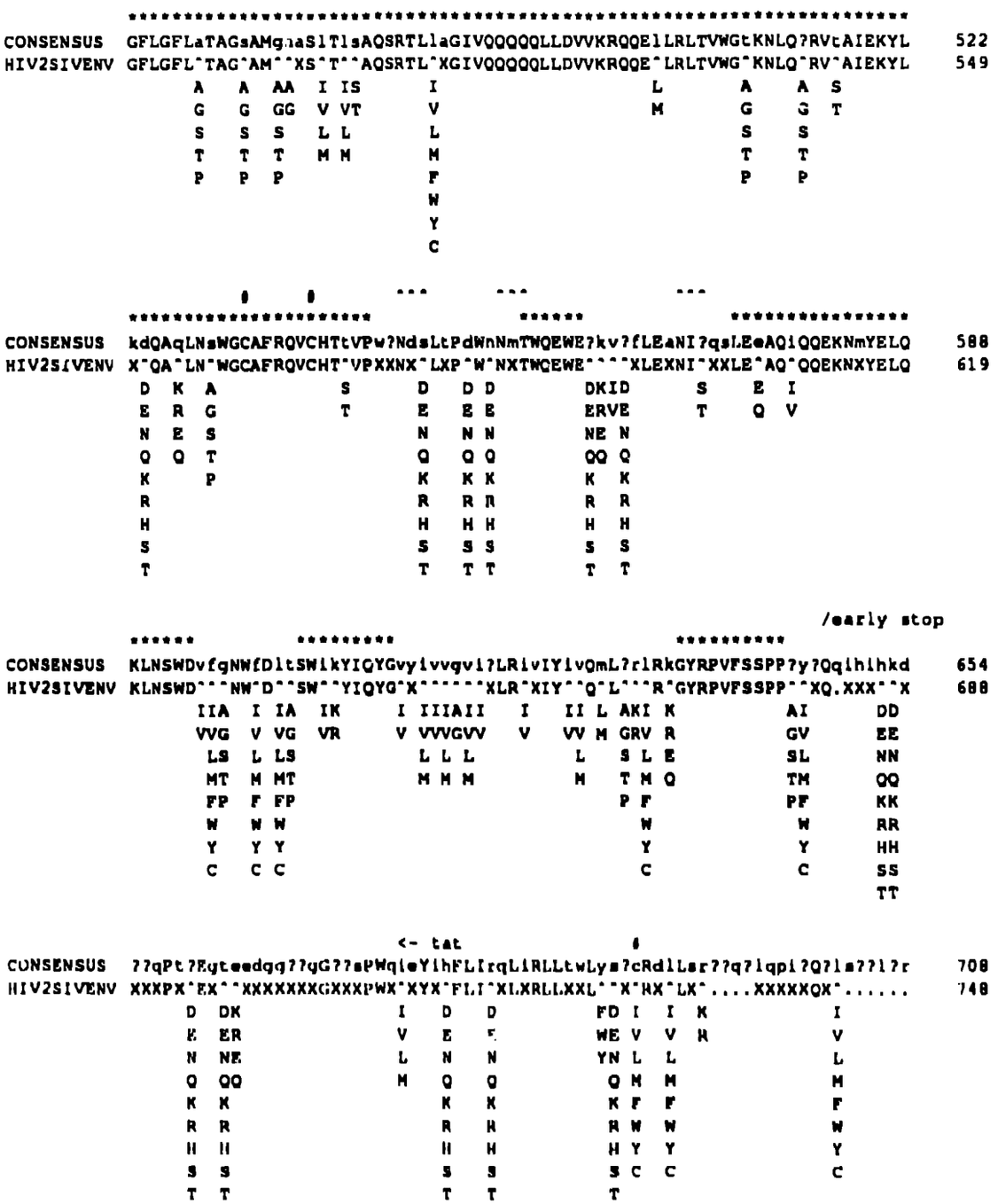
```
                 ^^^       *     *^^^      *              *           *^^^^^^^     ^*^       *
CONSENSUS  ??n?t????s?cymnhcntsvi?escdkhywdairfrycappgyallrcndtnysgf?pnCskvv?ssct  203
HIV2ROD    N.NST.NQTQ.CYMNHCNTSVITESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFAPN-SKVVASTCT  258
HIV2NIHZ   ..NNTSSQ.SKCYMNHCNTSVITESCDKHYWDAMRFRYCAPPGFALLRCNDTNYSGFAPN-SKVVAATCT  248
HIV2ISY    D.NST.D.RKRCYMNHCNTSVITESCDKHYWDAMKFRYCAPPGFVLLRCNDTNYSGFEPN-SKVVASTCT  249
HIV2ST     ..NDT.KKEKTCYMNHCNTSVITESCDKHYWDTMRFRYCAPPGFALLRCNDTNYSGFEPN-SKVVAATCT  252
HIV2BEN    ..NTT...AGTCYMRHCNTSIIKESCDKHYWDAMRFRYCAPPGFALLRCNDTNYSGFEPK-TKVVAASCT  258
HIV2D194   T.NGTG....TCYMRHCNTSVIKESCDKHYWDAMKFRYCAPPGFALLRCNDTNYSGFEPK-SKVVAASCT  247
HIV2GH1    ..NNTKDGKNRCYMNHCNTSVITESCDKHYWDAIKFRYCAPPGYALLRCNDTNYSGFEPK-SKVVASTCT  249
SIVMM251   G.NSTD.NESRCYMNHCNTSVIQESCDKHYWDTIRFRYCAPPGYALLRCNDTNYSGFMPK-SKVVVSSCT  266
SIVMM142   G.NSTG.NESRCYMNHCNTSVIQECCDKDYWDAIRCRYCAPPGYALLRCNDTNYSGFMPN-SKVVVSSCT  267
SIVMM239   G.NNTG.NESRCYMNHCNTSVIQESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPK-SKVVVSSCT  264
SIVMNE     G.NST.ENESRCYMNHCNTSVIQESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPN-SKVVVSSCT  264
SIVSMMH4   SANES...ESKCYMHHCNTSVIQESCDKHYWDAIRFRYCAPPGYALLRCNDSNYLGFAPN-SKVVV?3CT  265
SIVSMMPBJ  NSNET...DSKCYMNHCNTSVIQESCDKHYWDAIRFRYCAPPGYALLRCNDSNYSGFAPN-TKVVVSSCT  270
HIV2D205   PPAQEWKHSPLHGLASMVQGQRTGHIYIGMKKTIGPSAIHTIICQYTVRGQETRRLYQEP-QDYFSIH3L  280

                        ^^^   ^^^         ^^^         ^^^  *
CONSENSUS  rmmetqLstwfgfngtraenRtylywhgrdnrtiislnkyYnltm?crrPgNKtVlp?tiMSglvFHS??  269
HIV2ROD    RMMETQTSTWFGFNGTRAEN-TYIYWHGRDNRTIISLNKY-NLSLHCKR-G--I-KQIML--GHV---HY  328
HIV2NIHZ   RMMETQTSTWFGFNGTRAEN-TYIYWHGKDNRTIISLNNF-NLTMHCKR-G--T-LPITF--GFK---..  316
HIV2ISY    RMMETQPSTWLGFNGTRAEN-TYIYWHGRDNRTIISLNKY-NLTILCRR-E--T-VPITL--GRR---..  317
HIV2ST     RMMETQTSTWFGFNGTRAEN-TYIYWHGRDNRTIISLNKF-NLTVHCKR-G--T-VPITL--GLV---..  320
HIV2BEN    RMMETQTSTWFGFNGTRAEN-TYIYWHGRDNRTIISLNKY-NLTMRCKR-G--T-LPITL--GLV---..  326
HIV2D194   RMMETQTSTWFGFNGTRAEN-TYIYWHGKDNRTIISLNKY-NLTMHCKR-G--T-VPITL--GRR---..  315
HIV2GH1    RMMETQTSTWFGFNGTRAEN-TYIYWHGRDNRTIISLNKY-NLSIHCKR-G--T-VPITL--GLV---..  317
SIVMM251   RMMETQTSTWFGFNGTRAEN-TYIYWHGRDNRTIISLNKY-NLTMKCRR-G--T-LPVTI--GLV---..  334
SIVMM142   RMMETQTSTWFRFNGTRAEN-TYIYWHGRDNRTIISLNKH-NLTMKCRR-G--T-LPVTI--ALV---..  335
SIVMM239   RMMETQTSTWFGFNGTRAEN-TYIYWHGRDNRTIISLNKY-NLTMKCRR-G--T-LPVTI--GLV---..  332
SIVMNE     RMMETQTSTWFGFNGTRAEN-TYIYWHSKDNRTIISLNKY-NLTMKCRR-G--T-LPVTI--GLV---..  332
SIVSMMH4   RMMETQTSTWFGFNGTRAEN-TYIYWHGKSNRTIISLNKY-NLTMRCRR-E--T-LPVTI--GLV---..  333
SIVSMMPBJ  RMMETQTSTWFGFNGTRAEN-TYIYWHGRSNRTIISLNKY-NLTMRCRR-G--T-LPVTI--GLV---..  338
HIV2D205   SIRDPDKLGAGLRETGQKPR-KGPSNIPGIKEVQKISQAS-QNMEKVQI                       329

                       *                        ^^^                        *
CONSENSUS  qpi?N?rPkQAWCwF?G?WkeA?qEVKqTlvkHPRYtGtnd?t?kInftaPqgq?DPEVtfMMTNCRGEf  331
HIV2ROD    QPI.-KR-R----W-K-K-KD-MQ---E-LAK----R-TND.TRN-SFAA-GKGS----AY---------F  396
HIV2NIHZ   QPVI-KK-R----W-E-Q-KE-MQ---E-LAK----K-NRSRTEN-KFKA-GRGS----TY---------S  386
HIV2ISY    QKII-KK-R----R-K-E-RE-MQ---Q-LVK----K-TND.TNK-NFTA-EKDS----AY---------F  386
HIV2ST     QPI.-RR-R----W-K-E-KE-MK---L-LAK----K-TND.TEK-RFIA-GERS----AY---------F  388
HIV2BEN    QPI.-TR-R----R-G-R-RE-MQ---Q-LVQ----K-IND.TGK-NFTK-GAGS----AF---------F  394
HIV2D194   RPVY-KK-G----W-Q-N-IE-MR---Q-LAK----G-TND.TGK-NFTK-GIGS----TY---------F  384
HIV2GH1    QPI.-TR-R----W-K-K-RE-MQ---Q-LIK----K-TND.TKN-NFTK-GRGS----AY---------F  385
SIVMM251   QPI.-DR-K----W-G-K-KD-IK---Q-IVK----T-TNN.TDK-NLTA-GGG.----TF---------F  401
SIVMM142   QPV.-ER-K----R-G-N-KE-IK---Q-IVK----T-TNN.SDK-NLTA-RGG.----TF---------F  402
SIVMM239   QPI.-DR-K----W-G-K-KD-IK---Q-IVK----T-TNN.TDK-NLTA-GGG.----TF---------F  399
SIVMNE     QPI.-DR-K----R-E-N-KE-IK---Q-IVK----T-TNN.TDK-NLTA-GGG.----TF---------F  399
SIVSMMH4   QPI.-ER-K----W-E-S-KK-IQ---E-LVK----T-TND.TRK-NLTA-AGG.----TF---------F  400
SIVSMMPBJ  QPI.-ER-K----W-G-E-KK-IQ---E-LVK----T-TNK.TEQ-KLTA-GGG.----TF---------F  405
```


CONSENSUS LYC?M?WFLNWvEnr??????????????keq??rNYvpChIrQ!INTWHkVGkNvYLPPReGdLtCnSTVT 385
HIV2ROD ---N-T-----I-NK.....THR--AP-H-K-I-----K--R-V-----E-E-S-N---- 451
HIV2NIHZ ---N-T-----V-NR.....TGQKQR--AP-R-R-I-----R--K-L-----E-E-T-N---- 444
HIV2ISY ---N-T-----V-N.....KTGQQH--VP-H-E-I-----K--K-V-----E-E-S-E---- 443
HIV2ST ---N-T-----V-NR.....TNQTQH--VP-H-K-I-----K--K-V-----E-Q-T-N---- 446
HIV2BEN ---N-T-----V-D.....KNQTRR--..-H-K-I-----K--K-V-----E-E-A-E---- 449
HIV2D194 ---N-T-----V-NK.....TNQTHG--AP-H-R-I-----K--T-V-----E-E-T-N---- 442
HIV2GH1 ---N-T-----V-NR.....PNQTQH--AP-H-R-I-----K--K-V-----E-Q-T-N---- 443
SIVMM251 ---K-N-----V-DRD.....VTTQRPKERHRR--VP-H-R-I-----K--K-V-----E-D-T-N---- 466
SIVMM142 ---K-N-----V-DRS.....LTTQKPKERHKR--VP-H-R-I-----K--K-V-----E-D-T-N---- 467
SIVMM239 ---K-N-----V-DRN.....TANQKPKEQHKR--VP-H-R-I-----K--K-V-----E-D-T-N---- 464
SIVMNE ---K-N-----V-DKNL...TGTTQKPQERHKR--VP-H-R-I-----K--R-V-----E-D-T-N---- 466
SIVSMMH4 ---K-N-----V-DRDQKGGRWKQQNRKEQQKK--VP-H-R-I-----K--K-V-----E-D-T-N---- 470
SIVSMMPBJ ---K-N-----V-NI.QNGSRWTSQNQKERQRR--VP-H-R-I-----K--K-V-----E-D-T-N---- 474
CONSENSUS SiIAnIdw?dgnq??TnITfSAEvAELYRLELGDYKLvEITPIqfAPT?vkRY?????sRnkRGVfVLGF 446
HIV2ROD -I--N-DWQNNNQ..-N--F---V----------V-I---GF---KEK--SS.AHG-HT---F---- 518
HIV2NIHZ -I--N-D..AGDQ..-N--F---A----------V-I---GF---SVK--SS.AHQ-HT---F---- 509
HIV2ISY -I--N-D.VDGDNR.-N--F---V----------V-V---GF---AEK--SS.APG-HK---L---- 510
HIV2ST -I--N-D.GGENQ..-N--F---V----------I-V---GF---PVK--SS.APV-NK---F---- 512
HIV2BEN -I--N-D.IDKNRTH-N--F---V----------I-I---GF---DQR--SS.TPV-NK---F---- 517
HIV2D194 -I--N-D.SDGNQ..-N--F---V----------I-V---PF---KEK--SS.APV-NK---F---- 508
HIV2GH1 -I--N-D.VNSNQ..-N--F---V----------I-V---GF---REK--SS.APV-NK---F---- 509
SIVMM251 -L--N-DWTDGNQ..-S--M---V----------V-I---GL---DVK--TTGGTS-NK---F---- 534
SIVMM142 -L--N-NWTDGNQ..-S--M---V----------V-I---GL---NVK--TTGGTS-NK---F---- 535
SIVMM239 -L--N-DWIDGNQ..-N--M---V----------V-I---GL---DVK--TTGGTS-NK---F---- 532
SIVMNE -L--N-DWIDGNQ..-N--M---V----------V-I---GL---NVK--TTGGTS-NK---F---- 534
SIVSMMH4 -L--E-DWINSNE..-N--M---V----------I-I---GL---SVR--TTTGAS-NK---F---- 530
SIVSMMPBJ -L--E-DWINGNE..-N--M---V----------V-I---AF---SVK--TTTGAS-NK---F---- 542
CONSENSUS LGFLaTAGsAMqaaSlTlsAQSRTLlaGIVQQQQQLLDVVKRQQElLRLTVWGtKNLQ?RVtAIEKYLkd 515
HIV2ROD ----A---S--GAA-L-VS------LA----------L-------T----A--T------QD 588
HIV2NIHZ ----A---S--GAA-L-LS------LA----------M-------T----A--T------KD 579
HIV2ISY ----T---A--GAA-L-LS------FR----------M-------T----A--T------AD 580
HIV2ST ----T---A--GAA-L-LS------LA----------M-------T----A--T------KD 582
HIV2BEN ----A---S--GAR-L-LS------LA----------M-------T----A--T------KH 587
HIV2D194 ----A---S--GGA-L-LS------LA----------M-------T----A--T------KD 578
HIV2GH1 ----A---S--GAA-L-LS------LA----------M-------T----A--T------KD 579
SIVMM251 ----A---S--GAA-L-LT------LA----------L-------T----T--T------KD 604
SIVMM142 ----A---S--GAA-L-VT------LA----------L-------T----T--S------KD 605
SIVMM239 ----A---S--GAA-L-LT------LA----------L-------T----T--T------KD 602
SIVMNE ----A---S--GAA-L-LT------LA----------L-------T----T--T------KD 604
SIVSMMH4 ----A---S--GAA-V-LS------LA----------L-------T----T--T------KD 608
SIVSMMPBJ ----A---S--SAA-V-LS------LA----------L-------A----T--T------KD 612
CONSENSUS QAqLNsWGCAFRQVCHTtVPw?HdsLtPdWnNmTWQEWE?kv??fLEsNI?qsLEeAQlQQEKNmYELQKL 581
HIV2ROD --R--S----------T--WV-DS-A-D-D-M------KQVRY--A--S!S--Q--I-----M------ 658
HIV2NIHZ --Q--S----------S--WV-DT-T-D-N-M------QKVRY--A--SQS--Q--I-----M------ 649
HIV2ISY --R--S----------T--WV-DT-T-E-N-M------HKIRF--A--SES--Q--I-----M------ 650
HIV2ST --Q--S----------T--WV-DT-T-D-N-M------QRIRN--A--SES--Q--I-----M------ 652
HIV2BEN --Q--S----------T--WV-DS-S-D-K-M------KQVRY--A--SQS--E--I-----M------ 657
HIV2D194 --Q--S----------T--WV-DS-T-D-N-M------KRVHY--A--SQS--Q--I-----M------ 648
HIV2GH1 --Q--S----------T--WV-DS-S-D-N-M------KQVRY--A--SQS--Q--I-----M------ 649
SIVMM251 --Q--A----------T--WP-AS-T-D-N-D------RKVDF--K--TAL--E--I-----M------ 674
SIVMM142 --Q--A----------T--WP-AS-T-D-N-E------RKVDF--A--TAL--E--I-----M------ 675
SIVMM239 --Q--A----------T--WP-AS-T-K-N-E------RKVDF--E--TAL--E--I-----M------ 672
SIVMNE --Q--A----------T--WP-AN-T-N-N-E------RKVDF--E--TAL--E--I-----M------ 674
SIVSMMH4 --Q--S----------T--WP-ET-V-N-N-M------KQVDF--A--TQL--E--I-----M------ 670
SIVSMMPBJ --Q--S----------T--WP-DT-T-N-N-M------KQVNF--A--TQS--E--I-----T------ 682

```
                                                                  /early stop
CONSENSUS  NSWDvfgNWfDltSWikYIQYGvyivvgvi?LRivIYivQmL?rlRkGYRPVFSSPP?y?Qqihihkd??    645
HIV2ROD    ----IFG--F-LT--VK-----VLIIVAVIA--IV--VV-M-SRL-K----------GYI-QIHIHKDRG    728
HIV2NIHZ   ----VFT--L-FT--VR-----VYVVVGIVA--IV--IV-M-SRL-K----------GYI-QIHIHKDQE    719
HIV2ISY    ----VFG--F-LT--IK-----VMIVVGIVA--IV--VV-M-SRL-K----------GYI-QIHIHKDWE    720
HIV2ST     ----VFG--F-LT--IK-----VYIVVGIIV--IV--VV-M-SRL-K----------AYF-QIHIHKDRE    722
HIV2BEN    ----ILG--F-LT--VK-----VHIVVGIIA--IA--VV-L-SRF-K----------GYL-QIHIHKDRG    727
HIV2D194   ----VFG--F-LT--IK-----VYIVVGIIG--IA--IV-L-SRL-K----------GYL-QIHIHTDRG    718
HIV2GH1    ----VFG--F-LT--IK-----VYIVVGVIV--IA--IV-L-SRL-K----------GYL-QIHIHTDRG    719
SIVMM251   ----VFG--F-LA--IK-----IYVVVGVIL--IV--IV-M-AKL-Q----------SYF-*          735
SIVMM142   ----VFG--F-LT--IK-----IYIIVGVIL--IV--IV-M-ARL-Q----------SYF-*          736
SIVMM239   ----VFG--F-LA--IK-----VYIVVGVIL--IV--IV-M-AKL-Q----------SYF-QTHIQQDPA  742
SIVMNE     ----VFG--F-LA--IR-----VYIVVGVIL--IV--IV-M-AKL-Q----------SYF-XTHIRQDQA  744
SIVSMMH4   ----IFG--F-LT--IR-----VLIVLGVIG--IV--VV-M-ARL-Q----------AYV-QIPIHKGQE  748
SIVSMMPBJ  ----IFG--F-LT--IK-----VLIVLGVIG--IV--VV-M-ARL-Q----------AYV-QIPIQTGQE  752

                                <- tat cds          *
CONSENSUS  qp??egteed?g??gg??swpw?i?yihflirqlirllt?ly??crdllsr????lq?i?q?l???????    689
HIV2ROD    QPANEETEEDGGSNGGDRYWPWPIAYIHFLIRQLIRLLTRLYSICRDLLSRSFLTLQLIYQNL.......    791
HIV2NIHZ   QPAREETEEDVGSNGGDRSWPWPIAYIHFLIRLLIRLLTGLYNICRDLLSRISPILQPIFQSLQRALTAI    789
HIV2ISY    QPDREETEEDVGNDVGSRSWPWPIEYIHFLIRLLIRLLTRLYNSCRDLLSR....LYLILQPL.......    779
HIV2ST     QPAREETEEDVGNSVGDNWWPWPIRYIHFLIRQLIRLLNRLYNICRDLLSRSFQTLQLISQSLRRALTAV    792
HIV2BEN    QPANEGTEEDVGGDSGYDLWPWPINYVQFLIHLLTRLLIGLYNICRDLLSKNSPTRRLISQSL....TAI    793
HIV2D194   QPANEETEEDAGDDSGFGLWPWPLNYIQFLIHLLTRLLTGLYNSCRGLLSKNSPTRRLISQSL....TAI    784
HIV2GH1    QPANEGTEEDDRDDDGYDL*                                                     738
SIVMM239   LPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQILQPILQRLSATLQRI    812
SIVMNE     LPTKEGTEGDGGGSGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRAYQILQPIFQRFSTTLQRI    814
SIVSMMH4   PPTKEGEEGDGGDRGGSRSWPWQIEYIHFLIRQLIRLLTWLFSSCRDWLLRSYQILQPVLQSLSTTLQRV    818
SIVSMMPBJ  LPTKEGEEGDGGGRGGNRSWPWQIEYIHFLIRQLIRLLTWLFSSCRDWLLRNCQTLQPVLQSLSRTLQRA    822

                 nef ->
               <- rev cds
                    *
CONSENSUS  r??lr???aylqyg??wlqea?qa??r??retla?awr?lwe?l?r?grgila?prrirqg?el?ll*     736
HIV2ROD    RDWLRLRTAFLQYGCEWIQEAFQAAARATRETLAGACRGLWRVLERIGRGILAVPRRIRQGAEIALL-     858
HIV2NIHZ   RDWLRLKAAYLQYGCEWIQEAFQALARTTRETLAGAGRDLWRALQRIGRGILAVPRRIRQGAELALL-     856
HIV2ISY    RDWLRLKAAYLQYGCEWIQEAFQALARVTRETLTSAGRSLWGALGRIGRGILAVPRRIRQGAEIALL-     846
HIV2ST     RDWLRFNTAYLQYGGEWIQEAFRAFARATGETLTNAWRGFWGTLGQIGRGILAVPRRIRQGAEIALL-     859
HIV2BEN    RDWLRLKAAQLQYGCEWIQEAFQAFARTTRETLAGAWGWLWEAARRIGRGILAVPRRIRQGAELALL-     860
HIV2D194   RDWLRLKAAYLQYGCEWIQEAFRAFARTARETLAGAWRGLCEAAQRIGRGILAVPRRIRQGAEIALL-     851
SIVMM239   REVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETLRRGGRWILAIPRRIRQGLELTLL-     879
SIVMNE     REVLRTELTYLQYGWSYFQEAVQVAWRSATETLAGAWGDLWETLGRVGRWILAIPRRIRQGLELTLL-     881
SIVSMMH4   REVIRIEIAYLQYGWRYFQEAVQAWWKLARETLASAWGDIWETLGRVGRGILAIPRRIRQGLELTLL-     885
SIVSMMPBJ  REVIRVQIAYLQYGWRYLQEAAQAWWKFVRETLASAWRDLWETLGRVGRGILAIPRRIRQGLELTLL-     889
```


Information dense stretches (0.8 or greater) are indicated by ****.
CONSENSUS M?c??nqLliaill?sa?liyC?qyVTVFYGvPaWrNAtiPLfCaTkNRDTWGT?QClPDNdDY?EialN 62
HIV2SIVENV M.X..XXLX^XX^^X^XXXXXCXX^VTVFYG^PXW^NA^^PL^CXT^NRDTWGTXQCXPDNXDY^E^X^N 67
CONSENSUS vTEaFDAW?NtVTeQAieDVW?LFETSIKPCVKLtPLCvaM?C?????nktet???g?t??????????? 109
HIV2SIVENV ^TE^FDAW^NXVT^QA^^DVW^LFETSIKPCVKL^PLC^^M^C.....^X^..... 113
CONSENSUS ?????tttt???????????inetspci??dnCtGL??E?m??C?FnMtGL?rDkkk?YnetWYs?DlvCe 153
HIV2SIVENV .....XX^.X.....X^.^X^^^X.XX.^^C^GLXXEXX^XC^FXM^GL^XD^X^XY^^^WYXXD^XC. 167
CONSENSUS ???N?t????arCYmnhCNTsvlqEsCDKhYWDairfRYCAPPGyaLLRCNDtNYsGF?PnCsKVVasaC 214
HIV2SIVENV ...NX^.....CY^^^CNT^^I^EXCDK^YWD^^^^RYCAPPG^XLLRCND^NYXGFXP^C^KVVX^^C 228


CONSENSUS TRMMETQtSTWfgFNGTRAENRTYIYWHgrdNRTIISLNkyYNLtm?CrRPgNKtVlpit?mSglvFHS? 281
HIV2SIVENV TRMMETQ^STW^XFNGTRAENRTYIYWHX^^NRTIISLN^^YNL^^XC^RPXNKXVXX^XX^S^XXFHS. 297
CONSENSUS ?qpi?N?rP?QAWCwF?G?WkeAi?EVKqTlvkHPRY?G??tndt?kin?tapg?gsDPEVtfMWTNCRG 338
HIV2SIVENV .^X^.N^^PXQAWCXFXG^WXXA^^EVKXT^X^HPRYXG..X^X^XXXX^XXXXXX.DPEV^^MWTNCRG 362
CONSENSUS EfLYC?M?WFLNWvEnr????????????keq??rNYvpChIrQiINTWHkVGkNvYLPPReGdLtCnST 392
HIV2SIVENV EXLYC^M^WFLNW^E^.....^^XNY..C^I^Q^INTWH^VG^N^YLPPR^G^L^C^ST 414
CONSENSUS VTSiIAnIdw?dqnq??TnITfSAEvAELYRLELGDYKLvEITPIqfAPT?vkRY?????sRnkRGVfVL 453
HIV2SIVENV VTS^IA^I^..XX^^..T^ITXSAEXAELYRLELGDYKL^E^TPI^^APTXX^RY^^.^XXR^^RGV^VL 479


CONSENSUS GFLGFLaTAGsAMgnaSlTlsAQSRTLlaGIVQQQQQLLDVVKRQQElLRLTVWGtKNLQ?RVtAIEKYL 522
HIV2SIVENV GFLGFL^TAG^AM^^XS^T^^AQSRTL^XGIVQQQQQLLDVVKRQQE^LRLTVWG^KNLQ^RV^AIEKYL 549
CONSENSUS kdQAqLNsWGCAFRQVCHTtVPW?NdsLtPdWnNmTWQEWE?kv?fLEaNI?qsLEeAQiQQEKNmYELQ 588
HIV2SIVENV X^QA^LN^WGCAFRQVCHT^VPXXNX^LXP^W^NXTWQEWE^^^^XLEXNI^XXLE^AQ^QQEKNXYELQ 619
/early stop
CONSENSUS KLNSWDvfgNWfDltSWikYIQYGvyivvqvi?LRivIYivQmL?rlRkGYRPVFSSPP?y?Qqihihkd 654
HIV2SIVENV KLNSWD^^^NW^D^^SW^^YIQYG^X^^^^^^XLR^XIY^^Q^L^^^R^GYRPVFSSPP^^XQ.XXX^^X 688
<- tat
CONSENSUS ??qPt?Egteedgq??qG??sPWqieYihFLIrqLiRLLtwLys?cRdlLsr??q?lqpi?Q?ls??l?r 708
HIV2SIVENV XXXPX^EX^^XXXXXXXGXXXPWX^XYX^FLI^XLXRLLXXL^^X^HX^LX^....XXXXXQX^..... 748

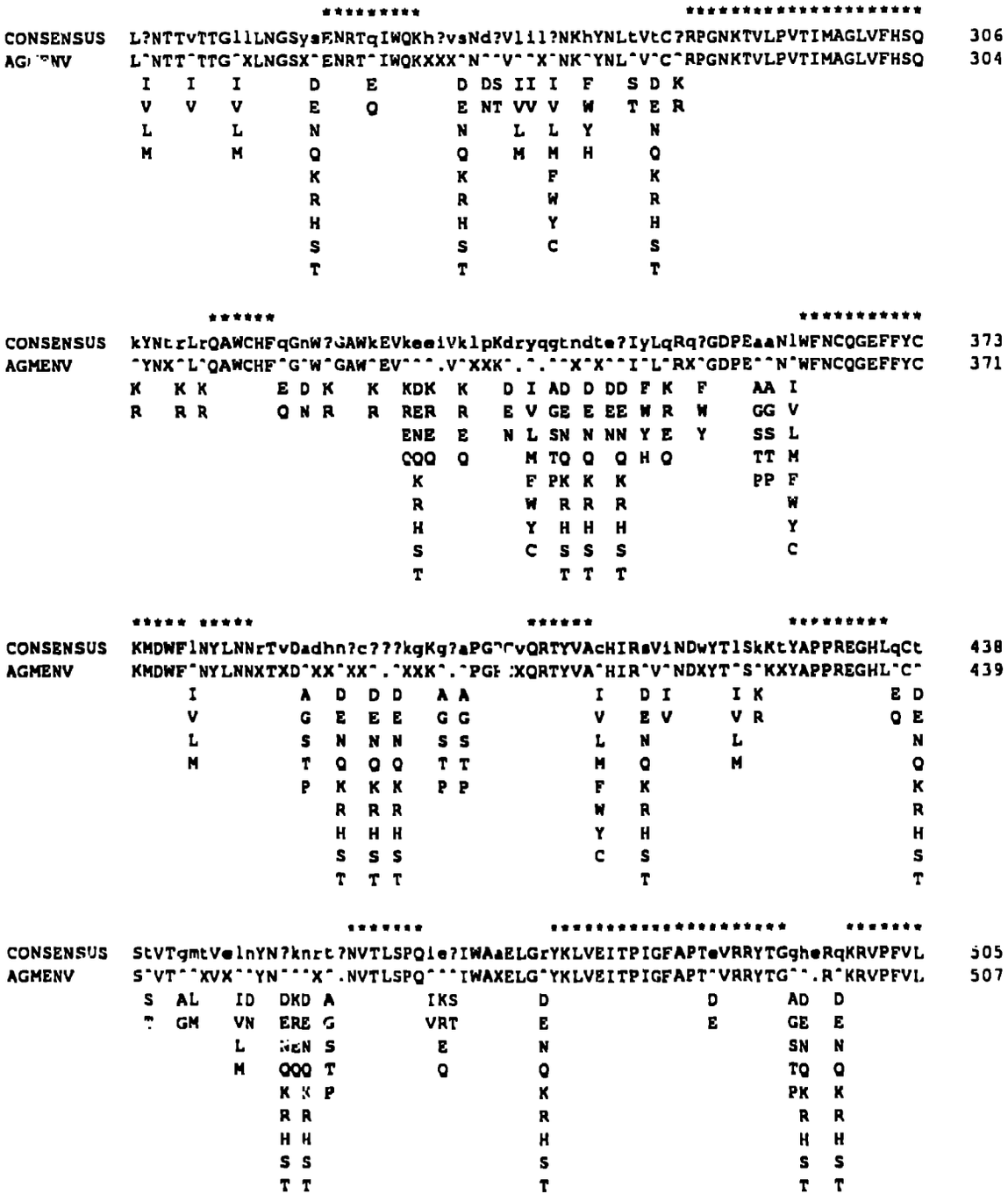
```
                  nef ->
               <- rev   #
                                                                  *********
CONSENSUS  iRevlRl??ayLQYG??wiqEA?qa??r?arETlagAwrdlwetl?r?GRgIlAiPRRIRQGlEltLL*     766
HIV2SIVENV .R^^^RXXX^XLQYGXX^^^EAX^XXX^XXXET^^XAXXX^^XXXX^XGRXIXA^PRRIRQGXE^^LL*     815
             DII    A       FID   K   K     IA     II    K       I         IA
             EVV    G       WVE   R   R     VG     VV    R       V         VG
              LL    S       YLN   E         LS     LL    E                 LS
              MM    T        MQ   Q         MT     MM    Q                 MT
              F     P        FK              P     FF                       P
              W              WR                    WW
              Y              YH                    YY
              C              CS                    CC
                              T
```

```
CONSENSUS  M??tli?igi??IGIgi??stkqqw?TVFYG?PVWKNSsVQAFCMTPtTrlWATTNcIPDDHDyTEVPLN   61
SIVAGMTYO  -RYTIITLGIIV---GIVLS.K.QWI-----I------S--------T-SL-----C------Y------  68
SIVAGM155  -TKFLGIFIVLG---GIGISTKQQWI-----V------S--------T-RL-----C------Y------  70
SIVAGM3    -KLTLL.IGILL---GVVLNTRQQWV-----V------S--------T-RL-----S------Y------  69
SIVAGM677  -GRLLIKILIIA---SIGI.G.NLYV-----I------T--------N-NM-----C------N------  68

CONSENSUS  ITEpFEAWaDrNPLvaQAgSNIHLLFEQT?kPCVKLSPlCIKM?CVELn??a?????ptttpksT?ta?t  120
SIVAGMTYO  ---P----G-R---IA--A----------MK-------L----N----NSTRERATTPTTTPKS-.....  133
SIVAGM155  ---P----A-R---VA--G----------LK-------L----N----KGSA..TSTPAT...S-.TAGT  134
SIVAGM3    ---P----A-R---VA--G----------LK-------L----S----NSSE.....PTTTPKS-.TAST  133
SIVAGM677  ---A----.-.---VK--E----------MR-------I----S----NGTA........TTKA-TTATT  128

CONSENSUS  t???????tlPCv?nkt??nlqsCn?tiIEkEmn?EpASNCTFAmAGYvRDqKKn?YSvvWnDaEi?Ck?  175
SIVAGMTYO  ........GL--VGPTSGENLQS-NASI--R-MED-P-------M---V--Q--NY--VV-N-A-IY-KN  195
SIVAGM155  K........L--VRNKTDSNLQS-NDTI--K-MND-A-------M---I--Q--N.--VV-N-A-IF-K.  194
SIVAGM3    TNITASTTTL--VQNKTSTVLES-NETI--K-LNE-P-------M---V--Q--K.--VV-N-A-IM-K.  201
SIVAGM677  T.MT....T.--.........QN-STEQ--G-MAE-P-------I---Q--V--N.--MT-Y-Q-LV-NN  182

CONSENSUS  ktns??gskeCYMIHCNDSVIKEACDKTYWDeLRlRYCAPAGyALLKCND?DYaGfkpnCsNVSVVHCT?  241
SIVAGMTYO  KTNS..TSKE--------------------Q--L-------Y-------E--N-YKQN-S--------G  263
SIVAGM155  RSTSHNGTKE--------------------E--L-------Y-------W--A-FKPE-S--------T  264
SIVAGM3    KGNN..SNRE--------------------E--L-------F-------Y--A-FKTN-S--------N  269
SIVAGM677  KTGSEKGSKD--------------------T--V-------Y-------K--R-FAPK-K--------R  252

CONSENSUS  L?NTTvTTGllLNGSysENRTqIWQKh?vsNd?Vlil?NKhYNLtVtC?RPGNKTVLPVTIMAGLVFHSQ  306
SIVAGMTYO  -M---V---LL----YH----Q----HRVN-NT-LILF--H---S-T-R--------------------  333
SIVAGM155  -M---V---LL----YS----Q----HGVS-DS-LILL--H---T-T-K--------------------  334
SIVAGM3    -I---V---LL----YS----Q----HRVS-DS-LVLF--H---T-T-K--------------------  339
SIVAGM677  -I---I---IG----RS----E----GGND-DT-IIKL--F---T-R-R--------------------  322

CONSENSUS  kYNtrLrQAWCHFqGnW?GAWkEVkeelVklpKdryqgtndte?IyLqRq?GDPEaaNlWFNCQGEFFYC  373
SIVAGMTYO  K--MK-R------E-N-R---R--KQKI-ELP-DRYKGTNNTEH-Y-Q-QW----AS-L----------  403
SIVAGM155  K--TR-R------Q-N-K---K--QEEI-KLP-ERYQGTNDTNK-F-Q-QF----AA-L----------  404
SIVAGM3    R--TR-R------Q-N-R---K--KNEI-KLP-DRYQGTNDTEE-Y-Q-LF----AA-L----------  409
SIVAGM677  K--TR-K------Q-D-K---K--REE.-KKV-N.L.TEVSIEN-H-R-IW----SA-F----------  389

CONSENSUS  KMDWFlNYLNNrTvDadhn?c???kgKg?aPGPCvqNTYVAcHIRsVlNDwYTlskKtYAPPREGHLqCt  4?8
SIVAGMTYO  -----L-----K-W-AYHNFCS.SKK-GHA----V------Y---S-I--S--L-K-T---------Q-R  472
SIVAGM155  -----L-----L-V-ADHNHCKNNAG-GRS----V------C---S-I--W--I-K-T---------Q-T  474
SIVAGM3    -----L-----R-V-PDHNPCNGTKG-GKA----A------C---S-I--W--L-R-T---------Q-T  479
SIVAGM677  -----I-----R-E-AEGTNRTCDKG-P.G----V------C---Q-V--W--V-K-V---------E-N  458

CONSENSUS  StVTqmtVelnYN?knrt?NVTLSPQle?IWAaELGrYKLVEITPIGFAPTeVRRYTGqheRqKRVPFVL  505
SIVAGMTYO  -T--GMT-ELN--SKNRT.-------IES---A---R--------------E------GHE-Q-------  541
SIVAGM155  -T--GMT-ELN--NQNRT.-------IET---A---R--------------E------GQE-Q------   543
SIVAGM3    -T--GMS-ELN--SKNRT.-------IET---A---R--------------E------GHD-T-------  548
SIVAGM677  -S--ALY-AID--NKSGPI-------VRS---Y---D--------------D------PT.-E-------  527

CONSENSUS  GFLGFLGAAGTAMGAAAtaLTVQS?HLLAGILQQQKNLLAAVeaQQQmLKLTIWGVKNLNARVTALEKYL  574
SIVAGMTYO  ----------------SS-----R----------------EA---M----------------------  611
SIVAGM155  ----------------TA-----Q----------------GA---M----------------------  613
SIVAGM3    ----------------TA-----Q----------------EA---M----------------------  618
SIVAGM677  ----------------TT-----R----------------EQ---L----------------------  597

CONSENSUS  eDQARLN?WGCAWKQVCHTTVpW?wnN?TPdWnNMTWLEWErQI??LEqNIT?qLeeAreQEeKNLDaYQ  638
SIVAGMTYO  E------S----------E-PWT-R--D-Q--------R--AD--S---GQ-VK-RE--E----A--  681
SIVAGM155  A------A----------P-TWN-.--E-N--------K--EG--G---KQ-EQ-RE--E----A--  682
SIVAGM3    E------A----------P-QWN-R--D-N--------R--SY--G---TQ-EE-RA--E----A--  688
SIVAGM677  E------S----------P-KYN-.--K-D--------R--NA--G---QL-EE-QN--G----L--  666
```

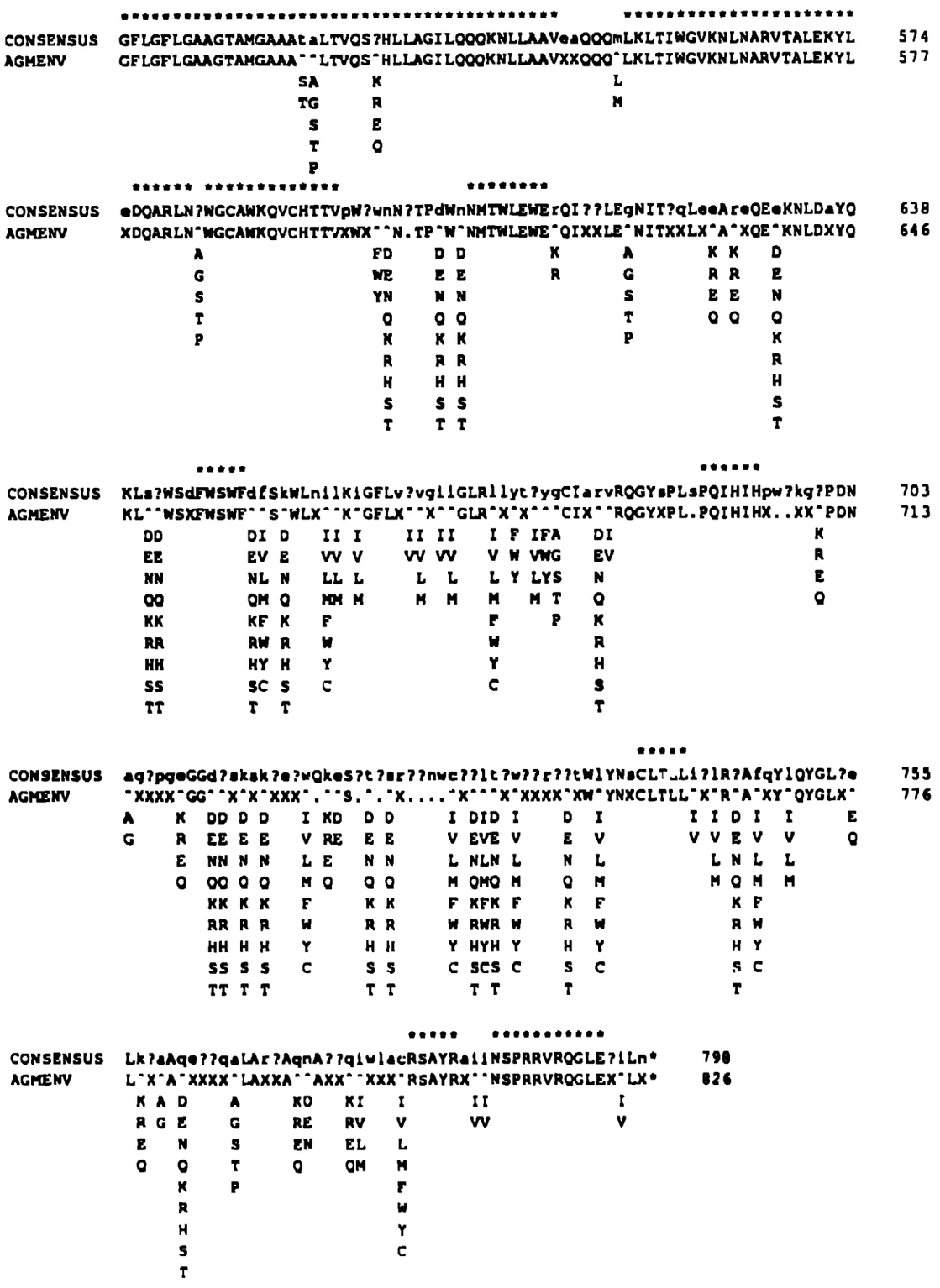
```
CONSENSUS  KLs?WSdFWSWFdfSkWLnilKiGFLv?vgiiGLRllyt?ygCIarvRQGYsPLsPQIHIHpw?kg?PDN   703
SIVAGMTYO  --TS--D-----DF-K--NIL-M---VIVGII---LLYTVYG--VRV----V--S------QVGKGR---   751
SIVAGM155  --SD--S-----DF-K--NIL-I---AVIGVI---LLYTLYT--ARV----S--S------PW.KGQ---   751
SIVAGM3    --SS--D-----DF-K--NIL-I---DVLGII---LLYTVYS--ARV----S--S------PW.KGQ---   757
SIVAGM677  --DD--G-----SL-T--GYV-I---VIVIIL---FAWVLWG--RNI----N--.------S.SA?R---   734

CONSENSUS  ag?pgeGGd?sksk?e?wQkeS?t?sr??nwc??lt?w??r??tWlYNsCLTLLi?lR?AfqYlQYGL?e   755
SIVAGMTYO  ADEPGE--DNSRIKLESW KD-.K.SR....CMQLTAWLTRLNT-L--S-----IQL-K-FQ-L----AE   814
SIVAGM155  AGEPEE--RTGKSKSTH*                                                      768
SIVAGM3    AEGPGE--DKRKNSSEPW-KE-GTAEWKSNWCKRLTNWCSISSI-L--S-----VHL-S-FQ-I----GE   827
SIVAGM677  GGGQDR--ESSSSKLIRL-EE-STPSRINNWWLNFKSCSLRIRT-C--I-----IFI-T-VG-L----QQ   804

CONSENSUS  Lk?aAqe??qaLAr?AqnA??qiwlacRSAYRaiiNSPRRVRQGLE?iLn*     798
SIVAGMTYO  -KTG-QEILQT--GV-QN-CHQIWLAC-----NIV----------EI-N-     864
SIVAGM3    -KAA-QEAVVA--RL-QN-GYQIWLAC-----AII----------GI-N-     877
SIVAGM677  -QEA-TGLAQA--RA-RE-WGRLGAIV-----AVI----------KV-G-     854
```

Information dense stretches (0.9 or greater) are indicated by ****.

```
                                    *********************    ***** ****** ******
CONSENSUS  M??tli?igi??IGIgi??stkqqw?TVFYG?PVWKNSsVQAFCMTPtTrlWATTNcIPDDHDyTEVPLN    61
AGMENV     MXXX^X.^X^^XIGI^^X^....X^^TVFYG^PVWKNS^VQAFCMTP^T^^WATTNXIPDDHDXTEVPLN    65
               I  I II    AI I     FI     I      S        D DL
               V  V VV    GV V     WV     V      T        E EM
               L  L  L    S  L     Y                      N N
               M  M  M    T  M                            Q Q
                  F       P                               K K
                  W                                       R R
                  Y                                       H H
                  C                                       S S
                                                          T T

           ***                 **********  *******
CONSENSUS  ITEpFEAWaDrNPLvaQAgSNIHLLFEQT?kPCVKLSPlCIKM?CVELn??a?????ptttpksT?ta?t   120
AGMENV     ITE^FEAW.D.NPL^XQAXSNIHLLFEQT^^PCVKLSP^CIKM^CVEL^^^X...........^T.....   117
              A          I              LK       I    D    DAS            A
              G          V              MR       V    E    EGT            G
              S                                  L    N    NS             S
              T                                  M    Q    QT             T
              P                                       K    KP             P
                                                      R    R
                                                      H    H
                                                      S    S
                                                      T    T

                                                 *******
CONSENSUS  t???????tlPCv?nkt??nlqsCn?tiIEkEmn?EpASNCTFAmAGYvRDqKKn?YSvvWnDaEi?Ck?   175
AGMENV     ..........PC.........^^C^X^XIEXE^X^E^ASNCTFA^AGYXRDXKK^.YS^XWXDXE^^C^.   166
                                ED D D     L D A       I         D   I      II D
                                QE E E     M E G       V         E   V      VV E
                                 N N N         S       L         N   L      LL N
                                 Q Q Q         T       M         Q   M      MM Q
                                 K K K         P                 K           F K
                                 R R R                           R           W R
                                 H H H                           H           Y H
                                 S S S                           S           C S
                                 T T T                           T             T

                     *********************    ***************          ********
CONSENSUS  ktns??gskeCYMIHCNDSVIKEACDKTYWDeLRlRYCAPAGyALLKCND?DYaGfkpnCsNVSVVHCT?   241
AGMENV     ^^X^..^^^^CYMIHCNDSVIKEACDKTYWD^LR^RYCAPAG^ALLKCNDXDYXG^XX^C^NVSVVHCTX   234
           KA D  ADKD                     D  I       F            F  D D
           RG E  GERE                     E  V       W            W  E E
            S N  SN                       N  L       Y            Y  N N
            T Q  TQ                       Q  M                       Q Q
            P K  PK                       K                          K K
              R   R                       R                          R R
              H   H                       H                          H H
              S   S                       S                          S S
              T   T                       T                          T T
```



AGMENV


CONSENSUS GFLGFLGAAGTAMGAAAtaLTVQS?HLLAGILQQQKNLLAAVeaQQQmLKLTIWGVKNLNARVTALEKYL 574
AGMENV GFLGFLGAAGTAMGAAA""LTVQS"HLLAGILQQQKNLLAAVXXQQQ"LKLTIWGVKNLNARVTALEKYL 577
CONSENSUS eDQARLN?NGCAWKQVCHTTVpW?wnN?TPdWnNMTWLEWErQI??LEgNIT?qLeeAreQEeKNLDaYQ 638
AGMENV XDQARLN"NGCAWKQVCHTTVXWX""N.TP"W"NMTWLEWE"QIXXLE"NITXXLX"A"XQE"KNLDXYQ 646
CONSENSUS KLa?WSdFWSWFdfSkWLnilKiGFLv?vgiiGLRllyt?ygCIarvRQGYsPLsPQIHIHpw?kq?PDN 703
AGMENV KL""WSXFWSWF""S"WLX""K"GFLX""X""GLR"X"X""CIX""RQGYXPL.PQIHIHX..XX"PDN 713
CONSENSUS aq?pqeGGd?skak?e?wQkeS?t?sr??nwc??lt?w??r??tWlYNsCLTLLi?lR?AfqYlQYGL?e 755
AGMENV "XXXX"GG""X"X"XXX"."S."."X...."X""X"XXXX"XW"YNXCLTLL"X"R"A"XY"QYGLX" 776
CONSENSUS Lk?aAqe??qaLAr?AqnA??qiwlacRSAYRaiiNSPRRVRQGLE?iLn* 798
AGMENV L"X"A"XXXX"LAXXA""AXX""XXX"RSAYRX""NSPRRVRQGLEX"LX* 826

```
CONSENSUS1 MGGKWSK?s??gWpavReRMrrae????????????EPAAdGVGAvSRD.....LeKHGaiTSSNTaat  49
HIVLAI     -------SSVVG-PTV-E--RRA.............----D----A---.....-E---AI-----AAT  51
HIVNL43    -------SSVIG-PAV-E--RRA.............----D----V---.....-E---AI-----AAN  51
HIVBRVA    -------..MAG-STV-E--RRAEPARERMRRAEPRA----D----V---.....-E---AI-----AAT  63
HIVSC      -------RSVVG-PTV-E--RKT.............----D----A---.....-E---AI-----PAN  51
HIVBAL1    -------SSIVV-PAV-K--RRT.............----D----V---.....-E---AI-----AAN  51
HIVJRCSF   -------HSVPG-STV-E--RRAEPATDRVRQ....T----V----V---.....-E---AI-----AAT  61
HIVOYI     -------CSMKG-PTI-E--KRAEL......Q...PP----E----A---.....-E---AI-----AAT  56
HIVSF2     -------RSMGG-SAI-E--RRAEP..........RA----D----V---.....-E---AI-----AAT  55
HIVSF162   -------R.MSG-SAV-E--KRAEP..........A----D----V---.....-E---AI-----AAN  53
HIVSF33    -------SKM.G-PAV-E--KRA.............----D----A---.....-E---AL-----AAT  50
HIVHAN     -------CS..G-PTV-E--KQAEP...........----D----A---.....-K---AI-----ATN  51
HIVRF      -------SKMGG-PAV-E--QKA.............----D----A---.....-E---TI-----AAN  51
CONSENSUS2 MGqKWSKss?vqWPaiReRiR?T??..........?ppaa?GVGAvSqD?????L?KhGAitSSntast  47
HIVELI     --G----SSIVG--AI-E-I-R-.............NPAAD----V-R-.....-E-H--IT--NTAST  51
HIVNDK     --G----SSLVG--AI-E-I-K-.............DPAAD----V-R-.....-E-H--IT--NTAST  51
HIVMAL     --G----SSIVG--KI-E-I-R-.............PPTET----V-Q-AVSQD-D-C--AA--SPAAN  56
HIVU455    --G----KSRVE--EV-K-M-E-.............PAAAK----V-Q-.....-D-Y--VT--NTSST  51
HIVZ321    --N----....G--AV-E-I-Q-PP...........APPAAE----A-Q-.....-A-H--IS--NTATN  50
SIVCPZ     --T----SSLVG--EV-R-I-EA.............PTAAE---EV-K-.....-ERH--IT-RNTPET  51

                                                        LTR ->
CONSENSUS1 NAdCaWLEAq??EdEevGFPVrPQVPLRPMTykaA?DlSHFlKekGGLeGll?SqkRQdILDLWiyHTQG 115
HIVLAI     --A-A----Q..-E-EV----T---------YKA-V-L---L-EK---E-LIH-QR--D-----IY---- 119
HIVNL43    --A-A----Q..-E-EV----T---------YKA-V-L---L-EK---E-LIH-QR--D-----IY---- 119
HIVBRVA    --D-A----Q..-D-EV----K---------YKA-V-L---L-EK---E-LIH-QQ--D-----VY---- 131
HIVSC      --D-A----Q..-E-EV----R---------YKA-V-L---L-EQ---E-LITPREDKIS-ICGSTTHKA 119
HIVBAL1    --D-A----Q..-D-EV----R---------RSA-I-L---F-KK---E-LIH-QK--D-----VY---- 119
HIVJRCSF   --D-A----Y..-D-EV----R---------YKA-I-L---L-EK---E-LIY-QK--D-----IY---- 129
HIVOYI     --D-A----Q..-D-EV----R---------YKG-L-L---L-EK---E-LIY-QK--D-----VY---- 124
HIVSF2     --D-A----Q..-E-EV----R---------YKA-L-I---L-EK---E-LIW-QR--E-----IY---- 123
HIVSF162   --D-A----Q..-D-DV----R---------YKA-L-L---L-EK---E-LIY-QK--D-----IH---- 121
HIVSF33    --D-A----Q..-D-EV----K---------YKA-L-L---L-EK---E-LVY-QK--D-----IY---- 118
HIVHAN     --A-A----Q..-E-EE----R---------YKG-L-L---L-EK---E-LIY-PK--E-----VY---- 119
HIVRF      --A-T----QED-D-EV----R---------FKA-V-L---L-EK---D-LVF-QK--D-----VY---- 121
CONSENSUS2 Na?CawlEaqe.EseeVGFPVRPQVPLRPMTyK?AfDLShFLKEKGGLdGLiwSkKRQEILDLWVYhTQG 114
HIVELI     -AD-AWL-AQE.-SDE---------------Y-E-L---H--------E--IW-K----------N--- 120
HIVNDK     -DT-AWL-AQE.-SEE---------------Y-E-V---H--------E--IW-K----------N--- 120
HIVMAL     -AS-...-PPE.-EEE---------------Y-G-F---H--------D--VW-P----------H--- 122
HIVU455    -AS-AWL-AQ..-EGD---------------Y-A-F---F--------D--IH-Q----------H--- 119
HIVZ321    -PD-AWL-AQE.-SEE---------------F-G-F---F--------D--IY-K----------H--- 119
SIVCPZ     -QTLAWL-EM..DNEE----------T----Y-A-F---H--------E--VY-RR----------H--- 119
```

HIV1 NEF

```
                -- max HIV2/SIV nef similarity --      -- LTR NFAT-1 region --
CONSENSUS1 YFPDWQNYTpGPGvRyPLtFGWCfKLVPvePekvEEanegENnsLlHPmslHGmdDpEkEVL?WkFDsrL    184
HIVLAI     ---------P---V-Y--T----Y----VE-DKV--ANKG--TS-L--VSL--MD-P-R---E-R--SR-    189
HIVNL43    ---------P---V-Y--T----Y----VE-DKV--ANKG--TS-L--VSL--MD-P-R---E-R--SR-    189
HIVBRVA    ---------P---V-Y--T----F----VE-EKI--ANEG--NS-L--MSQ--MD-P-R---Q-R--SR-    201
HIVSC      TSLIGRTTHQ-Q-SDI--C----F----VK-EKI--ANEG--NS-L--MSL--ME-P-R---E-R--NR-    189
HIVBAL1    ---------P---T-F--T----F----VE-EKV--ANEG--NC-S--MSL--MD-P-K---A-K--SS-    189
HIVJRCSF   ---------A---V-F--T----F----VD-EKV--ANEG--NC-L--MSQ--MD-P-K---V-K--SK-    199
HIVOYI     ---------P---I-Y--C----F----MD-DQV--ANEG--NS-L--ISL--MD-P-K---V-K--SR-    194
HIVSF2     ---------P---I-Y--T----F----VE-EKV--ANEG--NS-L--MSL--ME-A-K---V-R--SK-    193
HIVSF162   ---------P---I-Y--T----F----VD-DYV--ANAG--NS-L--MSQ--MD-P-K---V-R--SR-    191
HIVSF33    ---------P---V-F--T----F----VE-EKV--ANEG--NS-L--MSL--ME-P-K---V-K--SH-    188
HIVHAN     ---------P---V-Y--T----F----VE-D..--....--SS-L--ASL--TE-T-R---K-K--SH-    183
HIVRF      ---------P---T-Y--T----F----VE-DKV--ATEG--NS-L--ICL--MD-P-K---V-K--SR-    191
CONSENSUS2 ?FPDWqNYTPGPGiRyPLtFGWCfkLVPvdP?EVEEategEnNcLLHPicQhCmeD?EreVLmWkFdS?L    180
HIVELI     I----Q--------I-Y--T----YE---VD-Q----DTEG-T-S----IC-H-ME-P-RQ--K-R-N-R-    190
HIVNDK     I----Q--------I-Y--T----FQ---VD-Q----ATER-D-C----MC-Q-ME-P-RQ--M-R-N-R-    190
HIVMAL     Y----Q--------I-F--T----FK---MS-E----ANEG-N-C----IS-H-ME-A-RE--K-K-D-S-    192
HIVU455    F----Q--------I-Y--T----YK---VD-A----ATGG-N-S----IC-H-VD-E-KE--M-K-D-T-    189
HIVZJ21    F----H--------T-Y--C----FK---VD-R----ANTG-N-C----MS-H-MD-D-RE--M-K-D-S-    189
SIVCPZ     F----Q---T---T-F--C----FK---LTEEQ--QANEGDN-C----IC-H-ME-EDKE--V-R-D-R-    189

              -- NRF --
CONSENSUS1 AfhHmARelHPEyyKdC?      201
HIVLAI     -FH-V--EL---YF-N-*      206
HIVNL43    -FH-V--EL---YF-N-*      206
HIVBRVA    -FH-M--EL---YY-N-*      218
HIVSC      -FH-M--DL---YY-D-L      207
HIVBAL1    -FH-V--E*               197
HIVJRCSF   -LH-V--EL---YY-D-*      216
HIVOYI     -FR-M--EV---YY-D-*      211
HIVSF2     -FH-M--EL---YY-D-*      210
HIVSF162   -FH-M--EL---YY-D-*      208
HIVSF33    -FR-M--EL---YY-D-*      205
HIVHAN     -FH-K--EL---YY-D-*      200
HIVRF      -FH-V--EK---YY-D-*      208
CONSENSUS2 Al?H?ArE?HPEfYKd?*      193
HIVELI     -FE-K-R-M---F--N*       206
HIVNDK     -LE-K-R-L---F--DC*      207
HIVMAL     -LR-R-R-Q---Y--DC*      209
HIVU455    -LK-R-Y-L---F--D*       205
HIVZJ21    -RK-L-R-M---F--D*       205
SIVCPZ     -LR-I-R-Q---Y--D*       205
```

Information dense stretches (0.8 or greater) are indicated by ****.

```
CONSENSUS1 MGGKWSK?s??gWpavReRMrrae?????????????EPAAdGVGAvSRD.....LeKHGaiTSSNTaat     49
CONSENSUS2 --g----s--v--P-i---iR?T?-...........-ppaa?------q-?????-?-h-A-t--nt-s-     47
HIV1NEF    MGXKWSK....XW^X^R^R^^^^..............X^^XXGVGAXS^D.....LXKXG^X^SS^^^^^     47
                      A I K IKKA              AA        K           A A  DAAAD
                      G V R VRRG              GG        R           G G  EGGGE
                      S   E LEES              SS        E           S S  NSSSN
                      T   Q MQQT              TT        Q           T T  QTTTQ
                      P        P              PP                    P P  KPPPK
                                                                        R   R
                                                                        H   H
                                                                        S   S
                                                                        T   T

                                                  LTR ->
                            **************                      **************
CONSENSUS1 NAdCaWLEAq??EdEevGFPVrPQVPLRPMTykaA?DlSHFlKekGGLeGli?SqkRQdILDLWiyHTQG    115
CONSENSUS2 -a?--wl-a-e.-se-V----R----------K?-f-L-h-L-EK---d-L-w-kK--E-----VYh---    114
HIV1NEF    NXXC...E^X..E^X^XGFPV^PQVPLRPMTX^XA^D^S^F^K^^GGL^G^^XSX^RQ^ILDLW^^^TQG    112
                A    D D     D          D  I I F I KK   D II   K  D     IFD
                G    E E     E          E  V V W V RR   E VV   R  E     VWE
                S    N       N          N  L L Y L EE     L    E         YN
                T    Q       Q          Q  M M H M QQ     M    Q         HQ
                P    K       K          K  F     F                        K
                     R       R          R  W     W                        R
                     H       H          H  Y     Y                        H
                     S       S          S  C     C                        S
                     T       T          T                                 T

          -- max HIV2/SIV nef similarity --     -- LTR NFAT-1 region --
          *************      *********
CONSENSUS1 YFPDWQNYTpGPGvRyPLtFGWCfKLVPvePekvEEanegENnsLlHPmslHGmdDpEkEVL?WkFDsrL    184
CONSENSUS2 ?----q---P---i----------k----d-?EV---t---nNc-L--icQh--e-?-re--m---dS?-    180
HIV1NEF    ^FPDW^NYT^GPGXR^^LXFCWC^^LVP^^PX..EE....E^XXLXHPXXX^GX^DXE^^VLXW^F^^^L    176
           I    D   A     FA      FK   ID           D         D  D   KE    K DDD
           V    E   G     WG      WR   VE           E         E  E   RQ    R NEE
           L    N   S     YS      YE   LN           N         N               NN
           M    Q   T      T       Q   MQ           Q         Q               QQ
           F    K   P      P            K           K         K               KK
           W    R                       R           R         R               RR
           Y    H                       H           H         H               HH
           C    S                       S           S         S               SS
                T                       T           T         T               TT

          -- NRF --
CONSENSUS1 AfhHmARelHPEyyKdC?      201
CONSENSUS2 -l?-?-rF*---fY--?*      193
HIV1NEF    AX^HXAX^XHPE^^K^*       192
             D    D    FF D
             E    E    WW N
             N         YY
             Q
             K
             R
             H
             S
             T
```

HIV1NEF:64 GFPVrPQVPLRPMT
"QVPLRMT" is part of a cytotoxic T-cell epitope.
Koenig et al., J. Immunol. 145:127-135 (1990) and
Culmann et al. J. Immunol. 146: 1560-65 (1991).

HIV1NEF:103 RQkILDLWaerTQGfFPDWrNYTjGPG
"YHTQGYFPDWQ" can serve as a cytotoxic T-cell epitope,
and "NYTPGPG" is part of a cytotoxic T-cell epitope.
Culmann et al. J. Immunol. 146: 1560-65 (1991).

This fragment taken together with the following fragment
is of interest as a putative "G" site (Guy et al.,
Virology 176: 413-25, 1990).

This fragment contains a region similar to a conserved pattern
in HIV2, "GIipDWQ".

HIV1NEF:136 FGWCdpLVP
This fragment is part of a cytotoxic T-cell epitope.
Culmann et al. J. Immunol. 146: 1560-65 (1991).
Mutational analysis -- Guy et al., Virology 176:413-25,
1990.

This fragment contains a regior similar to a conserved pattern
in HIV2, "GWLWKLvpV".

```
                                                                <- env cds
CONSENSUS  MGasgSkkrskp???LrerLLrargetyg?l?????ggyaqsqegsgkgq?spSCEgq?Y?qG?fmNtPW      57
HIV2ROD    --ASG-KKHSRPPRG-QER--RARAGACGGYWNESGGEYSRFQEGSDREQKSP---GRQ-QQ-DFM-T--      70
HIV2NIHZ   --ASG-KKRSKPLQG-QER--QARGETCGGRCNESGGGYLQSHEGSGREQNSP---GQR-QQ-DFV-T--      70
HIV2ISY    --ASG-KKRSRPSRG-QER--RARGGACGGLWDESEGGYSQFHEGSGREQKLP---GQR-QQ-DFM-T--      70
HIV2ST     --ASG-KKRSEPSRG-RER--QTPGEASGGHWDKLGGEYLQSQEGSGRGQKSP---GRR-QQ-DFM-T--      70
HIV2BEN    --ASG-KKLSKHSRG-RER--RARGDGYGKQRDASGGEYSQFQEESGREQNSP---GQQ-QQ-EYM-S--      70
HIV2D194   --ASG-KKRSEHSQG-RER--RARGGGYVKQRNASGGESSQSQEGSGREQKSP---GQQ-QQ-EFM-T--      70
HIV2GH1    --ASG-KKHSKHSQR-RER--RAHGGGYVQQCNASGGEYSQSQEGSGKGQKSP---GQQ-RQ-DFM-T--      70
SIVMM251   --GAI-KRRSKPAGD-RQK--RARGETYGRLLGEVEDGSSQSLGGLGKGLSSR---GQK-NQ-QYM-T--      70
SIVMMP11   --GAI-MRRSKPTGD-RQK--RARGETYGRLLGEVEDGSSQSLGGLGKGLSSR---GQK-NQ-QYM-T--      70
SIVMM142   --GAI-KKRSKPPRD-RQR--RARGENYGRLFKGVEDGSSQSLGGLDKGLSSL---GQK-NQ-EYM-T--      70
SIVMNE     --GAT-KRRSKSPGD-RQR--RARGETYGRLWEELEDGYSQSLGGSDKGSSLL---GQK-NQ-QFM-T--      70
SIVSMMH4   --GAI-KKQYKRGGN-RER--QARGETYGRLWEGLEEGYSQSLGASGKGLSSL---PQK-SE-QYM-T--      70
SIVSMMPBJ  --GVT-KKQRRRGGN-YER--QARGETYGRLWEGLEGEYSQSQDASGKGLSSL---PQK-CE-QFM-T--      70

                                                                LTR ->
CONSENSUS  rnPA?ereKl?YrqQNmDDvDs?Ddd1vgvpv?prVPlR?mtyklAlDmShfIKekgGLeGlyYs?rRhr     121
HIV2ROD    KD--AERE-NL-RQ--M--V-SD-DDQVRVSVTPK--L-PMTHRL-I-M-HL--TRG--E-MF-SER-HK     140
HIV2NIHZ   RT--AERE-EL-KQ--M--V-LD-DDQVGFPVTPR--L-PMTFKL-V-M-HF--EKG--E-LF-SQR-HR     140
HIV2ISY    RT--TEKE-ES-RQ--M--V-SD-DDLVGVSDTSR--L-AMTYRM-V-M-DL--DKG--E-MY-SER-HR     140
HIV2ST     RA--.EGE-GS-KQ--M--V-SD-DDLVGVPVTPR--L-EMTYRL-R-M-HL--EKG--E-LY-SDR-RR     139
HIV2BEN    RN--TERQ-DL-RQ--M--V-SD-DDLIGVPVTPR--R-EMTYKL-I-M-HF--EKG--Q-MF-SRR-HR     140
HIV2D194   RT--AIGQ-NS-KQ--M--V-SD-DDLVGVPVMPR--L-EMTYKL-I-M-HF--EKG--E-IF-SRE-HR     140
HIV2GH1    RT--IEGQ-KL-KQ--M--I-SS-DDLVGVPVTPR--L-AMTYKL-V-M-HF--KR.--D-MF-SRD-HR     139
SIVMM251   RN--EEKE-LA-RK--M--I-EE-DDLVGVSVRPK--L-AMTYKL-I-M-HF--EKG--E-IY-SAR-HR     140
SIVMMP11   RN--EEKE-LA-RK--M--I-EE-DDLVGVSVRPK--L-AMTYKL-I-M-HF--EKG--E-IY-SAR-HR     140
SIVMM142   RN--EERK-LP-RK--I--I-EE-DDLVGIPVEAR--L-TMSYKL-I-M-HF--EKG--E-IY-SAR-HR     140
SIVMNE     KN--GERE-LA-RK--I--I-EE-NDLVGVPVRPR--L-IISYKL-V-M-HF--EKG--E-IY-SER-HK     140
SIVSMMH4   RN--TERA-LG-RQ--M--V-DE-DDLIGVSVHPR--L-AMTYKL-I-M-HF--EKG--E-IY-NER-HR     140
SIVSMMPBJ  RN--TERA-LD-RQ--M--V-SA-..LVGCPVSPR--V-IMTYKL-I-M-HF--EKG--E-IY-SDR-HK     138

                        -- max HIV1 NEF similarity --
CONSENSUS  iLd?yLekeeGIipDWQnyt?Gpq?RYpm?fGWLWKLvpV?vs?Eaqede?thyLvhpAQts?wDDpwGE     183
HIV2ROD    I-NIY-EKEE--IA---NYTH-PGV--PMFF------VP-DVPQ-GEDTE.THC-VHP--TSKF--PH--     209
HIV2NIHZ   I-DLY-DK.....................................................................  148
HIV2ISY    I-DIY-EKEE--IP---NYTH-LGV--PMFF------VP-TVPQ-GEDTE.TLC-MHS--VSRF--PH--     209
HIV2ST     V-DIY-EKEE--IG---NYTH-PGV--PKFF------VP-DVPQ-GDDSE.THC-VHP--TSRF--PH--     208
HIV2BEN    I-DIY-EKEE--IP---NYTH-PGV--PMYF------VS-ELSQ-AEEDEA.NC-VHP--TSRH--EH--     209
HIV2D194   I-DLF-EKEE--IP---NYTH-PGT--PMYF------VP-DISQ-AEEVE.TNC-VHP--TSRY--EH--     209
HIV2GH1    I-DLY-EKEE--IP---NYTH-PGV--PMCF------VP-DVSQ-AEDDE.TNY-THP--TSRH--EH--     208
SIVMM251   I-DMY-EKEE--IP---DYTS-PGI--PKTF------VP-NVSD-AQEDE.RHY-MQP--TSKW--PW--     209
SIVMMP11   I-DMY-E...--IP---DYTS-PRI--PKTF------VP-NVSD-AQEDE.RHY-MQP--TSKW--PW--     206
SIVMM142   I-DIY-EKEE--IP---.IHS-PGI--LKMF------IP-NVSD-AQEDE.EHY-VHP--TSQW--PW--     208
SIVMNE     I-DMY-EKEE--MP---NYTS-PGP--PKTF------VP-NVSD-AQEGE.ENY-LHP--TSQW--PW--     209
SIVSMMH4   I-DMY-EKEE--IP---NYTS-PGI--PMHY------VP-DVSD-AQEDE.THC-VHP--TYQW--PW--     209
SIVSMMPBJ  I-DLY-EKEE--IP---NYTA-PGI--PMFF------VP-NVSD-AQEDE.THY-MHP--TSQW--PW--     207
CONSENSUS  vLvWkFDp?LAy?Yeaf1ryPEKfq?hsqipe?ewkarlkARGip?s?dk?etr*          230
HIV2ROD    T-V-E--PL--YS-EAFIRY---FGHKSGLPEEEWKARLK---IPFS*                 256
HIV2NIHZ   ..............AFTLY---FGHNSGLPEKEWKARLK---IPFS*                 180
HIV2ISY    T-V-K--PM--HE-TTFILY---FGHKSGMEEDDWKAKLK---IPFS*                 256
HIV2ST     T-V-R--PT--FS-EAFIRY---FGYKSGLPEDEWKARLK---IPFS*                 255
HIV2BEN    T-V-Q--SM--YN-KAFTLY---FGHKSGLPEKEWKAKLK---IPYSE*               257
HIV2D194   T-V-R--PM--YS-KAFILH---FGHKSGLPEKEWKAKLK---IPYSE*               257
HIV2GH1    T-L-R--PT--YD-KAFILH---FGHKSGLPEKEWKAKLK---IPYS*                 255
SIVMM251   V-A-K--PT--YT-EAYARY---LEASQACQHKRLEEG*                          247
SIVMMP11   V-A-K--PT--YT-EAYVRY---FGSKSGLSEEEVRRRLT---LLNMADKRETR-         260
SIVMM142   V-A-K--PT--YT-EAYIRY---FGSKSGLSEKEVKRRLA---LLEMADRKETS-         262
SIVMNE     V-V-K--PT--YT-EAYIRY---FGSKSGLSEEEVRRRLT---LLKMADKRETS-         261
SIVSMMH4   V A K  PE  YS-KAFIKY  .  I...KSGLSEEEVKRRLT---LLKWLTRRKQAE  .  309
SIVSMMPBJ  V A K  PK--YN-KAFVEH---FGSQSGLSKEEVQRRLT---LLKMADKKKTS-         261
```

Information dense stretches (0.7 or greater) are indicated by ****.

```
                                                                  <- env cds
CONSENSUS  MGasgSkkrskp???LrerLLrargetyg?l?????ggysqsqegsgkgq?spSCEgq?Y?qG?fmNtPW     57
HIV2SIVNEF MG^XXSX^XX^XXXXLX^^LL^^X^XXXXXXXXXXXXXXXXXXXXX^XX^XXSCE^^^YX^G^^^N^PW     70
             A    K  K      EK  KA A                      K  D     AKK  E DFI S
             G    R  R      QR  RG G                      R  E     GRR  Q EWV T
                     E          ES                           N     SEE    NYL
                     Q          QT                           Q     TQQ    Q M
                                 P                           K     P      K
                                                             R            R
                                                             H            H
                                                             S            S
                                                             T            T

                                                                LTR ->
                      **********
CONSENSUS  rnPA?ereKl?YrqQNmDDvDs?Dddlvgvpv?prVPlR?mtyklAiDmShfIKekgGLeGiyYs?rRhr    121
HIV2SIVNEF ^XPA.XXXKXXY^^QN^DD^DXXD..X^X^^XX^^VPXRX^^^^^AXD^S^^IK^^.GL^G^^Y^X^R^^    136
           K           KK  I  I       I IA  AK     ISFKL   L DI  DK   D IF D D DK
           R           RR  V  V       V VG  GR     VTWRM   M EV  ER   E VW E E ER
                        E  L            LS  S      L Y       NL  N    N LY N N N
                        Q  M            MT  T      M H       QM  Q    Q M  Q Q Q
                                        FP  P                KF  K    K    K K K
                                        W                    RW  R    R    R R R
                                        Y                    HY  H    H    H H H
                                        C                    SC  S    S    S S S
                                                             T   T    T    T T T

                        -- max HIV1 NEF similarity --
                     *******              *********
CONSENSUS  iLd?yLekeeGIipDWQnyt?Gpg?RYpm?fGWLWKLvpV?vs?Eaqede?thyLvhpAQts?wDDpwGE    183
HIV2SIVNEF ^L^^^L^...GI^^DWQ.^^XGXXXRYXXX^GWLWKL^^V^^^^E^^^X^..X^LX^^AQXX^^DDX^GE    200
           I DIF D     IA    ID          F      IA DIAD ADD D   I  DA    KF   F
           V NVW E     VG    VE          W      VG EVGE GEE E   V  EG    RW   W
              LY       LS    LN          Y       S NLSN  N      L  NS    EY   Y
              M        MT    MQ                  T QMTQ  Q      M  QT    QH   H
                        P    FK                  P K PK  K      F  KP
                             WR                    R  R  R      W  R
                             YH                    H  H  H      Y  H
                             CS                    S  S  S      C  S
                              T                    T  T  T         T

CONSENSUS  vLvWkFDp?LAy?YeafiryPEEfq?ksglpe?ewkarlkARGip?s?dk?etr*    230
HIV2SIVNEF XLXW^FD^XLA^^Y^^^XX^PEE^XX^^^^X^^^^^XX*                    238
               K  A   FD DAF  F   I  DDAI KDDIK
               R  G   WE EGW  W   V  EEGV REEVR
               E  S   YN NSY  Y   L  NN L ENNLE
               Q  T   HQ QT   H   M  QQ M QQQMQ
                  P    K KP       F  KK F  KKF
                       R R        W  RR W  RRW
                       H H        Y  HH Y  HHY
                       S S        C  SS C  SSC
                       T T           TT    TT
```

```
CONSENSUS  MG?gnSKpqhkkqlklWraLhktr?trygLLaDPLIGqSST?QEECdKaL?kSlirKrnG?MTpegRrLQ     65
SIVAGMTYO  --SQN--PAHKKYSKL-QA-HKTHVTRYG--A-----T---V----D-A-RK-LIR-QN-N--EEE-R--     70
SIVAGM155  --LGS--PQHKKQLTI-RA-HATRHTRYG--A-----Q---L----D-G-RK-LIR-RN-N--PEG-R--     70
SIVAGM3    --LGN--PQHKKQLSL-HA-HKTRATRYG--A-----Q---L----D-A-KE-LIR-RN-K--PEG-K--     70
SIVAGM677  --SSN--RQQQGLLKL-RG-RGKPGADWV--S-----Q---V----G-A-KK-W.G-..-K--PDG-R--     67

CONSENSUS  eGDtwdEWsD?E?EVGFPVrPRVPLRQmTYKLAVDFSHFLKeKGGLDGIYYSdRRnkILNLYALNEWGII    133
SIVAGMTYO  E--TWE--S-D-E------R-------M----------E----------D--NK----------       140
SIVAGM155  D--QWD--S-E-D------R-------I----------E----------D--NK----------       140
SIVAGM3    E--KWD--S-E-D------R-------M----------E----------D--NQ----------       140
SIVAGM677  E--TFD--D-D-E------Q-------M----------S----------E--EK----------       137

CONSENSUS  DDWnAwSkGPGIRyPrcFGFCFKLVPVdLHEEAetCERHCLvHPAQmgEDPDGInHGE?LvWKFDPmLAv    202
SIVAGMTYO  ---N-W-K-----F-KC----------D-----QT------V----MG------S---I-V-----M--I    210
SIVAGM155  ---N-W-K-----Y-RC----------A-----ET------V----LH------N---I-A-----M--V    210
SIVAGM3    ---N-W-E-----Y-RC----------D-----ET------V----VR------N---V-V-----M--V    210
SIVAGM677  ---Q-Y-P-----Y-RV----------D-----RN------M----MG------D---V-V-----K--V    207

CONSENSUS  qYdPnreyftDmh?l?gkrk*    220
SIVAGMTYO  Q-D-NREYFT-MHGLV.KRK-    229
SIVAGM155  Q-D-SREYFT-LYSTVGTGN-    230
SIVAGM3    Q-D-NRKYLT-MHDL.GKRK-    229
SIVAGM677  E-R-..DMFK-MHEH.AKR*    223
```

```
   Information dense stretches (.9 or greater) are indicated by ****.

                                                *****
CONSENSUS  MG?gnSKpqhkkqlk!WraLhktr?trygLLaDPLIGqSST?QEECdKaL?kSlirKrnG?MTpegRrLQ     65
AGMNEF     MGXX^SKXX^^XXX^^W^^L^X^XX^^^XLL^DPLIG^SST^QEECXK^L^^S^.XK..G^MTX^XR^LQ     67
             D    DK   DI DA D D  ADF   A     D   I      A KK I      D   D  K
             E    ER   EV EG E E  GEW   G     E   V      G RR V      E   E  R
             N    NE   NL N  N N  SNY   S     N   L        E L      N
             Q    QQ   QM Q  Q Q  TQ    T     Q   M        Q M      Q
             K    K    K  K  K K  PK    P     K              F      K
             R    R    R  R  R R   R          R              W      R
             H    H    H  H  H H   H          H              Y      H
             S    S    S  S  S S   S          S              C      S
             T    T    T  T  T T   T          T                     T

                      ************************************************************
CONSENSUS  eGDtwdEWsD?E?EVGFPVrPRVPLRQmTYKLAVDFSHFLKeKGGLDGIYYSdRRnkILNLYALNEWGII     133
AGMNEF     ^GD^^^EW^D^E^EVGFPV^PRVPLRQ^TYKLAVDFSHFLK^KGGLDGIYYS^RR^^ILNLYALNEWGII     137
           D  DFD  D D D      K       I             D          D  DK
           E  EWE  E E E      R       V             E          E  ER
              NY   N           E       L             N             NE
              Q    Q           Q       M             Q              Q
              K    K                                 K
              R    R                                 R
              H    H                                 H
              S    S                                 S
              T    T                                 T

           ***     *****    ********** ***** ********** ******      *****
CONSENSUS  DDWnAwSkGPGIRyPrcFGFCFKLVPVdLHEEAetCERHCLvHPAQmgEDPDGInHGE?LvWKFDPmLAv     202
AGMNEF     DDW^A^SXGPGIR^P^^FGFCFKLVPVXLHESA^^CERHCL^HPAQ^XEDPDGI^HGE^LXWKFDPXLA^     207
             D F       F KI                KD      I    I       D   I          I
             E W       W RV                RE      V    V       E   V          V
             N Y       Y  L                EN      L    L       N
             Q            M                QQ      M    M       Q
             K            F                 K                    K
             R            W                 R                    R
             H            Y                 H                    H
             S            C                 S                    S
             T                              T                    T

CONSENSUS  qYdPnreyftDmh?l?qkrk*    220
AGMNEF     ^Y^P..^^^^D^^XX..^X*    222
           E D   DIID LF    D
           Q E   EVVE MW    E
             N   NLLN  Y    N
             Q   QMMQ  H    Q
             K   KFFK       K
             R   RWWR       R
             H   HYYH       H
             S   SCCS       S
             T   T  T       T
```

```
LOCUS       SIVMNDGB1    9215 BP SS-RNA
ACCESSION   M27470 X15781
FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT         450     1958   1 GAG POLYPROTEIN
LENGTH      503
        1 MGNGNSALLG TDLDKFEKIR LKRGGKKCYR LKHLCWCKGE LDRFGLSDKL LETQQGCEKI
       61 LSVCWPLYDQ GSDNLKALVG TVCVVACIHA GIEIKSTQDA LKKLKVITKK EEKQEDESKN
      121 FPVQRDAAGQ YQYTPISPRI IQTWVKTVEE KKWKPEVIPL FSALTEGAIS HDLNIMLNAV
      181 GDHQGAMQVL KDVINEQAAE WDLTHPQQQP AQPGGGLRTP SGSDIAGTTS TVEEQLAWMN
      241 MQQNAINVGT IYKSWIILGM NRLVKSHCPI SITDVRQGPK EAFKDYVDRF YNVMRAEQAS
      301 GEVKMWMQQH LLIENANPEC KQILRSLGKG ATLEEMLEAC QGVGGPQHKA RLMAEMMRTV
      361 VGQSQNFVQQ RGPQRGPVRQ PTGRKPICFN CNKEGHVARF FKAPRRKGCW NCGAMDHQKA
      421 QCPKPAQQQR VNFLGYGPWG PSKPGNYPAQ EVTPTAPPLE EKPLQKTLST YQKLGRGLRQ
      481 KMKEEKREED FHSLSTLFQE DQ*

FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT     <  1745     4774   1 POL (NH2-TERMINUS UNCERTAIN; AA AT 1745)
LENGTH      1010
        1 FFRVWPLGSL QTGELSGTRG DSNSSTIRGE TSAENSEHLS EIRERAQAED EGGEERGGFS
       61 FPEYSLSRRP IEEVSVDGVT IRALLDTGAD DTIFNERNIK LKGNWQPKII GGIGGNLRVK
      121 QYDNVYVEIR GKGTFGTVLI GPTPIDIIGR NIMEKLGGKL ILAQLSDKIP ITKVKLKPGV
      181 DGPRIKQWPL SKEKIVGLQK ICDRLEEEGK ISRVDPGNNY NTPIFAIKKK DKNEWRKLID
      241 FRELNKLTQD FHELQLGIPH PAGIKKCKRI TVLDIGDAYF SIPLDPDYRP YTAFTVPSVN
      301 NQAPGKRYMY NVLPQGWKGS PCIFQGTVAS LLEVFRKNHP TVQLYQYMDD LFVGSDYTAE
      361 EHEKAIVELR ALLMTWNLET PEKKYQKEPP FHWMGYELHP DKWKIEKVQL PELAEQPTVN
      421 EIQKLVGKLN WAAQLYPGIK TKQLCKLIRG GLNITEKVTM TEEARLEYEQ NKEILAEEQE
      481 GSYYDPNKEL YVRFQKTTGG DISFQWKQGN KVLRAGKYGK QKTAHSNDLM KLAGATQKVG
      541 RESIVIWGFV PKMQIPTTRE IWEDWWHEYW QCTWIPEVEF ISTPMLEREW YSLSPEPLEG
      601 VETYYVDGAA NRDSKMGKAG YITDRGFQRV EEYLNTTNQQ TELHAVKLAL EDSGSYVNIV
      661 TDSQYVVGIL ASRPTETDHP IVKEIIELMK GKEKIYLSWL PAHKGIGGNE QIDKLVSSGI
      721 RKVLFLQNIE PAQEEHEKYH SNEAQLREKF HLPALVAKQI VQSCSKCCHH GEPIKGQTDA
      781 SLGVWQIDCT HLENQIIIVA VHVASGFMKA EVITAETGKK TAEFLLKLLA QWPISKLHTD
      841 NGPNFTSQEV ETMCWWLGIE HTFGIPYNPQ SQGVVENKNK YLKELIEKIR EDCKELKTAV
      901 AMATFIHNFK GRGGLGGMTA GERIVNMINT ELEYQYQQNQ ISKNLNFKVY FREGRDQLWK
      961 GPGILLWKGE GAVVLKYQEE IKIVPRRKCK IIKDYGESGK NSQVNLESV*

FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT        4728     5246   1 VIF
LENGTH      173
        1 MERVERIVRL TWKVSSQRIE KWHWLVRRQM AWATANNEEG CWWLYPHFMA YNEWYTCSKV
       61 VIIINRDIRL IVRSYWHLQI EVGCLSTYAV SIEAVVRPPP FEKEWCTEIT PEVADHLIHL
      121 HFYDCFMDSA VMKAIRGEEV LKVCRFKAGH KAQGVLSLQF LCLRVIYGPE ER*

FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT        5227     5541   1 VPR PROTEIN
LENGTH      105
        1 MGQKRDEQVS EDQGPPREPY NQWLADTMEE IKEEARKHFP LIILNAVSEY CVQNTGSEEE
       61 ACEKFITLMN RAIWVHLAQG CDGTFRERRP QLPPSGFRPR GDRL*

FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT        5471     5730   1 TAT PROTEIN, EXON 2 (FIRST EXPRESSED EXON)
                7950     8037   2 TAT PROTEIN, EXON 3 (AA AT 7951)
LENGTH      116
        1 MEPSGKEDHN CPPQDSGQEE IDYKQLLEEY YQPLQACENK CWCKKCCFHC MLCFQKKGLG
       61 IRYHVYRKRV PGTNKKIPGS GEEAIRRAID LSFHRTASRT YTANGQTTEK KKATA*

FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT        5590     5677   1 REV PROTEIN, EXON 2 (FIRST EXPRESSED EXON)
                7950     8167   3 REV PROTEIN, EXON 3 (AA AT 7952)
LENGTH      102
        1 MLVQEMLLSL YALLSKEGER NKVPCLQETW LSTCLFTGLP AEPIPQTARQ RRRRKQQLRT
       61 RRAQLRELEG RILKQILDRG PDLLCQGVTN LALAEKSESS N*
```

SIVMND PROTEINS

```
FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT        5661     8126   1 ENV POLYPROTEIN
LENGTH      822
        1 MSTGNVYQEL IRRYLVVVKK LYEGKYEVSR SFSYTMFSLL VGIIGKQYVT VFYGVPVWKE
       61 AKTHLICATD NSSLWVTTYC IPSLPDYDEV EIPDIKENFT GLIRENQIVY QAWHAMGSML
      121 DTILKPCVKI NPYCVKMQCQ ETENVSATTA KPITTPTTTS TVASSTEIYL DVDKNNTEEK
      181 VERNHVCRYN IIGLCRDSKE EIVTNFRGDD VKCENNTCYM NHCNESVNTE DCQKGLLIRC
      241 ILGCVPPGYV MLRYNEKLVN NKLCSNISAV QCTQHLVATV SSFFGFNGTM HKEGELIPID
      301 DKYRGPEEFH QRKFVYKVPC KYGLKIECHR KGNRSVVSTP SATGLLFYHG LEPGKNLKKG
      361 MCTFKGRWGL ALWSLAKELN KLNDSIKVNQ TCKNFTSTGE ENKQNTDKQK EFAKCIKTLK
      421 IDNYTTSGDR AAEMMMMTCQ GEMFFCNVTR IMRAWNDPNE KKWYPYASCQ IRQIVDDWMQ
      481 VGRKIYLPPT SGFNNHIRCT HRVTEMYFEM QKIDSNETKM QIKFLPPSET SNQFVAYGAH
      541 YKLVKIMPIG IAPTDVKRHT LPEHHKEKRG AVILGILGLL SLAGSAMGSV SVALTVQSQS
      601 LVTGIVEQQK QLLKLIEQQS ELLKLTIWGV KNLQTRLTSL ENYIKDQALL SQWGCSWAQV
      661 CHTSVEWTNT SITPNWTSET WKEWETRTDY LQQNITEMLK QAYDREQRNT YELQKLGDLT
      721 SWASWFDFTN WVQYLKWGVF LVLGIIGLRI LLALWNTISR FRQGYRPVFS QDCQQNLYPK
      781 RPDNGEEESN SLELGEHNSE NLKEESLNRS LIEDLTSFAR E*

FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT        8170     8814   1 NEF
LENGTH      215
        1 MGSSQSKKRS EAWVRYSSAL RQLVGGPVTP DGYKQIESSQ GAEKQSLLRG RAYGTYSEGL
       61 DKVQNDPLTK DEKLDLTQQD PEEEEEVGFP VTRQVSLRVP SYKDLIDFSH FIKEKGGLGG
      121 IYYSRRREEI LDLYAENEWG FEPGWQQYTT GPGTRYPKTF GFLFKLEPVS RAIGDEYAAN
      181 NHLLHSSQLC PQEDPEGETL MWSGTLILPM TLQH*



LOCUS       SIVMDNS        710 BP SS-DNA
ACCESSION   M62883
FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT         177  +   436   1 TAT PROTEIN, EXON 2 (FIRST EXPRESSED EXON)
ACCESSION   M62884
FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT    +    351      441   2 TAT PROTEIN, EXON 3 (AA AT 352)
LENGTH      117
        1 MEPSGKEDHN CLPQDLGQEE IDYKQLLEEY YQPLQACENK CWCKKCCFHC MLCFHKKGLG
       61 IRYHVYRKRG PGTNKKIPGG GEEAIRRAID LCFFNRTCSR THTANGQTTE KKKATA*

ACCESSION   M62883
FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT         367  +   436   1 REV PROTEIN, EXON 2 (FIRST EXPRESSED EXON)
ACCESSION   M62884
FEATURES        FROM  TO/SPAN     DESCRIPTION
    PEPT    +    351      571   3 REV PROTEIN, EXON 3 (AA AT 353)
LENGTH      97
        1 MSTGNVDQEL IRRYLVVVKK LYEGLSTCAF STGLAALPIP QTARQRRRRK QQLRTRRAQL
       61 RELEGRILKQ ILDRGPDQLC QGVANLALAE KSESSN*
//
```

III

ANALYSIS

A. VPT Coding Sequences . . . . . III - 2

B. Protein Information Summary 4

C. V3 Loop (PND) Analyses and Summary 6

D. Gregory HIV-2/SIV gp120 model updated 23

C. Nucleotide Distance Relationships and Phylogenetic Trees 25

IIIA. VPT CODING SEQUENCE

E. Cohen *et al.*, *J. AIDS* 3, 601–8, 1990, have uncovered an orf called T. Conservation of a putative ribosomal frameshift site, typically "aaaaagtg", suggests that a 17 kD fusion protein could be made: the first 29 AA residues of the 107 AA (typically) T product would come from the *tat* protein. A summary of the potential coding sequences and the amino acid alignment follow. A preliminary search of the PIR database converted to AACC patterns, using the PLSEARCH tool, revealed no obvious protein similarity or homology.

SEQUENCE	NUCLEOTIDE COORDINATES	AA LENGTH	SHIFT SEQUENCE
LAI	5412-5731	106	aaaaagtg
HXB2R	5377-5699	107	aaaaagtg
HIVNL43	5830-6149	106	aaaaagtg
MN	5845-6167*	59	aaaaagtg
SC	541- 875	111	aaaaagtg
BAL1	255- 577*	62	aaaaagtg
JRCSF	5842-6164	107	aaaaagtg
JRFL	5147-5469	107	aaaaagtg
OYI	5373-5695*	62	aaaaggtg
SF2	5840-6162	107	aaaaagtg
NY5CG	5830-6108*	62	aaaaggtg
SF162	88- 440*	62	aaaaagtg
CDC42	88- 410	107	aaaaagtg
SF33	88- 410*	59	aaaaagtg
HAN	5296-5619	107	aaaaagtg
RF	5347-5669	107	aaaaagtg
ELI	5376-5698	107	aaaaagtg
Z2Z6	5833-6155	107	aaaaagtg
NDK	5365-5687	107	aaaaagtg
MAL	5405-5727	107	aaaaagtg
U455	5276-5600*	62	aaagtgtg

* internal stop

```
     tat ->                                  / new fragment
CONSENSUS1  MePVDPrLEPWKHPGSQPktactnCYCKkVLlSLpslfhnKrLrHlLWQEEaEtAtksSsrQsdSSsfsi     70
HIVLAI      -E----R-----------KTACTT----K--L--PSLFHN-S-R-L-----A-T-TKT-SR-SD--SFSI     70
HIVHXB2R    -E----R-----------KTACTN----K--L--PSLFHN-S-R-L-----A-T-TKS-SE-SD--SFSI     70
HIVNL43     -E----R-----------KTACTN----K--L--PSLFHD-S-R-L-----A-T-TKS-SE-SD--SFSI     70
HIVMN       -E----R-----------KTACTT----K--L--PSLFHK-S-R-L-----A-T-TKS-$R-SD--SFST     69
HIVSC       -D----R----.-------KAACTS----K--L--PSLFHN-R-R-L-----A-T-KKS-SR-SD--SFST     70
HIVBAL1     -E----K-----------KTACTT----K--L--PSLLHN-R-R-L-----A-T-TKS-SR-$D--SFSI     69
HIVJRCSF    -E----S-----------KTACTN----K--P--PSLFHN-R-R-L-----A-T-TKT-SR-SD--SFST     70
HIVJRFL     -E----S-----------KTACTT----K--P--PSLFHN-R-R-L-----A-T-TKS-SR-SD--SFSI     70
HIVOYI      -E----R-----------KTASNN----K--P--SSLFHK-R-R-F-----A-T-TKS-SR-$D--SFSI     69
HIVSF2      -E----N-----------RTACNN----K--L--LRVFHK-R-R-L-----A-T-TKS-SG-SD--SFSI     70
HIVNY5CG    -E----R-----------KTACTN----K--L--PSLFHN-S-R-L-----A-T-TKS-SR-$D--SFSI     69
HIVSF162    -E----R-----------KTACTN----K--L--PSLFHN-R-R-L-----A-T-TKS-SR-$G--SFST     69
HIVCDC42    -E----R-----------KTACTN----K--L--PSLFHE-S-R-L-----A-T-TKS-SR-SE--SISI     70
HIVSF33     -E----N-----------RTACTN----K--L--PSLLHN-R-R-I-----A-A-TES-$Q-SE--RFSI     69
HIVHAN      -E----R----.------RTPCTN----K--P--PSLFHK-R-K-L-----A-T-TKS-SR-SD--SFST     70
HIVRF       -E----R----.------KTACNN----K--L--PSLLLN-R-R-L-----E-T-TRT-SR-SD--SLFI     70
CONSENSUS2  MdPVDPNlEpWt.HPGSQPrTaCnkChCKkVLLsLPslLhneRLRHlLWQEEAETaTkTsSrrsqSS?SyT     69
HIVELI      -D-----L-P-N------R-P-NK-H--K---S--SL-LKQ----L--------A-R-S-RRSG--S-Y-     70
HIVZ2Z6     -D-----I-P-N------K-A-NR-H--K---S--SL-HNE----L--------E-K-F-RRSD--R-Y-     70
HIVNDK      -D-----L-S-N------R-A-NK-H--K---S--SL-HNE----L--------A-K-S-RRSG--S-Y-     70
HIVMAL      -D-----L-P-N------R-P-NK-Y--K---S--NV-HNE----L--------A-K-S-GQSG--R-S-     70
HIVU455     -E-----L-P-K------T-A-SN-Y--S---A--TL-SEK----F--------P-R-S-RQSG--N-Y-     69
```

```
CONSENSUS1  KAVSatcna?ftnisnSaiSSSsNnSNacvdhS?h?ri?*                    105
HIVLAI      ----STCNATYTN.SN-SI---N-N--SCVVH-NH.RI*                     106
HIVHXB2R    ----STCNATYTNSSN-SI---N-N--SCVVH-NH.RI*                     107
HIVNL43     ----STCNATY.NSSN-SI---N-N--SCVVH-NH.RI*                     106
HIVMN       ----STCNATFSNSSN-SI---R-N--SCVIH-IH.RI*                     106
HIVSC       ----NTCNATFTNIIN-SI---S-N--SCVVY-IHINIGKY*                  111
HIVBAL1     ----STCNASFTNISY-RI---S-N--SCVDH-IH.RI*                     107
HIVJRCSF    ----SACNATFTNISN-SI---R-N--NCVVH-TH.RI*                     107
HIVJRFL     ----SACNAIFIHISN-SI---N-N--SCVVH-TH.RI*                     107
HIVOYI      ----NTCNILFTNISN-SI---N-N--SCVDH-TL.RI*                     106
HIVSF2      ----SKCNAIFTNISN-II---S-N--SCVDH-TH.RI*                     107
HIVNY5CG    ----STCNAFFQ$LCGP$C-$NT                                      89
HIVSF162    ----SIFNATSSNISN-SI---S-N--SCVDH-VH.RIQENIKTKKNR*           116
HIVCDC42    ----STCHAIFRNSSN-RI---S-N--SCVDH-IH.RI*                     107
HIVSF33     ----STCNVIFNNISN-SN---N-N--SYMDH-IN.KI*                     106
HIVHAN      ----STCDVIFTNSSN-SI---S-N--SCVDH-SH.RI*                     107
HIVRF       ----SICNAIFRNISN-SI---S-T--SCVDH-WH.RN*                     107
CONSENSUS2  kAVSatCNaTFaDnSNSSISSanNnSNSCVDHSIh.Ri*                     106
HIVELI      K---ST--A--R-N-------SN-T---------H.-I*                     107
HIVZ2Z6     K---ST--A--T-N-------SS-N---------H.-I*                     107
HIVNDK      R---IP--A--S-N-------SN-N---------H.-I*                     107
HIVMAL      R---SI--T--S-I-------NA-N---------Y.-N*                     107
HIVU455     K---N$LM$$HLWK-GQ$QG$$$H$-$Q$$CGL$$V$N                       93
```

IIIB. PROTEIN INFORMATION SUMMARY

Information contents (IC) and information densities (ID), expressed as amino acid equivalents, can be calculated from the data sets presented in PART II; the equation for this calculation is discussed in Smith and Smith, *PNAS* 87, 118–122, 1990. In brief, an unvaried amino acid residue in an AACC pattern will have 1.0 AA equivalents of information and a wild-card will have 0.0 AA equivalents of information; variable but conserved residues, e.g., Asp and Glu, will have intermediate values. Thus an ID of 0.3 implies that, on the average, a protein pattern has 0.3 AA equivalents at every position. The IC of a protein, by this way of thinking, is the number of residues in the molecule; for an AACC pattern, this value must be corrected for the negative information resulting from gaps in an alignment. For example, the AACC pattern built from 62 myoglobin sequences has an IC of 68.3 and an ID of 0.45. Typical IDs for patterns derived from the PIR database are 0.2–0.3. Obviously these quantities depend upon the number of different sequences involved in the computation. While this approach is less than ideal, it is clearly preferable in our minds to off-the-cuff judgments about "conservation" and "homology."

The principal purpose of this analysis is to help clarify the problem of what is and what is not conserved. By comparing the fraction of unvaried residues from the consensus sequences in Part II to the IDs (c-a), some measure of the conservation over and above invariance is achieved. In the case of HIV-1s, inclusion of the SIV-cpz sequence in the AACC pattern but not in the ordinary consensus leads to a gross underestimation of this difference.

All of the figures that follow are approximate, based as they are upon changing data and averages; nevertheless, some tendencies can be provisionally inferred: i) the smaller, regulatory proteins are the more variable proteins (a possible exception being HIV-1 VPR); ii) while there are fewer samples of HIV-2/SIV and, especially, AGM, these show remarkable extents of variation (the AGM samples are for east African countries only); iii) the fraction of unvaried residues represents, typically, 70% to 85% of the actual information provided by the data set.

HIV-1 PROTEIN	# sequences	approximate length	# unvaried residues	fraction (a)	pat-IC (b)	pat-ID (c)	diff (c-a)
GAG	19	500	340	0.68	387	0.77	0.09
POL	17	1000	784	0.78	864	0.86	0.08
VIF	17	190	112	0.59	136	0.70	0.11
VPR	18	95	67	0.71	75	0.77	0.06
TAT	22	100	39	0.39	55	0.54	0.15
REV	24	115	47	0.41	61	0.56	0.15
VPU	19	75	30	0.40	43	0.51	0.11
ENV	31	800	400	0.50	452	0.55	0.05
NEF	19	200	98	0.49	123	0.62	0.13
T orf	22	105	39	0.37	51	0.47	0.10

HIV-2/SIV PROTEIN	# sequences	approximate length	# unvaried residues	fraction (a)	pat-IC (b)	pat-ID (c)	diff (c-a)
GAG	14	500	319	0.64	392	0.78	0.14
POL	15	1000	647	0.65	784	0.77	0.12
VIF	15	200	105	0.53	139	0.64	0.11
VPX	15	110	62	0.56	72	0.65	0.09
VPR	15	80	40	0.50	48	0.47	- 0.03
TAT	16	130	39	0.30	55	0.43	0.13
REV	15	90	36	0.43	53	0.50	0.07
ENV	16	815	441	0.54	534	0.64	0.10
NEF	14	240	72	0.3	115	0.47	0.17

AGM PROTEIN	# sequences	approximate length	# unvaried residues	fraction (a)	pat-IC (b)	pat-ID (c)	diff (c-a)
GAG	5	520	317	0.61	371	0.73	0.12
POL	4	1050	642	0.61	773	0.73	0.12
VIF	4				130	0.56	
VPX					91	0.76	
TAT					54	0.55	
REV					45	0.52	
ENV	4	830	523	0.63	605	0.71	0.08
NEF					168	0.74	

IIIC: Analysis of the HIV-1 V3 Loop

Because of its capacity to stimulate potent immune responses, the HIV-1 V3 loop has been the focus of intense interest over the last few years. The V3 loop is located in the third variable domain of the gp120 protein and is bounded by two cysteine residues capable of forming a disulfide bridge. Located in a highly variable region of gp120, insertions and deletions are frequent, so that the size of the V3 loop ranges from 33 to 40 amino acids (most have 35 amino acid residues), including the cysteines that form its boundaries. The V3 loop is also known as the principal neutralizing domain (PND), because peptides derived from this region are able to elicit type-specific neutralizing antibodies which can block viral infection in vitro by blocking fusion (1–6). Peptides from this region are also able to serve as epitopes for both cytotoxic T cells (7, 8) and helper T cells (9, 10). Finally, the PND has been implicated in determination of tropism (11).

Due to the biologically interesting features of the V3 loop, this region has often been included in PCR amplification sequence studies. In the current HIV database there are approximately 700–800 V3 sequences, although many of these have not yet been released.

The intent of this section is to provide analysis of the potential for variability within the V3 loop, and to present the published V3 loop sequences in alignments that facilitate comparisons of sequences obtained from different studies. This section includes:

1. A V3 loop consensus sequence and amino acid class covering pattern based on all of the sequences currently in the database.
2. Amino acid alignments of published V3 loop sequences.
3. The nucleic acid sequence alignments corresponding to the amino acid alignments of section C2.
4. Amino acid alignments of sets of sequences from the same individual.
5. Histograms showing the distribution of similarity comparisons between V3 nucleotide sequences.

References:

1. Palker *et al.*, *PNAS* 85:1932–36 (1988)
2. Goudsmit *et al.*, *PNAS* 85:4478–82 (1988)
3. Ho *et al.*, *J. Virol.* 61:2024–28 (1987)
4. Kenealy *et al.*, *AIDS Res. Hum. Retroviruses* 5:1973 (1989)
5. Rusche *et al.*, *PNAS* 85:3198–3202 (1988)
6. Javaherian *et al.*, *PNAS* 86:8768–72 (1989)
7. Takahashi *et al.*, *PNAS* 85:3105–09 (1988)
8. Takahashi *et al.*, *Science* 246:118–121 (1989)
9. Palker *et al.*, *J. Immunol.* 142:3612–19 (1989)
10. Hart *et al.*, *J Immunol.* 145:2677–85 (1990)
11. Westervelt *et al.*, *PNAS* 88:3097–101 (1991)

IIIC1. V3 loop consensus and amino acid class covering pattern

The analysis in this section is based on 527 V3 loop amino acid sequences from N. America and Europe. The 527 sequences include the sequences shown in the next section (IIIC2), the LaRosa data set (LaRosa *et al.*, *Science* 249: 932–35 (1990)), and many additional sequences graciously provided prior to publication by various groups. In particular many V3 sequences were provided by Steve Wolinsky's group at Northwestern University, Chicago, Ill., and by Jaap Goudsmit's group at the University of Amsterdam, the Netherlands. In cases where multiple identical sequences across the V3 loop occurred within a sequence set derived from a single individual, we included only one such representative sequence in this set of 527 sequences (see section IIIC3).

The consensus sequence shows the most common amino acid found in a given position. The consensus presented here is the same as the consensus derived by LaRosa *et al.*, as well as the consensus from all sequences excluding the LaRosa *et al.* set. The gaps included to maintain the alignment are not shown; insertions in these positions occur in less than 2% of the sequences (see section C2 for the consensus with gaps). Directly beneath the consensus sequence the frequency of the most common amino acid found in a given position is shown as a percentage.

A modification of the amino acid class covering (AACC) pattern, described at the beginning of section II, was used to generate an AACC pattern for the V3 loop. In AACC patterns, lower case letters represent related groups of amino acids. Because of the extensive variability of the V3 loop and the large volume of sequences available in this region, a standard AACC pattern yields very little information (see the AACC pattern generated when one considers 100% of the residues possible at a given position in the alignment, where "X" signifies a nonrelated set of amino acids and "g" signifies that a gap can occur at a given position). It is probable that sequence error or PCR amplification of biologically inactive mutant viruses account for some of the very rare and unusual amino acids found at a given position in an alignment of the V3 loop. Considerable limitations on the amino acid class typically found at a given position result when only the most common set of the amino acids at that position are considered. An AACC pattern was generated considering 95% of the total amino acids that can occur at a given position, excluding rare amino acids that yield a more divergent amino acid class. Directly under this pattern is the set of amino acids that were included to yield the amino acid class, listed in order of their frequency, with the most common at the top.

The 'GPGR' tip of the V3 loop is highly conserved, and is predicted to form a type II beta-turn (Javaherian *et al.*, *PNAS* 86:8768–72 (1989)). Among the 527 sequences analyzed here, 416 have GPGR at the tip (78.9%). GPGK is the second most frequently observed, present in 41 sequences (7.8%). GPGQ is the most common form found among the African isolates. Both GPGK and GPGQ are predicted structurally to make a type II beta-turn (Wilmot and Thorton, *J. Mol. Biol.* 203:221–232 (1988)). There are 24 variants of the four amino acid tip found in the 527 sequences that make up the N. American/European data set; most of these variants conserve the type II beta-turn predicted structure. The tip of the loop is important for the neutralizing antibody and T-cell epitopes described in the introduction to this section. It may have a functional role in fusion (Freed *et al. J. Virol.* 65:190–194 (1991)) and also as a protease cleavage site (Hattori *et al. FEBS Letts* 248:48–52 (1989) and Clements *et al.*, *AIDS Res. and Human Retroviruses* 7:3–16 (1991)).

V3 consensus sequence, N. America and Europe:

C	T	R	P	N	N	N	T	R	K	S	I	H	I	G	P	G	R	A	F	Y	T	T	G	E	I	I	G	D	I	R	Q	A	H	C
100	87	99	98	85	93	93	94	88	74	63	94	46	75	94	97	99	86	86	75	82	54	88	75	34	95	80	98	85	94	98	81	100	91	99

AACC pattern for ≥ 95% of sequences:

C	X	R	P	r	r	r	j	r	p	X	a	X	c	h	P	G	p	X	f	X	j	X	g	g	I	g	G	l	c	R	p	A	e	C
T				N	N	N	T	R	K	S	I	H	I	G			R	A	F	Y	T	T	G	E		I		D	I		Q		H	
I				S	T	T	A	K	Q	G	V	P	M	A			K	V	I	H	A	A	-	D		V		N	M		K		Y	
								S	R	R		N	L				Q	T	W	V		I	E	Q		T					R			
												T							V	F		R	K	R		-								
												R							L				R	K										
												S											D	A										
												Y												N										
																								H										
																								-										

AACC pattern for 100% of sequences:

C X n X X X X X g g X X g X X X X g X X g g g g g X g X X X X A X X

IIIC2. Peptide Alignments

The V3 loop amino acid alignments are divided into three sets: a North American/European set, and two smaller African sets.

The N. American/European set is aligned with the consensus described in detail in section C1. The aligned amino acid sequences then begin with V3 sequences extracted from HIV envelope sequences presented in Parts I and II of the database. These sequences are identifiable by the "HIV" prefix included in their names, and background information on these isolates can be found in Part I of the database. The next two sequences, labeled CAN0, were provided by Chin-Yi Ou's group at the CDC; these sequences are derived from one of the earliest N. American samples, taken from a Canadian airline steward (sometimes called "patient Zero"). The sequence FO is from an early isolate and was kindly provided by H. Guo, M. Reitz, and W. Blattner, NCI, LTCB. Following these sequences are 50% consensus sequences derived from PCR amplified sets of sequences from the same individual. More information on these sequences can be found in Part I, and in sections C4 and C5. The Epstein sequences are consensus sequences from samples taken from the brain (B) or spleen (S) of three HIV infected children (Epstein *et al.*, *Virology* **180**: 583–590 (1991)). The Wolfs sequences are consensus sequences from a set of children infected by transfusion from a single donor (Wolfs *et al.*, *PNAS* **87**: 9938–42 (1990); the corresponding nucleotide sequences for this set should be available in the near future). The "Edin" sequences are consensus sequences from a cohort of hemophiliacs from Edinburgh, some of whom were infected by a common source (Balfe *et al.*, *J. Virol.* **64**: 6221–6233 (1990) and Simmonds *et al.*, *J. Virol.* **64**:5840–50)). The next sequence is Guo.JB.con, a consensus derived from samples collected in 1983 from the patient BRU (Guo *et al.*, *Nature* **349**:745–6 (1991)). The final sequences in this set are consensus sequences derived from direct sequencing of samples taken in 1983 from three patients: BRU (sequences from BRU are called "JB"), LAI, and DL (Wain-Hobson *et al.*, *Science* **252**:961–64 (1991)). These sequences have the prefix "W-H". For comparison HIVHXB2 and HIVLAI (formerly called HIVBRU) V3s are included.

Presentation of the LaRosa *et al.* data set (*Science* **249**: 932–35 (1990)) in this database is pending further reevaluation of the sequences by the authors; the nucleotide sequences are available through GenBank, accession numbers M61417-M61661. The original data set was revised (LaRosa *et al.*, *Science* **251**: 811 (1991)). The uncorrected version, as originally provided by the authors, appeared in the 1990 compendium (pages III-54–55); a summary of the corrections was provided in a database mailing dated Jan. 1991.

The African V3 sequences are highly variable, and thus did not readily align with the N. American/European set. Therefore the alignments of the African sequences were generated using the PIMA program described in parts I and II of the database (a PIMA "cluster-score-cut-off" of 25 divided the African sequences into the two groups shown here). African sequence alignments are shown with a 50% consensus sequence at the top, and an amino acid class covering (AACC) pattern generated by the PIMA program at the bottom (see part II). Both the consensus and the AACC pattern are based on sequences which are shown here as well as an additional eight sequences kindly provided by several groups prior to publication. Most of the sequences shown here originated in Zaire, although sequences from Uganda (U455), Tanzania (CBL4), Gabon (OYI), and Rwanda (SF170) are also shown. Further information concerning the African sequences can be found in parts I and II of the database.

HIV-1 V3 LOOP ANALYSIS

V3 Loop Amino Acid Alignment, N. America and Europe:

```
CONSENSUS        CTRPNNN..TRKSIHI..GPGRAFY.TTGEIIGDIRQAHC     35
HIVSC            -------..-TR----..-------.A--D----------     35
HIVMN            -----Y-..K--R---..-------.--KN---T------     35
HIVJH32          ----SKT..--RR---..-------.--KQ-A--L-----     35
HIVWMJ12         -----DI..A-RR---..-------.-.-K---N------     34
HIVWMJ22         ----Y--..V-R-LS-..------R.-.R----I------     34
HIVWMJ32         -----DI..A-RR---..-------.-.-K---N------     34
HIVSF2           -------..-----Y-..------H.---R------K---     35
HIVRF            -------..-----TK..----VI-.A--Q------K---     35
HIVBRVA          -------..---R-TM..----VY-.---Q------R---     35
HIVJRCSF         ----S--..-------..-------.--------------     35
HIVJRFL          -------..-------..-------.--------------     35
HIVCDC42         ------H..---RVTL..----VW-.-----L-N------     35
HIVHAN           -------..---G---..-----V-.---R-V----L---     35
HIVNY5NEW        -------..---G-A-..----TL-.AREK----------     35
HIVJFL           -------..-----TL..-------.---D----------     35
HIVADA           -------..-------..-------.--------------     35
HIVSF162         -------..-----T-..-------.A--D----------     35
HIVSF33          -------..R-RR-TS..---KVL-.----------K-Y-     35
HIVALA1          -----IY..RKGR---..------H.--RQ--EN------     35
HIVBAL1          -------..-------..-------.--------------     35
HIVBAL2          -------..-----N-..-------.--------------     35
CANOA            ----H--..------M..---K---.---G----------     35
CANOB            ----H--..-----N-..-------.---Q----------     35
FO               -------..--R---F..-------.---Q---N------     35
Epstein.con.B1   -------..-SR--?-..------?.A-ER----------     35
Epstein.con.S1   -------..--R--TM..----VL-.---?----------     35
Epstein.con.B2   -------..-------..-------.--------------     35
Epstein.con.S2   -------..---G---..------F.--------------     35
Epstein.con.B3   -------..---GV--..-------.--.D-V--------     34
Epstein.con.S3   -------..---GV--..-------.--.D-V--------     34
Wolfs.con.donor  -------..-------..-------.--------------     35
Wolfs.con.1.I    ----S--..--R--N-..------F.-----------?-     35
Wolfs.con.2.I    -------..-----N-..-------.---A---N------     35
Wolfs.con.2.II   -------..-----N-..-------.A--A----------     35
Wolfs.con.3.I    ----S--..---G---..-------.--------------     35
Wolfs.con.3.II   ----S--..---G---..-------.-----V--------     35
Wolfs.con.4.I    -------..--R----..-------.--DR----------     35
Wolfs.con.4.II   -------..--R--YT..------H.--DR----------     35
Wolfs.con.5.I    -------..-------..-------.--------------     35
Wolfs.con.5.II   ----S--..-------..-------.---D-T--------     35
Wolfs.con.6.I    -------..-------..-------.---?---?------     35
Wolfs.con.6.II   -------..---G---..-------.---D----------     35
Edin.con.V77     ----S--..-------..----V-H.A-------------     35
Edin.con.V82     -------..---R-Y-..-----V-.--EQ---N------     35
Edin.con.V87     ----G--..--ER-S-..------I.AR-Q----------     35
Edin.con.V91     ----G--..-----P-..------I.A-SQ------K---     35
Edin.con.V12     ----S?-..I-R-?--..-S-----.-IEGV?--V-K-Y-     35
Edin.con.V74     -------..--RG---..-------.A--N----------     35
Guo.con.JB       -------..-------..-------.A-------------     35
W-H.con.DL       -------..--ERLS-..----P---.A-RR----------     35
W-H.con.JB       -------..-------..-------.?-------------     35
W-H.con.LA1      -------..-----R-QR------?.---K-.-NM-----     36
HIVLA1           -------..-----R-QR------V.-I-K-.-NM-----     36
HIVHXB2R         -------..--R-R-QR------V.-I-K-.-NM-----     36
```

V3 Loop Amino Acid Alignments, Africa:

```
Af1.con*  CTRPNNNTR..?SIRIGPGQAFYATG?IIGDIRQAHC      35
HVSF170   ---------..K-V----------D--------Y-      35
HIVZ321   -M--------..K--S----R--F---D----------      35
HIVOYI    ---------..NR-S----R--HT-KQ----------      35
HIVMAL    ----G----..RG-HF-----L-T-.G-V----R-Y-      34
HIVU455   -S--Y-TRKNIRRYS--S-----V--K----------      37
Af1.AACC* CXRPXXXrnggrXfrfhjGpXfeXrgggXGXXXpAeC
          CTRPNNNTR..KSIRIGPGQAFYATGDIIGDIRQAHC
           M  GSYRKNIRRVSLAS RTVFTKNA.V KTGK Y
           I  SITK IRNGYHV   KVLHV .E G I  E
           S  Y D    Q   F         KQ      R
                                    G
                                    K
                                    .
Af2.con   CTRP.YKNTRQRT...PIGLGQALYTTR???...?IIGQAHC      35
HIVELI    -A--.-Q------...------S-----SRS....-------      34
HIVZ6     ----.------S-...----------GRTK...-------      35
HIVZ2Z6   ----.-R-I----...S----------KTRS....-----Y-      34
HIVNDK    ----.--Y-----...S---R-S---ITGKKKKTGY------      38
HIVCBL4   ----Y---I-KG-HVGHV-P-R-I---NIKG..SGN-R----      40
HIVJY1    ----DN-I---S-...----------IK....GD-R--Y-      35
Af2.AACC  CjRPgXpXXRpXTgggXaGXXpjcYTXrXnggggggXIXQAeC
          CTRP.YKNTRQRT...PIGLGQALYTTRIRS....IIGQAHC
           A  YNQYI KS HVGSV PRRSI  IKGKTKKTGY R  Y
              D RI   G    H          TT K  S N
                                     NS G    D
                                        .
```

*The consensus sequence and amino acid class covering (AACC) pattern of the first set of African sequences (Af1) was derived from a total of 13 sequences. Eight of these sequences are not yet published so they cannot be presented in the database at this time. All sequences used for the consensus sequences and AACC pattern of the second set of African sequences are shown. Lowercase letters in the AACC patterns designate amino acid classes, or groups of related amino acids. For more information about AACC patterns, see part II of the HIV database. Aligned directly beneath the AACC pattern is the set of all amino acids found in a given position in the alignment. Af1 includes amino acids from the eight sequences that are not shown as well as the five that are. The amino acids are listed in order of their frequency, with the most frequent listed at the top.

IIIC3. Nucleic Acid Alignments:

This section presents a nucleic acid alignment of the sequences corresponding to the amino acid alignment of the previous section.

```
CONSENSUS*  TGTACAAGACCCAACAACAATACAAGAAAAAGTATACATATA......GGACCAGGGAGAGCATTTTATACAACAGGA?AAATAATAGGAGATATAAGACAAGCACATTGT   105
HIVSC       --------G----------------C--G-------------......----------------------G--------G-C-------------------------------   105
HIVMN       ---------------T------A---------G---------......-----------------------------AA-A-T---------AC-------------------   105
HIVJH3      -------------G---G-C--------G---G---------......-----------------------------AA-C-----GC-------T------------------   105
HIVWMJ12    ---------------------GT-----G---...--------CATATA---------------------...------G------G----A--------------------   105
HIVWMJ2     ------------T--------GT-----G----C--TC----......-------------------CG-----G--A-AT----...---AT-------------------   102
HIVWMJ32    ---------------G---T-G------G-C-----------......-----------------------GG-AA-AT----...---A--------------------   102
HIVSF2      ----------------------------------CT-----......-------------------C-----------AG-------------------A-----------   105
HIVRF       ------------------------------------AC--AG......-------------T-A-----G--------C--------------------A-----------   105
HIVBRVA     ----------------------------G--G---AC---G......-------------T--A-------------C---------------------G----------   105
HIVJRCSF    --------G----G-------------------------------......-------------------------G-------------------------------   105
HIVJRFL     ------------------------------------------......-----------------------T------G-------------- ----------------   105
HIVCDC42    ------------------C-------------GG--ACGC--......-------------T--GG------------G-----C-----A-------G------------   105
HIVHAN      -------------------- ---------G-----------......---------------G-------G------AG----G---------------T----------   105
HIVNY5NEW   -------------------- ---------G-----GC----......-----G------A--C-C---G---G--A-A-------------------------------   105
HIVJFL      -----------T-------- ---------------CAC-C--......-----------G-------------------G-C-----------C------------------   105
HIVADA      ---------------------------G---------------......--------A---------------------G-------------------------------C   105
HIVSF162    -----------T-----T-------------------AC----......-----G---------------G--------G-C-------------------------------   105
HIVSF33     ----------------------G-----G---G---AC--GT......----------A--T-C--------------G--------------------A-----T-----   105
HIVALA1     ----------------T-T---G--A-GG---G---------......-------------------C--------A--C---------A-A--------------------   105
HIVBAL1     ------------------------------------------......--------C---------------------G-------------------------------   105
HIVBAL2     ------------------------------------A-----......--------C---------------------G-------------------------------   105
HIVCAN0A    ------------C----------------------------G......----------A-------------------GG-------------------------------   105
HIVCAN0B    ------------C-----------------------A-----......-------------------------------GC-------------------------------   105
HIVFO       ---------------------------G------T---T-T......-------------------------------C-----------A----------G---------   105
B1.50CON    --------------------------T-G-------------......-------------------T-G------A-AG-------------------------------   105
S1.50CON    ---------------------------G-------AC---G......-------------T-C--------------G-------------------------------   105
B2.50CON    ------------------------------------------......--------C---------------------G-------------------------------   105
S2.50CON    ------------------------------G-----------......--------C----------T----------G-------------------------------   105
B3.50CON    --------------T---------------G--G--------......-----------------C----------...-C---G--------------------------   102
S3.50CON    ------------------------------G--G--------......-------------------------------...-C---G--------------------------   102
V77.50CON   -------------G------------------------C---......-----G-------T----C--G--------G-------------------------------   105
V82.50CON   ---------------------------------G---T-----......---------------G------------A-C------ ----A--------------------   105
V87.50CON   ------------GGT------------G----?---TCA---......-----G------------AT-G---G----C-------------------------------   104
V91.50CON   ------------GG-----------------------CA---......-----G------------AT-G-----A-CC--------------------A-----------   105
V12.50CON   -------------G---?----T---G-G------?------......---T---------------------T--A-GG-G--?C-------G-----A-----T-----   102
V74.50CON   ---------------------------G-G-----------......---------------------G--------A-C-------------------------------   105
Guo.con.JB  ------------------------------------------......---------------------G--------G-------------------------------   105
W.H.con.DL  -----------------T---------G----GC--AG----......-------------C--------G-----A--AG-------------------------------   105
W.H.con.JB  ------------------------------------------......---------------------?--------G-------------------------------   104
W.H.con.LAI ----------------------------------C-G---CCAGAGA-------------------------------A-----...---A----G---------------   108
HIVLAI      ----------------------------------C-G---CCAGAGG-----------------GT-----T----A-----...---A----G---------------   108
HIVHXB2R    --------------------------------A--C-G---CCAGAGA-----------------GT-----T----A-----...---A----G---------------   108
```

*The N. American/European consensus is a 50% consensus based on the sequences shown here.

A "?" in the consensus sequences means that no one base was present in a particular position in more than 50% of the sequences grouped together to create the consensus.

IIIC3. Nucleic Acid Alignments:

frican V3 Nucleotide Alignment 1:

```
f1.con* TGTAC?AGACCTAACAACAATACAAGA.....AAAAGGATA??TATAGGACCAGGACAAGCATTCTATGCAACAGGTGA?ATAATAGGAGATATAAGACAAGCACATTGT    105
IVSF170 -----C----------.....-----TG--CG-----------C--------G---------T-----    105
IVZ321  ----TG----------.....-----T---TC----------G--------T----------C----------G-----C---    105
IVOYI   -----A-----C----------.....--C------TCA----------GAG------TC--A-----AAAC-A----------    105
IVMAL   -----A--G---GGA----------.....-G-G-----CA-T-C--C-----G------C-----A-----...-GG---G----------AG-----T-----    102
IVU455  ---T-C------T----T-CA-G--A-AATATACGT---TATAG-------T----------T-------A-A--------G----------    111
```

frican V3 Nucleotide Alignment 2:

```
f2.con* TGTACAAGGCCC...TACAAAAATACAAGACAAAGTACA. .....CCTATAGGACTAGGGCAAGCACTCTATACAACAAGAA?AA?AA?A.....???A?AATAGGACAAGCACATTGT    105
IVELI   ---G---------...--TC----------A---.....----------T----------T------TC--G-TC-.....-T----------    102
IVZ6    ----------...----------.....----------G----------GGT-G--C-...AAA.....-T----------    105
IVZ226  ----------...----G-----T---------G---.....T---------T----------G----------A--C--G--GT.....-T----------T-----    102
IVMDK   ----------...-------T-------G-----G---.....T-G---------A-----T----------T--C-GG--A--AGAAGAAAACAGGATAC----------    114
IVCBL4  --C---------TAC----------T----A--G-----CATGTAGGA-A-G------C---A-G----A----------AC-T--A-GG-.....AGTGGA-AT---A----------    120
IVJY1   ----------GACA------T---------.....----------G-T--A-.....GGAGAT---A--------T-----    105
```

The consensus sequence for each set is based on a 50% consensus of the sequences shown here.

IIIC4. Intrapatient sequence sets

Sibling sequences are derived from viruses obtained from the same individual. Sets of sequences from the same individual are referred to in this section as sib1, sib2 ... In order to facilitate comparisons between data sets, we have aligned the sequences presented here to the N. American/European consensus, as in section C2.

The Edinburgh sequences (Balfe *et al., J. Virol.* 64:6221–6233 (1990) and Simmonds *et al., J. Virol.* 64:5840–50 (1990)) and the Epstein sequences (Epstein *et al., Virology* 180:583–590 (1991)) often contained more than one sequence that was identical within the V3 loop. Whenever there was identity in V3 amino acid sequences between two or more sequences from a single individual, we included the sequence in this section only once. A tag was included on the sequence reference name to indicate how many sequences were identical within a sibling set to the one shown. For example, B1.DIR.con3seqs, the first sequence in the Epstein set, was exactly the same as two additional B1 sequences in the V3 loop; the fact that it was present three times is indicated by the "con3seqs". Note that identity within the V3 loop does not imply identity in the surrounding sequences, nor at the nucleic acid level.

Also included in this section are sibling sequences from the patient BRU, designated "JB", and sequences from the patient LAI. Viral isolations from both of these patients were attempted at the Pasteur Institute in 1983 (Wain-Hobson *et al, Science* 252:961–64 (1991)). Apparently, virus from the patient LAI contaminated a culture which was thought to be from the patent BRU, and hence LAI was the source of the virus previously known as HIVBRU, now called HIVLAI. Due to the high degree of similarity between the sequences of HIVLAI and HIV-IIIB, these viruses are thought to have originated from the same isolate. For further clarification of the origins of these viral sequences, see Wain-Hobson *et al., Science* 252:961–64 (1991), Guo *et al., Nature* 349:745–6 (1991), Gallo, *Nature* 351:358 (1991).

The sibling sequence assignments previously made available for the LaRosa data set (LaRosa *et al., Science* 249:932–35 (1990)) in the database update dated Jan. 1991 and in GenBank, have been withdrawn by the authors, and are currently under reevaluation; therefore sibling assignments of the LaRosa data set are not presented here.

V3 sequences originating from the patient BRU; the virus is now designated HIVJB:
Wain-Hobson *et al.*, *Science* 252:961–64, (1991):

```
CONSENSUS*          CTRPNNN..TRKSIHI..GPGRAFY.TTGEIIGDIRQAHC     35
HIVJB.A046          -------..-------..-------.A----------        35
HIVJB.A051          -------..-----N-..-------.----------        35
HIVJB.A045          -------..-------..-------.------R-------     35
HIVJB.A049          ----S--..-------..-------.A----------        35
HIVJB.A053          -------..-------..-------.----------        35
HIVJB.A055          -------..-------..-------.A--K----------     35
HIVJB.A060          -------..-----P-..-------.---D----------     35
HIVJB.A064          -------..-------..-------.A---------H---     35
HIVJB.D24           -------..-------..-------.A----------        35
HIVJB.D29           ------D..-------..-------.A----------        35
HIVJB.D30           -------..-----R-..-------.A----------        35
HIVJB.B067          -------..-------..-------.A----------        35
```

The "A" series was sequenced directly from a stored lymph node biopsy from the patient BRU, the "D" series from a 1983 culture of BRU virus with JBB donor cells, and B067 from a 1990 culture of the original isolate from the man BRU.

Guo *et al.*, *Nature* 349:745–46 (1991):

```
HIVJB.con6seqs      -------..-------..-------.A----------        35
HIVJB31             -------..-------..--K----.A--E---R-------    35
HIVJBD221           -------..--RR-S-..------R.AA-Q------R---     35
```

The Guo sequences are from a 1983 sample of the BRU-JBB donor cell culture sent from the Pasteur Institute to the Laboratory of Tumor Cell Biology at the NIH.

*Sequences are aligned with the N. American/European consensus sequence.

V3 sequences originating from the patient LAI, formerly called HIVBRU and originally called LAV-1, but now called HIVLAI:

Wain-Hobson *et al.*, *Science* 252:961–64, (1991):

```
CONSENSUS*        CTRPNNN..TRKSIHI..GPGRAFY.TTGEIIGDIRQAHC      35
HIVLAI.L7         -------..-----R-QR------V.---K-.--M-----      36
HIVLAI.L3         -------..-----R-QR------V.-I-K-.-NM-----      36
HIVLAI.L2         -------..-----R-QR------V.---K-.-NM-----      36
HIVLAI.L6         -------..-----R-QR----T-V.-I-K-.-NM-----      36
HIVLAI.L9         -------..-----R-HR-------.-I-K-.-NM-----      36
HIVLAI.L4         -------..-----P-..-------.--------------      35
HIVLAI.L20        -------..-------..-------.-------N------      35
HIVLAI.L16        -------..-----R-QR-------.--------------      37
HIVLAI.E70        -------..-----R-QR------V.-I-K-.-NM-----      36
HIVLAI.E79        -------..-----R-QR------V.-I-K-.--M-----      36
HIVLAI.E69        -------..-----R-QR-----V.-I-K-.RNM-----      36
HIVLAI.E86        -------..-G---R-QR------V.-I-K-.-NM-----      36
HIVLAI.K140       -------..-----R-QR------V.-I-K-.-NM-----      36
HIVLAI.K139       -------..-----R-QR------V.-I-K-.-NV-----      36
HIVLAI.K138       -------..-----R-QEE-K---V.-IRK-.EN------      36
HIVLAI.K145       -------..-----C-QR------V.-I-K-.-NM-----      36
HIVLAI.K147       -------..-----R-QR------V.-I-K-.-N-----Y      36
HIVLAI.K155       -------..---G-R-QR------V.-I-K-.-NM-----      36
HIVLAI.K144       -------..-----R-QR------V.-I-K-.--M-----      36
HIVLAI.K152       -------..-------..-------.A-------VG----      35
HIVLAI.H117       -------..-----R-QR------V.-I-K-.-NM-----      36
HIVLAI.H116       -------..-----R-QR------V.-L-K-.-NM-----      36
HIVLAI.H127       -------..-----R-QR------V.-I-K-.ENM-----      36
HIVLAI.H130       ------D..-----R-QR------V.-I-K-.-NM-----      36
HIVLAI.H135       ------...-----R-QR------V.-I-K-.-NM-----      35
```

The "L" series above is from a peripheral blood sample from the patient LAI taken in 1983 and stored. The "E" series is from the original culture of LAI's virus with donor JBB cells. The "K" series is from a bone marrow culture (M2T-/B) from which supernatant was sent to the Laboratory of Tumor Cell Biology at the NIH. Originally M2T-/B was thought to have been derived from cultured virus from the patient BRU; apparently M2T-/B was contaminated by virus from the patient LAI. The "H" series is from a B-cell adapted culture that originated from M2T-/B.

For comparison, some of the early sequenced isolates that are very similar to the sequences from LAI are shown:

Wain-Hobson *et al.*, *Cell* 40:9–17 (1985):

```
LAV               -------..-----R-QR------V.-I-K-.-NM-----      36
```

Ratner *et al.*, *Nature* 313:277–284 (1985):

```
BH10              -------..-----R-QR------V.-I-K-.-NM-----      36
BH8               -------..---K-R-QR------V.-I-K-.-NM-----      36
```

*Sequences are aligned with the N. American/European consensus.

EDINBURGH SIBLING SEQUENCES:

```
CONSENSUS*          CTRPNNN..TRKSIHI..GPGRAFY.TTGEIIGDIRQAHC     35
sib1:
V77171.con2seqs     ----S--..-------..A---V-H.A-----------       35
V77173.con4seqs     ----S--..-------..----V-H.A-----------       35
sib2:
V872420.con4seqs    ----G-Y..--ER-S-..------I.ARRQ----------     35
V872421             ----G--..--ER-S-..------I.AR-Q----------     35
V872422.con4seqs    ----S--..-----S-..-------.A-----------       35
sib3:
V91605              ----G--..-----P-..------I.A-SQ------K---     35
V91609              ----GD-..-----P-..-------.A--D----------     35
V91611.con2seqs     ----G--..---R-S-..------I.A-SQ------K---     35
V91613              ----G--..-----P-..-------.A--D----------     35
sib4:
V74164.con3seqs     -------..--RG---..-------.A--N----------     35
V74167              --X----..--RG---..-------.AARK----------     35
V74169              ----S--..--RG---..-------.A--N----------     35
V74179              ----G--..--R----..WM-----.A-----------       35
sib5:
V82711.con2seqs     -------..---R-Y-..-----V-.--EQ---N------     35
V82712.con2seqs     -------..---G---..---S---.A--G----------     35
V82716              -------..---R---..-----V-.--EQ---X------     35
V82717.con2seqs     -I-----..---R-Y-..-----V-.--EQ---N------     35
sib6:
V12834              ----SK-..I-R----..-S-----.-IEGVA--V-K-Y-     35
V12836              ----S--..K-R-M--..-S-----.-IEGVT--V-K-Y-     35
V12837              ----SK-..I-R-M--..-S-----.-IEGVA--V-K-Y-     35
V12840              ----S-T..--R----..-------.----VT--VEK-Y-     35
```

The patient V12 is an HIV infected American hemophiliac. The other patients listed here are from a cohort of HIV infected hemophiliacs from Edinburgh, who had been exposed to an HIV contaminated batch of factor VIII in 1984. Patients V77, V87, and V91 (sibs 1, 2, and 3) had closely related sequences, and were thought to be infected by related viruses. Balfe *et al.*, *J. Virol.* 64:6221–33 (1990), and Simmonds *et al.*, *J. Virol.* 64:5840–50 (1990).

*The sequences are aligned with the N. American/European consensus.

EPSTEIN ET AL. SIBLING SEQUENCES:

```
CONSENSUS*        CTRPNNN..TRKSIHI..GPGRAFY.TTGEIIGDIRQAHC      35
sib1:
B1.DIR.con3seqs   -------..--R--N-..-----L-.---A----------      35
B1.IA1.con3seqs   ----S--..-S-----..-------.A-ER-T--------      35
B1.IA2.con5seqs   -------..-SRR-R-..------V.ARER----------      35
B1.IB3            -------..-KR--N-..-----L-.---A----------      35
B1.IB5            -------..-SRR-S-..------V.ARER----------      35
B1.IC4            -------..-------..------F.--------------      35
B1.I.B1           --K----..-SRR-R-..K--K--V.ARER---N------      35
S1.DIR            -------..--R--TM..----VL-.---A--------D-      35
S1.IA10           -------..--R--TM..----VL-.A--Q----------      35
S1.IA2            -------..--R--N-..-----L-.---A----------      35
S1.IA6.con4seqs   -------..--RR-TM..----VL-.---Q------K---      35
S1.IA9.con2seqs   -------..-------..-------.A--D----------      35
S1.IB2            -------..--R--N-..---X-L-.---A----------      35
S1.IB4            -------..--R--TM..----VL-.---A----------      35
S1.IB7            -------..--X---X..-------.A--D----------      35
S1.IIA9           -------..--RR-TM..----VP-.---Q------K---      35
S1.IIB10          -I-----..---G---..------F.--------------      35
sib2:
B2.DIR            -------..-------..-------.--------------      35
B2.IIB2           -I-----..-------..-------.--------------      35
B2.IIB7           -------..---G---..-------.----M---------      35
B2.I.B1           -A---S-..-------..-------.--------------      35
B2.I.B2.con2seqs  -------..-------..-------.---G----------      35
B2.I.B10          -------..---G---..-------.---D----------      35
B2.I.B7           -I-----..---G-Q-..------F.--------------      35
S2.DIR            -------..---G---..-------.--------------      35
S2.IA5            -------..-----P-..-------.--------------      35
S2.IA6            -I-----..-----P-..------F.--------------      35
S2.IA7            -------..-------..------F.--------------      35
S2.1B3            -------..---G---..------F.--------------      35
S2.I.A11          -------..---G-Q-..-----NF.--------------      35
S2.I.A3           -I-----..---G---..---G---.--------------      35
sib3:
B3.DIR.con5seqs   -------..---GV--..-------.--.D-V--------      34
B3.IB3            -------..---G--M..-------.A--D----------      35
S3.IIC1           -------..---GV--..R------.--.D-V--------      34
S3.IIC4.con3seqs  -I-----..---GV--..-------.--.D-V--------      34
S3.IID1           -------..-------..-------.A--D----------      35
S3.II.D7          -------..---GV--..-------.--.D----------      34
S3.II.C5          -------..---K---..R-K----.---Q--RN--H---      35
```

B stands for brain and S for spleen; the number following refers to patients 1, 2 and 3 (Epstein *et al.*, *Virology* 180: 583–90 (1991)). "DIR" indicates that the sequencing was done directly from the sample—the other sequences were from clones. A direct sequence was obtained for samples B1, S1, B2, S2, and B3.

*The sequences are aligned with the N. American/European consensus.

IIIC5. V3 Loop Similarity Distributions

The similarity distributions in this section reflect the range of variation found among V3 loop nucleotide sequences. First, the similarity between 500 V3 loop nucleotide sequences was calculated. For these initial comparisons, no distinctions were made between intra- versus interpatient sequences, or for the geographic origin of the sequences (N. American, European, and African sequences were included) (Fig. 1). This then was contrasted with intrapatient variation from non-cultured sequence sets (Fig. 2), and with the interpatient variation found among African sequences (Fig. 3). While the overall similarity (including both inter- and intra-patient sequence comparisons) ranged from identity (100%) to 50% similar, the average similarity was around 90%. In contrast, sets of sequences from a single individual tended to be more highly conserved, and comparisons of 17 African viral sequences revealed much greater heterogeneity than is usually found between N. American and European isolates. Pairwise sequence similarity scores were calculated by dividing the number of homologous bases in two aligned sequences by the total number of bases in the aligned sequences. For the V3 loop nucleotides there are typically 105 bases. The percent similarity was rounded off for each pairwise comparison (98.4 and 97.7 would both be considered 98% similar) and the number of comparisons with a given percent similarity within a set was calculated. This number was converted to the fraction of the total number of comparisons within the set to have a particular percent similarity, and is presented in a histogram. Generally for similarity comparisons of multiple aligned sequences, an entire column is eliminated from the alignment if any sequence included in the alignment has a gap at a given position (Li *et al.*, *J. Mol. Biol. Evol.* 5:313–330 (1988)). Because of the limited number of positions available in the V3 loop, and the high frequency of insertions and deletions, we modified this approach here and simply excluded positions in each pairwise sequence similarity comparison for which either of the two sequences under consideration had a gap inserted.

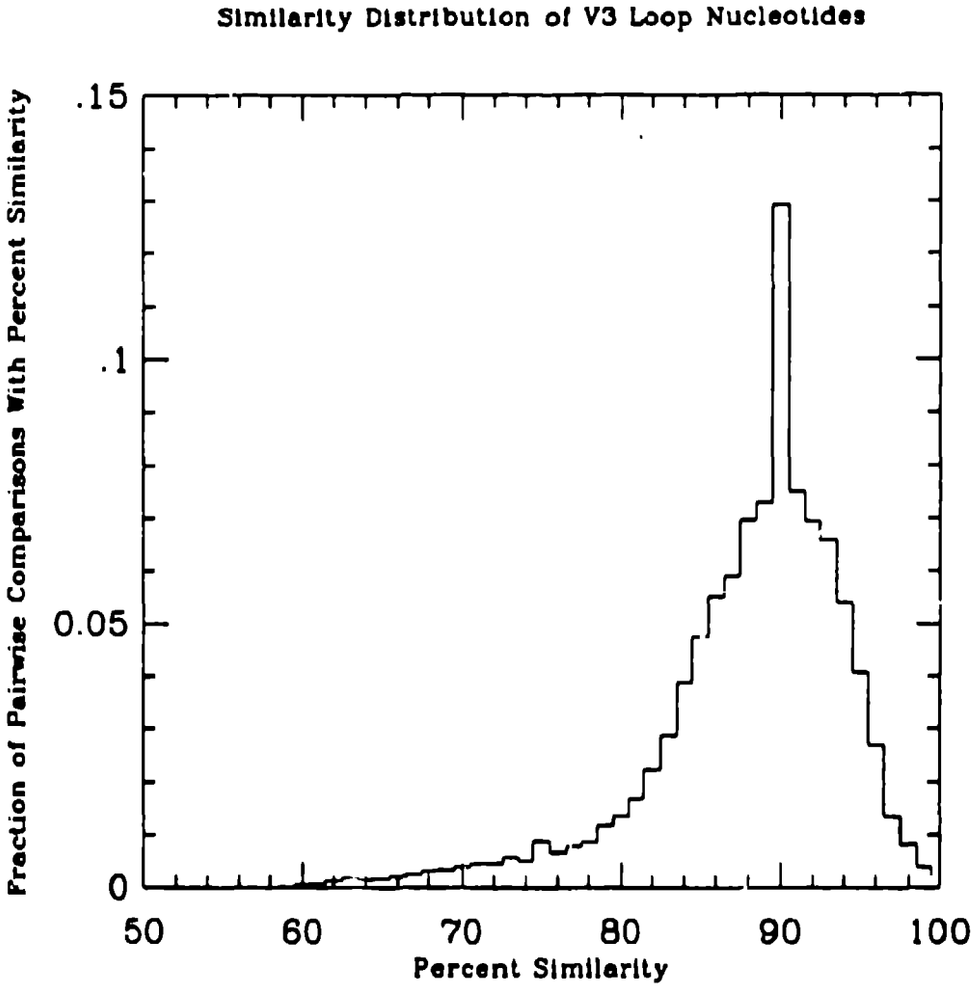


Figure 1. Similarity Distribution of Nucleotide Sequences Encoding the V3 Loop. Similarity comparisons were made between all nucleotide sequences encoding the sequences shown in section C2, including the N. American/European set and the African sets, as well as the nucleotide sequences from LaRosa *et al.*, *Science* 249: 932-35 (1990)), and nucleotide sequences from data that have not yet been published. No distinctions were made for inter- versus intrapatient comparisons, and this figure is meant only to show the overall potential variation within the sequences encoding the V3 loop. This distribution is based on 124,750 pairwise sequence comparisons of 500 sequences.

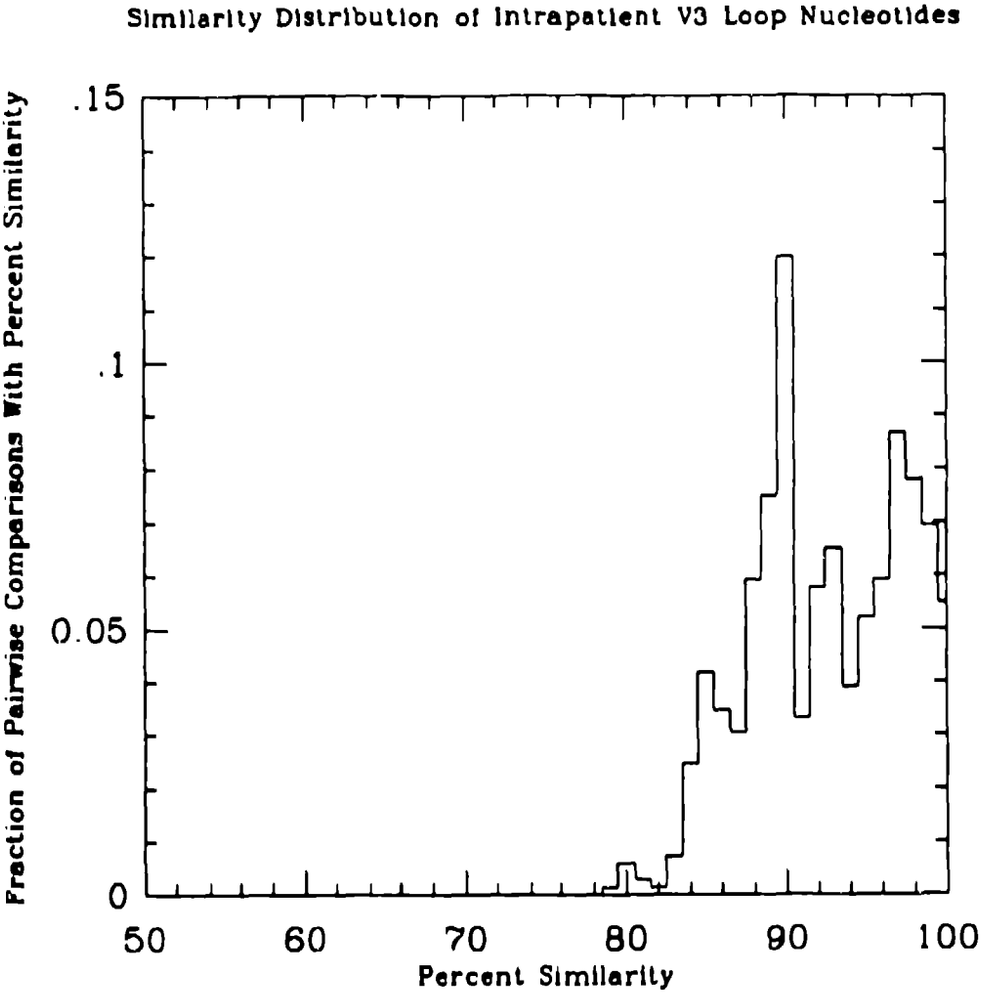


Figure 2. Similarity Distribution of Uncultured Intrapatient V3 Loop Nucleotide Sequences. This distribution is based on similarity comparisons of sibling sequences from each patient in the hemophiliac cohort (Balfe *et al.*, *J. Virol.* 64:6221–6233 (1990) and Simmonds *et al.*, *J. Virol.* 64:5840–50 (1990)), and from brain and spleen tissue samples (Epstein *et al.*, *Virology* 180:583–90 (1991)). The sibling sequence sets are shown in section C4; the corresponding nucleotide alignments are shown in part I of the HIV database. This distribution is based on 692 pairwise comparisons of sequences derived from sibling viruses, from the nine individuals described in the two studies cited above, with a range of 4–29 sequences available per sibling set.

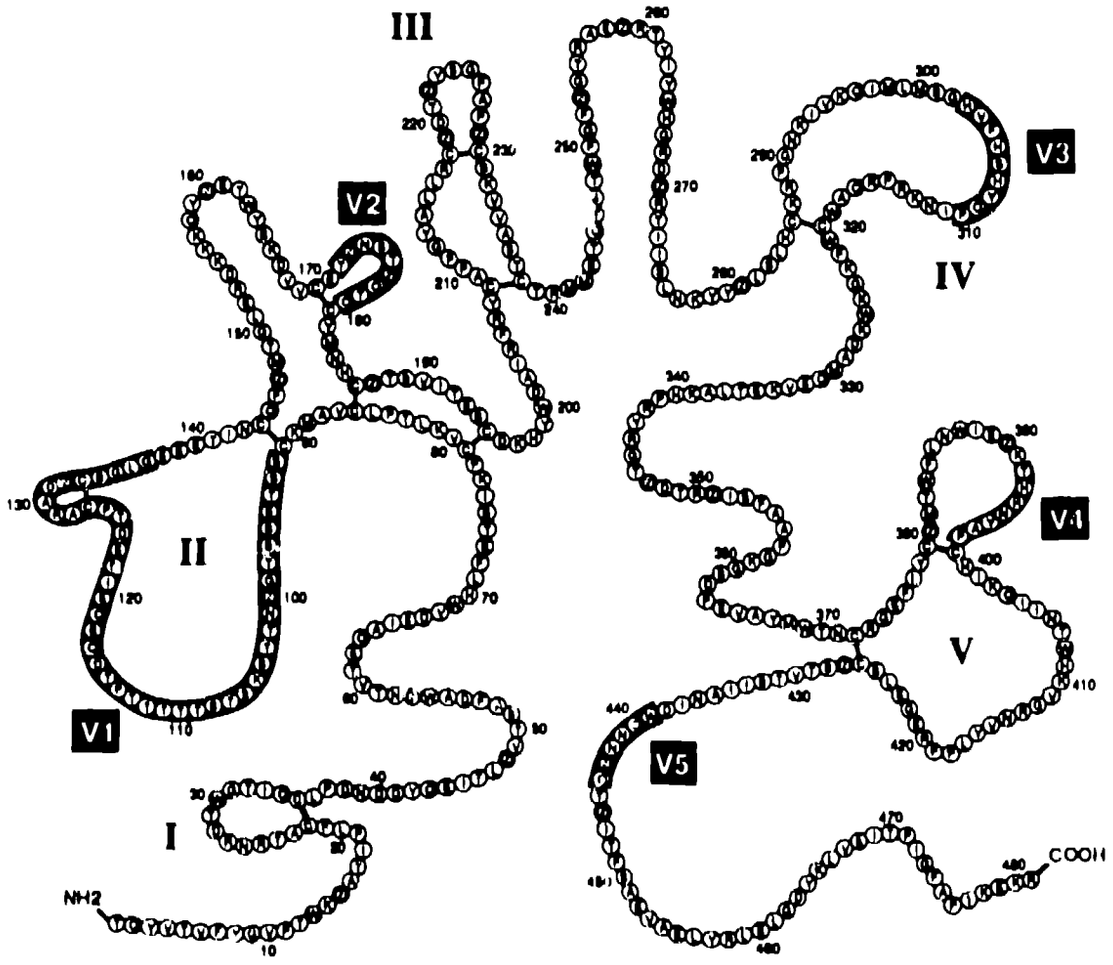
```
CONSENSUS  aaaTTcCActTaCCaGTaGAgAgaGA?atcTGGGAaCAgTGGTGGgcagAcTAtTGGCAaGTaACaTGGa  1764
HIV2ROD    AAA--T--CC-A--A--A--G-GA--AATC-----G--G------GATA-C--C-----A--G--A---A  1780
HIV2NIHZ   AAA--T--CC-A--A--A--G-GA--TACC-----G--G------GATA-C--C-----A--A--A---A  1777
HIV2ISY    AAA--T--CC-A--A--A--A-GA--GACC-----A--G------GATA-C--T-----A--G--A---A  1777
HIV2ST     AAA--T--CC-A--A--A--A-GA--TACC-----A--G------GATA-C--C-----A--G--A---A  1891
HIV2BEN    AAA--T--CC-G--A--G--G-GA--GACC-----G--G------GATA-C--C-----A--G--A---A  1837
HIV2D194   AAA--T--CC-A--G--A--G-GA--CACC-----G--A------GATA-C--C-----A--A--A---G  1891
HIV2GH1    AAA--C--CT-A--G--G--G-GA--GACC-----G--G------GATA-C--C-----A--G--A---A  1837
HIV2D205   GTG--C--TC-G--A--A--A-GA--GACA-----C--G------ACAG-T--C-----A--A--C---A  1837
SIVMM251   AAA--C--CT-A--A--T--G-GG--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1840
SIVMM142   AAA--C--CT-A--A--T--G-AG--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1840
SIVMM239   AAA--C--CT-A--A--T--G-AG--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1852
SIVMNE     AAA--C--CT-A--A--T--G-AA--TGTA-----A--G------ACAG-C--T-----G--A--C---A  1840
SIVSMMH4   AAA--C--TT-G--A--A--A-GA--AATT-----A--A------ACAG-T--T-----A--A--C---A  1729
SIVSMMPBJ  AGA--T--TT-G--A--A--G-GA--AATT-----A--A------ACAG-T--T-----A--A--T---A  1738

CONSENSUS  TcCCagA?TGGGAcTTtgT?TC?ACcCCaCCacTagTcAGatTagt?TttAAccTaGTaaaaGA?CCtaT  1829
HIV2ROD    -C--AG-C-----C--CG-G--T--C--A--AC-GG-C--GT-AGCG-TT--CC-G--AGGG--T--TA-  1850
HIV2NIHZ   -C--AG-C-----C--TG-A--T--C--A--AC-GG-C--GC-AGCA-TT--CC-G--AGGA--G--TG-  1847
HIV2ISY    -C--AG-C-----C--CG-A--C--C--A--GT-GG-C--GT-AGCA-TT--CC-G--AAAA--T--TA-  1847
HIV2ST     -C--AG-C-----C--CA-A--T--C--G--AC-GG-C--AT-AGTA-TT--CC-G--GAAA--T--CA-  1961
HIV2BEN    -C--AG-G-----C--TG-A--T--C--A--AC-GG-C--GT-AACA-TT--CC-A--AGGA--T--TA-  1907
HIV2D194   -C--AG-G-----C--CG-A--T--C--A--AC-GG-C--GT-GACA-TT--CT-G--AGGA--T--TA-  1961
HIV2GH1    -C--AG-G-----C--CG-A--T--C--A--AC-AG-C--GT-AACA-TT--CC-A--AGGA--T--TA-  1907
HIV2D205   -C--AG-G-----C--TG-C--G--C--A--AT-AA-A--AC-AGCC-AC--CC-A--CAAA--C--CC-  1907
SIVMM251   -A--GG-G-----T--TA-C--A--G--A--AC-AG-A--AT-AGTC-TC--TC-A--GAAG--C--TA-  1910
SIVMM142   -A--GG-A-----T--CA-C--A--A--A--AT-AG-A--AT-AGTC-TC--TC-A--GAAG--C--TA-  1910
SIVMM239   -A--GG-A-----T--TA-C--A--A--A--GC-AG-A--AT-AGTC-TC--TC-A--GAAG--C--TA-  1922
SIVMNE     -A--GA-A-----T--TA-C--A--A--A--AC-AG-A--AT-AGTC-TC--TC-G--AAAG--C--TA-  1910
SIVSMMH4   -A--AG-T-----C--TG-G--A--A--T--CT-AG-C--AT-AGTC-TC--CC-A--AAAA--G--TA-  1799
SIVSMMPBJ  -A--AG-A-----C--TG-G--A--A--T--CT-AG-C--AC-AGTT-TC--CC-A--AAAA--G--TA-  1808

CONSENSUS  acaaGG?gcAGAaACcTttTAcacaGATGGaTCaTGcAatAgacagTCaAaaGAaGGaAAaGCAgGaTAt  1898
HIV2ROD    ACCA--TGC---G--C-TC--CACA-----A--C--C-AT-GGCAA--A-AA--A--A--A---G-A--T  1920
HIV2NIHZ   ACCA--CGC---A--T-TC--CACA-----A--C--C-AT-GGCAG--A-AA--A--G--A---G-A--T  1917
HIV2ISY    ACCA--CGC---G--C-TC--CACG-----A--T--C-AT-GGCAA--A-AA--G--A--A---G-A--T  1917
HIV2ST     ACTA--CGC---A--C-TC--CACA-----A--C--C-AT-AGCAA--A-GA--A--A--A---G-A--C  2031
HIV2BEN    ACCA--CGC---G--C-TC--CACA-----A--A--C-AT-GACAG--A-AA--G--A--A---G-A--T  1977
HIV2D194   ACCA--CAC---G--C-TT--CACA-----A--A--C-AT-GACAG--A-AA--A--A--A---G-A--T  2031
HIV2GH1    ACCA--CGC---G--C-TC--CACA-----A--A--C-AT-GACAG--A-AG--A--A--A---C-A--T  1977
HIV2D205   AGAA--GAG---A--C-AC--CACA-----G--C--C-AT-GAACC--A-AG--A--A--A---G-A--T  1977
SIVMM251   AGAG--AGA---A--C-AT--TACA-----A--A--T-AT-AACAG--A-AA--A--G--A---G-A--T  1980
SIVMM142   AGAG--AGA---A--C-AT--TGTA-----A--A--T-GT-AACAG--A-AA--A--A--A---G-A--T  1980
SIVMM239   AGAG--AGA---A--C-AT--TACA-----A--A--T-AT-AACAG--A-AA--A--G--A---G-A--T  1992
SIVMNE     AAAG--AGA---A--C-AT--TGTA-----A--A--T-AT-AACAG--A-AA--A--G--A---G-A--T  1980
SIVSMMH4   ACAG--AGC---A--A-TT--TGTA-----A--C--T-AT-GGCAG--A-GA--A--A--A---G-C--T  1869
SIVSMMPBJ  ACAG--AGC---A--A-TT--TGTA-----A--C--T-AT-GGCAG--A-GA--A--A--G---G-C--T  1878
```

IIID. GP120 MODELS

In the SEP 90 release, on page III-52, a model for the HIV-2 gp120 protein structure was presented by Tim Gregory and colleagues. Since that time, those authors have improved their model (Gregory *et al.*, "Structure and Function in Recombinant HIV-1 gp120 and Speculation About the Disulfide Bonding in the gp120 Homologs of HIV-2 and SIV" in Advances in Experimental Medicine and Biology, in press, 1991), and those changes can be seen herein.

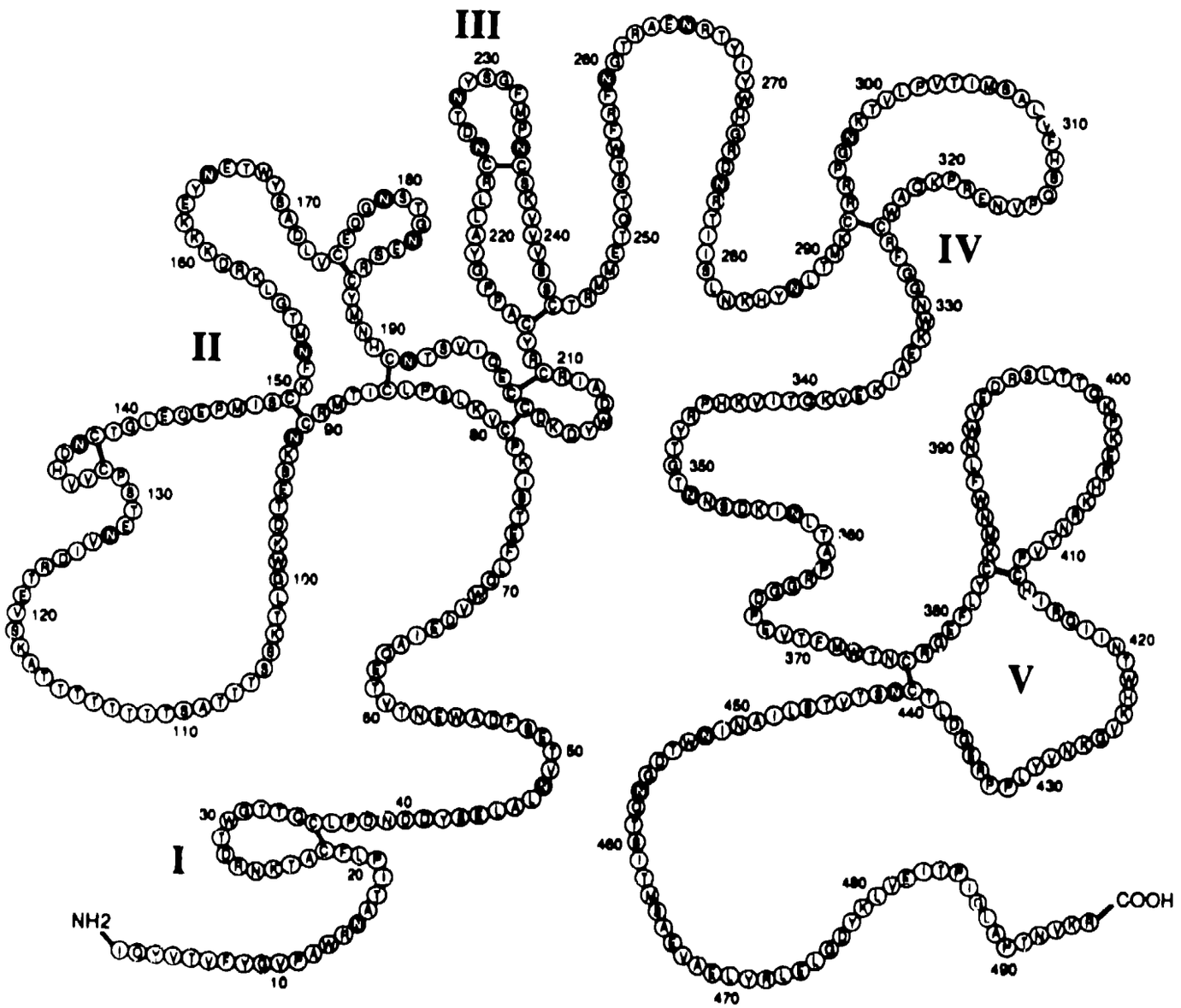
Again, these models have not been experimentally verified in the way that the HIV-1 gp120 structure was verified (see III-51 of SEP 90 and Leonard *et al.*, *J. Biol. Chem.* 265:10373–10382, 1990). For further information, contact Dr. Timothy Gregory, Genentech, (415)-266-1511.

HIV-2/ROD gp120 (Proposed)



● potential glycosylation site

SIV-MM142 gp120 (Proposed)



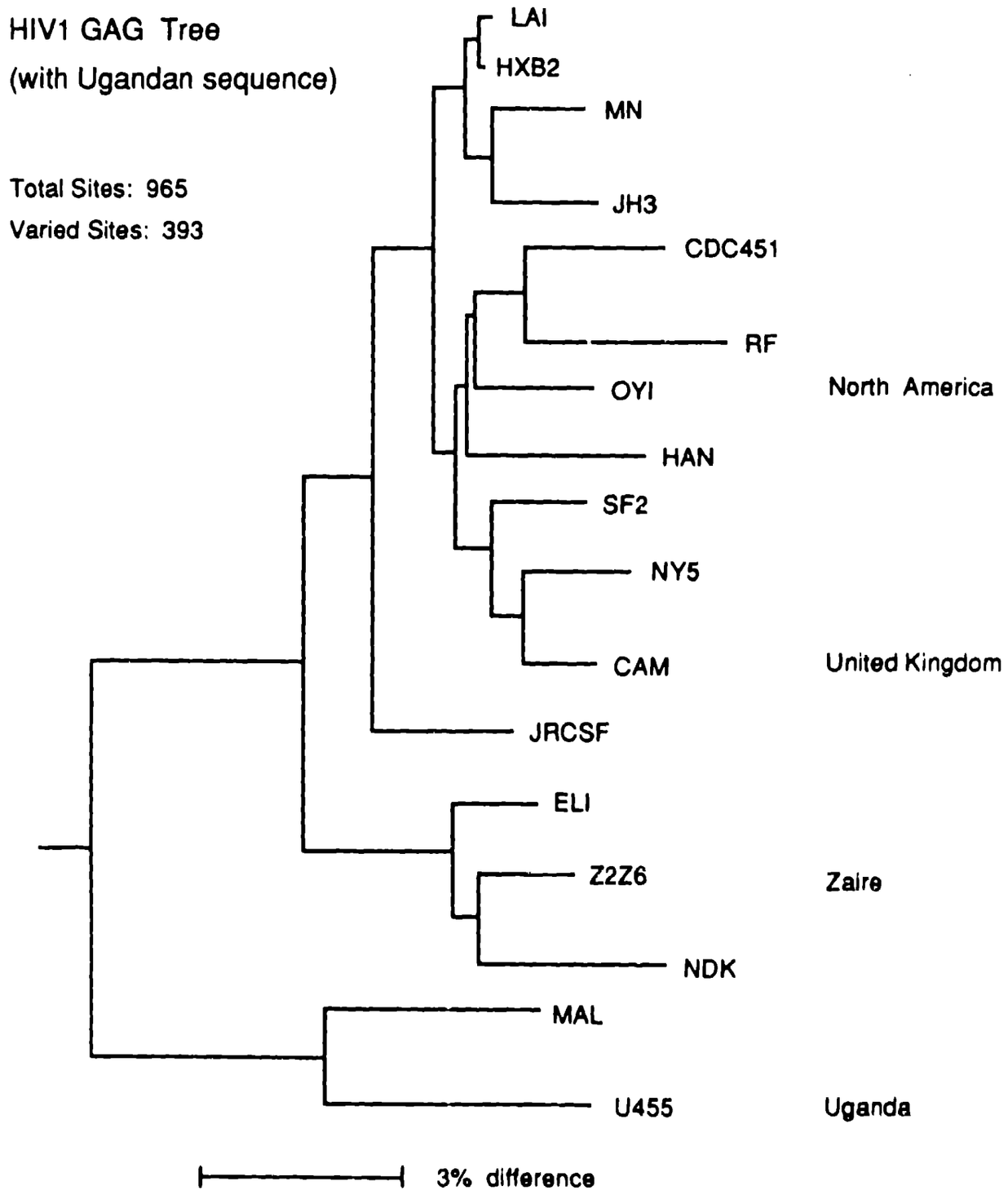
SEQUENCE DISTANCES and PHYLOGENETIC TREES

In the 1989 compendium (III-65 APR 89), a general discussion of HIV-1 nucleotide sequence distance relationships was initiated. At that time, sibling sequences representing viral forms taken from the same patient were found to differ from one another by up to 3.3% in the *env* cds. *Env* sequences representing different HIV-1 infected persons in North America, sampled between 1983 and 1986, differed by 6% to 12%, and comparable samples from Zaire covering the same time period differed by typically 16%. The first result, representing intrapatient variation, was underestimated because it was based upon short-term or long-term cultured viral samples. In recent studies it is more commonly the case that sibling sequences differ by 4% to 6% when viral samples are directly amplified and sequenced (see Meyerhans *et al.*, *Cell* 58, 901, 1989; Balfe *et al.*, *J. Virol.* 64, 6221, 1990; Delassus *et al.*, *J. Virol.* 65, 225, 1990; Vartanian *et al.*, *J. Virol.* 65, 1779, 1991; Wain-Hobson *et al.*, *Science* 252, 961, 1991). Mullins and colleagues at Stanford have found a three-fold difference in the extent of variation, typically, between cultured and uncultured samples (in press, 1991); in contrast to these results, exceptionally homogeneous sets of sibling and close cousin (linked) sequences have been reported by McNearny *et al.*, *PNAS* 87, 1917, 1990 and by Kleim *et al.*, *AIDS Res. Hum. Retroviruses* 7, 417, 1991.

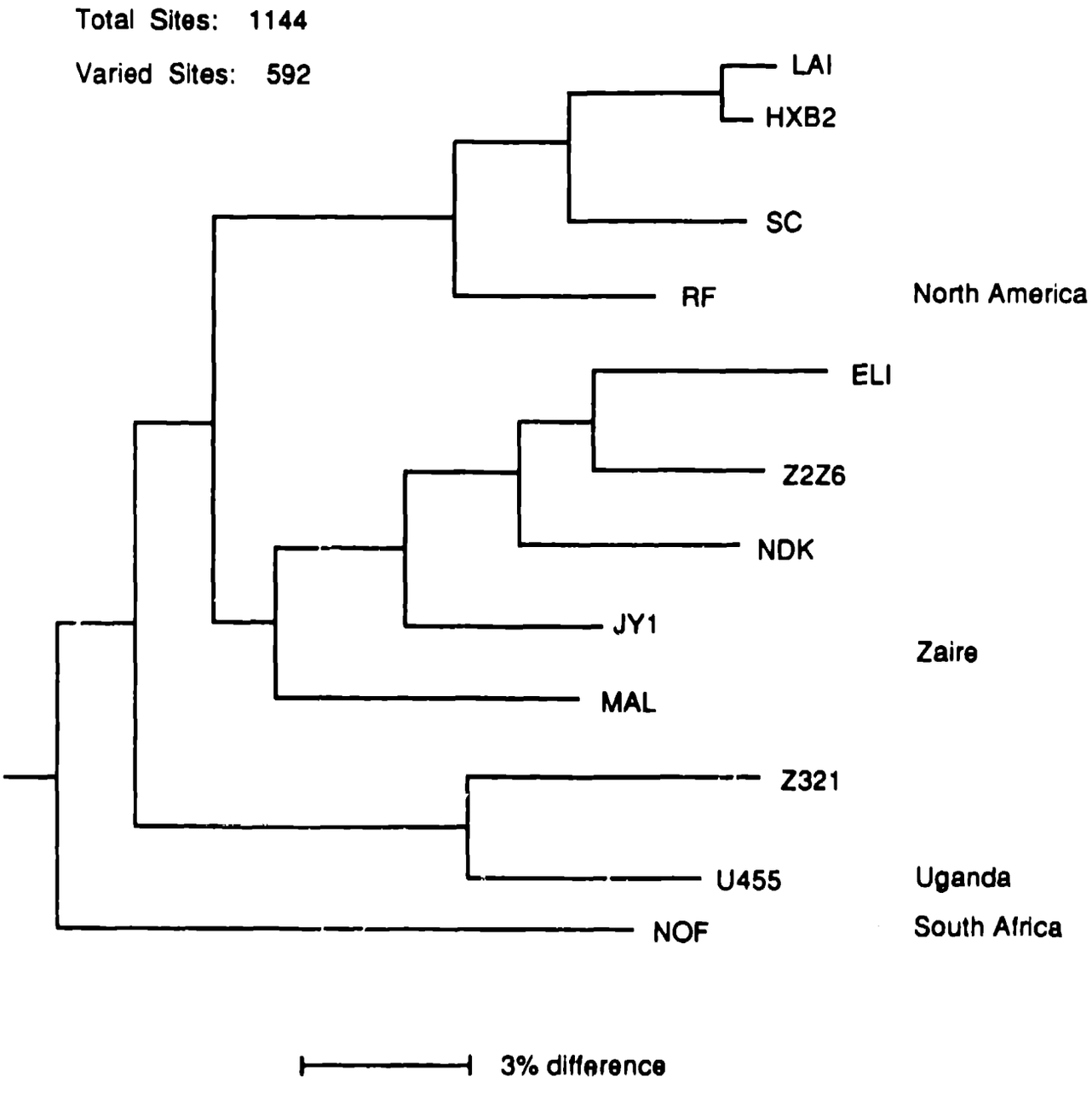
The second sampling result, representing interpatient variation, is not so acutely affected by *ex vivo* selection, but it is subject to change in time, as a result of mutation, recombination and viral migration. Thus *env* V3 and V4V5 sequences from HIV-1 infected persons in a single community of Florida in 1990 were found to differ by 13% on the average (see Morbidity and Mortality Weekly Report of January 18, 1991, Vol. 40, p.21). The extent of variation observed in this single study was nearly as great in 1990 as what had been seen in 1985 for North America and Zaire taken together.

These estimates are minimal estimates for several reasons: when derived from phylogenetic tree analyses, they represent only single base changes—insertions and deletions are not considered in most tree algorithms. Accordingly, when simple distance measurements are made in this database, we have elected not to score gaps in order to achieve close comparison with the tree distances; but also to avoid ambiguous alignments that gaps often entail and to sidestep any artifactually inserted (PCR-generated) stretches. If "gaps" are scored, all distance relationships must be asserted on the basis of similarly scored sequences ("apples" should not be compared with "oranges"). Moreover, it is most prudent with HIV to score each gap as the equivalent of a single base substitution (Reitz *et al.*, *Nature* 351, 358, 1991).

The tree analyses shown on the following pages were performed as described earlier (III-7 APR 90). They are of principal interest because they include the newly determined and relatively divergent U455 sequence from Uganda (Oram *et al.*, *AIDS Res. Hum. Retroviruses* 6, 1073, 1990). The NOF sequence was kindly provided prior to publication by Dr. Michael Becker, University of Stellenbosch, South Africa. The sequence designated CAM (Cambridge, U.K.) was kindly provided by Allison McIntosh, National Institute for Medical Research, Mill Hill. Both HIV-1 trees shown—for *gag* and *env*—were rooted upon the SIV-cpz sequence (not shown). Rather than present branch lengths (only horizontal branch lengths are relevant to distances), a bar-scale is given.


HIV1 GAG Tree
(with Ugandan sequence)
Total Sites: 965
Varied Sites: 393
LAI
HXB2
MN
JH3
CDC451
RF
OYI
HAN
SF2
NY5
CAM
JRCSF
ELI
Z2Z6
NDK
MAL
U455
North America
United Kingdom
Zaire
Uganda
3% difference

HIV1 ENV Tree (with Ugandan sequence)



IV

Related Sequences

PART Related Sequences

HSPULTR (X05592 Y00070) 2

HSPUENV (X05591 Y00070) 3

SIMSP1 (M55279) 5

SIMSP2 (M33561) 6

HSPULTR

```
LOCUS       HSPULTR      2206 bp ss-RNA
DEFINITION  Human spumaretrovirus, HFV, DNA with 3' LTR; bel-1, -2, -3.
ACCESSION   X05592 Y00070
SOURCE      Human spumaretrovirus, HFV.
REFERENCE   1  (bases 1 to 2206; enum. 4550 to 6755)
  AUTHORS   Fluegel,R.M., Rethwilm,A., Maurer,B. and Darai,G.
  TITLE     Nucleotide sequence analysis of the env gene and its flanking
            regions of the human spumaretrovirus reveals two novel genes
  JOURNAL   EMBO J. 6, 2077-2084 (1987)
COMMENT     Recombinant clones that represent the 3' part of the genome of the
            human spumaretrovirus (foamy virus) were established from cDNA
            (clone C55) and viral DNA (clone B52).  See X05591 for C55
            sequence.  There are 12 octamer sequences, 6 nonamers and one
            decamer that form direct repeats within the LTR. The indirect
            repeats include a decamer with a loop size of 78bp and two octamers
            with a relatively small loop size of 34 and 5bp, respectively.
            *source: cell type=HSRV-infected HEL fibroblasts; clone=C55;
            Data kindly reviewed (02-SEP-1987) by Fluegel R.M.
BASE COUNT      691 a    406 c    484 g    625 t
ORIGIN
        1 atcctgaagt tggaatttgg gTAAaatata aacccctcag aggaattgtt ggatctgctg
                          bel-2 orf ->
       61 tgtttattat gcataaacat caaagaaact gttctcttgt taaaccttct accagttgct
      121 cagaaggtcc aaaaccaaga cctaggcacg atcctgtcct tcgctgtgac atgttgaaa
      181 agcatcacaa gcctcggcag aaacgaccca ggagacgatc catcgataat gagtcatgtg
      241 cttccagtag tgacaccatg gtcaatgagc caTGAtcact atgcaccaac cctctttgga
                                                 <- bel-1
      301 atccttgacc gctactatca gggctacttg aagagtccag caacctacca aacttggaag
      361 ttcacatgtc aggtggaccc ttctgggaag aggtttatgg agactcaatt ttgggtcccc
      421 cctctgggtc aggtgaacat tcagttttat aagaattatc agattctaac ttgctgtcag
      481 gctgtagacc catttgctaa tatttttcat ggtactgatg aagaaatgtt tgacattgat
      541 tcaggtcctg atgtttggtg ttctccctct ttgtgtttca aggtaatttt tgaaggggca
      601 atgggccaaa agcaagaaca aaaaacatgg ctgtgcagac taggacatgg tcatcgtatg
      661 ggagcatgcg attaccgtaa agtagatctg tatgcaaTGA gacaaggaaa agaaaaccct
                                          bel-3 orf ->
      721 tatggagata ggggtgatgc agctttgcaa tatgcttatc aggttaaaag gggctgtaaa
      781 gcagggtgct tggcatcacc tgtacttaac tacaaagctt tgcagtttca tagaaccatt
      841 atggcagact tcaccaatcc tagggattgga gaaggacatc ttgctcatgg ttaccaagca
      901 gctatggaag ctttatggacc tcagagaggg agtaacgagg agagggtgtg gtggaatgtc
      961 actagaaacc agggaaaaca agggagagag tattacaggg aaggaggtga agaacctcat
     1021 tacccaaata ctcctgctcc tcatagacgt acctgggatg agagacacaa ggttcttaaa
     1081 ttgtcctcat tcgctactcc ctctgacatc caacgcTGAa ctactaaagc attgccttat
                                                    <- bel-2
     1141 ggctggaaag tggtcacgga aagcggaaat gattatacta gccgcagaaa gatcagaaca
     1201 TTGAcagaga tgactcagga tgaaattaga aaaaggtggg aaagtggata ttgtgacccc
            <- bel-3
     1261 ttcattgact cagggaagtg ctcagatgga ccctctaaaa agccacagac agtaaaaatg
     1321 tgttagcact ttatacaata ttatatctgc ttaagctatt caagctttca catactcagt
     1381 agctgtttca caatcaacaa aacaatgatg atgtaatcat aaggaagtag tttaaaatag
     1441 gttaagtaag tttactgcag tagataatcc ctgggggaga tctggctctg taagctggaa
     1501 cagcaatgtt ttcagttcca atcctctcaa aggagaacca gagggatgat gtgttagttc
     1561 aaatcccatt atcctcatgg ttccctttic catagtttac tatattaatt taaggataag
     1621 gtataaggat taaggtatga ggtgtgtggc tcaacacgta gggtgacaag aaaatctact
     1681 gtaataggac acaacacctc taaagttgcc cgtgggaagg tgaagtgaga tcgaatcttt
     1741 cctttaacgca gacagcttit tatccactag ggataatgtt ttaaggaata ctatagtaat
     1801 agattgatag ttttaacaat gatagtaaata gtatataagg atagttttta gattgtacgg
     1861 gaggctcttc actactcgct gcgtcgagag tgtacgagac tctccaggtt tggtaagaaa
     1921 tatttatat tgttataatg ttactatgat ccatttaaac tctgcttata gattgtaagg
     1981 gtgattgcaa tgctttctgc atataaactt ggttttcttg ttaatcaata aaccgacttg
     2041 attcgagaac caactcctat attattgtct ctttatact ttattaagta aaaggatttg
     2101 tatattagcc ttgctaagag agacatctag tgatataagt gtgaactaca ctatcttaa
     2161 atgatgtaac tcctaggat aatcaatata caaaattcca tgacaa
```

```
LOCUS       HSPUENV      4549 bp ss-RNA
DEFINITION  Human spumaretrovirus (HFV or foamy virus) genomic RNA with env
            gene.
ACCESSION   X05591 Y00070
SOURCE      Human spumaretrovirus HFV.
REFERENCE   1  (bases 1 to 4549; enum. 1 to 4549)
  AUTHORS   Fluegel,R.M., Rethwilm,A., Maurer,B. and Darai,G.
  TITLE     Nucleotide sequence analysis of the env gene and its flanking
            regions of the human spumaretrovirus reveals two novel genes
  JOURNAL   EMBO J. 6, 2077-2084 (1987)
COMMENT     Recombinant clones that represent the 3' part of the genome of the
            human spumaretrovirus (foamy virus) were established from viral DNA
            (clone C55; see X05592) and cDNA (clone B52). The region common to
            both viral inserts comprises 817 bp (from the BamHI at bp 4548 in
            X05591 to the HINDIII site at bp 816 in X05592 corresondig to bp
            5365 in a fusioned viral sequence.  The env precursor polypeptide
            contains 14 pot. N-linked glycosylation sites.  *source: cell
            type=HSRV-infected HEL fibroblasts; clone=B52;
            Data kindly reviewed (02-SEP-1987) by Fluegel R.M.
BASE COUNT     1525 a    812 c    885 g   1327 t
ORIGIN
        1 aagcttgcca cccaaggaag ttatgtggtt aattgtaata ccaaaaaacc aaacctggat
pol ->
       61 gcagagttgg atcaattatt acagggtcat tatataaaag gatatcccaa acaatataca
      121 tattttttag aagatggcaa agtaaaagtt tccagacctg aaggggttaa aattattccc
      181 cctcagtcag acagacaaaa aattgtgctt caagcccaca atttggctca caccggacgt
      241 gaagccactc ttttaaaaat tgccaacctt tattggtggc caaatatgag aaaggatgtg
      301 gttaaacaac taggacgctg tcaacagtgt ttaatcacaa atgcttccaa caaagcctct
      361 ggtcctattc taagaccaga taggcctcaa aaaccttttg ataaattctt tattgactat
      421 attggacctt tgccaccttc acagggatac ctatatgtat tagtagttgt tgatggaatg
      481 acaggattca cttggttata ccccactaag gctccttcta ctagcgcaac tgttaaatct
      541 ctcaatgtac tcactagtat tgcaattcca aaggtgattc actctgatca aggtgcagca
      601 ttcacttctt caacctttgc tgaatgggca aaggaaagag gtatacattt ggaattcagt
      661 actccttatc acccccaaag tggtagtaag gtggaaagga aaaatagtga tataaaacga
      721 cttttaacta aactgctagt aggaagaccc acaaagtggt atgacctatt gcctgttgta
      781 caacttgctt taaacaacac ctatagccct gtattaaaat atactccaca tcaactctta
      841 tttggtatag attcaaatac tccatttgca aatcaagata cactTGActt gaccagagaa
                                                     S1 orf ->
      901 gaagaacttt ctcttttaca ggaaattcgt acttctttat accatccatc cacccctcca
      961 gcctcctctc gttcctggtc tcctgttgtt ggccaattgg tcaggagagg gtggctaggc
     1021 ctgcttcttT GAgacctcgt tggcataaac cgtctactgt acttaaggtg ttgaatccaa
                    <- pol
     1081 ggactgttgt tattttggac catcttggca acaacagaac tgtaagtata gataatttaa
     1141 aacctacttc tcatcagaat ggcaccacca ATGacactgc aacaatggat catttggaaa
                                       env ->
     1201 aaaatgaaTA Aagcgcatga ggcacttcaa aatacaacaa ctgtgactga acagcagaag
                    <- S1 orf
     1261 gaacaaatta tactggacat tcaaaatgaa gaagtacaac caactaggag agataaattt
     1321 agatatctgc tttatacttg ttgtgctact agctcaagag tattggcctg gatgttttta
     1381 gtttgtatat tgttaatcat tgttttggtt tcatgctttg tgactatatc cagaatacaa
     1441 tggaataagg atattcaggt attaggacct gtaatagact ggaatgttac tcaaagagct
     1501 gtttatcaac ccttacagac tagaaggatt gcacgttccc ttagaatgca gcatcctgtt
     1561 ccaaaatatg tggaggtaaa tatgactagt attccacaag gtgtatacta tgaaccccat
     1621 ccggaaccca tagtggtgaa ggagagggtc ctaggtcttt ctcaaattct gatgattaat
     1681 tcagaaaaca ttgctaataa tgctaatttg acacaagaag taaagaagtt gttaactgaa
     1741 atggttaatg aagaaatgca aagtttgtca gatgtaatga ttgactttga aattccttta
     1801 ggagaccctc gtgatcaaga acaatatata catagaaaat gctatcaaga atttgcaaat
     1861 tgttatttag taaaatataa agaacccaaa ccgtggccta aggagggcct tatagctgat
     1921 caatgcccat taccaggtta ccatgctgga ttaacctata atagacagtc tatttgggat
     1981 tactatatta aagtggagag tattagacct gcaaattgga caacaaagag taaatatgga
     2041 caagctagac taggaagttt ttatattcct agcagcctga gacaaatcaa tgttagtcat
     2101 gtactattct gtagtgatca attatattct aaatggtata atatagaaaa taccatagaa
     2161 caaaacgagc ggtttctgct taataaacta aataacctta catctggaac ctcagtattg
     2221 aagaaaagag ctcttccgaa ggattggagt tctcaaggta aaaatgctct gtttagagaa
```

```
2281 atcaatgtgt tagatatctg cagtaaacct gaatctgtaa tactattgaa tacttcatac
2341 tattccttct ctttatggga aggagattgt aattttacta aagatatgat ttctcagttg
2401 gttccagaat gtgatggatt ttataacaat tctaagtgga tgcatatgca tccatatgct
2461 tgtagattct ggagaagtaa gaagaatgaa aaagaagaaa ctaaatgtag agatggggaa
2521 actaagagat gtctgtatta tcctttatgg gacagtcccg aatctacata tgattttggt
2581 tatttagcat accaaaaaga ttttccttcc cctatctgta tagaaca.ca gaaaattaga
2641 gatcaagatt atgaagtcta ttctttgtat caagaacgca aaatagc.tc taaagcatat
2701 ggaattgata cagttttatt ctctctaaag aattttctta attatacagg aactcctgta
2761 aatgaaatgc ctaatgcaag agcttttgta ggcctaatag atcccaagtt tcctccttcc
2821 tatcccaatg ttactaggga acattatact tcctgtaata ataggaaaag aagaagtgtt
2881 gataataact atgctaagtt aaggtctatg gggtatgcac ttacaggagc agtgcaaacc
2941 ttatctcaaa tatcagatat taatgatgaa aacttacagc aaggaatata tttattaagg
3001 gatcatgtaa taaccttaat ggaagctaca ttgcatgata tatctgttat ggaaggaatg
3061 tttgctgtac aacatttgca tacacatttg aatcatttga agacaatgct tctagaaaga
3121 agaatagact ggacctatat gtctagtact tggctacaac aacaattaca gaaatctgat
3181 gatgagatga aagtaataaa gagaattgct agaagtttgg tatattatgt taaacaaacc
3241 catagttctc ccacagctac agcctgggag attggattat attatgaatt ggttatacct
3301 aaacacattt acttgaataa ttggaatgtt gtcaatatag gtcacttagt taaatcagct
3361 ggacaattga ctcatgtaac tatagctcat ccttatgaaa taatcaataa ggaatgtgta
3421 gagactatat atctgcatct tgaggactgc acaagacaag attatgtcat atgtgatgtg
3481 gtaaagatag tgcagccttg tggcaatagc tcagacacga gtgattgtcc tgtctgggct
3541 gaagctgtaa aagaaccatt tgtgcaagtc aatcctctga aaaacggaag ttatctggtt
3601 ttggcaagtt ccacagactg tcagatccca ccatatgttc ctagcatcgt gactgttaat
3661 gaaacaacgt catgctttgg actggacttt aaaaggccac tggttgcgga agaaagattg
3721 agctttgagc cacgactgcc aaatctacaa ctaagattac cacatttgat tggaattatt
3781 gcaaaaatca aagggataaa aatagaagtc acatcctctg gagaaagtat aaaagagcag
3841 attgaaagag caaaagctga gctccttcga ctggacattc acgagggaga tactcctgcc
3901 tggatacaac agctagctgc agcaacaaag gacgtctggc cagcagcagc ttctgctcta
3961 caaggaattg gtaacttttt atctgggact gcccaaggaa tatttggaac tgcctttagt
4021 ctcttgggat acttaaagcc tatcctaata ggagtagggg tcattctctt ggttattctt
4081 atatttaaga ttgtatcatg gattcctacg aaaaagaaga atcagTAGcc tccacctctg
                                                          <- env
4141 gaattcagga cctgcagact ctgagtgagc ttgttggccc tgaaaatgcc ggagagggag
4201 agcTGActat tgctgaggaa cctgaaagaa aatcctcgac gccccagacg atataccaaa
 bel-1 orf ->
4261 agagaagtca aatgtgtgtc ttatcatgca tataaagaaa ttgaggacaa acatcctcaa
4321 catattaaac tgcaggatty gatccccaca ccagaggaaa tgagtaagtc actctgtaaa
4381 agacttattt tatgtggatt gtatagtgca gaaaaggcct cagagatttt aaggatgcct
4441 tttacagttt ctggggaaca atcagatact gaccctgact gttttattgt aagctataca
4501 tgtatatttt gtgatgctgt aatacatgat cccatgccca taagatggg
                                                  bel-1 orf ->
```

```
LOCUS       SIMSP1      1665 bp ss-RNA
DEFINITION  Simian spumavirus, SFV-1, LTR and flanking sequence.
ACCESSION   M55279
SOURCE      Simian spumavirus (SFV-1).
SEGMENT     1 of 2.
REFERENCE   1  (bases 1 to 1665)
  AUTHORS   Mergia,A., Shaw,K.E.S., Pratt-Lowe,E., Barry,P.A. and Luciw,P.A.
  TITLE     Simian foamy virus type 1 is a retrovirus which encodes a trans-
            criptional transactivator
  JOURNAL   J. Virol. 64, 3598-3604 (1990)
COMMENT     Kindly submitted in computer readable form by K.Shaw and P.Luciw,
            University of California at Davis.
FEATURES        from  to/span     description
    LTR            1     1645
BASE COUNT      506 a    303 c    347 g    509 t
ORIGIN
        1 aatatggtag ttgcgaggag agggtgtggc agacagccac taaatgtata ggaccagagg
       61 aggaaaatta ctggtgtgaa tatgatcacc gtggattttt tcctatggtt ccaaacaagc
      121 tgtctcctac ctgggtgaga catgctgccc cctactgtat ccagaggttc gcaacaccat
      181 atgatctcca gtattttgca aatgagcttc tacccccctgg tttcagtata acgactccta
      241 agggagtgag ttataccagt gatcgaaggc tccattatgg aaatgaagga actctccagg
      301 agtataatga gaactgtgat aaggttaaga gaggatatga tgaaatatct tccagtgatt
      361 actcagatga agattaatag aaatttaccc agcaactgct tatgcttgct tatgattcat
      421 gcctttgttt aggataagag aagctttaaa catagttaat ccttaggaag catttggtaa
      481 attctactaa gcagacctgt tcatttacta ccgtgcttcc gatggagaac ctagggacga
      541 ggctgtgagt tcaatatcat ccttatctcg attgtctccc ttttcctttt cttttaaata
      601 agaattatgc aataggttta agtataaaat tataataagt taatcctaag ggagggagtg
      661 gaacgtcctg atagaaaaca gacatgacgc tctcccatcc ctccttttct catattcaaa
      721 tctaaagtaa gcttgttgac tttctgccag ctgttagcat agaagtatga taaaacagga
      781 aaccacaaac aggtgaaggc tagctcactg aataaattga ctagtctttg ctcaagaacc
      841 cagggagcag tgttgtgtgt tcaaatctca ataatgcatc ctggtcgttc ttta aaaat
      901 tatgtcattg taaacatcat atgaaagcaa gaaatgactt tgtaagacca cagagggaaa
      961 tagctaatgt gcaaagtatc aaccttgtac ttggctgttc tcctttggta ttcaagttca
     1021 aatcctgtaa actgtgctac caccgatgta atattgtgta ataatgtgtt ttctatatga
     1081 gtgtaaaatt aactcctgat gactcacgag gtgaatggct cacagtgaac gacgactgaa
     1141 tcttccttac gctgcgttgc caccacctcc aggaatgcag tgggtataga gcagtgaatt
     1201 ccagaatctc ttcatactaa ctacattctt ttgtattcaa agataggaac tagggaaata
     1261 gtttggaatg ttgtattagc tatttagaaa gatataaaaa cataggaata gattgtacgg
     1321 gagctcttca ctactcgctg cgccgagagt gtttgagact ctccaggctt ggtaagaaat
     1381 attataatct tgttattctg atccttatt gctctgctat ttagattgta atgggtaatg
     1441 gcaatgctta atcagatttg gttcaataaa ccgacttaat tcgagaacca tacttattta
     1501 ttgtctcttt caatacttta tgtaaagtga aaggagttgt gtgttagcct tgcttagggg
     1561 accatctagt gggataagtg ggtactacac ttatcataag aggtgttagt tcctaaggat
     1621 aatcaatata caatattcca cgacaattgg cgcccaacgt gggggc
```

```
LOCUS       SIMSP2      3534 bp ss-RNA
DEFINITION  Simian spumavirus SFV-1, partial pol cds, complete env cds.
ACCESSION   M33561
SOURCE      Simian spumavirus (SFV-1).
SEGMENT     2 of 2.
REFERENCE   1  (bases 1 to 3534)
  AUTHORS   Mergia,A., Shaw,K.E.S., Lackner,J.E. and Luciw,P.A.
  TITLE     Relationship of the env genes and the endonuclease domain of the
            pol genes of simian foamy virus type 1 and human foamy virus
  JOURNAL   J. Virol. 64, 406-410 (1990)
COMMENT     Kindly submitted in computer readable form by K.Shaw and P.Luciw,
            University of California at Davis. The human (HFV) and simian (SFV)
            pol and env amino acid sequences are, respectively, about 84% and
            64% similar.
BASE COUNT     1165 a    648 c    674 g   1047 t
ORIGIN
        1 gaattcagta ctccttacca cccccaaagt agtggtaaag tggaaaggaa aaatagtgac
pol ->
       61 attaaacgac ttttaactaa actgctaatt gggagacctg ctaagtggta tgatctacta
      121 cctgttgtac aattggcctt aaataattct tatagtccct cttctaaata tactcctcat
      181 caactcttgt ttggtgtaga ttccaacaca ccgtttgcaa attctgatac acttgactta
      241 tccagagaag aggaactgtc tcttttacag gaaattagat cttctctaca ccagccaacc
      301 tcccctcctg cctcctctcg ttcctggtct ccttctgttg gccaactagt ccaggagagg
      361 gtagctcgcc ctgcttcact tcgaccacgc tggcataagc ctacagctat tttggaggtc
      421 gtgaatcctc ggacagtgat aattttggac catcttggca acagacgtac tgtaagtgtt
      481 gacaacctta agttaacagc ttatcaggat aatggcacct ccaATGactc tggaacaatg
                                                 env start ->
      541 gctcttatgg aagaagatga gtcaagcaca tcaagcactT GAaaatgtaa ccaccttgac
                                                      <- pol
      601 tgaggaacag aagcaacaag ttataataga cattcagcat gaagatgttg ttcctactag
      661 gatggacaaa ttgaaatatc tggcctattc atgctgcgct actagcacac gtgtattgtg
      721 ctggatagtg ttagtttgcg tcttgctatt agttgtattt atatcctgct ttgtgacaat
      781 gtccaggata caatggaata aggatattgc tgtttttggt ccagtcattg actggaatgt
      841 tagccaacaa gctgtgattc aacaaataag agctaaaaga ttagcaagat caattagggt
      901 ggaacatgct actgagacat atgtagaggt caatatgacc agtatacctc aaggggtgtt
      961 atatgtgcct catccagaac caataattct caaggagagg gttcttggtt tatctcaggt
     1021 cataatgata aactctgaaa atattgctaa tactgctaac cttactcaag aaactaaggt
     1081 actgttagca gacatgatta atgaagagat gaatgattta gctaatcaaa tgatagattt
     1141 tgaaatccca ttaggagatc ccagagatca aaaacaatac cagcatcaaa aatgttttca
     1201 agaatttgca cattgttatt tagtaaaata taaaactact aaaggatggc ctagttctac
     1261 tgttatagca gatcaatgcc ctttgcctgg taaccatcct acagtacaat atgcacatca
     1321 aaatatatgg gattattatg tcccctttga acaaattcgg ccagaaggat ggaactcaaa
     1381 aagttattat gaagatgcta gaataggagg gttttatata ccaaaatggt tacgaaataa
     1441 ttcctatacc catgtcttat tttgttctga tcaaatttat ggaaaatggt ataatattga
     1501 tctcacagcc caggagaggg aaaatttatt agtccaaaaa ttaattaatt tagctaaagg
     1561 aaattcatca caattaaagg atagagctat gccagctgaa tgggataaac aaggaaaagc
     1621 tgatctattt agacaaatta atactttaga tgtttgtaat agaccagaaa tggtattttt
     1681 gttaaattcc tcatattatg aattttccct atgggaagga gattgtggtt ttaccagaca
     1741 gaatgttaca caggctaatt ccttatgtaa agatttctat aataactcaa aatggcaaaa
     1801 attacatcca tattcgtgta gattttggag atataaacaa gagaaagaag aaactaaatg
     1861 tagtaatggt gaaaagaaaa aatgtcttta ttacccacaa tgggatactc ctgaagcttt
     1921 atatgacttt gggttcctag catatttaaa ttcttttcct tctccaatct gtataaaaaa
     1981 tcagactata agggaacctg agtatgaaat ctcttcttta tacctagaat gcatgaatgc
     2041 ttcagacaga catggtatag atagtgcttt attagctttg aagacatttt taaactttac
     2101 tggtcagtct gtaaacgaaa tgccattagc tagagccttt gtaggcctta ctgaccctaa
     2161 atttccacca acatatccca acattacaag ggaatcttct ggttgtaata ataacaaaag
     2221 aaaaaggaga agtgttaata attatgaaag acttagatct atgggatatg ctttaactgg
     2281 agctgttcaa actttatctc aaatatctga tattaatgat gagaggctgc aacacggagt
     2341 atatttactc cgggatcatg tggtaaccct gatggaagct gcccttcatg atgtttcgat
     2401 tatggaagga atgttagcaa ttcaacatgt gcatactcat ctcaatcatc tcaagacc
     2461 acttttgatg agaaagattg attggacatt catcagaagt gactggattc aacagcaatt
     2521 acagaagaca gatgatgaaa tgaaattgat acgaagaact  acgaagtc tagtctacta
     2581 tgtcacacaa acctccagtt ctcctacagc tacttcctgg  gattggaa tatattatga
     2641 aatagtaatt cctaaacata tatatttaaa taattggcaa gtaatcaa g taggtcattt
```

```
2701 attggagtca jctggtcatc tgactcatgt aaaggttaag catccttatg aaataattaa
2761 taaggaatgt agtgacactc aatatttaca tcttgaggaa tgcattagag aggattatgt
2821 gatttgtgac atagtacaaa tagttcaacc atgtggaaat gcaacagaat tgagtgattg
2881 tccagtagca gcattaaagg tgaagactcc atatattcaa gtgtctcccc tgaagaatgg
2941 aagttattta gttttatcta gtactaagga ttgttctata cctgcatatg tacctagtgt
3001 ggtcacagtc aatgaaacag ttaagtgctt tggagtagag tttcacaaac cactttatgc
3061 tgaaacaaaa accagctatg aaccacaagt tccgcatttg aagcttcgtt taccccactt
3121 gactgggatt attgccagct tgcaatcact ggaaatagaa gttacttcta cacaagagaa
3181 tataaaagac cagatcgaaa gggccaaagc acagcttctc cggctggaca ttcacgaagg
3241 agactttcct gactggctga aacaagtcgc ctctgcaacc agggacgttt ggcctgctgc
3301 agcttccttt atacaaggag taggtaactt cttatctaat actgcccagg ggatattcgg
3361 ctcagcggta agcctcctat cctatgcaaa acctattttg ataggaatag gagttatact
3421 gcttattgcc ctttctttta agataaatatc atggcttcct gggaagctca agaagaatTG
3481 Agagaacttc tacatcatct accagaggac gatccaccag cagatctaac tcat
    <- env
```

V

Communications

The 1989 and 1990 releases of the Human Retroviruses and AIDS database contained a list of supplemental references captured from the molecular subset of MEDLINE as communicated through GENINFO at the National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, Maryland. The reference set published in the 1990 compendium has not been updated at the time of publication and will not be included in this issue of the 1991 compendium. An updated list will be published later this year. The complete list of references published in the 1989 and 1990 compendia can be made available upon request.

The enclosed diskettes comprise the electronic portion of Part V, providing HIV and SIV nucleic acid and protein sequences in ASCII characters. As a general rule, American and non-European users receive nucleic acid files in GenBank format and protein files in a reduced PIR format. European users receive these files in EMBL and Swiss-Prot formats. Both PC-DOS and Apple Macintosh formatted diskettes are available; which format you receive depends upon what you have requested. If you wish to change the diskette format or the file formats, simply contact us at 505-665-1356 or fax a message to us at 505-665-3493. If you wish to contact us via the postal service, our address is: HIV Sequence Database, T-10, MS K710, LANL, Los Alamos, NM, 87545, U.S.A. The diskettes contain READ.ME files to provide you with the most recent information regarding their contents.

GenBank, the EMBL Data Library, and the DNA Data Bank of Japan have begun to implement a new, standard FEATURE Table format which is intended to increase the amount and type of machine-parsible information contained in the FEATURE table, and which will enhance data-sharing between the databases. As part of a gradual conversion to the new format, old entries from the AIDS database will be presented with the new FEATURE table format and new or revised entries will be presented in the old FEATURE table format. By the end of 1991, all nucleotide sequence entries will be in the new format.