

5-2009

Evolutionary analysis of host proteins CD4, CXCR4 and CCR5, and HIV/SIV gp12

Lana Bunning

University at Albany, State University of New York

Follow this and additional works at: https://scholarsarchive.library.albany.edu/honorscollege_biology



Part of the [Biology Commons](#)

Recommended Citation

Bunning, Lana, "Evolutionary analysis of host proteins CD4, CXCR4 and CCR5, and HIV/SIV gp12" (2009). *Biological Sciences*. 2. https://scholarsarchive.library.albany.edu/honorscollege_biology/2

This Honors Thesis is brought to you for free and open access by the Honors College at Scholars Archive. It has been accepted for inclusion in Biological Sciences by an authorized administrator of Scholars Archive. For more information, please contact scholarsarchive@albany.edu.

**Evolutionary analysis of host proteins CD4, CXCR4
and CCR5, and HIV/SIV gp120**

An honors thesis presented to the
Department of Biological Sciences,
University at Albany,
State University of New York
in partial fulfillment
of the Honors Program Requirements

**Lana Bunning
2009**

Abstract

The acquired immune deficiency syndrome, AIDS, is a growing epidemic in the United States and the world. Since its discovery in 1981, the virus that causes AIDS, human immunodeficiency virus (HIV), has escalated. Certain African ape (*i.e.*, chimpanzees and gorillas) and monkey species are known to harbor forms of the virus termed SIV (simian immunodeficiency virus). Chimpanzees are the natural hosts of the SIV strains from which HIV-1 evolved, but do not rapidly progress to AIDS, unlike their human relatives. In the wild, gorillas have been observed to harbor SIV, but this species' disease progression is currently unknown. As the closest living species to humans, the chimpanzee genome is over 95% identical to the human genome, yet genetic differences between the species are known to exist that are thought to play a role in their different responses to SIV/HIV infection. It is postulated that African apes and monkeys have co-evolved with SIV for a few million years, and thus have been able to adapt to, and co-evolve with, this deadly virus. By contrast, the recent cross-over of HIV to humans would suggest that such adaptive changes are missing from the human genome.

Previous work by this and other labs has identified the T cell surface proteins CD4, CCR5, and CXCR4 — which are involved in HIV infection of these cells — as potential targets of selection in the viral-host response. This past year I analyzed the protein-coding exons of the CD4, CCR5 and CXCR4 genes and their inferred proteins from a variety of primate species. In addition to the analysis of these host genes, I gathered numerous sequences for the HIV/SIV surface protein gp120 and scanned the translated amino acid sequences for unique changes at sites of interaction with the host CD4 protein. I found strong evidence for rapid evolution of CD4 on the chimpanzee lineage, and found no change on the human lineage. Two of the amino acid replacements on the chimpanzee lineage create two potential N-linked glycosylation sites which, if glycosylated, would likely interfere with gp120-binding. This finding supports the thesis that chimpanzees have adapted genetically to SIV.

Introduction

The human immunodeficiency virus (HIV) is known to attack the immune system of its host, and is a serious threat to the human species. The virus preferentially targets and destroys CD4⁺ T cells, severely crippling the host's ability to coordinate a successful immune response. AIDS (acquired immune deficiency syndrome) occurs when the CD4⁺ T cell count drops below 200 cells per microliter of blood. The virus does not discriminate, as males, females, heterosexuals, and homosexuals from all populations globally are affected by HIV (Wessner 2006).

Chimpanzees (Corbet *et al.* 2000), and perhaps gorillas (Takehisa *et al.* 2009), are the natural hosts of the SIV strains that gave rise to the major HIV-1 strains M, N, and O. Since chimpanzees typically do not progress to AIDS (Hvilsom *et al.* 2008), it may be informative to investigate the genetic differences between humans and these African primates to further understand SIV and HIV infections. We do know that HIV-1 and HIV-2 have entered the human race from two different African primate species. HIV-1 came from cross-species transmission from chimpanzees and HIV-2 came from the sooty mangabey (Wessner 2006). Recent analysis has shown that gorillas also harbor SIV in the wild (Takehisa *et al.* 2009). The SIV of gorillas is most closely related to the O strain of HIV. SIVcpz (*Pan troglodytes*) is most closely related to the M strain, which is the most infectious, and the N strain found in central Africa. Although many believe that sexual intercourse spread SIV to humans, the most likely cause of the cross-species infection was from the butchering and consuming of bushmeat (Peeters *et al.* 2002).

The evolutionary tree in Figure 1 shows the relationship of humans to the African apes, chimpanzee and gorilla, and the Asian apes, orangutan and gibbon. As indicated on this tree, we hypothesize that chimpanzees and gorillas were independently infected by SIV a few million years ago, which would allow them sufficient time to adapt to the virus at multiple loci. In contrast, it appears that the human lineage somehow avoided this infection until quite recently. If

so, the expectation would be that chimpanzee and gorilla proteins such as the HIV receptors CD4, CXCR4 and CCR5 would show evidence of adaptation to the virus, whereas the human proteins would not. Based on this hypothesis, careful phylogenetic analysis was done on CD4, CXCR4 and CCR5 of the host genomes, as well as gp120 of HIV and SIV.

The CD4, CXCR4 and CCR5 proteins are necessary for the acquisition of HIV and SIV. CD4 is a major player in the immune system and can be found on T-helper cells. It increases interactions between the helper T-cells and MHC class II cells by forming the ternary complex with T-cell receptors (Clapham *et al.* 2002). CD4 is a member of the immunoglobulin superfamily, which includes molecules that share structural features with variable (V) or constant (C) immunoglobulin domains (Brunet *et al.* 1987). Conformational changes occur within the four domains (D1, D2, D3, D4) that make up CD4. These changes are important for the binding between CD4 and gp120, discussed below (Clapham *et al.* 2002).

The differences between HIV and SIV infection and diseases progression between the human and non-human primates depend on pathogenic properties of the viruses and host-specific factors such as virus-receptor/co-receptor interactions (Hvilsom *et al.* 2008). CD4 is a main receptor for HIV infection and is found on the surface of immune system cells (T-cells). The HIV virion uses its gp120 to attach itself to CD4 and spill its contents into the host cell with the help of a chemokine co-receptor such as CCR5 and CXCR4 (Kwong *et al.* 1998).

The binding of HIV/SIV to CD4 occurs through a number of steps; the initial binding of the virion to CD4 causes a conformational change in gp120 allowing for a binding site for the chemokine receptors CCR5 and CXCR4. Evidence for these conformational changes is in the crystal structure's cavity laden CD4-gp120 interface where conserved binding sites were found for CXCR4 and CCR5 (Kwong *et al.* 1998). The third variable (V3) region of gp120 loop determines which chemokine receptor is necessary (Kwong *et al.* 1998). Specificity is also based upon time of entry into the cells; CCR5 and CXCR4 are used in the early and late stages of

HIV/SIV infection, respectively (Bleul *et al.* 1997). The three amino acids and four sulphated tyrosines that make up the N terminus of CCR5 are negatively charged. This allows the positive amino acids on gp120 to interact with CCR5, as well as the negatively charged E2 loop of CXCR4 (Clapham *et al.* 2002).

The gp120 protein has five variable regions when compared among the primate immunodeficiency viruses (Kwong *et al.* 1998). The first four of these variable regions form surface-exposed loops that contain disulphide bonds at their bases (Leonard *et al.* 1990). As stated before, analysis in the Stewart lab found the evolution of two potential N-linked glycosylation sites in CD4 of chimpanzee which are unique to this species. Leonard *et al.* (1990) found that the conserved and variable regions of gp120 contained glycosylation sites as well. These regions of glycosylation may provide important clues as to why HIV and SIV infect humans and chimpanzees so differently. The interaction of CD4 and gp120 is at the forefront of HIV/SIV research. The crystal structure of these interacting proteins described by Kwong *et al.* (1998) allows us to visualize the interaction between the two, and provides insight to the potential consequences of glycosylation of CD4.

Just recently, portions of the gorilla genome have been sequenced and released onto the public databases. Based on the available sequences and analyses by lab members C.B. Stewart and S. Bandla, I will discuss the similarities of the amino acid replacements and polymorphisms that gorilla CD4 appears to share with chimpanzee CD4.

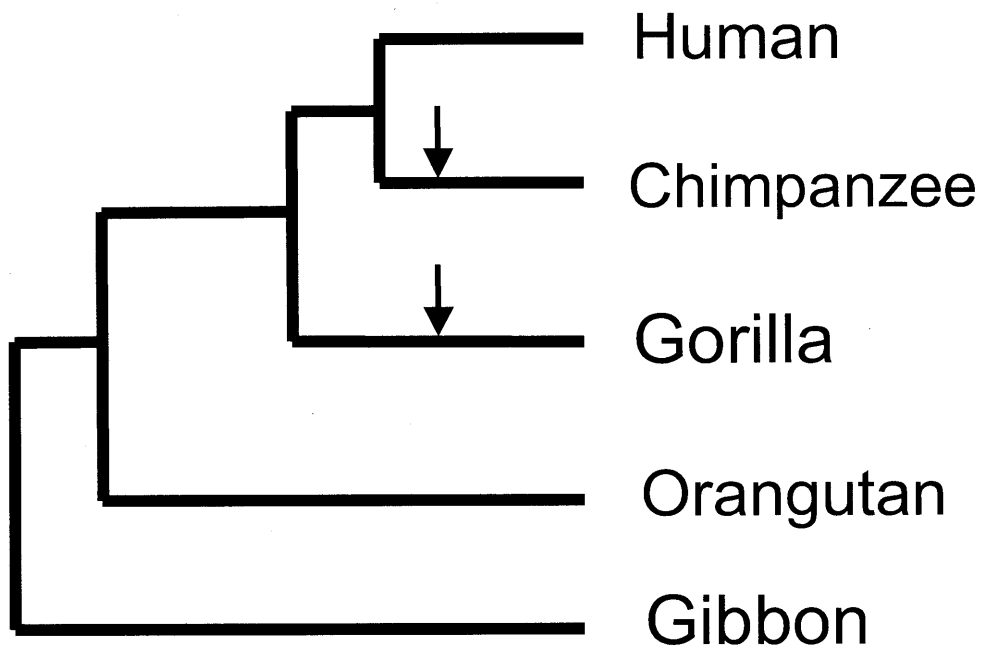


Figure 1: Phylogenetic relationship of the hominoids. This tree shows the relationship between the human and non-human apes. As seen here, the humans are most closely related to the chimpanzee, followed by gorilla, orangutan, and gibbon. The arrows indicate the lineages that appear to have been infected with SIV for a few million years, which would allow them to currently harbor SIV without rapidly progressing to AIDS.

Materials and Methods

Creation of the CD4 Alignment

Using the databases on the NCBI website (<http://www.ncbi.nlm.nih.gov/>) and Ensembl Genome Browser (<http://www.ensembl.org/index.html>), I searched for each protein's nucleotide sequence. Sequences for human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), orangutan (*Pongo pygmaeus*), and macaque (*Macaca mulatta*) were obtained from Ensembl Genome Browser. Partial sequence of gorilla (*Gorilla gorilla*) CD4 was mined from the available gorilla genome sequence by Santhoshi Bandla. The sequences that I was able to find were exported into the FASTA format and placed into a *Se-AL* [<http://tree.bio.ed.ac.uk/software/seal/>] document. *Se-AL* is a multiple sequence alignment software package (<http://tree.bio.ed.ac.uk/software/seal/>). In the *Se-AL* program I formatted the nucleotide sequences into amino acids and aligned them accordingly. After all of the sequences were aligned, I imported them as NEXUS files into *MacClade* [Maddison and Maddison 1992], which was used for inference of average numbers of amino acid replacements per lineage, as well as for identifying the specific amino acid replacements that likely occurred on each lineage.

Phylogenetic Analysis of CD4

The nucleotide sequences were translated into amino acid sequences (see Appendix 1) and a phylogenetic tree was built and rooted by the macaque sequence. Within the *MacClade* program I was able to measure the minimum, average, and maximum number of amino acid replacements per lineage. In doing so, I was able to visualize the evolutionary changes that occurred on each lineage of the tree.

Creation of the gp120 Alignment

Takehisa *et al.* (2009) produced new SIVcpz and SIVgor sequences, which I used to analyze gp120. I retrieved these sequences from the databases on the NCBI website (<http://www.ncbi.nlm.nih.gov/>). The sequences were exported into the FASTA format and imported into *Se-Al*. I viewed and aligned the sequences as amino acids. I obtained a sequence of the protein used to create the crystal structure of gp120 by Kwong *et al.* [1998] on the PDB website [<http://www.rcsb.org/pdb/home/home.do>]. Using amino acid sequence of this engineered gp120 protein as a guide, I was able to align the HIV, SIVcpz and SIVgor gp120 sequences accordingly.

Results

Phylogeny of CD4

Figure 2 shows the phylogeny of CD4 scaled to average number of amino acid replacements per lineage. I used macaque as the outgroup to root the tree, and found the following average of amino acid replacements. The orangutan lineage had 15.6 inferred amino acid replacements. There is an average of 6.9 amino acid replacements leading to the chimpanzee and human lineage. There is an additional 6 amino acid replacements on the chimpanzee lineage. In contrast, no amino acid replacements are inferred upon the human lineage. Thus, all of the observed sequence differences between the human and chimpanzee CD4 sequences (Hvilsom *et al.* 2008) are due to derived amino acid replacement on the chimpanzee lineage. The sequences can be seen in Appendix 1.

Phylogeny of CXCR4

Figure 3 shows the phylogeny of CXCR4 scaled to average number of amino acid replacements. The orangutan lineage has an additional 2 amino acid replacements, whereas the orangutan lineage has none. Gorillas do not have any additional amino acid replacements. Humans and chimpanzees share 1 amino acid replacement, and neither have any additional replacements after their split. Thus, the protein sequence of CXCR4 is highly conserved within the hominoids, in contrast to CD4. The sequences can be seen in Appendix 2.

Phylogeny of CCR5

The phylogenetic tree of CCR5 is shown in Figure 4, scaled to amino acid replacements. The orangutan branch does not have any inferred amino acid replacements. Gorilla is shown to have an average of one more amino acid replacement on its lineage. Chimpanzees and humans share one amino acid replacement. The chimpanzee lineage does not show any further amino

acid replacements. The human lineage shows an additional 2 amino acids changes on its lineage. These changes are occur throughout the sequence and will be discussed further below. The sequences can be seen in Appendix 3.

Alignment of gp120

The alignment of gp120 can be seen in Appendix 4. On the top line of this alignment is the protein sequence of the engineered gp120 used in producing the crystal structure (Kwong *et al.* 1998). Aligned to this sequence are examples of sequences of the M and N strains of HIV-1. The next are sequences of SIVcpzPtt (from *Pan troglodytes troglodytes*, Ptt), followed by SIVcpzPts (from *P. t. schweinfurthii*, Pts). The HIV-1 O strains are next in the alignment followed, by the SIVgor strains. A majority of these sequences were the same as those used by Takehisa *et al.* (2009), as seen Figure 5; however, I included additional examples of M and N strains to the alignment in Appendix 4 to increase the power of our analysis for detecting adaptive amino acid replacements. This alignment of gp120 reveals many conserved regions as well as many variable regions (Appendix 4).

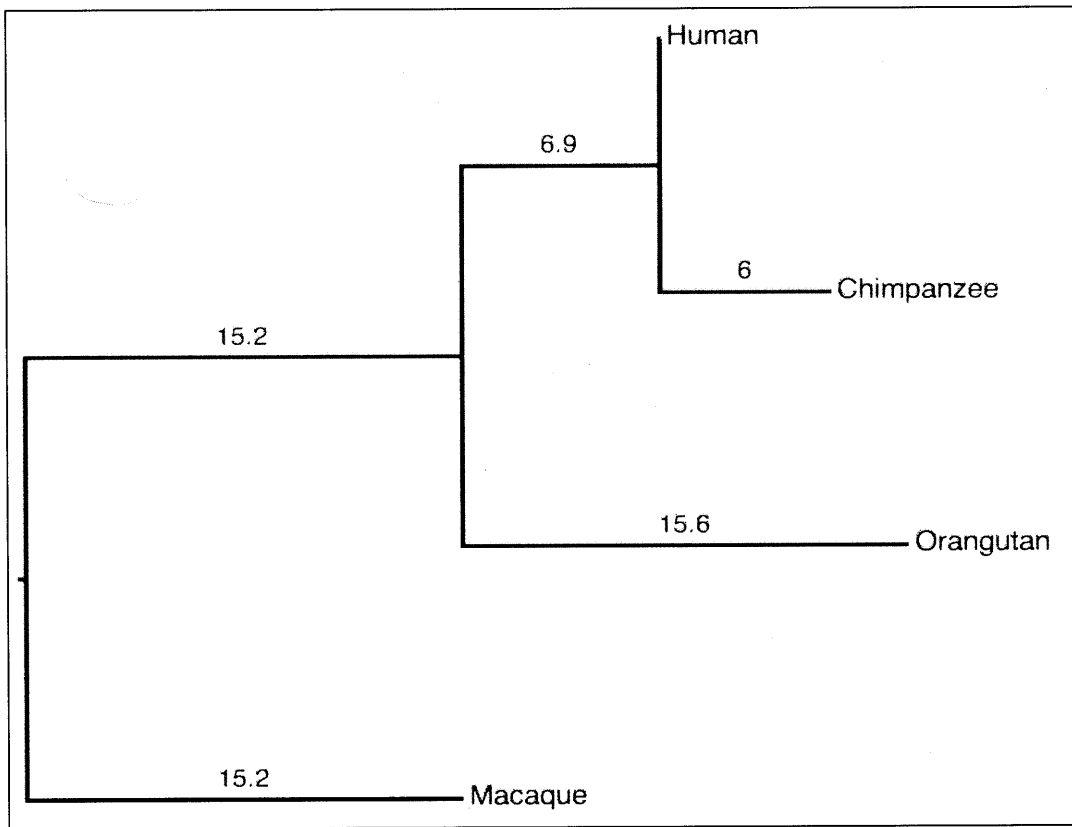


Figure 2: Phylogenetic analysis of CD4 based on amino acid replacements.

The average number of amino acid replacements was determined for the CD4 protein for the available hominoid sequences with a parsimony approach. Human and chimpanzee share an average of 6.9 amino acid replacements, but we see that chimpanzees show an additional 6 amino acid replacements that humans do not share.

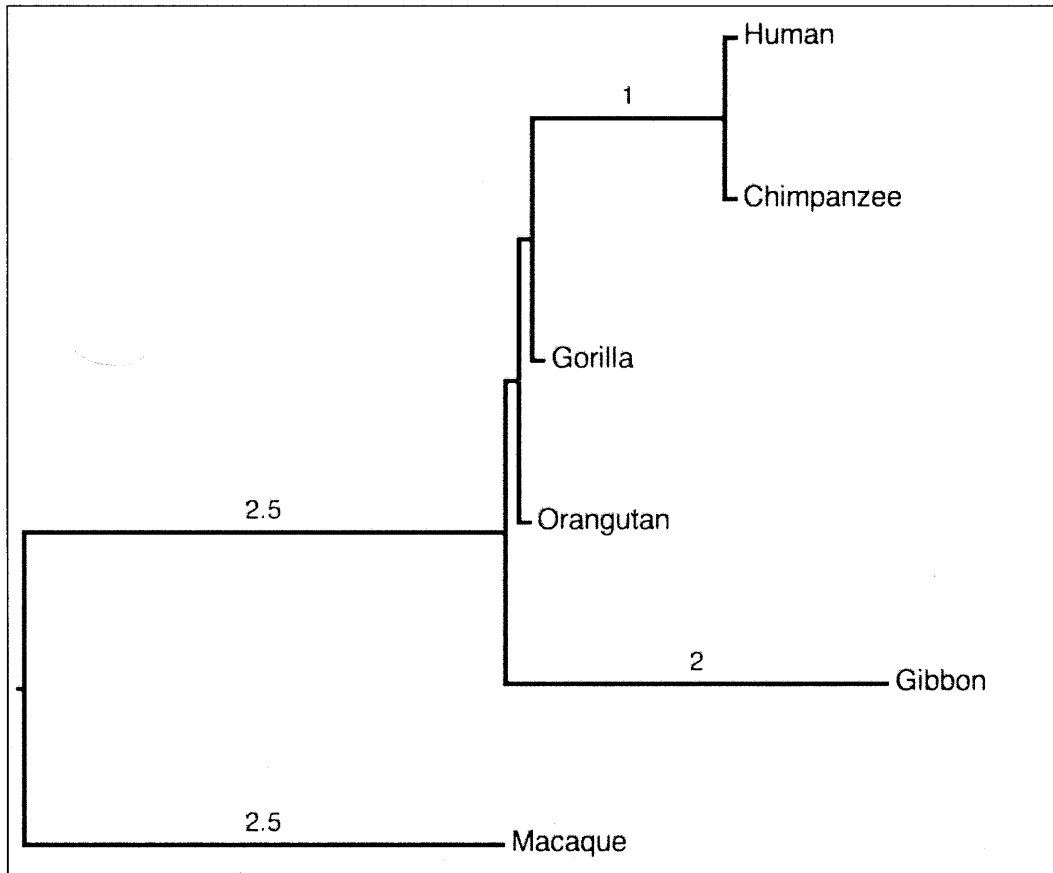


Figure 3: Phylogenetic analysis of CXCR4 based on amino acid replacements.

The average number of amino acid replacements was determined for the CXCR4 chemokine co-receptor for the available hominoid sequences with a parsimony approach. Human and chimpanzee share an average of 1 amino acid replacement. For CXCR4, we do not see additional replacements on the chimpanzee lineage as we saw for CD4 sequence.

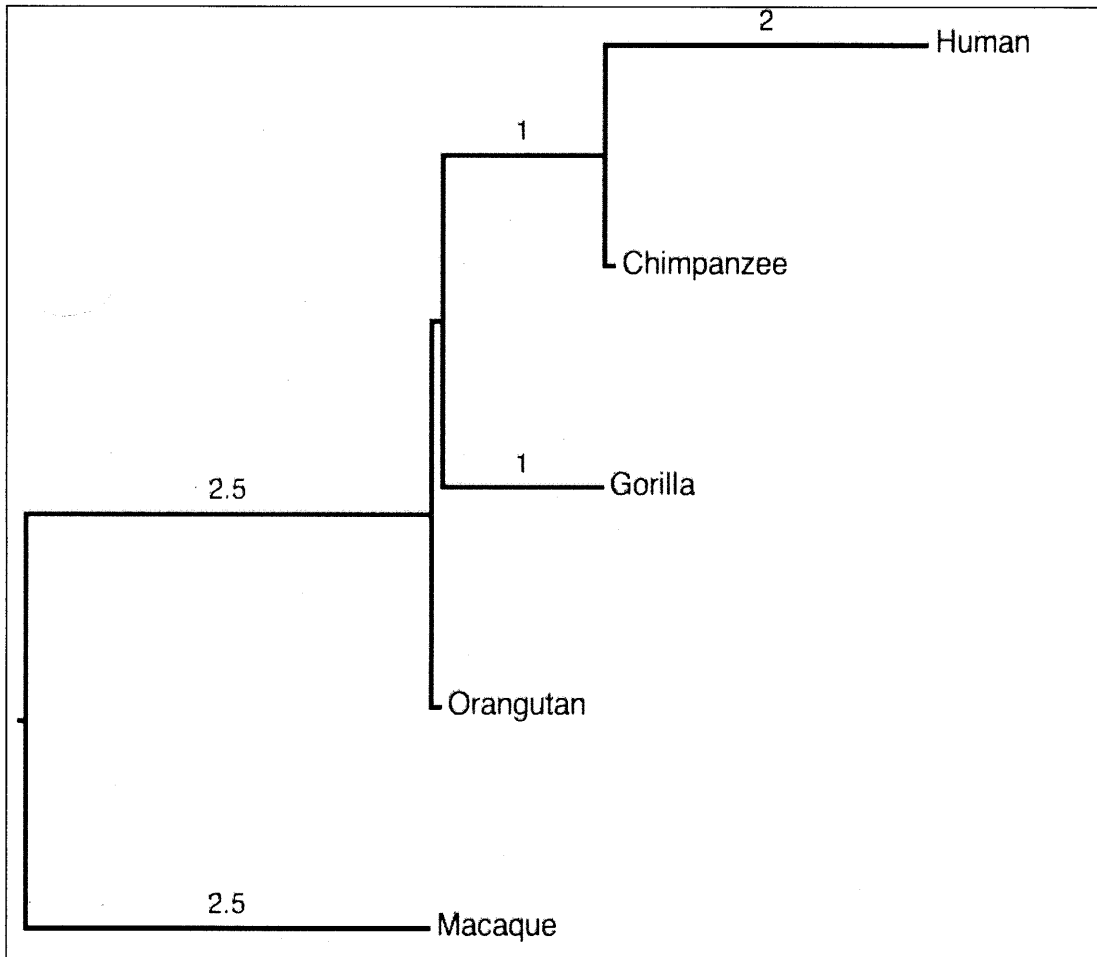


Figure 4: Phylogenetic analysis of CCR5 based on amino acid replacements.

The average number of amino acid replacements was determined for the CCR5 chemokine co-receptor for the available hominoid sequences with a parsimony approach. Human and chimpanzee share an average of 1 amino acid replacement. Gorilla also expressed 1 amino acid replacement. Human had an additional 2 amino acid replacements, while the chimpanzee lineage showed none.

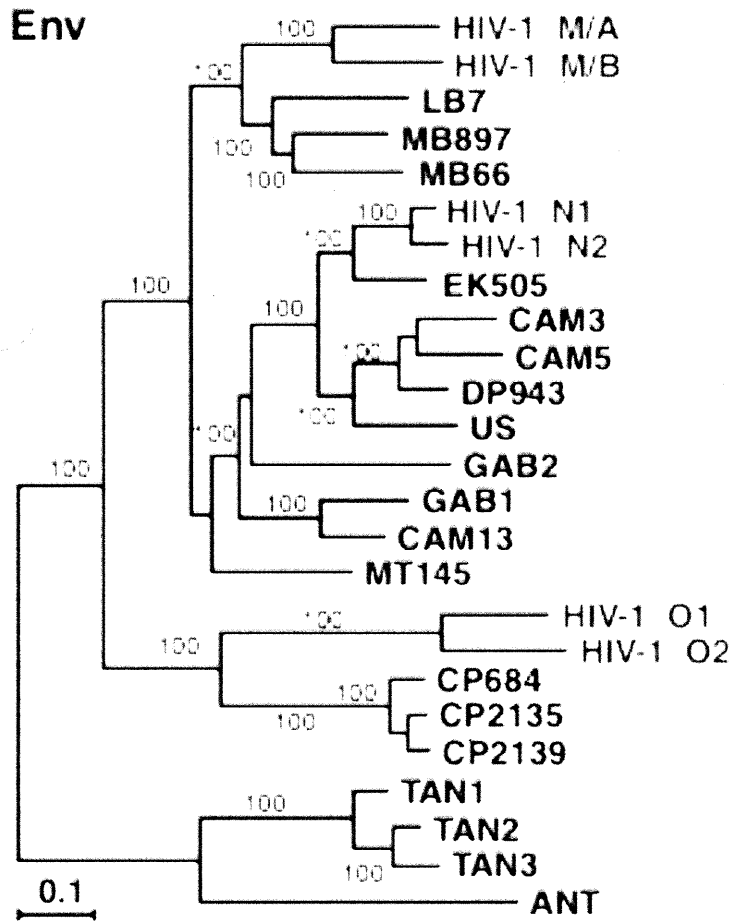


Figure 5: Molecular phylogeny of envelope protein from various SIV and HIV strains. This phylogeny illustrates how it is known that humans got HIV-1 from chimps. The numbers above the lineages indicate the statistical bootstrap support for the clades. The black clades represent SIVcpz, the blue clades represent HIV-1 strains M, N, and O. The red clade represents SIVgor. Note that the three human HIV-1 strains are nested within the chimp and gorilla SIV strains, providing the evidence that HIV jumped from these apes to humans. SIV strains TAN1, TAN2, TAN3 and ANT are found in the *Pan troglodytes schweinfurthii* sub-species, whereas the remainder of the SIVcpz strains are from *Pan troglodytes troglodytes* (see Figure 6). [Figure taken from Takehisa *et al.* 2009]

Discussion

As shown in Figures 3, phylogenetic analysis revealed no evidence of rapid protein sequence evolution on the chimpanzee or gorilla CXCR4 lineages. Interestingly, there seem to be possible amino acid replacements and/or polymorphisms on the gorilla and human lineages of CCR5. Humans seem to have 2 amino acid replacements. If these are found to be in the regions that interact with HIV, they might be allowing humans to be more susceptible to HIV infection than chimpanzees. Perhaps CCR5 plays a bigger role in allowing HIV infection in humans rather than changes in chimpanzees and gorillas that prevent it. We can see that chimpanzees have 6 amino acid changes while humans have none as illustrated in Figure 2. This shows that CD4 has evolved very rapidly on the chimpanzee, but not human, lineage.

Then we must ask the question: are these amino acid replacements fixed in chimpanzees? To answer this question requires understanding the demographics of the chimpanzees and sequencing the genes of different populations and sub-species.

The map in Figure 6 shows the chimpanzee sub-species and their estimated dates of divergence. The chimpanzee genome that has been sequenced is from a *P.t. verus*. Each of the sub-species has differences in infection rates for SIV. *P.t. verus* is from western Africa as are the majority of chimps that are kept in captivity in the USA and used in HIV/AIDS research. This subspecies has not been found to harbor SIV like chimps from central and eastern Africa (Figure 6). They are extremely difficult to infect with SIV or HIV, and they do not progress to AIDS if infected experimentally. As the map shows, further eastern and central chimps such as *P. t. troglodytes* and *P. t. schweinfurthii* are infected with SIV, but there has been little or no indication that they die from AIDS. However, recent evidence suggests that, in the wild, Eastern chimpanzees infected with SIV are less fit than uninfected chimps. Taken together, these data suggest to us that the differences in SIV infection rates of the different populations is likely due to differences in levels of adaptation to the virus.

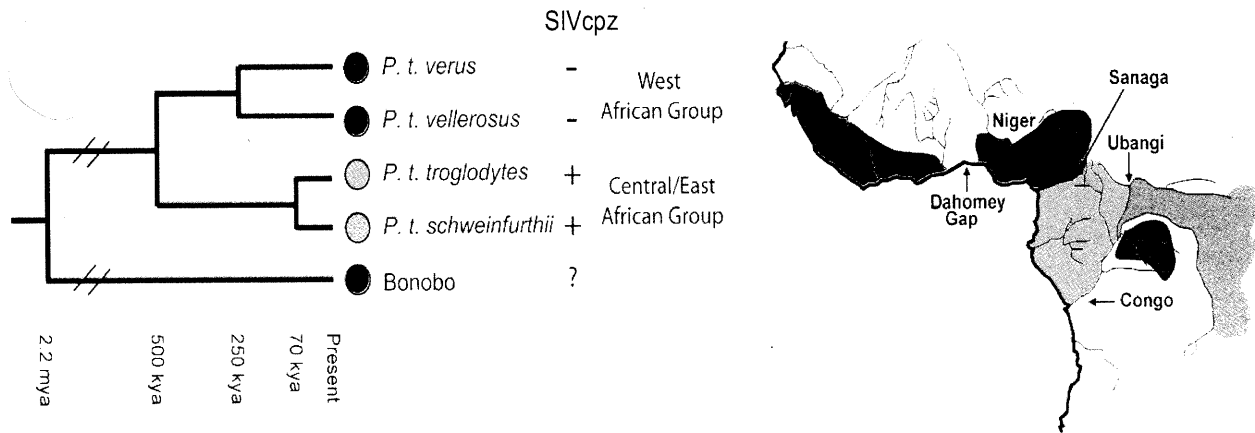


Figure 6: Phylogeny of chimpanzee sub-species indicating SIV infection and geographic location.

The left side of the figure shows the molecular phylogeny of the chimpanzees and bonobo (*Pan paniscus*). The colored circles next to their names indicate their geographical location as shown on the map to the right. SIVcpz status of these sub-species is indicated by (+) or (-). Note that only *P. t. troglodytes* and *P. t. schweinfurthii* are shown to harbor SIV to this date. The chimpanzee genome that was sequenced, and thus the one that was used in this study was from *P. t. verus*. [Figure made by K. Gonder].

If this were correct, we would expect to see differences in the sequences of host proteins involved in SIV infection, such as CD4. Indeed, Hvilsom *et al.* (2008) characterized the genetic diversity of CD4 in all four recognized subspecies of chimpanzees and found variation among them. They discovered that amino acid replacements in CD4 are conserved in individuals belonging to the *P. t. verus* subspecies and divergent from the other three subspecies, which harbored highly variable CD4 receptors (Hvilsom *et al.*, 2008). However, these researchers did not analyze their data in phylogenetic framework, as we have done in Figure 7. Phylogenetic analysis reveals that there are 4 amino acid replacements shared by all of the subspecies, and thus were ‘fixed’ on the ancestral chimpanzee lineage. One of these is a threonine that replaces an isoleucine and creates a potential N-linked glycosylation site. For glycosylation to occur the motif N–X–T/S is needed. *P. t. verus* has fixed a proline to threonine change, which would create a second N-linked glycosylation site; this site is polymorphic in the other subspecies. Human

CD4 has neither of these glycosylation sites. If glycosylation occurs, large and bulky carbohydrates have the opportunity to be added to CD4, which could hinder the binding of SIV to the host CD4. *P.t. verus* has fixed an additional asparagine to aspartic acid replacement not seen in the other subspecies. Thus, all six replacements inferred to have happened on the chimpanzee CD4 lineage (Figure 2) do appear to have been fixed in *P.t. verus*, but not the other subspecies. The other subspecies have additional polymorphic sites not seen in *P.t. verus*, however.

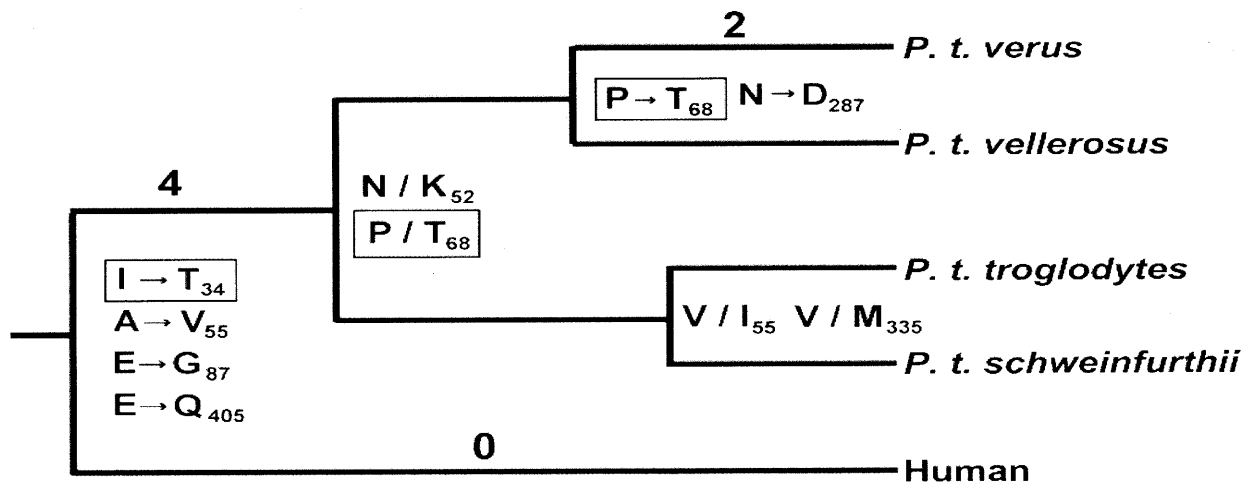


Figure 7: Mapping of the CD4 amino acid replacements on a phylogeny of the chimpanzee sub-species.

This tree illustrates the 6 amino acid replacements on the chimpanzee lineage. The amino acid replacements (arrows) and polymorphisms (slashes) are shown under the lineages in which they are inferred to have happened, with the exception of the two polymorphisms shown at the ancestral chimpanzee node (which are found in all subspecies except for *P.t. verus*). Four of the amino acid replacements are found in all four of the sub-species, and thus were likely fixed on the lineage leading to this species. Two additional amino acid replacements are found only in *P.t. verus*. The boxes around some of the amino acid replacements indicate those that result in potential N-linked glycosylation sites.

Now that we are aware of the 6 amino acid replacements in chimpanzee CD4, it is important to see where these amino acid replacements occur in the 3D structure of the protein. Figure 8a is a representation of the interaction of Domains 1 and 2 of human CD4 (in yellow) with a genetically engineered version of gp120 from HIV-1 (in purple). Note that this gp120 molecule only interacts with Domain 1. Figure 8b illustrates the amino acid replacements of the chimpanzee lineage modeled onto the human CD4 structure. Note that 4 of the 6 replacements found in *P.t. verus* CD4 occurred in Domain 1. These include the replacement of an otherwise conserved glutamate (negative charge) with a glycine (no charge) at the interacting face. They also include the gain of two threonine residues that create N-glycosylation sites that are not found in human CD4. If long-chain N-linked carbohydrates were added onto CD4 at these positions, they likely would hinder the binding of gp120 to the host cell's CD4. If so, this could help explain why *P.t. verus* appears difficult to infect with HIV or SIV experimentally, and perhaps is part of the reason that this subspecies currently lacks endemic SIV.

As discussed above, recent research has shown that wild gorillas harbor SIV and there is no evidence that they progress to AIDS (Takehisa *et al.*, 2009). This led us to wonder whether CD4 might also be the target of selection by SIV, and to ask what amino acid changes have occurred on the gorilla CD4 lineage. A recent search of the gorilla genome database revealed a newly-available partial sequence of the CD4 gene. Santhoshi Bandla analyzed this gene, and some of her findings are discussed next.

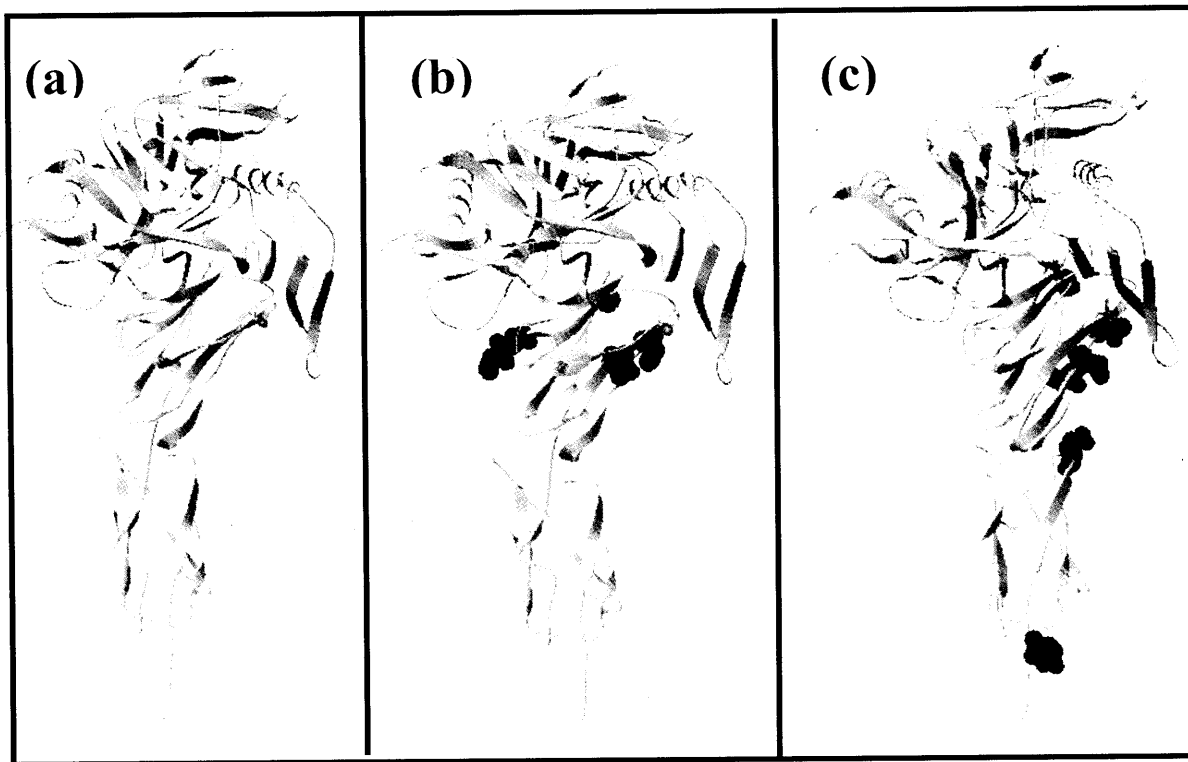


Figure 8: Host CD4 interacting with SIV/HIV gp120.

(a) This figure shows the interaction of Domains 1 and 2 of human CD4 (in yellow) with a genetically engineered version of gp120 from HIV-1 (in purple). This figure was produced in VMD (molecular visualization program for displaying, animating, and analyzing large biomolecular systems) by Santhoshi Bandla, using the PDB file IG9M from Kwong *et al.* (1998).

(b) Chimpanzee amino acid replacements mapped onto human CD4.

In this figure, Bandla mapped the amino acid replacements of the chimpanzee CD4 protein onto human CD4. The bright pink residue is a glycine, which replaced a glutamate. The green one is a valine, which replaced an alanine. The two threonine replacements are represented in blue. The asparagine residues that would be glycosylated as a result of these threonine gains (which create N-X-T sites) are modeled in purple. This structure shows that the possible N-linked glycosylation sites occur in Domain 1 where CD4 of the host interacts with gp120 of the virus.

(c) Gorilla amino acid replacements mapped onto human CD4. Gorilla amino acid replacements mapped onto human CD4, as in (b). The threonine replacement is represented in blue, and the asparagine residue that would be glycosylated is modeled in purple. The bright pink residue is a glycine, which in this case replaced an aspartate.

Seen in Figure 8c, gorillas have similar amino acid replacements as chimpanzees. This figure shows the amino acid replacements of gorilla CD4 modeled onto human CD4. Once again, the purple amino acid replacements represent asparagines and blue represents threonine. The pink amino acids in these figures represent the replacements that caused an independent loss of a charged residue at the interface. Gorillas have a gained an N-linked glycosylation site on the same side of CD4 as has chimpanzee. Also, the glycosylation site is near the same place on the crystal structure but at a different position. Thus, this is a case of convergent evolution of glycosylation sites in these two SIV-endemic species. It is tempting to think that these represent independent adaptations to SIV infection.

The fact that chimpanzee and gorilla CD4 appear to have adapted to SIV is striking information and makes us wonder what protein adaptations humans need to evolve in order to allow us to avoid HIV infection or to harbor HIV without AIDS. We know that HIV evolves rapidly, allowing the virus to dodge medications like azidothymidine (AZT). With this rapid evolution we expect to see co-evolution of the virus especially in HIVgp120 adapting to humans. Using the HIV/SIV envelope protein sequences of Takehisa *et al.* (2009) seen in Figure 5, and some additional HIV sequences, I aligned the sequences of gp120. In doing so, I was looking for evidence of adaptive evolution of HIV-1 gp120 following the 'jumps' of SIV from chimpanzees to humans. I was also looking for patterns of shared changes among the sequences of the HIV strains, M, N and O. I found some sites to be conserved and others to be highly variable (Appendix 4). Some regions were so variable that they were very hard to align. This illustrates the well-known fact that gp120 is evolving extremely rapidly.

The amino acid replacements on chimpanzee CD4 in comparison to human CD4 helps to explain why chimpanzees successfully live without progressing to AIDS, while humans fail to do so. With new gorilla sequences available, there is more evidence that these African primates have evolved molecularly to survive with the SIV. The rapid rate of evolution and the

'convergence' of the chimpanzee and gorilla CD4 proteins suggest that they have adapted under positive selection to SIV gp120. Importantly, both the rapid evolution and the convergence occurred in the regions of chimpanzee and gorilla CD4 that interact with gp120. Unfortunately, it is still unclear whether HIV-1 gp120 has adapted to human CD4 following the cross-species transmissions.

References Cited

- Bleul, C.C., Wu, L., Hoxie, J.A., Springer, T.A., Mackay, C.R., 1997. **The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes.** *Proceedings of the National Academy of Sciences, USA*, 94: 1925-1930.
- Brunet, J. Denizot, F., Luciani, M., Roux-Dosseto, M., Suzan, M., Mattei, M. & Golstein, P., 1987. **A new member of the immunoglobulin superfamily—CTLA-4.** *Nature* 328: 267 – 270.
- Clapham P.R., McKnight, A., 2002. **Cell surface receptors, virus entry and tropism of primate lentiviruses.** *Journal of General Virology*. 83, 1009 - 1829.
- Hvilsom, C., Carlsen, F., Siegismund, H., Corbet, S., Nerrienet, E., Fomsgaard, A., 2008. **Genetic subspecies diversity of the chimpanzee CD4 virus-receptor gene.** *Genomics* 92: 322 – 328.
- Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A. 1998. **Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.** *Nature* 393: 648 – 6559.
- Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N., and Gregory, T.J.. 1990. **Assignment of intrachain disulfide bonds and characterization of potential of potential glycosylation sites of the type 1 recombinant immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cell.** *Journal of Biological Chemistry* 265: 10373 - 10382.
- Maddison, W. P. and Maddison, D. R. (1992). *MacClade*. Sinauer Associates, Inc., Sunderland, MA.
- Peeters, M., Courgnaud, V., Abela, B., Auzel, P., Pourrut, X., Bibollet-Ruche, F., Loul, S., Liegeois, F., Butel, C., Koulagna, D., Mpoudi-Ngole, E., Shaw, G.M., Hahn, B.H., Delaporte, E. 2002. **Risk to Human Health from a Plethora of Simian Immunodeficiency Viruses in Primate Bushmeat.** *Emerging Infectious Diseases* 8: 451 - 457.
- Takehisa, J. Kraus, M.H., Ayouba, A., Bailes, E., Van Heuverswyn, F., Decker, J.M., Li, Y., Rudicell, R.S., Learn, G.H., Neel, C., Mpoudi Ngole, E., Shaw, G.M., Peeters, M., Sharp, P.M. Hahn. B.H. 2009. **Origin and Biology of Simian Immunodeficiency Virus in Wild-Living Western Gorillas.** *Journal of Virology* 83: 1635 – 1648.
- Wessner, D. 2006. **HIV and AIDS**, The Benjamin Cummings Topics in Biology Series. M.A. Palladino, Ed. 35 pp. *Pearson Education Inc.*, Benjamin Cummings. San Francisco.
- Se-Al. <http://tree.bio.ed.ac.uk/software/seal/>.

Appendix 1: CD4 Alignment

	281	290	300	310	315
Human	H L T L P Q A L P Q Y A G S G N L T L A L E A K T G K L H Q E V N L V				
Chimpanzee				
Orangutan R				
Macaque				

	316	320	330	340	350
Human	V M R A T Q L Q K N L T C E V W G P T S P K L M L S L K L E N K E A K				
Chimpanzee				
Orangutan E				
Macaque F E I G T T				

	351	360	370	380	385
Human	V S K R E K A V W V L N P E A G M W Q C L L S D S G Q V L L E S N I K				
Chimpanzee				
Orangutan Q .. V Q				
Macaque Q A				

	386	390	400	410	420
Human	V L P T W S T P V Q P M A L I V L G G V A G L L L F I G L G I F F C V				
Chimpanzee				
Orangutan P A				
Macaque V P T				

	421	430	440	450	455
Human	R C R H R R R Q A E R M S Q I K R L L S E K K T C Q C P H R F Q K T C				
Chimpanzee Q K V W P S R R				
Orangutan				
Macaque				

	456	458			
Human	S P I				
Chimpanzee				
Orangutan				
Macaque				

Appendix 2: CXCR4 Alignment

	1	10	20	30	40																																							
Human	M	S	I	P	L	P	L	Q	I	Y	T	S	D	N	Y	T	E	E	M	G	S	G	D	Y	D	S	M	K	E	P	C	I	R	E	E	N	A	N	I					
Chimpanzee			
Gorilla			
Orangutan			
Gibbon			
Macaque			
	41	50	60	70	80																																							
Human	N	K	I	F	L	P	T	I	Y	S	I	F	L	T	G	I	V	G	N	G	L	V	I	L	V	M	G	Y	Q	K	K	L	R	S	M	T	D	R	Y					
Chimpanzee		
Gorilla	
Orangutan	
Gibbon	
Macaque	
	81	90	100	110	120																																							
Human	R	L	H	L	S	V	A	D	L	L	F	V	I	T	L	P	F	W	A	V	D	A	V	A	N	W	Y	F	G	N	F	L	C	K	A	V	H	V	I	Y				
Chimpanzee	
Gorilla
Orangutan
Gibbon
Macaque
	121	130	140	150	160																																							
Human	I	V	N	L	Y	S	S	V	L	I	L	A	F	I	S	L	D	R	Y	L	A	I	V	H	A	T	N	S	Q	R	P	R	K	L	L	A	E	K	V	Y				
Chimpanzee
Gorilla
Orangutan
Gibbon
Macaque
	161	170	180	190	200																																							
Human	Y	V	G	V	W	I	P	A	L	L	L	T	I	P	D	F	I	F	A	N	V	S	E	A	D	D	R	Y	I	C	D	R	F	Y	P	N	D	L	W	Y				
Chimpanzee
Gorilla
Orangutan
Gibbon
Macaque
	201	210	220	230	240																																							
Human	V	V	F	Q	F	Q	H	I	M	V	G	L	I	L	P	G	I	V	I	L	S	C	Y	C	I	I	I	S	K	L	S	H	S	K	G	H	Q	K	R	K				
Chimpanzee	
Gorilla	
Orangutan	
Gibbon	
Macaque	

	240	250	260	270	280																																				
Human	A	L	K	T	T	V	I	L	L	A	F	F	A	C	W	L	P	Y	Y	I	G	I	S	I	D	S	F	I	L	L	E	I	K	Q	C	E	F				
Chimpanzee
Gorilla	
Orangutan	
Gibbon	
Macaque	

	280	290	300	310	320																																				
Human	E	N	I	V	H	K	W	S	I	T	E	A	L	A	F	F	H	C	C	L	N	P	I	L	Y	A	F	L	G	A	K	F	K	T	S	A	Q	H	A		
Chimpanzee
Gorilla
Orangutan
Gibbon
Macaque

	320	330	340	350	357																																			
Human	L	T	S	V	S	R	G	S	S	L	K	I	L	S	K	G	K	R	G	G	H	S	S	V	S	T	E	S	E	S	S	S	F	H	S	S	T			
Chimpanzee
Gorilla
Orangutan
Gibbon
Macaque

Appendix 3: CCR5 Alignment

	1	10	20	30	35
Human	MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPP				
ChimpanzeeD.....				
OrangutanT.....D.....				
GorillaT.....D.....T.....				
MacaqueT.....D.....				
	36	40	50	60	71
Human	L Y S L V F I F G F V G N M L V I L I L I N C K R L K S M T D I Y L L				
Chimpanzee				
Orangutan				
Gorilla				
MacaqueI.....V.....				
	72	80	90	100	105
Human	N L A I S D L F F L L T V P I W A H Y A A A Q W D F G N I M C Q L L T				
Chimpanzee				
Orangutan				
Gorilla				
MacaqueL.....				
	106	110	120	130	140
Human	G L Y F I G I F S G I F F I L L T I D R Y L A V V H A V F A L K A R				
ChimpanzeeI.....				
OrangutanI.....				
GorillaI.....				
MacaqueI.....				
	141	140	160	170	175
Human	T V T F G V V T S V I T W V V A V F A S L P G I I F T R S Q K E G L H				
Chimpanzee				
Orangutan				
Gorilla				
MacaqueR.....				
	176	180	190	200	210
Human	Y T C S S H F P Y S Q Y Q F W K N F Q T L K I V I L G L V L P L L V M				
Chimpanzee				
Orangutan				
Gorilla				
MacaqueM.....				

210 220 230 240 245
 Human V I C Y S G I L K T L L R C R N E K K R H R A V R L I F T I M I V Y F
 Chimpanzee
 Orangutan
 Gorilla
 Macaque

246 250 260 270 280
 Human L F W A P Y N I V L L L N T F Q E F F G L N N C S S S N R L D Q A M Q
 Chimpanzee
 Orangutan
 Gorilla
 Macaque

281 290 300 310 315
 Human V I E T L G M T H C C I N P I I Y A F V G E K I R N Y L L V F F Q K H
 Chimpanzee
 Orangutan
 Gorilla
 Macaque

316 320 330 340 350
 Human I A K R F C K C C S I F Q Q E A P E R A S S V Y T R S T G E Q E I S V
 Chimpanzee
 Orangutan
 Gorilla
 Macaque

351 360 370 380 385
 Human G L I M D Y Q V S S P I Y D I N Y Y T S E P C Q K I N V K Q I A A R L
 Chimpanzee D
 Orangutan I D
 Gorilla I D I
 Macaque I D

386 390 400 410 420
 Human L P P L Y S L V F I F G F V G N M L V I L I L I N C K R L K S M T D I
 Chimpanzee
 Orangutan
 Gorilla
 Macaque I V

420 430 440 450 455
 Human Y L L N L A I S D L F F L L T V P F W A H Y A A A Q W D F G N I M C Q
 Chimpanzee
 Orangutan
 Gorilla
 Macaque L

460 480 470 450 490
 Human L L T G L Y F I G F F S G I F F I L L T I D R Y L A V V H A V F A L
 Chimpanzee I
 Orangutan I
 Gorilla I
 Macaque I

490 500 510 520 525
 Human K A R T V T F G V V T S V I T W V V A V F A S L P G I I F T R S Q K E
 Chimpanzee
 Orangutan
 Gorilla
 Macaque R

530 540 550 560 565
 Human G L H Y T C S S H F P Y S Q Y Q F W K N F Q T L K I V I L G L V L P L
 Chimpanzee
 Orangutan
 Gorilla
 Macaque M

570 570 580 590 595
 Human L V M V I C Y S G I L K T L L R C R N E K K R H R A V R L I F T I M I
 Chimpanzee
 Orangutan
 Gorilla
 Macaque

600 600 610 620 630
 Human V Y F L F W A P Y N I V L L L N I F Q E F F G L N N C S S S N R L D Q
 Chimpanzee
 Orangutan
 Gorilla
 Macaque

	630	640	650	660	670	680
Human	A M Q V T E T L G M T H C C I N P I I Y A F V G E K F R N Y L L V F F					
Chimpanzee					
Orangutan					
Gorilla					
Macaque					

	690	700	710	720	730	740
Human	Q K H I A K R F C K C C S I F Q Q E A P E R A S S V Y T R S T G E Q E					
Chimpanzee					
Orangutan					
Gorilla					
Macaque					

	750	760
Human	I S V G L T	
Chimpanzee	
Orangutan	
Gorilla	
Macaque	

Appendix 4: gp120 Alignment

	1	10	20	30	35																															
(crystal structure)	G	A	R	S	E	V	L	V	N	V	T	E	N	F	N	M	W	K	N	D	M	V	E	Q	M	H	E	D	I	I	S	L	W	D		
(HIV M M62320)	P	N	P	Q	.	I	D	E	N	.	D
(HIV M U46016)	P	S	P	Q	.	L	G	.	E
(HIV M AF190127)	P	N	P	Q	.	M	.	E
(HIV M K03454)	P	N	P	Q	.	I	A	.	E
(HIV M AF286238)	P	N	P	Q	.	.	N	.	A
(HIV M U88824)	P	N	P	Q	.	I	K	.	E
(HIV M AY371158)	P	D	P	Q	.	I	F	.	D
(HIV M AY371158)	P	D	P	Q	.	I	F	.	D
(HIV M AY173951)	P	N	P	Q
(HIV M U54771)	P	N	P	Q	.	I	H	.	E
(HIV M AF077336)	P	N	P	Q	.	.	F	.	K
(HIV M AY084936)	P	S	P	Q	.	I	N	M	R
(SIVcpzPtt CAM5-AJ271369)	P	N	P	Q	.	.	Y	M	P
(HIV_1 N env complete genome (N env))	P	N	P	Q	.	.	P	.	D
(HIV N YBF30)	P	N	P	Q	.	.	L	.	P
(HIV N YBF16)	P	N	P	Q	.	.	L	.	P
(SIVcpzPtt MB66)	P	N	P	E	.	.	I	E
(SIVcpzPtt EKS05)	P	N	P	Q	.	.	L	I	P
(SIVcpzPtt MB897)	P	N	P	E	.	.	I	T	.	E
(SIVcpzPts U42720)	P	D	P	I	.	V	.	R
(SIVcpzPtt GAB)	P	S	P	E	.	.	I	.	I
(SIVcpzPtt CAM3)	P	N	P	Q	.	.	Q	I	P
(SIVcpzPtt MT145)	P	I	P	Q	.	.	L	.	P
(SIVcpzPtt CAM13)	P	N	P	Q	H
(SIVcpzPtt GAB1)	P	S	P	Q	.	.	F	.	P
(SIVcpzPtt US-AF103818)	P	N	P	Q	.	.	H	.	P
(SIVcpzPtt LB7-DQ373064)	P	D	P	K	.	A	I	E	.	S
(SIVcpzPts AF447763)	P	N	P	Y	.	.	T	.	A
(SIVcpzPts DQ374658)	P	N	P	Y	.	.	P	I	A
(SIVcpzPts DQ374657)	P	N	P	Y	.	.	P	.	A
(HIV O L20571)	P	N	P	H	.	.	F	.	P
(HIV O L20587)	P	I	P	Y	.	.	Y	.	P
(gorilla CP2135, FJ424863)	P	S	P	I	.	.	F	.	P
(gorilla CP684, fj424871)	P	S	P	I	.	.	F	.	L
(gorilla CP2139 AJ424864)	P	R	P	I	.	.	F	.	P
(gorilla CP2139.2 FJ424869)	P	R	P	I	.	.	F	.	P
(gorilla pBR-)	P	R	P	I	.	.	F	.	P

	30	40	50	60	70																																					
{crystal structure}	Q	S	L	K	P	C	V	K	L	I	P	L	C	V	---	S	C	N	T	S	V	I	T	Q	A	C	P	K	V	S	F	E										
{HIV M M62320}																N								I																		
{HIV M U46016}						G										N								I									L	D								
{HIV M AF190127}																N																										
{HIV M K03454}																N																										
{HIV M AF286238}																N																										
{HIV M U88824}																N																		M	T							
{HIV M AY371158}																									I										D							
{HIV M AY371158}																									I										D							
{HIV M AY173951}																																										
{HIV M US4771}																	N																		D							
{HIV M AF077336}																	N																		W	D						
{HIV M AY084936}	E																N								V										I							
{SIVcpzPtt CAM5-AJ271369}																	N								I										I							
{HIV_1 N env complete genome (N env)}																	N																									
{HIV N YBF30}																	N																									
{HIV N YBF16}																	N																									
{SIVcpzPtt MB661}	V																N																									
{SIVcpzPtt EK505}																	N																									
{SIVcpzPtt MB897}																	N																									
{SIVcpzPts U42720}	H																																									
{SIVcpzPtt GAB1}	R																N																									
{SIVcpzPtt CAM3}																	N																									
{SIVcpzPtt MT145}																	N																									
{SIVcpzPtt CAM13}																	N																									
{SIVcpzPtt GAB1}																	N																									
{SIVcpzPtt US-AF103818}																	N																									
{SIVcpzPtt LB7 DQ373064}																	N																									
{SIVcpsPts Af447763}	F																N																			E	D	S	T			
{SIVcpzPts DQ374658}	F																N																					E	D	S	T	
{SIVcpzPts DQ374657}	F																N																						E	D	S	T
{HIV O L20571}																		N																					S	T	I	K
{HIV O L20587}																		N																						S	T	I
{gorilla CP2135, FJ424863}																		N																							V	M
{gorilla CP684, fj424871}																		N																							V	M
{gorilla CP2139 AJ424864}																		N																							V	M
{gorilla CP2139.2 FJ424869}																		N																							V	M
{gorilla pBR }																		N																							V	M

```

(crystal structure) P I P I H Y C A P A G F A I L K C N N K I F N - - G I G P C I N V S I
(HIV M M62320) . . . . . K D P E . . . . . K . . . . . R . . . . .
(HIV M U46016) . . . . . Y . . . . . R D . . . . . T . . . . . H . . . . .
(HIV M AF190127) . . . . . G . . . . . K . . . . . . . . . . .
(HIV M K03454) . . . . . R D . . . . . K . . . . . . . . . . .
(HIV M AF286238) . . . . . Y . . . . . K D . . . . . E . . . . . K . . . . . S . . . . . S . . . . . S
(HIV M U88824) . . . . . E . . . . . K . . . . . . . . . . . K . . . . .
(HIV M AY371158) . . . . . Y . . . . . Y . . . . . D . . . . . R . . . . . K . . . . . L . . . . .
(HIV M AY371158) . . . . . Y . . . . . Y . . . . . D . . . . . R . . . . . K . . . . . L . . . . .
(HIV M AY173951) . . . . . . . . . . . D . . . . . K . . . . . S . . . . . N . . . . . K . . . . .
(HIV M U54771) . . . . . T . . . . . Y . . . . . D . . . . . N . . . . . . . . . . . K . . . . . S . . . . .
(HIV M AF077336) . . . . . Y . . . . . D . . . . . R . . . . . . . . . . . K . . . . .
(HIV M AY084936) . . . . . . . . . . . V D . . . . . K . . . . . . . . . . . T . . . . . N . . . . .
(SIVcpzPtt-CAM5-AJ271369) . . . . . P . . . . . M . . . . . D Q K . . . . . . . . . . . I . . . . .
(HIV-1 N env complete genome (N env) . . . . . P . . . . . M . . . . . E A N . . . . . . . . . . . E . . . . . K . . . . .
(HIV N YBF30) . . . . . P . . . . . M . . . . . E G N . . . . . S . . . . . N . . . . . S . . . . .
(HIV N YBF16) . . . . . P . . . . . M . . . . . E G N . . . . . S . . . . . N . . . . . S . . . . .
(SIVcpzPtt-MB66) . . . . . . . . . . . L . . . . . D . . . . . N . . . . . T . . . . . I . . . . . Q . . . . . K . . . . .
(SIVcpzPtt-EK505) . . . . . Y . . . . . P . . . . . I . . . . . D Q D . . . . . . . . . . . E . . . . . N . . . . .
(SIVcpzPtt-MB897) . . . . . Y . . . . . L . . . . . D Q N . . . . . K . . . . . K . . . . . T . . . . . R . . . . .
(SIVcpzPts-U42720) . . . . . Y . . . . . R . . . . . E D E D . . . . . T . . . . . V . . . . . M . . . . . K . . . . . V . . . . .
(SIVcpzPtt-GAB) . . . . . . . . . . . L . . . . . D . . . . . D Y P . . . . . K . . . . . K . . . . . K . . . . .
(SIVcpzPtt-CAM3) . . . . . T P . . . . . Y . . . . . M . . . . . M P N . . . . . G T . . . . . R . . . . . N . . . . . I . . . . .
(SIVcpzPtt-M1145) . . . . . . . . . . . L . . . . . D I D Y K . . . . . N E T . . . . . K . . . . .
(SIVcpzPtt-CAM13) . . . . . . . . . . . D V N . . . . . S . . . . . K . . . . . K . . . . . R . . . . .
(SIVcpzPtt-GAB1) . . . . . . . . . . . D . . . . . D . . . . . S . . . . . K . . . . . K . . . . .
(SIVcpzPtt-US-AF103818) . . . . . P . . . . . . . . . . . E . . . . . D . . . . . K . . . . . K . . . . . E . . . . . K . . . . .
(SIVcpzPtt-LB7-DQ373064) . . . . . . . . . . . L . . . . . R . . . . . Y . . . . . S . . . . . N . . . . .
(SIVcpsPts-AF447763) . . . . . R . . . . . L . . . . . R D Q N . . . . . T . . . . . K . . . . . Q . . . . . S . . . . . V . . . . .
(SIVcpzPts-DQ374658) . . . . . R . . . . . Y . . . . . M . . . . . K D Q N Y T . . . . . N . . . . . I . . . . . Y . . . . . T V . . . . .
(SIVcpzPts-DQ374657) . . . . . R . . . . . Y . . . . . M . . . . . D Q G Y T . . . . . N . . . . . T . . . . . S . . . . . T V . . . . .
(HIV O L20571) . . . . . T . . . . . Y . . . . . F . . . . . D T D . . . . . . . . . . . L . . . . . H . . . . . I . . . . . V . . . . .
(HIV O L20587) . . . . . Y . . . . . F . . . . . S T E . . . . . . . . . . . T . . . . . R . . . . . I T V . . . . .
(gorilla CP2135, FJ424863) . . . . . L . . . . . Q . . . . . Y . . . . . L M . . . . . D . . . . . Q . . . . . . . . . . . E . . . . . K . . . . . I . . . . .
(gorilla CP684, fj424871) . . . . . L . . . . . Q . . . . . Y P L . . . . . D E Q . . . . . . . . . . . E . . . . . N . . . . . I . . . . .
(gorilla CP2139 AJ424864) . . . . . L . . . . . Q . . . . . Y . . . . . L M . . . . . D I Q . . . . . . . . . . . E . . . . . R . . . . . I . . . . .
(gorilla CP2139.2 FJ424869) . . . . . L . . . . . Q . . . . . Y . . . . . L M . . . . . D I Q . . . . . . . . . . . E . . . . . R . . . . . I . . . . .
(gorilla pBR ) . . . . . L . . . . . Q . . . . . Y . . . . . L M . . . . . D I Q . . . . . . . . . . . E . . . . . R . . . . . I . . . . .
    
```

	100	110	120	130	140																																	
(crystal structure)	V	Q	C	T	H	G	I	R	P	V	V	S	T	Q	L	L	N	G	S	L	A	E	E	E	V	V	I	R	S									
(HIV M M62320)								K																			R	I	R									
(HIV M U46016)								K																			G	T	I			F						
(HIV M AF190127)																											V											
(HIV M K03454)																																						
(HIV M AF286238)								A																			G	K	M									
(HIV M U88824)																																						
(HIV M AY371158)																																						
(HIV M AY371158)																																						
(HIV M AY173951)																																						
(HIV M U54771)																																						
(HIV M AF077336)																																						
(HIV M AY084936)																																						
(SIVcpzPtt CAM5 AJ271369)																												T	V				E	G	P	A	A	
(HIV-1 N env complete genome)								K																														
(N env)								M																														
(HIV N YBF30)								K																														
(HIV N YBF16)								K																														
(SIVcpzPtt MB66)	H																																					
(SIVcpzPtt EK505)								K																														
(SIVcpzPtt MB897)	H							K																														
(SIVcpzPtt U42720)	H							S	M	A	W																											
(SIVcpzPtt GAB)	H							K	T	T																												
(SIVcpzPtt CAM3)																																						
(SIVcpzPtt MT145)	H							K	A	T																												
(SIVcpzPtt CAM13)	H							K	T																													
(SIVcpzPtt GAB1)	H							K	T																													
(SIVcpzPtt US AF103818)								K	T																													
(SIVcpzPtt LB7 DQ373064)	H							K	T																													
(SIVcpsPtt AF447763)	H							Y	M	A	A	H																										
(SIVcpzPtt DQ374658)	H							F	M	A	A	H																										
(SIVcpzPtt DQ374657)	H							F	M	A	A	H																										
(HIV O L20571)	T							K	T																													
(HIV O L20587)	T																																					
(gorilla CP2135 FJ424863)	H																																					
(gorilla CP684 FJ424871)	H																																					
(gorilla CP2139 AJ424864)	H																																					
(gorilla CP2139.2 FJ424869)	H																																					
(gorilla pBR)	H																																					

	140	150	160	170	175
(crystal structure)	V N F T D N A K T I I V	--- Q L N T S V E I N C T	--- H C N I S R A		
(HIV M M62320)	E L L N S I F T L T T	--- V N P L K L S T	--- V L R		
(HIV M U46016)	E L L N S I F T L T T	--- E L T L T	--- E E		
(HIV M AF190127)	K L I L T N S L T L T	--- E P L Q L T	--- Y L G K		
(HIV M K03454)	E L L N S I N L A T H L	--- E L K T L A	---		
(HIV M AF286238)	E L L N S I N L T L T	--- F L K P L P L	---		
(HIV M U88824)	E L L N S I F T L T T	--- E L P L T L	--- G L		
(HIV M AY371158)	E L L N S I F T L T T	--- F L E L K L T	--- I N G T		
(HIV M AY371158)	E L L N S I F T L T T	--- F L E L K L T	--- I N G T		
(HIV M AY173951)	E L L N S I F T L T T	--- K E P L T L	--- S T		
(HIV M U54771)	E D L L N S I F T L T	--- H L K L T L	--- Y L E N G T		
(HIV M AF077336)	Q L S N S I F T L T T	--- H L E L Q L T	--- G T		
(HIV M AY084936)	E L L S I F T L T L T	--- H L R L T L	--- V L W T		
(SIVcpzPtt-CAM5 AJ271369)	E D L L R H L T L T T	--- F K E G I N L T	--- M L V L K L		
(HIV-1 N env complete genome (N env))	--- N N S G G N L L	--- W L E I T M L	--- V L N		
(HIV N YBF 30)	--- S R G N L L	--- W L E T S L K L	--- V L E		
(HIV N YBF 16)	--- N L S H S N L L	--- W L E T L P L	--- Y L V L K E		
(SIVcpzPtt MB66)	--- H L H G G N L L	--- W L E T L S L	--- Y L V L E		
(SIVcpzPtt EK505)	--- S N D M D L	--- K L E T L R L	--- F L Q L N K T		
(SIVcpzPtt MB897)	--- S K D L L	--- E L P L T L	--- Y L E V N L T		
(SIVcpzPts U42720)	K S L T E L T L T	--- A L A I Y L T	--- E L R N G T		
(SIVcpzPtt GAB)	--- E S V L R F G K E F E N L T	--- T L F L F L T V N K T			
(SIVcpzPtt CAM3)	M T A P G L S D L V L T	--- K R A I P L S L S L	--- H L N G T		
(SIVcpzPtt MT145)	--- G Y N F L L T L T	--- F Q K L S L T L	--- Y L Q N L T		
(SIVcpzPtt CAM13)	--- P S E N L A L T L T	--- K D P L K L T R L	--- Y L E N G T		
(SIVcpzPtt CAM13)	A S K L T H D W L T L T	--- S L A N L T L K L F L	--- E L N G T		
(SIVcpzPtt GAB1)	K S K L T D V W L T L T	--- V E A S L L H L Y L	--- K N G T		
(SIVcpzPtt US AF103818)	--- R S K N F L D I L L T L T	--- F S E G L N M T L	--- F L K L		
(SIVcpzPtt LB7-DQ373064)	D K L L L V L L T L T	--- I H L L L L L	--- E L N Q T		
(SIVcpsPts AF447763)	N T P L L L L L T L T	--- K F L V L I N L T L	--- E L Y L S V N A T		
(SIVcpzPts DQ374658)	N P P L L L L L T L T	--- K F L E L P L T L	--- G L Y L W V N A T		
(SIVcpzPts DQ374657)	N P P L L L L L T L T	--- K F L E L S L T L	--- E L Y L H V N A T		
(HIV O L20571)	--- L E S L N L L T L T	--- P I N M T L	--- Y L T Y N K T		
(HIV O L20587)	--- D I L E G G L N L T L T	--- S T L N M T L	--- E L Y L K Y N A T		
(gorilla CP2135. FJ424863)	--- S S L S G L D L L T L T	--- K L A R P L T L	--- E L F L K Y N K T		
(gorilla CP684. fj424871)	--- S S L S L T D L L T L T	--- K L A N P L T L	--- E L F L K Y N T T		
(gorilla CP2139 AJ424864)	--- S S L S G L D L L T L T	--- K L A R P L T L	--- E L F L K Y N E T		
(gorilla CP2139.2 FJ424869)	--- S S L S G L D L L T L T	--- K L A R P L T L	--- E L F L K Y N E T		
(gorilla pBR-)	--- S S L S G L D L L T L T	--- K L A R P L T L	--- E L F L K Y N E T		

	170	180	190	200	210																														
(crystal structure)	K	W	N	T	L	K	Q	I	A	S	K	L	R	E	Q	F	G	N	N	K	T	I	I	-	F	K	Q	-	-	S	S	G	G		
(HIV M M62320)	D	-	R	-	I	Q	-	V	-	E	Q	-	K	K	K	-	N	-	-	-	-	-	-	-	-	F	A	-	-	S	-	-	-	-	
(HIV M U46016)	-	-	-	-	K	-	-	Q	K	V	K	E	-	Q	K	H	-	P	-	-	-	-	-	-	-	E	-	-	P	-	-	-	-	-	
(HIV M AF190127)	Q	-	E	-	-	H	K	V	I	T	-	-	G	S	Y	-	D	-	-	-	-	-	-	-	-	L	Q	P	-	-	P	A	-	-	
(HIV M K03454)	Q	-	S	K	-	Q	-	V	-	R	-	-	G	T	L	-	-	-	-	-	-	-	-	-	-	K	-	P	-	-	-	-	-	-	
(HIV M AF286238)	E	-	A	-	-	-	K	V	V	E	Q	-	-	H	-	P	-	-	-	-	-	-	-	-	-	F	N	-	-	S	-	-	-	-	
(HIV M U88824)	G	-	K	-	Q	-	V	-	E	-	-	G	N	L	L	N	Q	-	T	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	
(HIV M AY371158)	L	-	A	-	-	N	R	V	-	A	E	V	K	N	L	T	N	I	-	-	-	-	-	-	-	K	-	E	P	-	-	-	-	-	
(HIV M AY371158)	L	-	A	-	-	N	R	V	-	A	E	V	K	N	L	T	N	I	-	-	-	-	-	-	-	K	-	E	P	-	-	-	-	-	
(HIV M AY173951)	-	-	-	-	-	-	G	-	V	K	-	-	K	-	-	N	-	-	-	-	-	-	-	-	T	I	V	F	K	K	-	-	-	-	
(HIV M U54771)	-	-	-	-	K	-	V	L	K	Q	V	T	E	K	L	K	E	H	F	-	K	T	I	-	-	F	Q	P	-	-	-	-	-	-	-
(HIV M AF077336)	Q	-	-	-	E	Y	V	K	A	E	-	K	S	H	-	P	-	-	-	-	-	-	-	-	T	A	L	K	I	N	Q	-	-	-	
(HIV M AY084936)	-	-	-	-	E	-	R	D	V	Q	A	-	Q	-	Y	-	I	-	K	S	I	E	I	N	S	-	-	-	-	-	-	-	-	-	
(SIVcpzPtt-CAMS_AJ271369)	Q	-	M	K	-	N	E	T	-	L	A	I	A	K	Y	D	K	S	I	A	K	V	I	F	T	P	N	-	-	-	-	-	-	-	
(HIV-1 N env complete genome (N env))	E	-	R	S	M	W	N	K	T	K	E	-	I	K	S	L	L	-	-	-	-	-	-	-	N	I	T	F	K	A	Q	E	K	N	-
(HIV NYBF30)	N	-	R	I	M	W	N	R	T	K	E	-	I	G	N	K	-	V	Q	-	-	-	-	-	-	A	R	E	K	N	-	-	-	-	
(HIV NYBF16)	L	-	E	P	M	W	N	R	T	R	E	E	I	K	K	I	L	-	K	-	N	I	T	F	R	A	R	E	R	N	E	-	-	-	-
(SIVcpzPtt-MB66)	E	-	G	S	M	W	N	K	T	K	K	-	I	K	R	L	L	-	-	-	-	-	-	-	N	-	T	F	K	A	Q	D	K	N	-
(SIVcpzPtt-EK505)	V	-	T	-	-	F	-	K	V	R	H	A	-	N	-	I	Y	K	G	Y	L	C	N	E	I	I	T	F	-	G	P	-	-	-	-
(SIVcpzPtt-MB897)	-	-	E	Q	I	W	N	T	T	K	Q	I	I	E	N	-	R	K	-	I	-	-	-	-	-	F	P	N	-	-	-	-	-	-	
(SIVcpzPts U42720)	A	-	E	-	-	R	-	V	K	E	A	-	N	N	T	Y	R	-	L	N	L	S	L	T	E	I	N	F	E	G	A	-	-	-	-
(SIVcpzPtt-GAB1)	L	-	E	Q	A	R	N	K	T	E	H	V	-	A	-	H	W	K	K	V	D	N	K	I	N	A	-	T	I	W	T	F	Q	D	-
(SIVcpzPtt-CAM3)	L	-	-	-	I	-	N	R	-	K	Q	-	I	K	N	S	T	I	W	H	R	G	D	-	I	-	T	K	H	-	-	-	-	-	
(SIVcpzPtt-MT145)	V	-	D	E	R	-	N	E	T	G	Q	A	-	-	L	-	T	-	L	T	Q	V	N	F	I	V	-	-	-	-	-	-	-	-	
(SIVcpzPtt-CAM13)	Q	-	A	K	A	-	N	E	T	K	E	V	-	-	N	I	L	R	K	-	I	S	F	M	V	-	-	-	-	-	-	-	-	-	
(SIVcpzPtt-GAB1)	T	-	E	-	-	Q	K	V	R	E	S	-	I	K	E	I	K	A	-	A	N	G	T	Y	N	I	T	F	E	P	-	-	-	-	-
(SIVcpzPtt-US_AF103818)	T	-	R	-	V	E	E	V	K	K	A	-	A	T	S	S	N	R	I	A	A	N	-	I	L	N	R	-	-	-	-	-	-	-	-
(SIVcpzPtt-LB7_DQ373064)	T	-	E	K	Q	R	-	Y	T	L	E	I	I	K	K	E	A	N	L	-	V	E	L	I	P	N	-	-	-	-	-	-	-	-	-
(SIVcpsPts AF447763)	H	-	Y	K	I	-	N	-	V	K	R	E	-	I	T	V	-	N	E	S	N	K	T	V	H	-	A	N	-	-	-	-	-	-	-
(SIVcpzPts DQ374658)	T	-	Y	R	N	-	D	W	A	M	A	A	I	N	I	M	R	A	R	N	E	T	V	Q	T	F	Q	W	Q	R	-	-	-	-	-
(SIVcpzPts DQ374657)	L	-	Y	K	S	-	D	N	A	M	A	I	N	K	T	L	K	K	S	G	L	N	V	I	R	R	F	D	W	H	K	-	-	-	-
(HIV O L20571)	L	-	D	Q	S	-	S	R	A	M	E	A	I	N	K	T	L	A	L	Y	N	K	A	V	N	E	-	F	E	W	S	K	-	-	-
(HIV O L20587)	V	-	E	-	A	-	Q	-	T	-	I	R	Y	L	N	L	V	N	Q	T	E	N	V	T	I	I	F	S	R	-	-	-	-	-	-
(gorilla CP2135_FJ424863)	D	-	G	K	I	-	-	T	-	E	R	Y	L	-	L	V	N	-	T	G	S	-	N	M	I	F	N	H	-	-	-	-	-	-	
(gorilla CP684_fj424871)	N	-	Q	-	A	-	-	D	T	V	R	A	-	K	A	N	Q	I	R	F	Q	N	-	-	-	-	-	-	-	-	-	-	-	-	-
(gorilla CP2139_AJ424864)	A	-	Q	K	A	-	Q	A	T	V	R	A	-	K	T	N	L	T	H	Y	F	N	K	-	-	-	-	-	-	-	-	-	-	-	-
(gorilla CP2139.2_FJ424869)	N	-	Q	G	A	-	-	D	T	V	Q	A	-	N	-	S	L	I	H	F	R	N	-	-	-	-	-	-	-	-	-	-	-	-	-
(gorilla pBR)	N	-	Q	G	A	-	-	D	T	V	Q	A	-	N	-	S	L	I	H	F	R	N	-	-	-	-	-	-	-	-	-	-	-	-	-

```

(crystal structure) D P E I V T H S F N C G G E F F Y C N S T Q L E N S I W F N S I W S T
(HIV M M62320) . . . . . T S G . . . . . I . N G . . . . .
(HIV M U46016) . . . . . T S N . . . . .
(HIV M AF190127) . . . . . T . K . . . . . T . . . S Y T N
(HIV M K03454) . . . . . T S G . . . . . N I . A . N N
(HIV M AF286238) . . . . . T . G . . . . . E . G . N K Q
(HIV M U88824) . . . . . T . R . . . . . K R N N S E W
(HIV M AY371158) . . . . . S . . . . . D T . A . . . T . L L . T . M D N
(HIV M AY371158) . . . . . S . . . . . D T . A . . . T . L L . T . M D N
(HIV M AY173951) . . . . . M . . . . . F . . . . . K . . . . . N S T E G N D
(HIV M U54771) . . . . . T M . H . . . . . R . . . . . T . K . . . . . N . C L G N E T M A
(HIV M AF077336) . . . . . T M . . . . . R . . . . . D T S G . . . . . D . G S . N G T . .
(HIV M AY084936) . . . . . T . . . . . . . . . . T S G . . . . . N S I L K . N I . E
(SIVcpzPtt CAM5 AJ271369) . . . . . T N M M . . . . . . . . . . T . P . . . . . K F N Y T C D N Q N
(HIV-1 N env complete genome (N env) . . . . . V T H L M . . . . . . . . . . T S R . . . . . E S M N T N . . . . .
(HIV N YBF30) . . . . . V T H F M . . . . . . . . . . T S K . . . . . E . G D T . . . . .
(HIV N YBF16) . . . . . L . V T H L M . . . . . . . . . . T S R . . . . . E S E N K T N K
(SIVcpzPtt MB66) . . . . . L . V T N L H L I . . . . . . . . . . T S I . . . . . T S I I F N E T K
(SIVcpzPtt EK505) . . . . . L . V T N L M I . . . . . . . . . . T S . . . . . T N Q N G . . . . .
(SIVcpzPtt MB897) . . . . . L . V T . . . . . Y . . . . . . . . . . T S N . . . . . Q S T . . . . .
(SIVcpzPts U42720) . . . . . V K V . W . . . . . Q . . . . . . . . . . D I . P W . . . . . A . Y T G N L I . . . . .
(SIVcpzPtt GAB) . . . . . V . N F M . . . . . . . . . . . . . . T S R . . . . . I C . . . . .
(SIVcpzPtt CAM3) . . . . . V T N M M . . . . . . . . . . . . . . T . T . . . . . Y . . . K . N N . . . . .
(SIVcpzPtt MT145) . . . . . V T N . H . . . . . . . . . . . . . . T S E I I . . . . . I . K . . . . .
(SIVcpzPtt CAM13) . . . . . A N . M . . . . . . . . . . . . . . D T R K M . . . . . E S E P . . . . .
(SIVcpzPtt GAB1) . . . . . V T H . M . . . . . . . . . . . . . . T S . . . . . I . T D N I . . . . .
(SIVcpzPtt US AF103818) . . . . . V . N M M L . . . . . . . . . . . . . . T I P . . . . . M . Y N . . . . .
(SIVcpzPtt LB7 DQ373064) . . . . . V A N L H . . . . . . . . . . . . . . T S A . . . . . D . L L R N K . . . . .
(SIVcpsPts AF447763) . . . . . V T S F W . . . . . Q . . . . . . . . . . L . N W T . . . . . T . . . I A N R I N N
(SIVcpzPts DQ374658) . . . . . L . V Q N F W . . . . . Q . . . . . . . . . . L . T W T G N . . . . . T R N R T H Q
(SIVcpzPts DQ374657) . . . . . V S S F W . . . . . Q . . . . . . . . . . L . G W T . . . . . P . . . . . N . T N H R N
(HIV O L20571) . . . . . A . V S H L H . . . . . H . . . . . . . . . . T S G M . . . . . Y . F I . C . K . G
(HIV O L20587) . . . . . L . V T H L H . . . . . H . . . . . . . . . . T A K M . . . . . Y . F S . C N G T . .
(gorilla CP2135 FJ424863) . . . . . V T F L H . . . . . H . . . . . . . . . . D T . K M . . . . . Y N . . . . . C . K E S
(gorilla CP684 fj424871) . . . . . V T S L H . . . . . H . . . . . . . . . . T . K M . . . . . Y K . . . . . C . G E H
(gorilla CP2139 AJ424864) . . . . . S . V T F L H . . . . . H . . . . . . . . . . T . K M . . . . . S Y N . . . . . C . T E G
(gorilla CP2139.2 FJ424869) . . . . . S . V T F L H . . . . . H . . . . . . . . . . T . K M . . . . . S Y N . . . . . C . T E G
(gorilla pBR ) . . . . . V T F L H . . . . . H . . . . . . . . . . T . K M . . . . . S Y N . . . . . C . T E G
    
```

	240	250	260	270	280
(crystal structure)	...KGSN	NTEGSD	TIILPC	RIRKQI	INMWQKVGK
(HIV M M62320)	...SMS	DM PNG	...Q	...R	...Q
(HIV M U46016)	...LELE	FNSSTINL	...Q	...G	...R
(HIV M AF190127)	DTYNSN	...IT	...DITGN	...Q	K...V...R...Q
(HIV M K03454)	ITES	...NNS	TNTN	...Q	...K VAGR
(HIV M AF286238)	NY	...TE	ND	...Q	...R...R
(HIV M U88824)	RS DN	...TP	E	...Q	...E
(HIV M AY371158)	NG	...I	...V	V...R...R	
(HIV M AY371158)	NG	...I	...V	V...R...R	
(HIV M AY173951)	DGEER	...N	...V	...E	
(HIV M U54771)	GCND	...K	...G	A...Q	
(HIV M AF077336)	...V	...G	...R		
(HIV M AY084936)	NND	...N	K...V	R...R...Q	
(SIVcpzPtt CAM5 AJ271369)	NCTR	NNT...D	NSDNS	L...IS	...R...V...S...MR
(HIV_1 N env complete genome (N env))	...TNG	...R	V...L...TR		
(HIV N YBF30)	...G	...R	V...L...TR		
(HIV N YBF16)	...TGE	P...R	V...L...TR		
(SIVcpzPtt MB66)	...DN	I...R	VRL...R...R		
(SIVcpzPtt EKS05)	...TGN	Q...R	V...L...TR		
(SIVcpzPtt MB897)	...INNI	THI...R	V...Q...G		
(SIVcpzPts U42720)	...TNG	ALIAH...V	H...GI...S		
(SIVcpzPtt GAB)	...NSS	DTSEYI...K	R...V...S...MR		
(SIVcpzPtt CAM3)	...ITK	G...N...TFF	...R...V...S...MR		
(SIVcpzPtt MT145)	...INK	TENMTII...R	V...S...MR		
(SIVcpzPtt CAM13)	...FHEN	M...I...R	V...S...MR...R		
(SIVcpzPtt GAB1)	...TNG	I...R	V...S...MR...R		
(SIVcpzPtt US AF103818)	...TDN	I...K...R	V...Q...MR		
(SIVcpzPtt LB7 DQ373064)	...TLL	N...TLH	NNL...R	V...L...LR	
(SIVcpsPts AF447763)	THG	...LVA	...LR	V...H...GI...S	
(SIVcpzPts DQ374658)	TS	...ALNK	THLLAQ...LR	V...H...GI...S	
(SIVcpzPts DQ374657)	...G	LVA...LR	V...H...GI...S		
(HIV O L20571)	CQEI	...E	TNKNG...I...KLR	L...VRS...M...GES	
(HIV O L20587)	CSVS	NV...Q...GN	NG...KLR	V...RS...IRGQS	
(gorilla CP2135 FJ424863)	CDC	...ITG...C...N...YI	...HL...VMS...MR...S		
(gorilla CP684 FJ424871)	CNCT	DKELCN	KTAEM...EYI...L...VV...S...MR...S		
(gorilla CP2139 AJ424864)	CNTT	ET...T...TK...YI	...HL...VRS...MR...S		
(gorilla CP2139.2 FJ424869)	CNTT	ET...T...TK...YI	...HL...VRS...MR...S		
(gorilla pBR)	CNTT	ET...T...TK...YI	...HL...VRS...MR...S		

	310	320	330	340	347
(crystal structure)	... F R P G G D M R D N W R S E L Y K Y K V V K T E				
(HIV M M62320) K				
(HIV M U46016)	I ... E				E . K
(HIV M AF1901271) Q ..				
(HIV M K03454) Q ..				
(HIV M AF286238)	T				
(HIV M U88824)	I ... N				L .
(HIV M AY371158)	... L . I . E				Q . Q
(HIV M AY371158)	... I L . I . E				Q . Q
(HIV M AY173951)	E ... I ... N				
(HIV M U54771)	... T ... N I K				Q .
(HIV M AF077336)	... T ... E . N . K				E .
(HIV M AY084936) K N				T . . K
(SIVcpzPtt - CAM5 - AJ271369)	I T ... V Y . S . E . . L . . L . . N R . . . S . .				
(HIV - 1 N env complete genome (N env))	... V Y . T . . N . V N L . . Q T . S . .				
	... V Y . S . . N . V N L . . H . . . Q . . . S . .				
(HIV N YBF 30)	- K E T - I V Y . S . . N . V N L . . Q S . .				
(HIV N YBF 16)	- K - - - T V Y . S . . N . V N L . . Q S . .				
(SIVcpzPtt - MB66)	S A - - - M . T . V . . E . . N R . .				
(SIVcpzPtt - EKS05)	Y . . . T I Y . T . . . T N L . . Q S . .				
(SIVcpzPtt - MB897)	... T . . T . . . N				
(SIVcpzPts U42720)	... K V S . - A A R V A . Q . . A . . S R . Q . . E . ?				
(SIVcpzPtt - GAB)	... V Y L S I I				
(SIVcpzPtt - CAM3)	T ... L Y . T . . N . V . L . . L . . H . . . S . .				
(SIVcpzPtt - MT145)	D Q N N T Y V C L T . . N . K . I I . E . Q				
(SIVcpzPtt - CAM13) I K N I R . .				
(SIVcpzPtt - GAB1)	L T T . . N . K . I R . .				
(SIVcpzPtt - US - AF103818)	I T ... V M . T T A L . K N . . H . . . S . .				
(SIVcpzPtt - LB7 - DQ373064)	I T T N . . I				
(SIVcpsPts AF447763)	... T P Q F S A V V E . Y . K V . . A R . . . E . Q				
(SIVcpzPts DQ374658)	... T P T F S S K V E . Y . K V . . A R . . . E . Q				
(SIVcpzPts DQ374657)	... V P T F S . K V E . Y . K V . . A R . . . E . Q				
(HIV O L20571)	T ... L . V . . . K . I . . T K . . N . . Q . K				
(HIV O L20587)	T I . . . K . I . . T . . F N . . . R V K				
(gorilla CP2135, FJ424863)	I D Y N - I L . . I . . . T N I P V K				
(gorilla CP684, FJ424871)	T ... L . . L . . E . . N I P V K				
(gorilla CP2139 AJ424864)	I ... L . . I . . . T N I P R V K				
(gorilla CP2139.2 FJ424869)	T ... L . . I . . . T N I P R V K				
(gorilla pBR -)	T ... L . . I . . . T N I P R V K				