12-2014

Parallel Adaptation of CD4 in SIV-Endemic Gorilla and Pan Lineages

Katie E. Brown
University at Albany, State University of New York

Parallel adaptation of CD4 in SIV-endemic *Gorilla* and *Pan* lineages

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

Katie E. Brown

2014

The Honors College

University at Albany

BIOLOGICAL SCIENCES

Certification of the Completion of a Senior Honors Thesis

- Successful completion of a senior honors thesis is required for graduation from The Honors College.
- The determination of whether a student's work satisfies the criteria for a senior honors thesis is left to each department.
- Each honors student must return this completed certification, along with an electronic copy of his or her senior honors thesis, to the honors college office.
- This certification must be signed by the student's thesis advisor and by either the Director of the department's honors program or the Chair of the department.

This Honors Thesis has been read and approved by the signatures below and, is hereby recommended for acceptance.

Student Name: Katie Elizabeth Brown **Student ID: 001075517**

Thesis Title: Parallel adaptation of CD4 in SIV-endemic *Gorilla* and *Pan* lineages.

Department in which Thesis was completed: **BIOLOGICAL SCIENCES**

Honors Thesis **Presentation Date** **April 18th, 2014**

***Senior Honors Thesis Research Advisor:** Caro-Beth Stewart 12/10/14

Print name Signature Date

***Honors Thesis Committee Member:** Albert Millis 12/10/14

Print name Signature Date

Honors Thesis Committee Member (optional):

Print name Signature Date

***Required**

CHOOSE ONE: ❐ **Department Chair** OR ☒ Director of Department Honors Committee

Biology Honors Committee Director & Department Chair RICHARD S. ZITOMER 12/11/14

Print name Signature Date

Original	– Dr. Jeffrey Haugaard, Associate Vice Provost and Honors College Director LC-31
Copy	– Department of Biological Sciences

ABSTRACT

The simian immunodeficiency virus (SIV) is homologous to the human immunodeficiency virus (HIV), and naturally infects chimpanzees and gorillas in the wild. Some African primate species appear to have evolved resistance to SIV, in that the virus no longer is found in the species or, if infected, individuals within the species show no serious symptoms of simian AIDS (SAIDS). In contrast, Asian primate species do not appear to naturally harbor SIV and, like humans, often progress to AIDS following infection. CD4 is the primary T cell receptor that SIV/HIV interacts with to infect host T cells. Domain 1 (D1) of CD4 holds the main interaction with the viral envelope protein, gp120.

During my course in Dr. Stewart's lab, I analyzed 77 primate CD4 gene sequences in comparison to each other and to the *Homo* lineage in search of fixed changes on lineages, as well as sequence variation within species. I found that D1 in the known SIV resistant species, chimpanzee and gorilla, had several notable amino acid replacements on ancestral lineages, as well as variation within the species. These amino acid replacements likely have the potential to prevent SIV gp120 from binding to CD4. In contrast, the *Homo* sequence had no amino acid replacements and little variation within humans. Perhaps these results shed light on why humans suffer with HIV/AIDS today, with little to no resistance to infection.

ACKNOWLEDGEMENTS

I would like to thank all of the members of the Stewart laboratory, both previous and current. Dr. Caro-Beth Stewart, Dr. Randall Collura, Fahd Sheikh, Jennifer Giza, and Steven Nevins: without you this would not have been possible. Thank you Dr. Stewart for welcoming me into your lab with open arms. Thank you for your patience, kindness, and wisdom. You have been a mentor to me from the moment I met you, and you have taught me so much that I would not have learned otherwise. You have played a large role in my success, and I cannot thank you enough for your support. Dr. Collura, you were always there to help us with the difficult procedures. I appreciated your helpful explanations as well as your jokes and sarcasm. You always kept Dt. Stewart informed about what needed to be done. Fahd, you passed on your knowledge to help me learn the ropes of the lab. Thank you for making time for me and thank you for the laughter. Jennifer, you always knew how to light up a room, and you were always there to help. Steven, you are amazing at the technical work, and you do a great job of explaining your methods. Thank you for your help and for always being there to listen.

I would like thank the Research Foundation for providing me with STEM support this past summer. Without that support, I would not have been able to pay for my summer stay to do research. You helped me to connect with my fellow undergraduate researchers, and supported my love for research. In addition, thank you for the FRAP-B grant and endowed fellowship that you granted our lab. Your support is greatly appreciated. I would like to thank Dr. Albert Millis for seeing my potential and agreeing to be on my honors committee, despite his partial retirement. Thank you for your kindness, patience, and support.

Lastly, I would like to thank my parents, boyfriend, and close friend. Edith Fundaro, Dean Brown, Miguel Moran, and Stephanie Grant: thank you for your support and for always believing in me, no matter how tough things seemed at the time. I love you all dearly. Stephanie and Miguel, thank you for making the time to help me make the figures I needed for this thesis, as well as for previous papers. Miguel, thank you for always being willing to bring me lunch when I did not have time to pack one for myself. Thank you both for cooking for me before I arrived home so that I could eat right away after my long day at lab. You guys are the absolute best.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

HIV and SIV

HIV is a relatively new virus for humans, believed to have emerged within the past 100 years following transmission from chimpanzees and sooty mangabeys (Gifford *et al.*, 2012; Compton *et al*., 2013; Locatelli *et al.*, 2014). Immunodeficiency viruses are estimated to have infected African monkeys, such as the colobus monkeys, mangabeys, mandrills, and green monkeys, for at least 14 million years, making several jumps to African apes (Compton & Emerman, 2013). Today, most African monkey species harbor strains of SIV (Castro-Nallar *et al*., 2012).

SIV has co-evolved with host species and has made several species jumps (Castro-Nallar *et al.*, 2012). Importantly for this project, phylogenies of SIV genes suggest that African apes acquired SIV from African monkeys through at least one jump, and that SIV jumped from monkeys and apes to humans several times independently, resulting in HIV-2 and HIV-1, respectively. SIV from sooty mangabeys jumped to humans at least three times, leading to HIV-2, while gorilla SIVs and chimpanzee SIVs jumped to humans at least four times recently resulting in HIV-1. HIV-1 has become a global human pandemic in the past 30 years, killing 33 million people worldwide (Locatelli *et al.*, 2014). These observations raise the question of why this virus is so lethal to humans, yet is endemic and not as lethal to African primate hosts. Why humans progress to AIDS more rapidly than African primate hosts is an unsolved mystery. A major place where one might expect to see differential selection would be on the HIV receptor molecules, CD4 being the primary one. The purpose of my research was to look for evidence of selection on CD4 in African apes.

CD4 Background

The typical function of CD4 in a healthy individual is to bind to MHC class II molecules to enhance the activity of T cells (Harrison, 1993). CD4 is encoded by 10 exons that span over 31,000 basepairs (bp) of DNA. The extracellular portion of the protein is comprised of four immunoglobulin-like domains. Domain 1 (D1) is encoded by two exons that are separated by over 10,000 bp. CD4 is the primary T cell receptor for HIV/SIV (Sharp & Hahn, 2011).

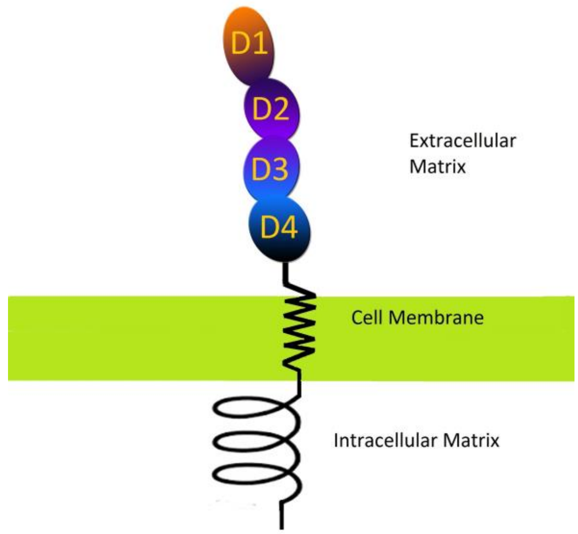


Figure 1. Diagram of CD4. CD4 is a T-cell receptor that consists of four immunoglobulin-like domains (D1, D2, D3, and D4), as well as transmembrane and intracellular segments.

SIV/HIV infects T cells through the binding of gp120, the viral envelope glycoprotein, with the host glycoprotein CD4 and a chemokine co-receptor, such as CCR5 or CXCR4. The viral gp120 protein binds preferentially to the most amino-terminal of the four immunoglobulin-like domains that CD4 possesses, domain 1 (D1), but may also interact with domain 2

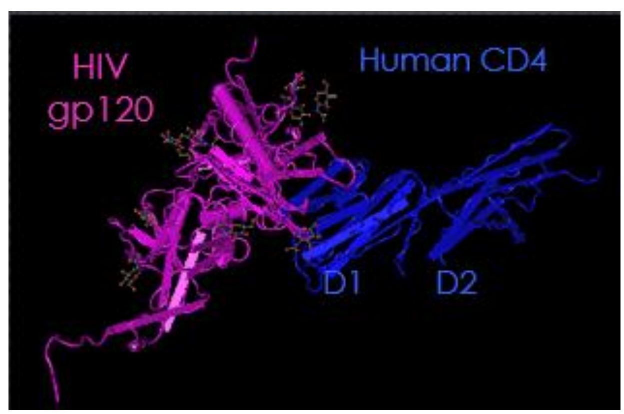


Figure 2. CD4/GP120 Ribbon Diagram. A ribbon diagram of HIV gp120 bonded to Human CD4 domain 1. (Source: http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?uid=79081)

(D2) (Harrison, 1993). Previous research has shown that there is notable variation across species in D1 of CD4 (Hvilsom *et al*., 2008; Koito *et al*., 1994; Wyatt *et al*., 1995). The interaction of SIV/HIV interaction with CD4 primarily involves the D1 region, as well as the complementary determining region (D2) of CD4 (Koito *et al*., 1994; Wyatt *et al*., 1995). D1 possesses the highest species-specific variation seen so far in CD4 (Hvilsom *et al*., 2008).

The binding of CD4 and gp120 causes conformational changes in gp120 that exposes the binding site for the chemokine co-receptor, typically either CCR5 or CXCR4. This conformational change enables the cell membrane and viral membrane to fuse. SIV/HIV is then able to implant its genetic information into the host cell (Freeman & Heron, 2007). When HIV/SIV is bound to CD4 in a host, T-cell activation is reduced. In highly susceptible hosts, SIV and HIV cause the destruction of CD4+ lymphocytes, which leads to the onset of acquired immunodeficiency syndrome (AIDS) (Kwong *et al*., 1998).

Analysis of the Great Apes CD4 Gene Sequences

For this thesis project, I examined the CD4 sequences of the great apes to see how they differ from each other and from the *Homo* (human, Neanderthal, and Denisovan) lineage. We hypothesized that part of the mystery of SIV resistance in African apes, and the lack of such resistance in humans and Asian apes, was buried within CD4.

The four recognized subspecies of chimpanzee are the western (*Pan troglodytes verus*), Cameroonian (*P.t. ellioti*), central *(P.t. troglodytes*), and eastern (*P.t. schweinfurthii*) chimpanzees (Sharp & Hahn, 2011; Prado-Martinez *et al.*, 2013). Certain central and eastern chimpanzee populations harbor endemic SIV in the wild, whereas the western and Cameroonian chimpanzees do not (reviewed in Sharp & Hahn, 2011). The bonobos, also known as pygmy chimpanzees (*Pan paniscus*), appear to lack endemic SIV (Dooren *et al.*, 2002; Li *et al*., 2012).

Early estimates of the age of SIV in chimpanzees were as low as 500-1000 years (Wertheim & Worobey, 2009), but these dates were recalibrated to a minimum of about 77,000 years using the separation of Bioko Island from mainland Africa (Worobey *et al.*, 2010). Worobey *et al.* reasoned that since each Bioko SIV lineage shares ancestry with mainland virus that was found in monkeys of the same genus, the mainland and Bioko SIV variants must have been evolving independently since the isolation of Bioko Island (10,000 years before present [B.P.]). They used the ancestor date of 10,000 years B.P. as well as amino acid sequences to recalibrate the molecular clock, and found that the most was recent common ancestor of SIV was 76,794 years B.P. Because of the limited distribution of SIV in the wild, it is widely believed among HIV researchers that central and eastern chimpanzees likely acquired SIV after their divergence from western and Cameroonian chimpanzees (Sharp & Hahn, 2011). Based upon population divergence estimates by Prado-Martinez *et al.* (2013), this would suggest that SIV infected the ancestor of the central and eastern chimpanzees between about 300-800 thousand years ago (kya). Other researchers have suggested that SIV is much older than this in the *Pan* lineage, with estimates at over 2 million years ago (mya) (Walter *et al.*, 2005; de Groot & Bontrop, 2013). As will be explained in this thesis, my results on CD4 support this older date of initial infection by SIV of the *Pan* and *Gorilla* lineages.

CD4 is conserved in individuals of western chimpanzees but variable in central and eastern chimpanzees (Hvilsom *et al*., 2008). The CD4 receptor of chimpanzees (Hvilsom *et al.*, 2008), as well as other apes (as shown herein), differs genetically from that of humans, and it is possible that these differences may play roles in SIV resistance in the western and Cameroonian chimpanzee and may explain why humans have less resistance to HIV than do other primates.

The two major species of gorillas are the western (*Gorilla gorilla gorilla*) and eastern (*Gorilla beringei graueri*) gorillas. Certain populations of western lowland gorillas have SIV in the wild, but the virus has yet to be found in other populations of gorillas (Neel *et al.*, 2010), although only a few populations of eastern gorillas have been examined to date. Phylogenetic analysis of SIV genes suggests that western gorillas acquired the virus from central chimpanzees (*P.t.t.*) living in the same region in Africa (reviewed in Sharp & Hahn, 2011; Neel *et al.* 2010). This would imply that the strain of SIV currently infecting gorillas is less than 300,000 years old, based upon the same logic used above. However, our analyses suggest that the common ancestor of the two gorilla species was infected by SIV (or some other pathogen that uses CD4 as a receptor), which will later be discussed.

There are two species of orangutans, the Sumatran (*Pongo abelii*) and Bornean (*P. pygmaeus*). Both of these species appear to lack SIV endemically, like other Asian primates

Figure 3. The Great Apes. This image shows a map of equatorial Africa on the upper left, and a map of Malaysia and Indonesia in the box on the bottom right. These maps show the current known locations of the great apes: the Western chimpanzee, Nigeria-Cameroon chimpanzee, Central chimpanzee, Eastern chimpanzee, Western lowland gorilla, Eastern lowland gorilla, bonobo, Sumatran orangutan, and Bornean orangutan. The plus signs (+) denote a species or subspecies in which SIV is endemic. The minus signs (-) denoted the lack of endemic SIV in the species.

(Compton *et al*., 2013; Locatelli *et al*., 2014). Thus, the orangutan CD4 sequences should serve as a natural control for the African ape CD4 sequences in our studies.

There have been suggestions that the common ancestral lineage leading to bonobos and chimpanzees shows signs of genetic adaptation to an SIV-like retrovirus (Walter *et al*., 2005; reviewed in de Groot & Bontrop, 2013). Extensive studies on captive *P.t. verus* have shown this species to be resistant to infection by HIV or SIV, and if infected, show resistance to progression to AIDS (Gordon *et al.*, 2008). Moreover, extensive surveys of wild *P.t. verus* have not revealed any natural infections by chimpanzee or local monkey SIVs (Reviewed in Sharp *et al.*, 2005; Santiago *et al.*, 2003). Together, these findings suggest that the western chimpanzees likely have natural resistance to SIV infection, but the genetic basis for this resistance is not known. In contrast, the central and eastern chimpanzees are naturally infected with SIV, and eastern chimpanzee populations show evidence of significant impact by this pathogen (Locatelli *et al.* 2014). In contrast to the chimpanzees, experiments have not been reported on captive gorillas, so little is known about their resistance or susceptibility to SIV and AIDS (Locatelli *et al.*, 2014).

The pattern seen in chimpanzees and gorillas is in stark contrast to that seen in humans and orangutans. Being of Asian origin, orangutans are not naturally exposed to SIV, and no SIV has been reported in wild populations. The current consensus appears to be that SIV is restricted to African primates, although a few studies of Asian primate SIV status have been reported in the literature. Macaques, which are Asian monkeys, have been confirmed to not harbor SIV as a natural pathogen. (Sharp & Hahn, 2011). Why humans — which are originally of African origin — are more susceptible to HIV than are African apes is currently unknown.

Focus of this Study

The great apes and humans share much of their genome, especially chimpanzees, humans' closest relatives, who share 98% of their genome with us. The Stewart laboratory suspected than somewhere along the genome, there was enough variation between apes and humans to make a difference in HIV/SIV susceptibility between species. My role in the SIV project was to study the CD4 mRNA gene sequences of the hominoids, and to understand the molecular basis for the potential adaptive evolution of CD4 to SIV. Previous research in the

Stewart laboratory has suggested adaptive evolution of the chimpanzee and gorilla CD4 genes, as well as other adhesion proteins involved in HIV infection and disease progression (Bandla, 2009). However, these studies were limited because sequences were available from only western chimpanzee and western gorilla. That is, we had no information about bonobos, the other subspecies of chimpanzees, or gorillas. Thus, we were not able to infer when the adaptation occurred on the African ape lineages. Furthermore, sequences were not available from multiple individuals within species, and thus information was not available about ongoing adaptation in populations.

The publication in 2013 of whole genome sequences from 79 great ape individuals representing all known subspecies (Prado-Martinez *et al.*, 2013) provided these missing data, and allowed us to study the evolution of CD4 in the great apes at both the phylogenetic and population genetic levels. We obtained CD4 sequences for human, Neanderthal, Denisovan, and the great apes through a variety of public databases, including: GenBank (http://www.ncbi.nlm.nih.gov), the 1000 Genomes Project (http://www.1000genomes.org), Ensembl (http://neandertal.ensemblgenomes.org/index.html), and the UCSC Genome Browser (http://genome.ucsc.edu/). We assembled the ape CD4 gene sequences from short sequence reads, and inferred the haplotypes for each. Analysis of these new data revealed that there were two ancestral episodes of adaptation in D1 of CD4, one on the ancestral *Pan* lineage and the other on the ancestral *Gorilla* lineage. No such episodes were seen on the ancestral *Pongo* or

Homo lineages. Moreover, we found evidence of ongoing selection on D1 of CD4 in populations of gorillas and chimpanzees, but not in humans or orangutans.

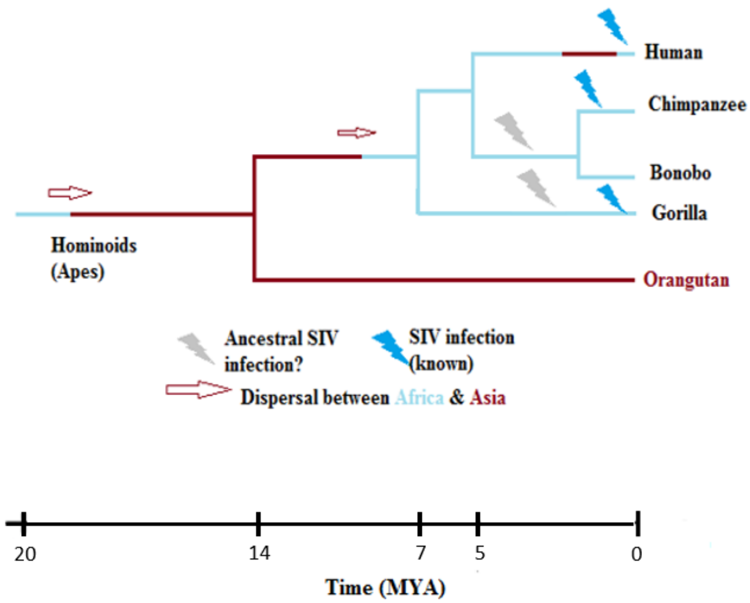


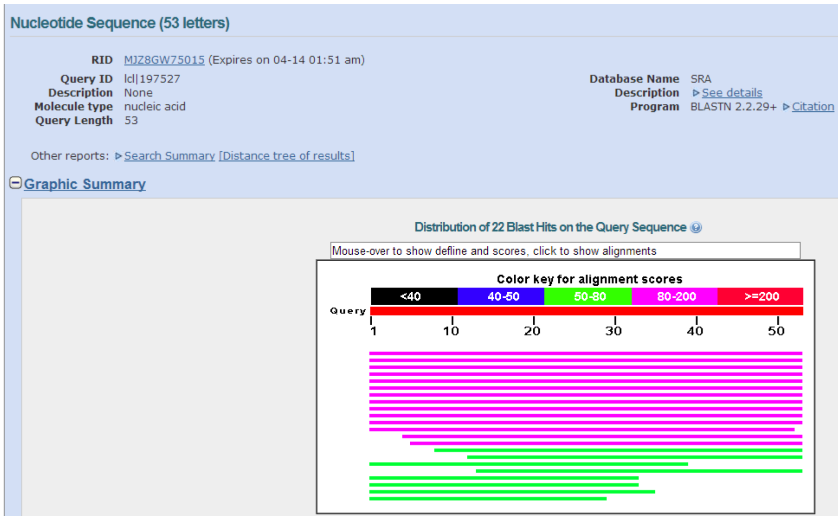
Figure 4. Hominoid Phylogeny. A phylogenetic tree depicting SIV infection in hominoids. The red lineages represent Asian primates, whereas the blue lineages represent African primates. The blue bolt depicts known SIV infection, while the gray bolt depicts hypothesized SIV infection. Phylogeny based on Prado-Martinez *et al*. (2013).

MATERIALS AND METHODS

Primary Analysis

The short sequence reads of the following great ape genomes were made available by Prado-Martinez *et al.* (2013) in the Sequence Read Archive (SRA) database on the National Center for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/sra) for the following species: *Pan troglodytes verus* (Western chimpanzee), *Pan troglodytes ellioti* (Nigerian-Cameroonian chimpanzee), *Pan troglodytes troglodytes* (Central chimpanzee), *Pan troglodytes schweinfurthii* (Eastern chimpanzee), *Pan paniscus* (bonobo), *Gorilla gorilla gorilla* (Western lowland gorilla), *Gorilla beringei graueri* (Eastern lowland gorilla), *Pongo abelii* (Sumatran orangutan), and *Pongo pygmaeus* (Bornean orangutan). A list of all of the individuals used in this study, along with their SRA Experiment Set (SRX) file numbers, is given in Appendix A. For each individual, all of their SRX files were jointly searched for CD4 homologies using the Basic Local Alignment Tool (BLAST) available on the SRA site. The RefSeq mRNA sequence for the given species was used as the query in most searches. For searches that spanned intron-exon boundaries, the human genome sequence was used as the query. The maximum number of aligned sequences selected to display for these BLAST searches was 20,000, and the program was selected to optimize for highly similar sequences to avoid turning up gene paralogs and artifacts. The alignment files returned for each individual were saved as PDF files and archived in the Stewart laboratory. These files were visually examined for polymorphic sites, and annotated accordingly. A site was considered to be polymorphic if a nucleotide sequence difference existed in about half of the short reads, provided that the reads in question were 'clean' and the apparent polymorphism did not show hallmarks of being an artifact of the sequencing method (see below for further discussion). In those cases when more than one polymorphism existed within a given short sequences read, this information was used to help phase the haplotypes in that individual. The full haplotypes of the CD4 coding regions were inferred using the program BEAGLE by Dr. Randall Collura. The *Homo sapiens* CD4 mRNA sequences analyzed were made available through the 1000 Genomes Project (http://www.1000genomes.org), while the Neanderthal and Denisovan sequences were provided by Ensembl (http://neandertal.ensemblgenomes.org/index.html) and the UCSC Genome Browser (http://genome.ucsc.edu/), respectively.

Sequence variation was scored conservatively, and further analyzed in a master alignment of the sequences of all available individuals across species. CD4 mRNA sequences from 27 Western lowland gorillas, 3 Eastern lowland gorillas, 4 Western chimpanzees, 10 Cameroonian chimpanzees, 6 Eastern chimpanzees, 4 Central chimpanzees, 13 bonobos, 5 Sumatran orangutans, 5 Bornean orangutans, 1 Denisovan, 5 Neanderthals, and over 1000 humans were analyzed. A master alignment of CD4 genes was created from the all individuals' allelic sequences, which had been organized into FASTA files. These FASTA files were imported into *Se-AL* v2.a11 (Rambaut, 2002), a program that can be used to manually align multiple DNA sequences according to codons, and can translate the DNA sequences into protein sequences. The alignments were then imported into the program *MacClade* v. 4.04 (Maddison & Maddison, 2004), which was then used to most parsimoniously assign nucleotide substitutions and amino acid replacements upon the lineages of the CD4 tree (See Figure 9), using the known primate phylogeny as a guide. The gibbon CD4 sequence was used as an outgroup to aid in the assignment of changes to the various great ape lineages.




```
Query                  1    GAATGATCGCGTTGACTCAAGAAGAAGCCTTTGGGACCAAGGAAACTTTACCC  53
SRR726407.42553904.1   27   ...A.................................................  79
SRR726407.9373181.1    62   ...A.................................................  10
SRR726406.36395089.2   65   ...A.................................................  13
SRR726405.89739263.2   80   .................................................C...  28
SRR726405.84825027.1   10   ...A.................................................  62
SRR726405.78128420.2   12   ...A.................................................  64
SRR726405.34364036.1   86   ...A.................................................  34
SRR726405.20073864.1   12   ...A.................................................  64
SRR726405.15832381.1   59   ...A.................................................  7
SRR726404.48718037.2   66   ...A.................................................  14
SRR726404.304481.2     80   .................................................C...  28
SRR726405.27112659.2   52   ...A................................................   1
SRR726404.47151050.2   97       .................................................  49
SRR726407.73840298.2   100       ............................................C...  53
SRR726404.65326948.1   1           .........................................C...  45
SRR726406.100103405.1  1                .........................................  41
SRR726404.17877564.1   39   .......................................                1
SRR726404.93990999.2   1                 ....................................C...  40
SRR726407.66113829.1   33   .................................                      1
SRR726407.66113825.1   33   .................................                      1
SRR726405.10961017.2   35   ...A...............................                    1
SRR726404.21055738.2   29   .............................                          1
```

Figure 5. BLAST return example. Shown above is a clean example of a short span of a chimpanzee CD4 D1 gene sequence from a BLAST search. Two polymorphisms can be seen in this image: T/A and A/C. In this case, the two polymorphisms are on the same runs, so the haplotypes can be determined as “A…A” and “T….C”. Note that the polymorphic sites are found on reads in both orientations, a further sign that they are real sequence differences.

Hidden Alleles

During the analysis of the ape CD4 sequences, a notable concern arose. Some of the BLAST returns lacked sequences at the ends of exons, near the exon-intron boundaries, in a manner that looked like the runs had been truncated. A considerable number of sequence runs had this problem. We suspected that there was a possibility that polymorphisms present in these truncated runs were not being revealed due to the sequence reads being cut off due to mismatches near the ends. To see if this was the case, I redid some of the BLAST searches using query sequences that spanned intron-exon boundaries by about 100 basepairs. This analysis revealed that, indeed, there were several places were apparently real polymorphisms had been hidden by the mRNA-based queries. Thus, I redid all of the BLAST searches including about 100 base pairs of intronic sequence flanking both upstream and downstream from each exon, except for the first and last exons, which had intronic sequence flanking the downstream, and upstream sides of the exons, respectively. This second batch of hominoid CD4 BLAST searches was done using a *Homo sapiens* CD4 DNA query instead of the respective species' mRNA sequences. This analysis revealed that there had been some missing polymorphisms within the leader sequence of the gene and in domains 3 and 4 in certain individuals, but these did not change the major findings about fixed changes and polymorphism in D1 and D2 of the SIV-endemic species. Since the BLAST search setting was set for 'highly similar sequences', the deviating ends of runs at the intron-exon boundary were simply not entirely included in the return. This potential problem is certainly an issue to keep in mind when doing such bioinformatics analyses.

Heterozygosity

Prado-Martinez *et al*. (2013) measured genome-wide heterozygosity of the great ape species and subspecies. To compare the heterozygosities of CD4, particularly D1, between groups and to the genome averages, I measured the heterozygosity of the CD4 gene and D1of each individual. Prado-Martinez and colleagues did not indicate how they calculated genomic heterozygosities, but the values were reported as heterozygotes/bp. I therefore conducted the

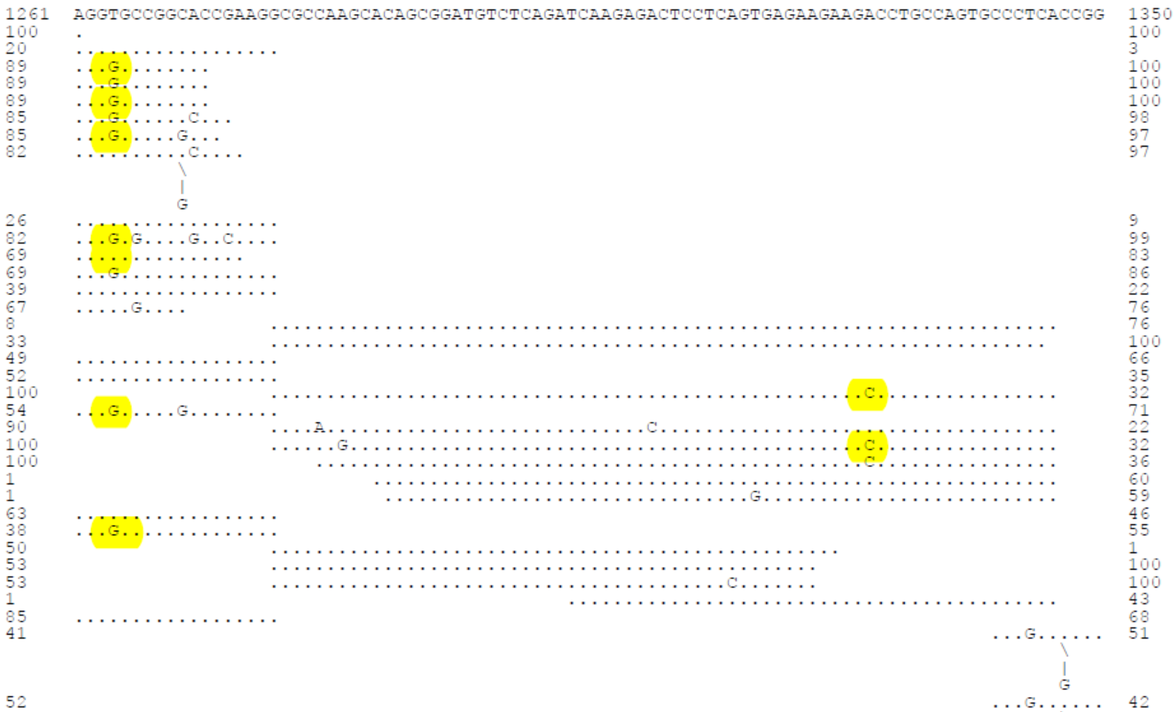
CD4 measurements by counting the number of polymorphisms in each individual, and then dividing that number by the length of the gene (1377 bp). I thereby calculated the average heterozygosity for each species or subspecies. These values can be found in Appendix C.

RESULTS

Sequence Artifacts

A noteworthy concern that arose was there appeared to be many artifacts riddled throughout the sequence reads in the alignments from the BLAST searches. Some of these were the common G artifacts found in this type of genomic sequencing, which is illustrated in Figure 6 for PTT Doris. Additional problems we encountered were sites that looked like potential polymorphisms, but under closer evaluation seemed likely to be artifacts. Two such examples are also shown in Figure 6. The string of G's and C's depicted in both chimpanzee individuals shown, in nucleotide positions 1264 and 1330, were present in almost all individuals for chimpanzees, bonobos and gorillas. In addition, all of the runs they were located on were in the reverse orientation, and many of the individual runs with these G's and C's on them were also riddled with other 'dirt' (additional nucleotide differences). Lastly, in most of the apes, these nucleotide differences were present in low frequency. For example, PTV Jimmie, as shown in Figure 6, is the only chimpanzee that had such a surplus amount of G's in position 1264. Thus, based on this reasoning, these sites were no longer considered polymorphic in any species, whether or not there were numerous runs with the presumed artifactual nucleotide. This kind of careful scrutiny and visual analysis was applied to all other apparently polymorphic sites to ensure the utmost accuracy in base calls. When in question, we tried to be conservative in calling polymorphisms or substitutions.

***P.t.v.* Jimmie**



***P.t.t.* Doris**

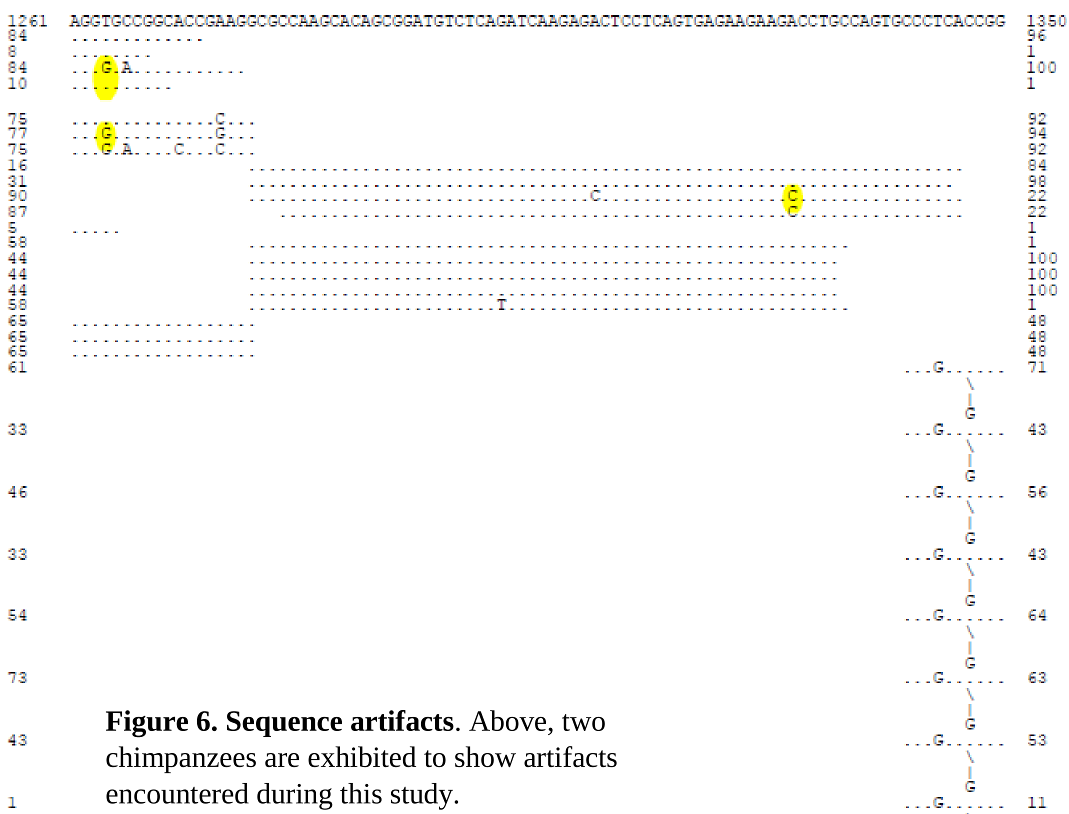


Figure 6. Sequence artifacts. Above, two chimpanzees are exhibited to show artifacts encountered during this study.

Hidden Alleles

An additional concern arose when BLAST searching the primates sequences. Some of the returned sequence runs appeared to be cut off near the end of certain exons. I was not sure why this was so, and was concerned that fixed changes or polymorphism could be hiding within the sequence that was not being shown. To remedy this issue, I re-BLASTED all the apes with a *Homo sapiens* CD4 DNA FASTA sequence, in which I patched together the exons with 100 bp of intron flanking either side. In doing so, some regions of CD4 did reveal polymorphic sites, such as the leader sequence, which contained an A/G nucleotide polymorphism within the first hundred base pairs. These polymorphic sites were of no importance in our major results, and this site in question caused a synonymous substitution. An individual who had this hidden allele was *Pan troglodytes schweinfurthii* Bwambale. This site appeared to be an A against a *Pan troglodytes* mRNA query, but upon re-BLASTing it with a *Homo* mRNA sequence, which contained a G in this position, it appeared to be a G. Including the introns in the BLAST search revealed the polymorphic site to contain both A's and G's (see Figure 7).

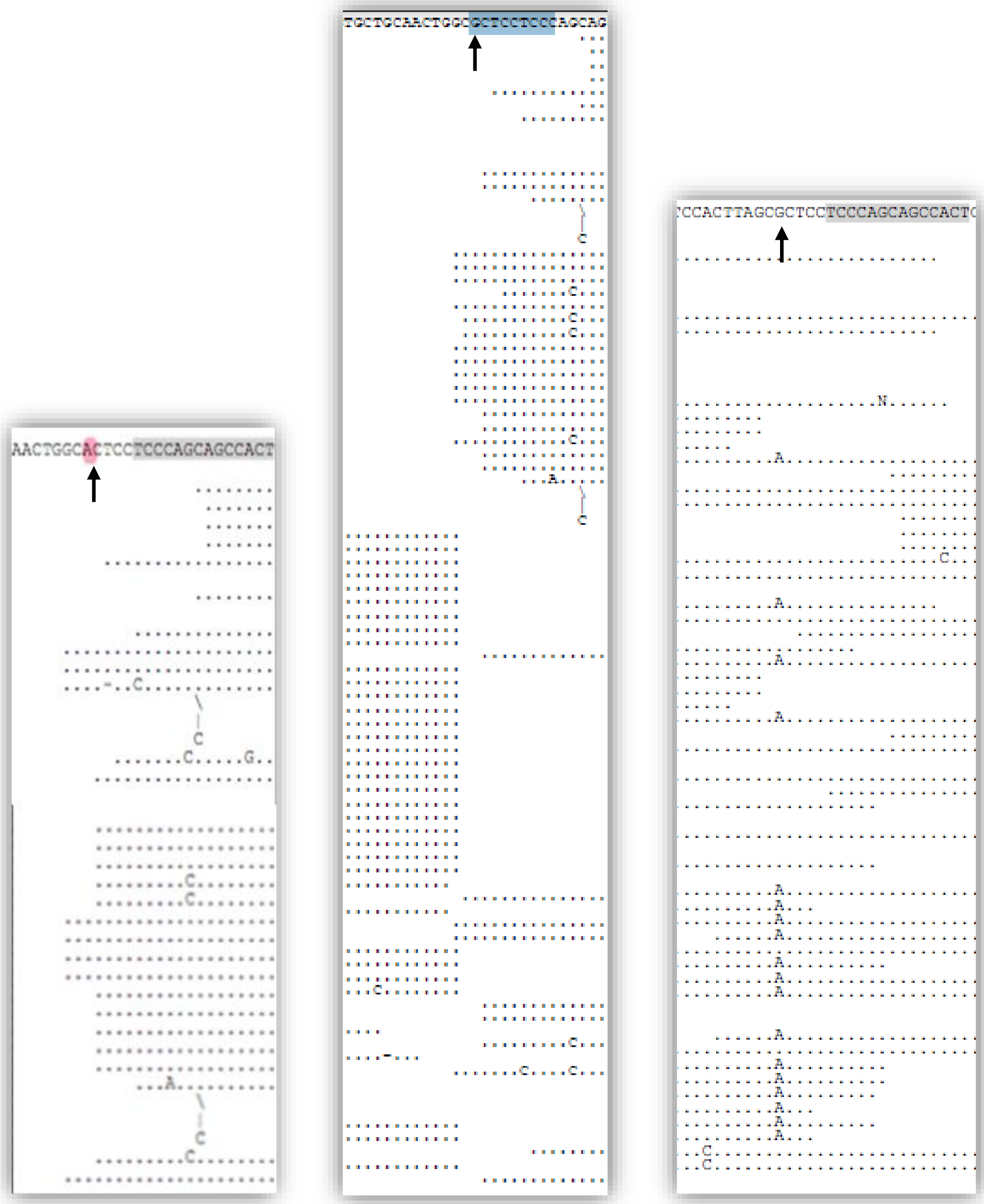


Figure 7. Hidden Alleles. From left to right, the data sequences show BLAST returns of PTS Bwambale against a *Pan troglodytes* mRNA query, a *Homo sapien* mRNA query, and a *Homo* DNA query. The last BLAST run reveals the polymorphic site (denoted by the arrow), and shows the most coverage.

Heterozygosity

Upon measuring the heterozygosity, as described in the Materials and Methods section, I found that the heterozygosity of D1 was higher than that of the rest of CD4, and that the African apes who have SIV in the wild (Western lowland gorillas and central chimpanzees) had higher levels of heterozygosity in D1 than in the rest of their genome (Prado-Martinez *et al.*, 2013). The mutations that occurred on these lineages do not seem to be affecting the entire CD4 gene on average, but are instead targeting D1. This suggests ongoing adaptation in these African apes to the SIV virus. For calculated heterozygosity values for CD4, please see Appendix C. For genome averages, please see Figure 1b in Prado-Martinez *et al.* (2013).

Lineage Based Analysis

The hominoid CD4 mRNA sequences that were analyzed had much to reveal. Remarkably, the *Pan* and *Gorilla* ancestral lineages each had several fixed amino acid replacements, most of which reside in D1. Remarkably, similar amino acid replacements happened independently on the ancestral *Pan* and *Gorilla* lineages: both had replacements that created N-glycosylation sites in D1 near the gp120 binding surface, and both had charge changes at the gp120-CD4 interacting face. In addition, chimpanzee and gorilla populations have highly polymorphic CD4 proteins, again with most of the polymorphisms in D1. These changes likely have the potential to prevent gp120 from binding to CD4. This analysis revealed what appears to be parallel adaptation in the chimpanzees and the gorillas in CD4, presumably due to SIV infections over prolonged periods of evolutionary time.

There were notable amino acid changes in D1 among the chimpanzees, bonobos, and gorillas. The chimpanzees and gorillas also contained significant amount of polymorphism in D1 compared to the rest of CD4. In the western lowland gorillas, amino acid site 40 consisted of a fixed change from a Threonine (T) to an Asparagine (N), resulting in the formation of an N-glycosylation site (Asparagine-Cysteine-Threonine). Site 113 in these gorillas contained a change from Aspartic Acid (D) to Glycine (G), resulting in a charge change. The chimpanzees and bonobos shared fixed changes that were just as notable. Amino acid site 59 contained an Isoleucine (I) to Threonine (T) change resulting in an N-glycosylation site (Asparagine-Glutamine-Threonine). Site 93, a polymorphic site in some chimpanzees, forms an N-glycosylation site that is created from the change of a Proline (P) to Threonine (T) (Asparagine-Phenylalanine-Threonine). Site 112 revealed a D to G change, resulting in an alteration of charge. Such structure and binding altering fixed mutations were not found on the *Homo* and Asian lineages. The detailed visualization of all fixed changes and polymorphisms discovered can be found in the sequences shown in Appendix B. Much of the polymorphism and change was located on the CD4-gp120 interface. Figure 8 shows CD4 models, created by Jennifer Giza, of some individual gorillas and chimpanzees that contain polymorphism, as well as fixed amino acid changes in D1 of CD4.

In contrast to the African apes, the ancestral *Pongo* lineage had many fixed changes, but these changes were spread over all domains of CD4 instead of being concentrated in D1. Similarly, there were virtually fixed differences between the two orangutan species, and little polymorphism within the two species. Importantly for understanding HIV infection in humans, there were no inferred amino acid replacements on the ancestral *Homo* lineage, and no significant amount of polymorphism within *Homo* species. That is, in stark contrast to the African apes, the human CD4 protein shows no evidence of adaptation to ongoing immunodeficiency virus infection. This contrast is likely highly attributed the very recent of transmission of SIV/HIV to humans.

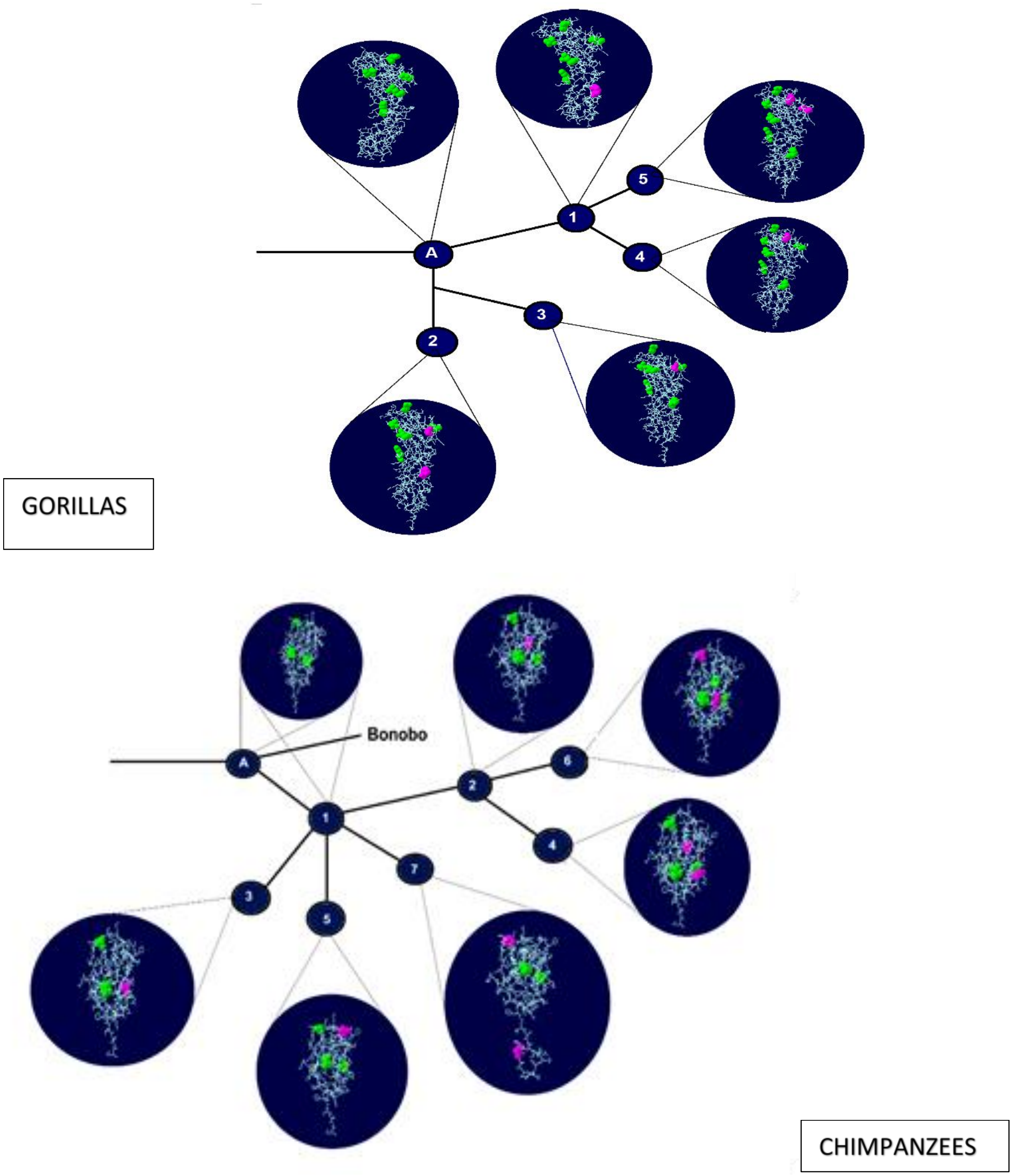


Figure 8. Chimpanzee and Gorilla fixed changes and polymorphisms. These webs depict D1 of CD4 in individuals from the *Gorilla* and *Pan* lineages, respectively. The green represents fixed amino acid replacements while the pink represents polymorphism. Much of the adaptation is right at the CD4-gp120 interface. These images were used with the permission of Jennifer Giza, who created these models using Swiss Pdb.

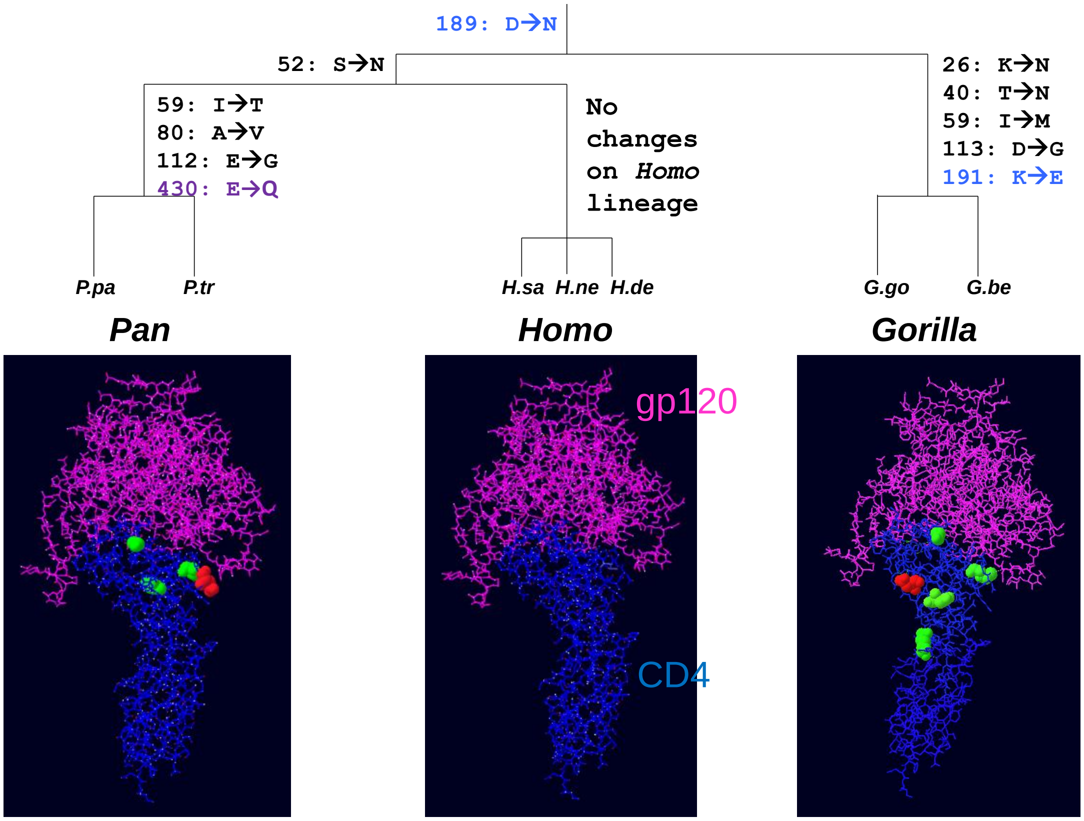


Figure 9. Amino acid substitution tree. This figure shows the most parsimonious assignment of the amino acid replacements on the *Pan*, *Homo*, and *Gorilla* lineages. *P.pa* is bonobo, *P.tr* is *Pan troglodytes*, *H.sa* is *Homo sapiens*, *H.ne* is *Homo neanderthalensis*, *H.de* is *Homo denisovan*, *G.go* is *Gorilla gorilla*, and *G.be* is *Gorilla beringei*. The numbers represent amino acid change sites; black text indicates an amino acid change in D1, blue text denotes a change in D2, and purple text denotes a change beyond the first two domains. Note that sites 59 and 40 represent addition of N-glycosylation sites in the *Pan* and *Gorilla* lineages, respectively, while sites 112 and 113 are charge change sites (from negative to neutral) in these lineages, respectively. No changes were found on the *Homo* lineage. Structures of CD4 (blue) bound to gp120 (pink) and their corresponding changes are shown under each lineage.

CONCLUSIONS

The additions of the N-glycosylation sites near the CD4-gp120 interface means that a large complex carbohydrate structure will attach to the asparagine side chain. This attachment would likely be an obstruction that would block gp120 from being able to bind properly to CD4. The fixed charge changes shared by the *Pan* and *Gorilla* lineages are also likely to block gp120 form being able to bind properly to CD4 due to incompatibility. In addition to this study, others have found notable sequence changes in the chimpanzees, as well. Hvilsom and colleagues (Hvilsom *et al.*, 2008) stated that most of the amino acid differences between human and chimpanzee CD4 sequences were located in domain 1, which holds the most interaction with the SIV/HIV virus. Meyerson *et al*. (2014) found that CD4, among three other HIV host factor genes (NUP153, RANBP2/NUP358), showed strong signatures of positive selection in D1 using a sliding window dN/dS analysis across CD4. However, these authors did not perform phylogenetic analysis of the sequences changes in CD4 in the hominoids, and thus did not discover the parallel adaptive episodes and ongoing selection that we report here.

In addition to the fixed changes on the *Pan* and *Gorilla* lineages, within the gorillas and chimpanzees, but not the bonobos, there were numerous polymorphisms in D1 of CD4, suggesting ongoing adaptation against SIV in certain populations of African apes. In contrast, the *Homo* lineage shows no amino acid replacements in CD4, and very few polymorphisms are found within modern humans. Again, this suggests that the ancestral human lineage was likely not infected by SIV, at least not on an ongoing basis like for chimpanzees and gorillas. The lack of exposure of ancestral humans to SIV would have prevented adaptive co-evolution, so that human lack HIV resistance today.

However, it is possible that humans are currently undergoing selection to HIV. One possible form of adaptation is the existence of different isoforms of CD4 in humans and possibly in other primates. We found evidence in various mRNA databases that at least one of the predicted transcript variants for CD4 is likely expressed in humans, gorillas, and macaques, and is perhaps found in New World monkeys, as well. This transcript variant codes for a short version of CD4 that lacks D1 and D2, and thus would not bind gp120. The details of the isoforms have not been looked at in depth by the Stewart lab, nor are these isoforms mentioned in the scientific literature. These variants imply that there are currently uncharacterized 'short' forms of the CD4 protein made in humans (and other primates) that may exhibit unique functions in the immune system.

The higher levels of heterozygosity that were seen in D1 of some species were expected based on all the existing variability in D1 among species. Based upon the parallel amino acid replacements seen in D1 of CD4 on the ancestral Pan and Gorilla lineages, we are led to hypothesize that the ancestors of the chimpanzees and bonobos, as well as the ancestors of the two gorilla species, were infected by SIV. Based upon the phylogeny of the hominoids, this would mean that they were infected for initially over 1-2 million years ago, meaning that SIV is much older in apes than most researchers appear to believe at this time. This long period of co-evolution of apes and SIV has allowed the hosts to adapt genetically to the virus. Humans, in contrast, do not appear to have had this long period of co-evolution with the virus, and are not genetically adapted to it. We believe that the chimpanzee and gorillas ancestors stayed in the region of the retro-viral infections in the rainforests of Africa, while the human ancestors migrated away from the area of exposure to SIV. It is also possible that those hominins who lived in Africa at the time died from the virus. This hypothesis might explain why African apes have built up a degree of resistance to the SIV virus, while humans and Asian apes have little to no resistance to HIV, and progress to AIDS. Our findings may help explain why the African apes are more resistant to SIV/HIV infection than are humans and Asian primates, as well as help identify regions of CD4 that might be safely targeted by drugs or antibodies.

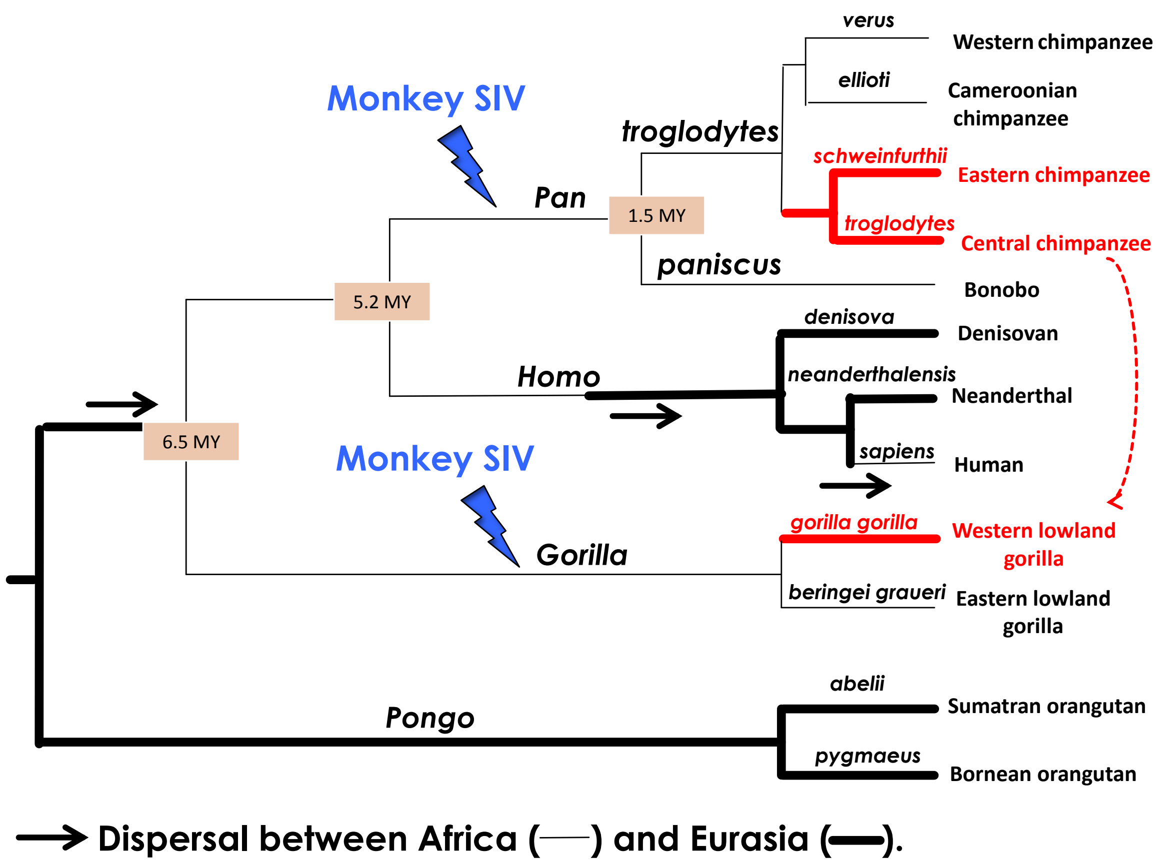


Figure 10. Hominoid dispersal and estimated infection. This figure shows the phylogenetic relationship between chimpanzee, bonobo, human, and orangutan lineages. The red branches represent the apes that harbor SIV in the wild, while the red downward arrow shows that the Central chimpanzee SIV jumped to the Western lowland gorillas. We hypothesize that monkey SIV spread to the *Pan* and *Gorilla* lineages over 2 million years ago. The divergence dates for the ape lineage was adapted from Prado-Martinez *et al.* (2013).

LITERATURE CITED

Bandla, S. 2009. Molecular evolution of genes underlying phenotypic differences between humans and chimpanzees. Ph.D dissertation. State University of New York at Albany.

Bianco-Melo, D., Venkatesh, S., and P.D. Bieniasz. 2012. Intrinsic cellular defenses against human immunodeficiency viruses. *Immun*ity 37(3): 399-411.

Castro-Nallar, E., Pérez-Losada, M., Burton, G.F., and K.A. Crandall. 2012. The evolution of HIV: inferences using phylogenetics. *Molecular phylogenetics and evolution* 62(2): 777-92.

Compton, A.A. and M. Emerman. 2013. Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. *PLoS Pathogens* 9(1).

Compton, A.A., Malik, H.S., and M. Emerman. 2013. Host gene traces the evolutionary history of ancient primate lentiviruses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1626).

Dooren, S.V., Switzer, W.M., Heneine, W., Goubau, P., Verschoor, E., Parekh, B., de Meurichy, W., Furley, C., Ranst M.V., and A. Vandamme. 2002. Lack of evidence for infection with simian immunodeficiency virus in bonobos. *Aids Research and Human Retroviruses* 18(3): 213-6.

Freeman, S. and J.C. Herron. 2007. *Evolutionary Analysis: 4th edition*. Pearson Prentice Hall, Upper Saddle River, NJ.

Gifford, R.J., Katzourakis, A., Tristem, M., Pybus, O.G., Winters, M., and R.W. Shafer. 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proceedings of the National Academy of Sciences* 105(51): 20362-7.

Gordon, S., Pandrea, I., Dunham, R., Apetvel, C., and G. Silvestri. (2005) The call of the wild: what can be learned from studies of SIV infection of natural hosts? *Theoretical Biology and Biophysics Group, Los Almos National Laboratory*.

de Groot, N.G. and R.E. Bontrop. 2013. The HIV-1 pandemic: does the selective sweep in chimpanzees mirror humankind's future? *Retrovirology* 10:53.

Guex, N., Diemand, A., Peitsch, M.C., and T. Schwede. 2012. Swiss-PdbViewer DeepView v4.1

Harrison, S.C. 1993. CD4: Structure and interactions of an immunoglobulin superfamily adhesion molecule. *The American Chemical Society.* 26(8): 449–53.

Hvilsom, C., Carlsen, F., Siegismund, H.R., Corbet, S., Nerrienet, E., and A. Fomsgaard. 2008. Genetic subspecies diversity of the chimpanzee CD virus-receptor gene. *Genomics* 92(5): 322-8.

Koito, A., Harrowe, G., Levy, J.A., and C. Cheng-Mayer. 1994. Functional role of the V1/V2 region of human immunodeficiency virus type 1 envelope glycoprotein gp120 in infection of primary macrophages and soluble CD4 neutralization. *Journal of Virology* 68(4): 2253-9.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and W.A. Hendrickson. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393(6686): 648-59.

Li, Y., Ndjango, J.B., Learn, G.H., Ramirez, M.A., Keele, B.F., Bibollet-Ruche, F., Liu, W., Easlick, J.L., Decker, J.M., Rudicell, R.S., *et al*. 2012. Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. *Journal of Virology* 86(19): 10776-91.

Locatelli, S., McKean, K.A., Sesink-Clee, P.R., and M.K. Gonder. 2014. The evolution of resistance to simian immunodeficiency virus (SIV): a review. *International Journal of Primatology* 35: 349-75.

Maddison, W.P. and D.R. Maddison. 2004. *MacClade V4.04*.

Meyerson, N.R., Rowley, P.A., Swan, C.H., Le D.T., Wilkerson, G.K., and S.L. Sawyer. 2014. Positive selection of primate genes that promote HIV-1 replication. *Journal of Virology* 454-455:291-8.

Neel, C., Etienne, L., Li, Y., Takehisa, J., Rudicell, R.S., Bass, I.N., Moudindo, J., Mebenga, A., Esteban, A., Van Heuverswyn, F., *et al*. 2010. Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. *Journal of Virology* 84(3):1464-76.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. 2013. Great ape genetic diversity and population history. *Nature* 499: 471-5.

Rambaut, A. 2002. *Se-Al v2.0a11 Carbon*. Oxford University.

Santiago, M.L., Lukasik, M., Kamenya, S., Li, Y., Bibollet-Ruche, F., Bailes, E., Muller, M.N., Emery, M., Goldenberg, D.A., Lwanga, J.S., *et al.* 2003. Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglogytes schweinfurthii*). *Journal of Virology* 77(13): 7545-62.

Sharp, P.M. and B.H. Hahn. 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine* 1(1).

Sharp, P.M., Shaw, G.M., and B.H. Hahn. 2005. Simian immunodeficiency virus infection of chimpanzees. *Journal of Virology* 79(7): 3891-902

Walter, N.A., Stebbing, J., and W. Messier. 2005. The potential significance of adaptive evolution and dimerization in chimpanzee intercellular cell adhesion molecules (ICAMS). *Journal of Theoretical Biology* 232(3): 339-46.

Wertheim, J.O., and M. Worobey. 2009. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Computational Biology* 5(5).

Worobey, M., Telfer, P., Souquiere, S., Hunter, M., Coleman, C.A., Metzger, M.J., Reed, P., Makuwa, M., Hearn, G., Honarvar, S., *et al*. 2010. Island biogeography reveals the deep history of SIV. *Science* 329: 1487

Wyatt, R., Moore, J., Accola, M., Desjardin, E., Robinson, J., and J. Sodroski. 1995. Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *Journal of Virology* 69(9): 5723-33.

APPENDIX A

All data displayed and that was used in this appendix was made publicly available by Prado-Martinez et al. (2013).

Accession Number	Name	Species
SRX243449	Bosco	*Pan troglodytes verus*
SRX243448	Bosco	*Pan troglodytes verus*
SRX243527	Clint	*Pan troglodytes verus*
SRX243488	Jimmie	*Pan troglodytes verus*
SRX243487	Jimmie	*Pan troglodytes verus*
SRX243499	Koby	*Pan troglodytes verus*
SRX243510	Akwaya-Jean	*Pan troglodytes ellioti*
SRX243511	Banyo	*Pan troglodytes ellioti*
SRX243512	Basho	*Pan troglodytes ellioti*
SRX243513	Damian	*Pan troglodytes ellioti*
SRX360476	Damian	*Pan troglodytes ellioti*
SRX243514	Julie_LWC21	*Pan troglodytes ellioti*
SRX360477	Julie_LWC21	*Pan troglodytes ellioti*
SRX243515	Kopongo	*Pan troglodytes ellioti*
SRX243516	Koto	*Pan troglodytes ellioti*
SRX360478	Koto	*Pan troglodytes ellioti*
SRX243517	Paquita	*Pan troglodytes ellioti*
SRX243518	Taweh	*Pan troglodytes ellioti*
SRX360479	Taweh	*Pan troglodytes ellioti*
SRX243519	Tobi	*Pan troglodytes ellioti*
SRX243496	Clara	*Pan troglodytes troglodytes*
SRX243495	Clara	*Pan troglodytes troglodytes*
SRX243492	Doris	*Pan troglodytes troglodytes*
SRX243491	Doris	*Pan troglodytes troglodytes*
SRX243494	Julie_A959	*Pan troglodytes troglodytes*
SRX243493	Julie_A959	*Pan troglodytes troglodytes*
SRX243490	Vaillant	*Pan troglodytes troglodytes*
SRX243489	Vaillant	*Pan troglodytes troglodytes*
SRX237492	Andromeda	*Pan troglodytes schweinfurthii*
SRX237526	Bwambale	*Pan troglodytes schweinfurthii*
SRX237524	Bwambale	*Pan troglodytes schweinfurthii*
SRX243451	Harriet	*Pan troglodytes schweinfurthii*
SRX243450	Harriet	*Pan troglodytes schweinfurthii*
SRX237539	Kidongo	*Pan troglodytes schweinfurthii*
SRX237527	Kidongo	*Pan troglodytes schweinfurthii*
SRX237583	Nakuu	*Pan troglodytes schweinfurthii*
SRX237541	Nakuu	*Pan troglodytes schweinfurthii*
SRX237455	Vincent	*Pan troglodytes schweinfurthii*

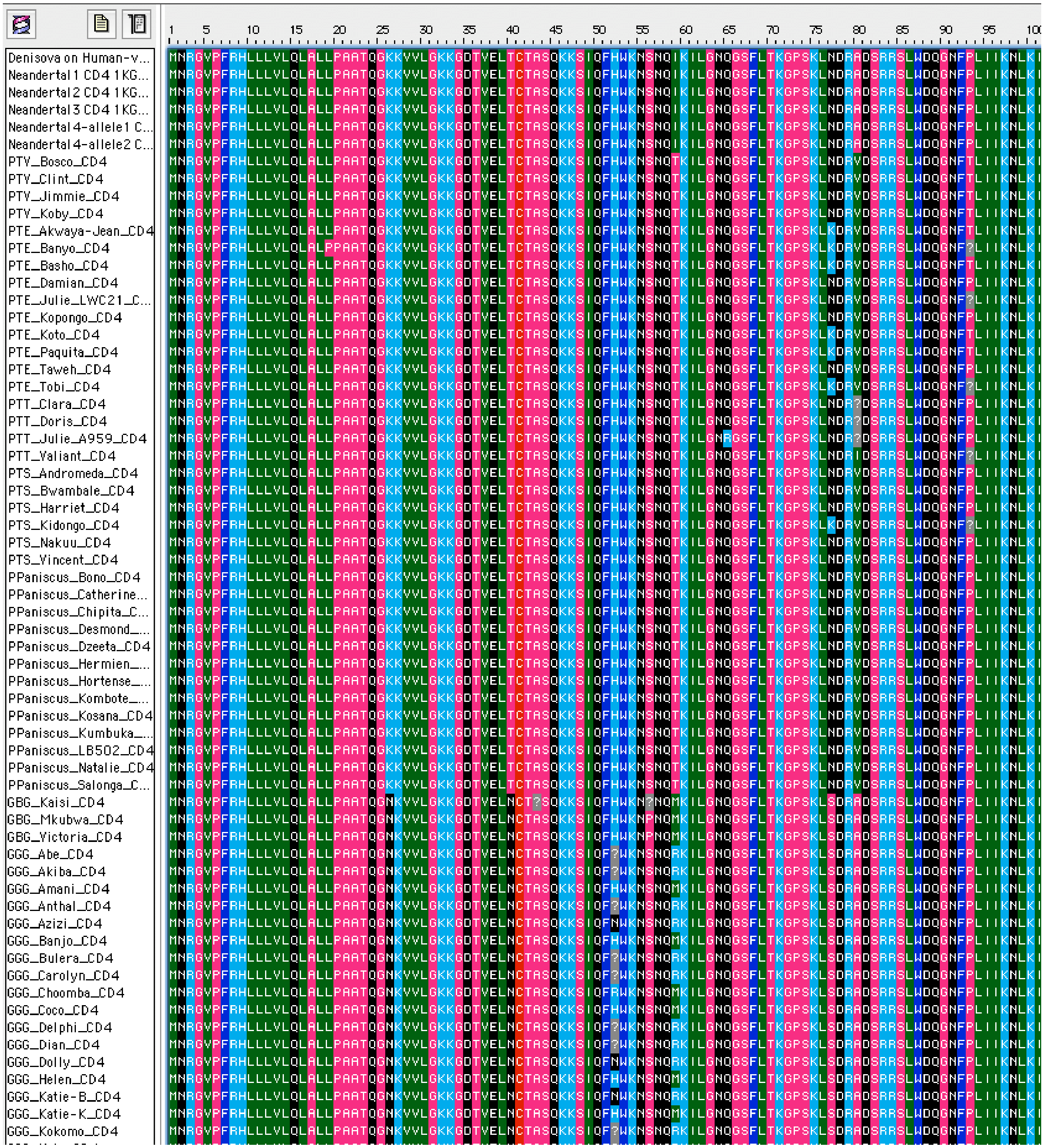
Accession Number	Name	Species
SRX243484	Babu	*Pongo abelii*
SRX243483	Babu	*Pongo abelii*
SRX243486	Buschi	*Pongo abelii*
SRX243485	Buschi	*Pongo abelii*
SRX243482	Dunja	*Pongo abelii*
SRX243481	Dunja	*Pongo abelii*
SRX243478	Elsi	*Pongo abelii*
SRX243477	Elsi	*Pongo abelii*
SRX243480	Kiki	*Pongo abelii*
SRX243479	Kiki	*Pongo abelii*
SRX243476	Napoleon	*Pongo pygmaeus*
SRX243475	Napoleon	*Pongo pygmaeus*
SRX243468	Nonja	*Pongo pygmaeus*
SRX243467	Nonja	*Pongo pygmaeus*
SRX243472	Sari	*Pongo pygmaeus*
SRX243471	Sari	*Pongo pygmaeus*
SRX243470	Temmy	*Pongo pygmaeus*
SRX243469	Temmy	*Pongo pygmaeus*
SRX243474	Tilda	*Pongo pygmaeus*
SRX243473	Tilda	*Pongo pygmaeus*

Accession Number	Name	Species
SRX242686	Kaisi	*Gorilla beringei graueri*
SRX242685	Kaisi	*Gorilla beringei graueri*
SRX242688	Kaisi	*Gorilla beringei graueri*
SRX242687	Kaisi	*Gorilla beringei graueri*
SRX243453	Mkubwa	*Gorilla beringei graueri*
SRX243452	Mkubwa	*Gorilla beringei graueri*
SRX243533	Victoria	*Gorilla beringei graueri*
SRX243532	Victoria	*Gorilla beringei graueri*
SRX243531	Victoria	*Gorilla beringei graueri*
SRX243530	Victoria	*Gorilla beringei graueri*
SRX243529	Victoria	*Gorilla beringei graueri*
SRX243528	Victoria	*Gorilla beringei graueri*

Accession Number	Name	Species
SRX243502	Abe	*Gorilla gorilla gorilla*
SRX243504	Akiba	*Gorilla gorilla gorilla*
SRX243498	Amani	*Gorilla gorilla gorilla*
SRX243497	Amani	*Gorilla gorilla gorilla*
SRX243508	Anthal	*Gorilla gorilla gorilla*
SRX243443	Azizi	*Gorilla gorilla gorilla*
SRX243442	Azizi	*Gorilla gorilla gorilla*
SRX243456	Banjo	*Gorilla gorilla gorilla*
SRX243455	Banjo	*Gorilla gorilla gorilla*
SRX243441	Bulera	*Gorilla gorilla gorilla*
SRX243440	Bulera	*Gorilla gorilla gorilla*
SRX243520	Carolyn	*Gorilla gorilla gorilla*
SRX243505	Choomba	*Gorilla gorilla gorilla*
SRX243464	Coco	*Gorilla gorilla gorilla*
SRX243463	Coco	*Gorilla gorilla gorilla*
SRX243462	Delphi	*Gorilla gorilla gorilla*
SRX243461	Delphi	*Gorilla gorilla gorilla*
SRX243460	Dian	*Gorilla gorilla gorilla*
SRX243459	Dian	*Gorilla gorilla gorilla*
SRX243538	Dian	*Gorilla gorilla gorilla*
SRX243537	Dian	*Gorilla gorilla gorilla*
SRX243525	Dolly	*Gorilla gorilla gorilla*
SRX243523	Helen	*Gorilla gorilla gorilla*
SRX243509	Katie_B650	*Gorilla gorilla gorilla*
SRX243526	Katie_KB4986	*Gorilla gorilla gorilla*
SRX243501	Kokomo	*Gorilla gorilla gorilla*
SRX243500	Kokomo	*Gorilla gorilla gorilla*
SRX243466	Kolo	*Gorilla gorilla gorilla*
SRX243465	Kolo	*Gorilla gorilla gorilla*
SRX243439	Kowali	*Gorilla gorilla gorilla*
SRX243438	Kowali	*Gorilla gorilla gorilla*
SRX243458	Mimi	*Gorilla gorilla gorilla*
SRX243457	Mimi	*Gorilla gorilla gorilla*
SRX243536	Mimi	*Gorilla gorilla gorilla*
SRX243535	Mimi	*Gorilla gorilla gorilla*
SRX243524	Oko	*Gorilla gorilla gorilla*
SRX243506	Paki	*Gorilla gorilla gorilla*
SRX243521	Porta	*Gorilla gorilla gorilla*
SRX243454	Sandra	*Gorilla gorilla gorilla*
SRX242689	Sandra	*Gorilla gorilla gorilla*
SRX243445	Suzie	*Gorilla gorilla gorilla*
SRX243444	Suzie	*Gorilla gorilla gorilla*
SRX243503	Tzambo	*Gorilla gorilla gorilla*
SRX243522	Vila	*Gorilla gorilla gorilla*

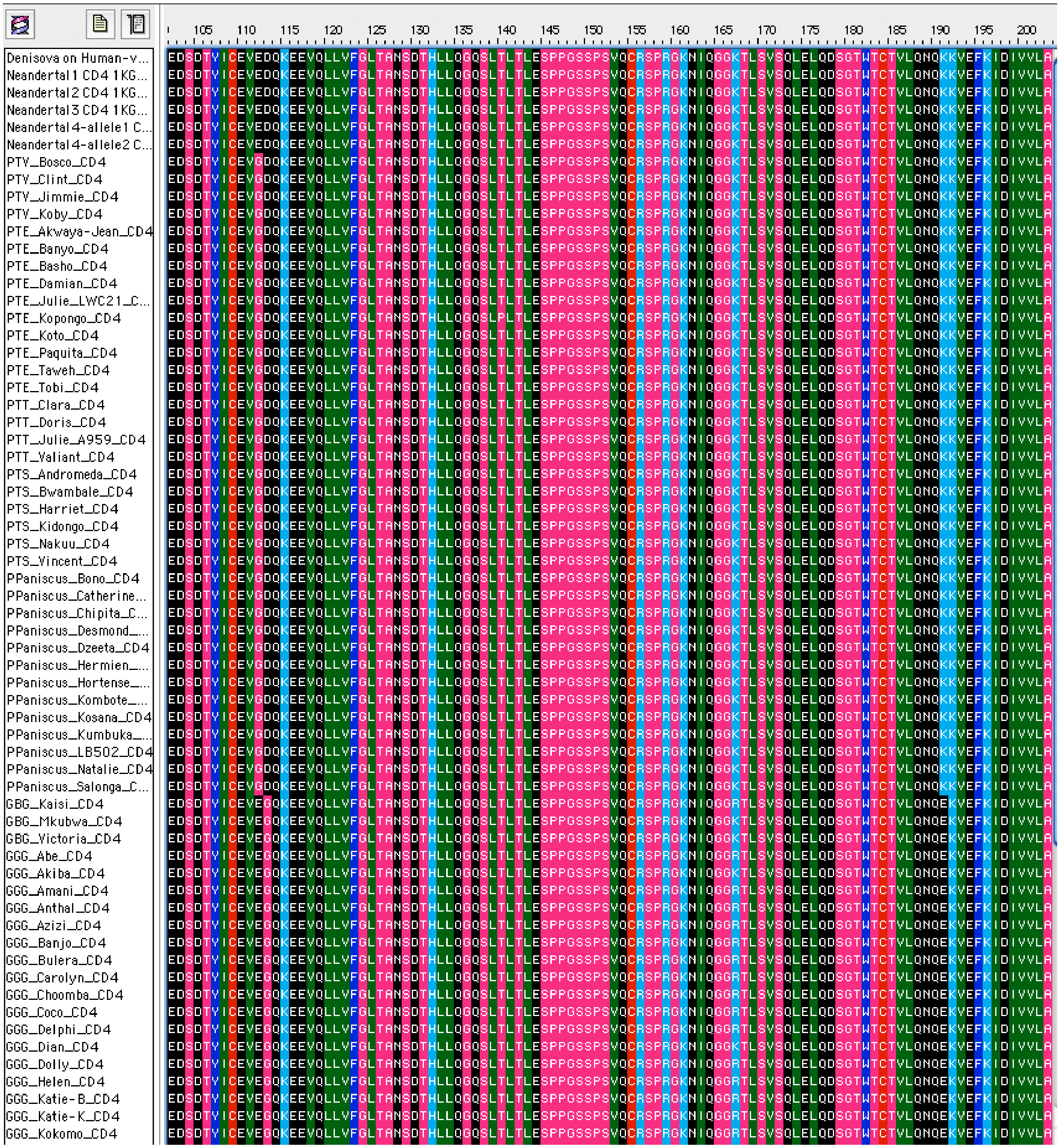
Accession Number	**Name**	**Species**
SRX241378	Bono	*Pan paniscus*
SRX241367	Bono	*Pan paniscus*
SRX241358	Bono	*Pan paniscus*
SRX241352	Bono	*Pan paniscus*
SRX241416	Bono	*Pan paniscus*
SRX241399	Bono	*Pan paniscus*
SRX241303	Catherine	*Pan paniscus*
SRX241302	Catherine	*Pan paniscus*
SRX241312	Chipita	*Pan paniscus*
SRX241311	Chipita	*Pan paniscus*
SRX241295	Desmond	*Pan paniscus*
SRX241294	Desmond	*Pan paniscus*
SRX241291	Dzeeta	*Pan paniscus*
SRX241290	Dzeeta	*Pan paniscus*
SRX241293	Hermien	*Pan paniscus*
SRX241292	Hermien	*Pan paniscus*
SRX237623	Hortense	*Pan paniscus*
SRX237602	Hortense	*Pan paniscus*
SRX241308	Kombote	*Pan paniscus*
SRX241307	Kombote	*Pan paniscus*
SRX241305	Kombote	*Pan paniscus*
SRX241304	Kombote	*Pan paniscus*
SRX241310	Kombote	*Pan paniscus*
SRX241309	Kombote	*Pan paniscus*
SRX241289	Kosana	*Pan paniscus*
SRX241288	Kosana	*Pan paniscus*
SRX242682	Kumbuka	*Pan paniscus*
SRX242681	Kumbuka	*Pan paniscus*
SRX242684	Kumbuka	*Pan paniscus*
SRX242683	Kumbuka	*Pan paniscus*
SRX243437	LB502	*Pan paniscus*
SRX243436	LB502	*Pan paniscus*
SRX241481	Natalie	*Pan paniscus*
SRX241477	Natalie	*Pan paniscus*
SRX241461	Natalie	*Pan paniscus*
SRX241441	Natalie	*Pan paniscus*
SRX241541	Natalie	*Pan paniscus*
SRX241528	Natalie	*Pan paniscus*
SRX241545	Salonga	*Pan paniscus*
SRX241544	Salonga	*Pan paniscus*

APPENDIX B

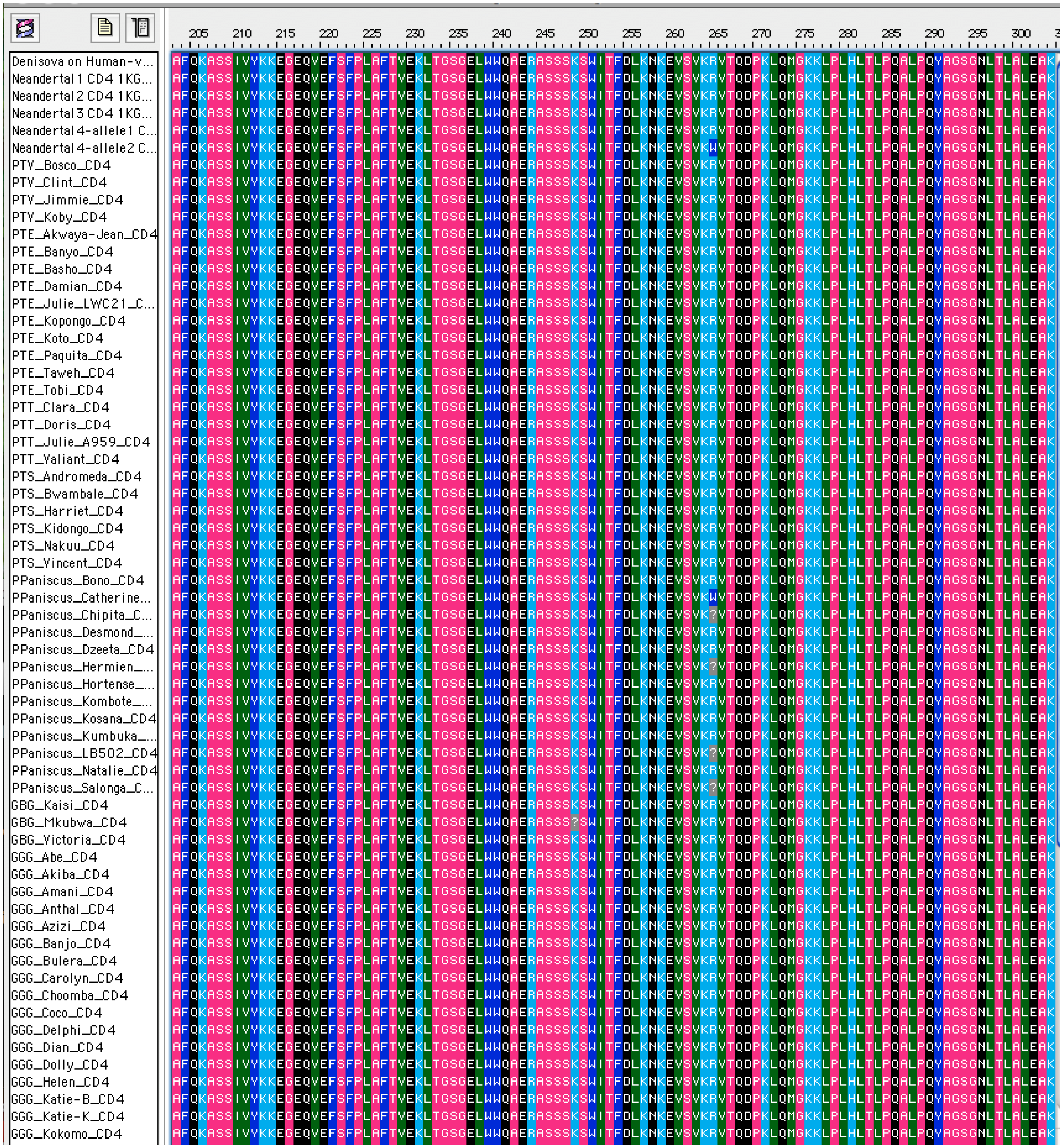


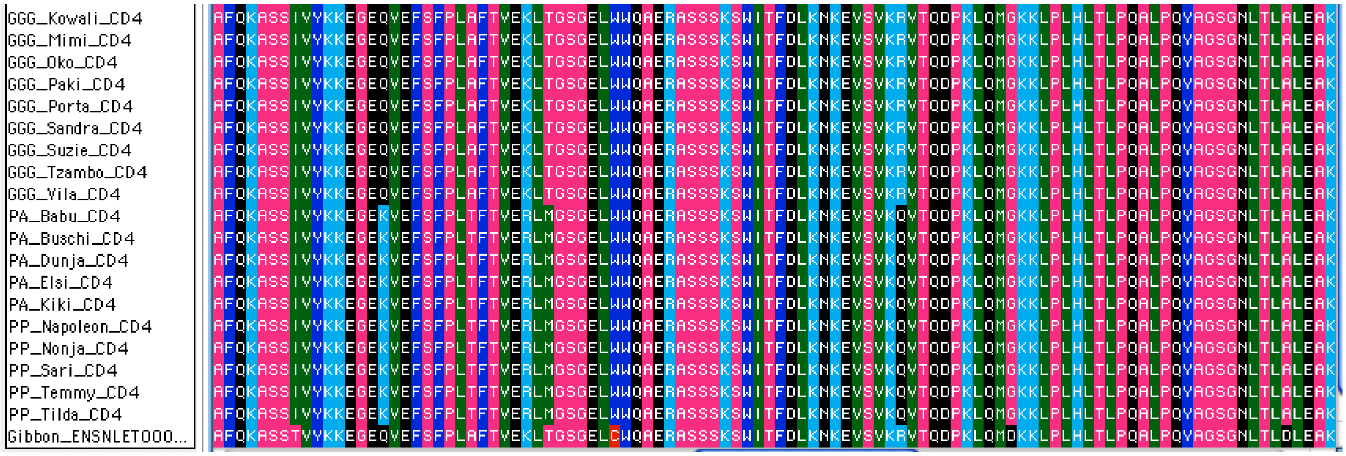
This appendix contains the obtained hominoid sequences used in this project. These Illumina HiSeq runs were obtained as mRNA sequences, and were then analyzed and organized into this master alignment in *Se-Al v2.0a11 Carbon* (Rambaut, 2002). This alignment shows the amino acid sequences in color coding to aid in visualization. A question mark (?) denotes an ambiguity in the amino acid sequence.

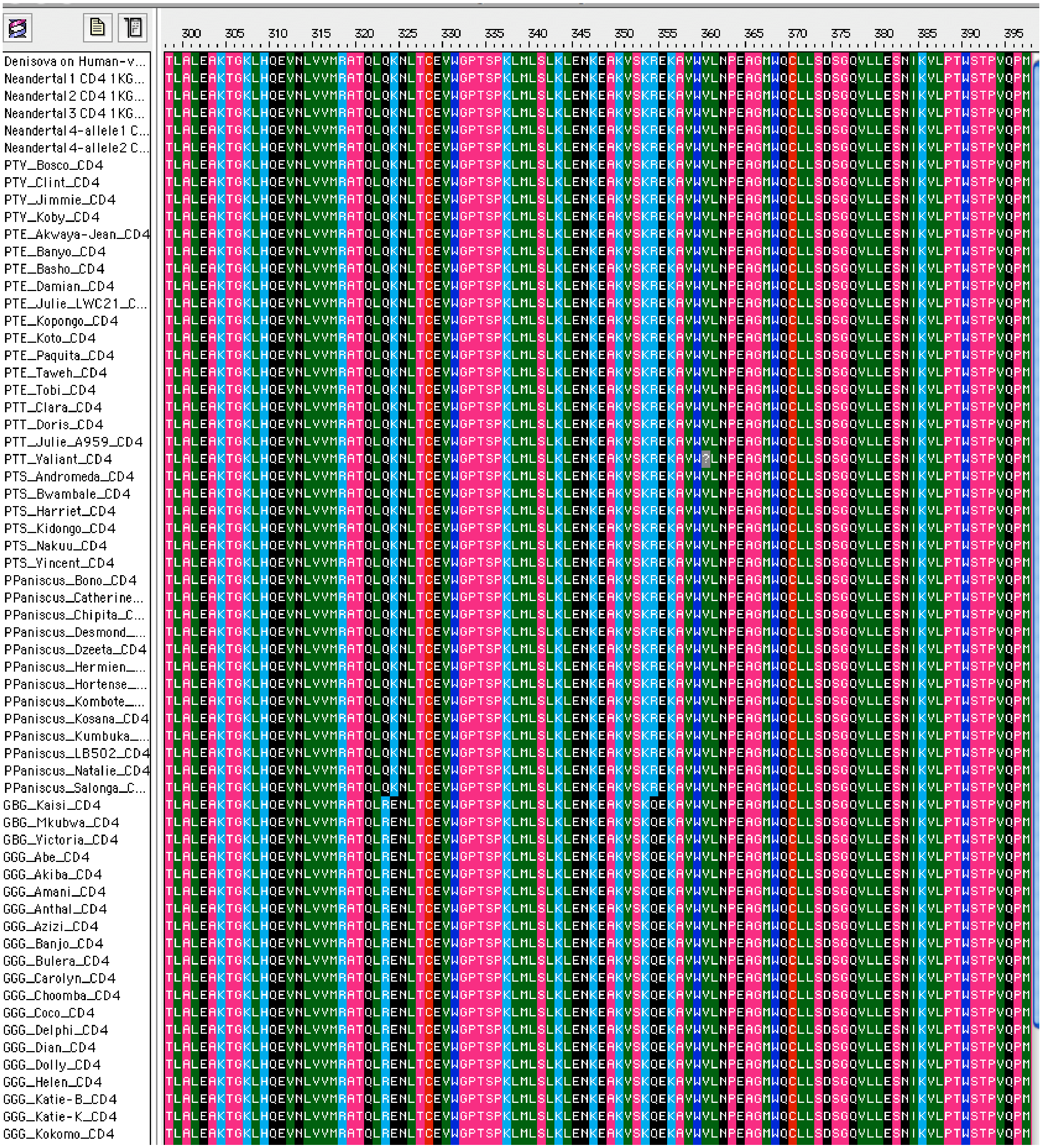

GGG_Kolo_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQF?WKNSNQRKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Kowali_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQF?WKNSNQRKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Mimi_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQFHWKNSNQMKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Oko_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQFHWKNSNQMKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Paki_CD4 MNRGVPFRHLLLVLQLALPPAATQGNKVVLGKKGDTVELNCTASQKKSIQF?WKNSNQRKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Porta_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQFHWKNSNQMKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Sandra_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQFHWKNSNQMKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Suzie_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQF?WKNSNQRKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Tzambo_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQFHWKNSNQMKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
GGG_Vila_CD4 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELNCTASQKKSIQFHWKNSNQMKILGNQGSFLTKGPSKLSDRADSRRSLWDQGNFPLIIKNLKIE
PA_Babu_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PA_Buschi_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PA_Dunja_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PA_Elsi_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PA_Kiki_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PP_Napoleon_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PP_Nonja_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PP_Sari_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PP_Temmy_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
PP_Tilda_CD4 MNQGIPFRHLLLVLQLVLLPAATPGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLSNRADSRRSLWDQGNFPLIIKNLKIE
Gibbon_ENSNLET000... MNPGIPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASPKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLSDRADSRKSLWDQRNFPLIIKNLKIE

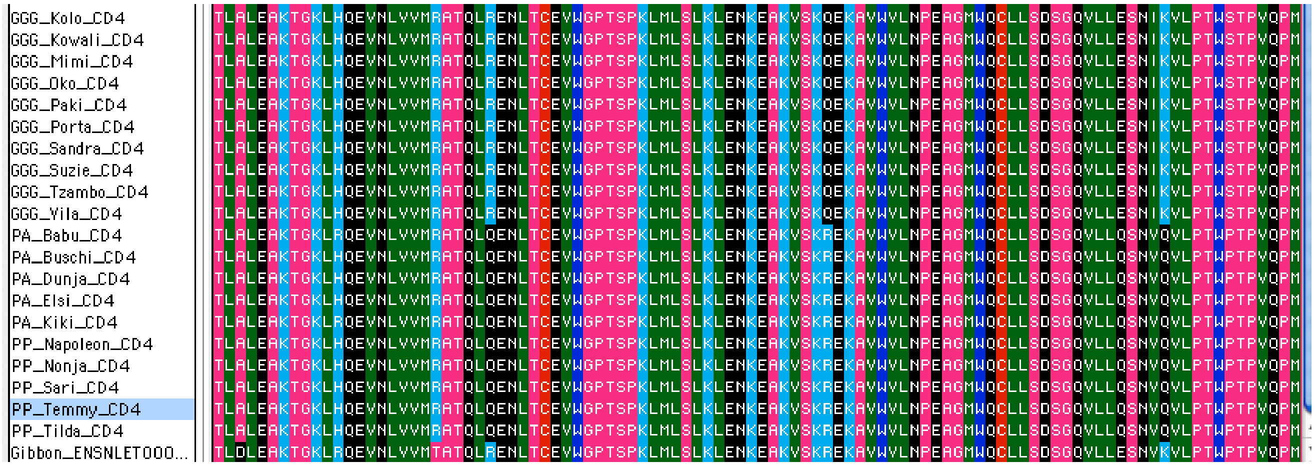

105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200
Denisova on Human-v...
Neandertal1 CD4 1KG...
Neandertal2 CD4 1KG...
Neandertal3 CD4 1KG...
Neandertal4-allele1 C...
Neandertal4-allele2 C...
PTV_Bosco_CD4
PTV_Clint_CD4
PTV_Jimmie_CD4
PTV_Koby_CD4
PTE_Akwaya-Jean_CD4
PTE_Banyo_CD4
PTE_Basho_CD4
PTE_Damian_CD4
PTE_Julie_LWC21_C...
PTE_Kopongo_CD4
PTE_Koto_CD4
PTE_Paquita_CD4
PTE_Taweh_CD4
PTE_Tobi_CD4
PTT_Clara_CD4
PTT_Doris_CD4
PTT_Julie_A959_CD4
PTT_Valiant_CD4
PTS_Andromeda_CD4
PTS_Bwambale_CD4
PTS_Harriet_CD4
PTS_Kidongo_CD4
PTS_Nakuu_CD4
PTS_Vincent_CD4
PPaniscus_Bono_CD4
PPaniscus_Catherine...
PPaniscus_Chipita_C...
PPaniscus_Desmond_...
PPaniscus_Dzeeta_CD4
PPaniscus_Hermien_...
PPaniscus_Hortense_...
PPaniscus_Kombote_...
PPaniscus_Kosana_CD4
PPaniscus_Kumbuka_...
PPaniscus_LB502_CD4
PPaniscus_Natalie_CD4
PPaniscus_Salonga_C...
GBG_Kaisi_CD4
GBG_Mkubwa_CD4
GBG_Victoria_CD4
GGG_Abe_CD4
GGG_Akiba_CD4
GGG_Amani_CD4
GGG_Anthal_CD4
GGG_Azizi_CD4
GGG_Banjo_CD4
GGG_Bulera_CD4
GGG_Carolyn_CD4
GGG_Choomba_CD4
GGG_Coco_CD4
GGG_Delphi_CD4
GGG_Dian_CD4
GGG_Dolly_CD4
GGG_Helen_CD4
GGG_Katie-B_CD4
GGG_Katie-K_CD4
GGG_Kokomo_CD4

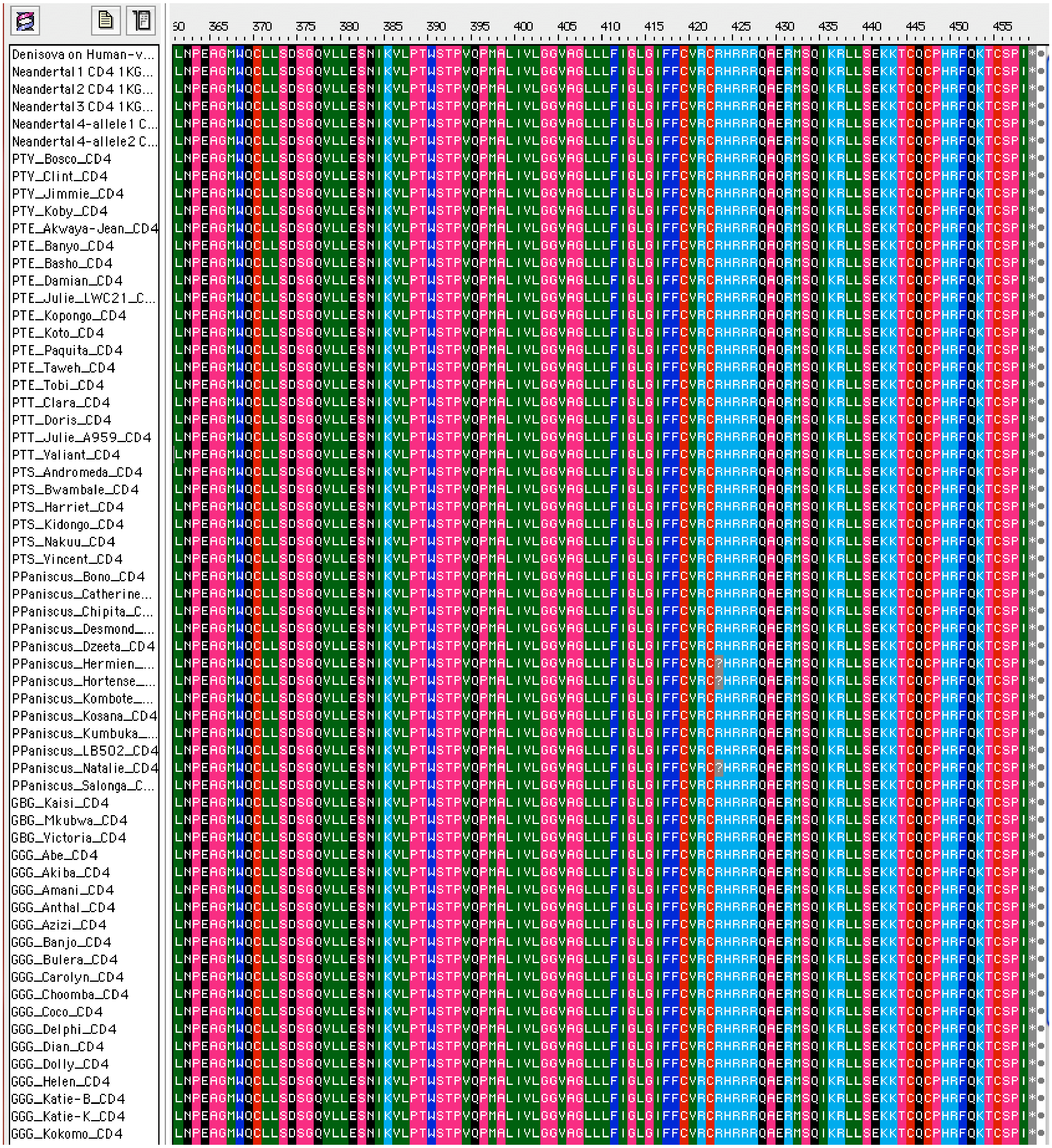

GGG_Kolo_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Kowali_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Mimi_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Oko_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Paki_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Porta_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Sandra_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Suzie_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Tzambo_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
GGG_Vila_CD4 EDSDTYICEVEGQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGRTLSVSQLELQDSGTWTCTVLQNQEKVEFKIDIVVLA
PA_Babu_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PA_Buschi_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PA_Dunja_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PA_Elsi_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PA_Kiki_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PP_Napoleon_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PP_Nonja_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PP_Sari_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PP_Temmy_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
PP_Tilda_CD4 EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPTGKNIQAGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA
Gibbon_ENSNLET000... EDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLEGPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQDQKKVEFKIDIVVLA


205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300
Denisova on Human-v... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
Neandertal 1 CD4 1KG... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
Neandertal 2 CD4 1KG... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
Neandertal 3 CD4 1KG... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
Neandertal 4-allele1 C... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
Neandertal 4-allele2 C... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKWVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTV_Bosco_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTV_Clint_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTV_Jimmie_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTV_Koby_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Akwaya-Jean_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Banyo_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Basho_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Damian_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Julie_LWC21_C... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Kopongo_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Koto_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Paquita_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Taweh_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTE_Tobi_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTT_Clara_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTT_Doris_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTT_Julie_A959_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTT_Valiant_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTS_Andromeda_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTS_Bwambale_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTS_Harriet_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTS_Kidongo_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTS_Nakuu_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PTS_Vincent_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Bono_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Catherine... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKWVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Chipita_C... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVK?VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Desmond_... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Dzeeta_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Hermien_... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVK?VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Hortense_... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Kombote_... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Kosana_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Kumbuka_... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_LB502_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVK?VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Natalie_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PPaniscus_Salonga_C... AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVK?VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GBG_Kaisi_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GBG_Mkubwa_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSS?SWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GBG_Victoria_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Abe_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Akiba_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Amani_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Anthal_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Azizi_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Banjo_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Bulera_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Carolyn_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Choomba_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Coco_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Delphi_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Dian_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Dolly_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Helen_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Katie-B_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Katie-K_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Kokomo_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK


GGG_Kowali_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Mimi_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Oko_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Paki_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Porta_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Sandra_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Suzie_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Tzambo_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
GGG_Vila_CD4 AFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PA_Babu_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PA_Buschi_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PA_Dunja_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PA_Elsi_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PA_Kiki_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PP_Napoleon_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PP_Nonja_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PP_Sari_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PP_Temmy_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
PP_Tilda_CD4 AFQKASSIVYKKEGEKVEFSFPLTFTVERLMGSGELWWQAERASSSKSWITFDLKNKEVSVKQVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK
Gibbon_ENSNLET000... AFQKASSTVYKKEGEQVEFSFPLAFTVEKLTGSGELCWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMDKKLPLHLTLPQALPQYAGSGNLTLDLEAK


300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395
Denisova on Human-v... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
Neandertal 1 CD4 1KG... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
Neandertal 2 CD4 1KG... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
Neandertal 3 CD4 1KG... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
Neandertal 4-allele1 C... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
Neandertal 4-allele2 C... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTV_Bosco_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTV_Clint_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTV_Jimmie_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTV_Koby_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Akwaya-Jean_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Banyo_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Basho_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Damian_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Julie_LWC21_C... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Kopongo_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Koto_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Paquita_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Taweh_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTE_Tobi_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTT_Clara_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTT_Doris_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTT_Julie_A959_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTT_Valiant_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVW?LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTS_Andromeda_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTS_Bwambale_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTS_Harriet_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTS_Kidongo_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTS_Nakuu_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PTS_Vincent_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Bono_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Catherine... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Chipita_C... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Desmond_... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Dzeeta_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Hermien_... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Hortense_... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Kombote_... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Kosana_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Kumbuka_... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_LB502_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Natalie_CD4 TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PPaniscus_Salonga_C... TLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GBG_Kaisi_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GBG_Mkubwa_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GBG_Victoria_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Abe_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Akiba_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Amani_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Anthal_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Azizi_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Banjo_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Bulera_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Carolyn_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Choomba_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Coco_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Delphi_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Dian_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Dolly_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Helen_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Katie-B_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Katie-K_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Kokomo_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM


GGG_Kolo_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Kowali_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Mimi_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Oko_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Paki_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Porta_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Sandra_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Suzie_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Tzambo_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
GGG_Yila_CD4 TLALEAKTGKLHQEVNLVVMRATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKQEKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPM
PA_Babu_CD4 TLALEAKTGKLRQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PA_Buschi_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PA_Dunja_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PA_Elsi_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PA_Kiki_CD4 TLALEAKTGKLRQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PP_Napoleon_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PP_Nonja_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PP_Sari_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PP_Temmy_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
PP_Tilda_CD4 TLALEAKTGKLHQEVNLVVMRATQLQENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPM
Gibbon_ENSNLET000... TLDLEAKTGKLRQEVNLVVMTATQLRENLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNVKVLPTWPTPVQPM



60 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455
Denisova on Human-v...
Neandertal 1 CD4 1KG...
Neandertal 2 CD4 1KG...
Neandertal 3 CD4 1KG...
Neandertal 4-allele1 C...
Neandertal 4-allele2 C...
PTV_Bosco_CD4
PTV_Clint_CD4
PTV_Jimmie_CD4
PTV_Koby_CD4
PTE_Akwaya-Jean_CD4
PTE_Banyo_CD4
PTE_Basho_CD4
PTE_Damian_CD4
PTE_Julie_LWC21_C...
PTE_Kopongo_CD4
PTE_Koto_CD4
PTE_Paquita_CD4
PTE_Taweh_CD4
PTE_Tobi_CD4
PTT_Clara_CD4
PTT_Doris_CD4
PTT_Julie_A959_CD4
PTT_Valiant_CD4
PTS_Andromeda_CD4
PTS_Bwambale_CD4
PTS_Harriet_CD4
PTS_Kidongo_CD4
PTS_Nakuu_CD4
PTS_Vincent_CD4
PPaniscus_Bono_CD4
PPaniscus_Catherine...
PPaniscus_Chipita_C...
PPaniscus_Desmond_...
PPaniscus_Dzeeta_CD4
PPaniscus_Hermien_...
PPaniscus_Hortense_...
PPaniscus_Kombote_...
PPaniscus_Kosana_CD4
PPaniscus_Kumbuka_...
PPaniscus_LB502_CD4
PPaniscus_Natalie_CD4
PPaniscus_Salonga_C...
GBG_Kaisi_CD4
GBG_Mkubwa_CD4
GBG_Victoria_CD4
GGG_Abe_CD4
GGG_Akiba_CD4
GGG_Amani_CD4
GGG_Anthal_CD4
GGG_Azizi_CD4
GGG_Banjo_CD4
GGG_Bulera_CD4
GGG_Carolyn_CD4
GGG_Choomba_CD4
GGG_Coco_CD4
GGG_Delphi_CD4
GGG_Dian_CD4
GGG_Dolly_CD4
GGG_Helen_CD4
GGG_Katie-B_CD4
GGG_Katie-K_CD4
GGG_Kokomo_CD4


GGG_Kolo_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Kowali_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Mimi_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Oko_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Paki_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Porta_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Sandra_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Suzie_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Tzambo_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
GGG_Vila_CD4 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PA_Babu_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PA_Buschi_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PA_Dunja_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PA_Elsi_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PA_Kiki_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PP_Napoleon_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PP_Nonja_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PP_Sari_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PP_Temmy_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
PP_Tilda_CD4 LNPEAGMWQCLLSDSGQVLLQSNVQVLPTWPTPVQPMALIALGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*
Gibbon_ENSNLET000... LNPEAGMWQCLLSDSGQVLLESNVKVLPTWPTPVQPMALIVLGGVAGLLLFIGLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTCQCPHRFQKTCSPI*

APPENDIX C

CD4 D1	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
PTV Bosco							288	0.00E+00	0.00E+00
PTV Clint							288	0.00E+00	
PTV Jimmie							288	0.00E+00	
PTV Koby							288	0.00E+00	
PTE Akwaya-Jean							288	0.00E+00	1.74E-03
PTE Banyo			1			1	288	6.94E-03	
PTE Basho							288	0.00E+00	
PTE Damian							288	0.00E+00	
PTE Julie_LWC21			1				288	3.47E-03	
PTE Kopongo							288	0.00E+00	
PTE Koto							288	0.00E+00	
PTE Paquita							288	0.00E+00	
PTE Taweh							288	0.00E+00	
PTE Tobi			1			1	288	6.94E-03	
PTT Clara							288	0.00E+00	2.60E-03
PTT Doris		1					288	3.47E-03	
PTT Julie		1					288	3.47E-03	
PTT Valliant			1				288	3.47E-03	
PTS Andromeda							288	0.00E+00	1.16E-03
PTS Bwambale							288	0.00E+00	
PTS Harriet							288	0.00E+00	
PTS Kindongo			1			1	288	6.94E-03	
PTS Nakuu							288	0.00E+00	
PTS Vincent							288	0.00E+00	

This appendix contains the heterozygosity analysis and calculation values for all the great apes analyzed in this study for D1 of CD4, CD4 minus D1, and CD4 as a whole.

CD4 D1	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
Bonobo Bono							288	0.00E+00	0.00E+00
Bonobo Catherine							288	0.00E+00	
Bonobo Chipita							288	0.00E+00	
Bonobo Desmond							288	0.00E+00	
Bonobo Dzeeta							288	0.00E+00	
Bonobo Hermien							288	0.00E+00	
Bonobo Hortense							288	0.00E+00	
Bonobo Kombote							288	0.00E+00	
Bonobo Kosana							288	0.00E+00	
Bonobo Kumbuka							288	0.00E+00	
Bonobo LB502							288	0.00E+00	
Bonobo Natalie							288	0.00E+00	
Bonobo Salonga							288	0.00E+00	
GBG Kaisi	1	1			1		288	1.04E-02	3.47E-03
GBG Mkubwa							288	0.00E+00	
GBG Victoria							288	0.00E+00	

CD4 D1	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
GGG Abe			1	1			288	6.94E-03	4.22E-03
GGG Akiba			1	1			288	6.94E-03	
GGG Amani		1					288	3.47E-03	
GGG Anthal			1	1			288	6.94E-03	
GGG Azizi							288	0.00E+00	
GGG Banjo							288	0.00E+00	
GGG Bulera		1	1	1			288	1.04E-02	
GGG Carolyn			1	1			288	6.94E-03	
GGG Choomba							288	0.00E+00	
GGG Coco		1					288	3.47E-03	
GGG Delphi			1	1			288	6.94E-03	
GGG Dian			1	1			288	6.94E-03	
GGG Dolly							288	0.00E+00	
GGG Helen		1					288	3.47E-03	
GGG Katie B							288	0.00E+00	
GGG Katie K							288	0.00E+00	
GGG Kokomo		1	1	1			288	1.04E-02	
GGG Kolo			1	1			288	6.94E-03	
GGG Kowali		1	1	1			288	1.04E-02	
GGG Mimi							288	0.00E+00	
GGG Oko		1					288	3.47E-03	
GGG Paki			1	1			288	6.94E-03	
GGG Porta							288	0.00E+00	
GGG Sandra							288	0.00E+00	
GGG Suzie		1	1	1			288	1.04E-02	
GGG Tzambo	1	1					288	6.94E-03	
GGG Vila							288	0.00E+00	
PA Babu							288	0.00E+00	0.00E+00
PA Buschi							288	0.00E+00	
PA Dunja							288	0.00E+00	
PA Elsi							288	0.00E+00	
PA Kiki							288	0.00E+00	
PP Napoleon							288	0.00E+00	0.00E+00
PP Nonja							288	0.00E+00	
PP Sari							288	0.00E+00	
PP Temmy							288	0.00E+00	
PP Tila							288	0.00E+00	

CD4-D1	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
PTV Bosco							1377	0.00E+00	3.63E-04
PTV Clint		1					1377	7.26E-04	
PTV Jimmie				1			1377	7.26E-04	
PTV Koby							1377	0.00E+00	
PTE Akwaya-Jean							1377	0.00E+00	2.90E-04
PTE Banyo	1		2				1377	2.18E-03	
PTE Basho							1377	0.00E+00	
PTE Damian							1377	0.00E+00	
PTE Julie_LWC21			1				1377	7.26E-04	
PTE Kopongo							1377	0.00E+00	
PTE Koto							1377	0.00E+00	
PTE Paquita							1377	0.00E+00	
PTE Taweh							1377	0.00E+00	
PTE Tobi							1377	0.00E+00	
PTT Clara	1	2	1				1377	2.90E-03	1.63E-03
PTT Doris				1			1377	7.26E-04	
PTT Julie	1		1				1377	1.45E-03	
PTT Valliant		2					1377	1.45E-03	
PTS Andromeda							1377	0.00E+00	2.42E-04
PTS Bwambale		1					1377	7.26E-04	
PTS Harriet							1377	0.00E+00	
PTS Kindongo		1					1377	7.26E-04	
PTS Nakuu							1377	0.00E+00	
PTS Vincent							1377	0.00E+00	

CD4-D1	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
Bonobo Bono							1377	0.00E+00	5.59E-04
Bonobo Catherine							1377	0.00E+00	
Bonobo Chipita	1						1377	7.26E-04	
Bonobo Desmond							1377	0.00E+00	
Bonobo Dzeeta							1377	0.00E+00	
Bonobo Hermien	2	1					1377	2.18E-03	
Bonobo Hortense	1	1					1377	1.45E-03	
Bonobo Kombote							1377	0.00E+00	
Bonobo Kosana							1377	0.00E+00	
Bonobo Kumbuka							1377	0.00E+00	
Bonobo LB502	1						1377	7.26E-04	
Bonobo Natalie	1	1					1377	1.45E-03	
Bonobo Salonga	1						1377	7.26E-04	
GBG Kaisi							1377	0.00E+00	1.45E-03
GBG Mkubwa	1	2	1	2			1377	4.36E-03	
GBG Victoria							1377	0.00E+00	

CD4-D1	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
GGG Abe							1377	0.00E+00	6.22E-04
GGG Akiba		1					1377	7.26E-04	
GGG Amani							1377	0.00E+00	
GGG Anthal		1					1377	7.26E-04	
GGG Azizi							1377	0.00E+00	
GGG Banjo	1	2					1377	2.18E-03	
GGG Bulera	1						1377	7.26E-04	
GGG Carolyn	1						1377	7.26E-04	
GGG Choomba							1377	0.00E+00	
GGG Coco		1					1377	7.26E-04	
GGG Delphi			1				1377	7.26E-04	
GGG Dian		1					1377	7.26E-04	
GGG Dolly							1377	0.00E+00	
GGG Helen		1					1377	7.26E-04	
GGG Katie B							1377	0.00E+00	
GGG Katie K							1377	0.00E+00	
GGG Kokomo		1					1377	7.26E-04	
GGG Kolo		1					1377	7.26E-04	
GGG Kowali							1377	0.00E+00	
GGG Mimi		1		1			1377	1.45E-03	
GGG Oko		1					1377	7.26E-04	
GGG Paki	1	1					1377	1.45E-03	
GGG Porta		1					1377	7.26E-04	
GGG Sandra		2					1377	1.45E-03	
GGG Suzie	1						1377	7.26E-04	
GGG Tzambo	1						1377	7.26E-04	
GGG Vila		1					1377	7.26E-04	
PA Babu							1377	0.00E+00	2.90E-04
PA Buschi		1					1377	7.26E-04	
PA Dunja							1377	0.00E+00	
PA Elsi		1					1377	7.26E-04	
PA Kiki							1377	0.00E+00	
PP Napoleon							1377	0.00E+00	1.45E-04
PP Nonja	1						1377	7.26E-04	
PP Sari							1377	0.00E+00	
PP Temmy							1377	0.00E+00	
PP Tila							1377	0.00E+00	
PP Tila							1377	0.00E+00	

CD4	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
PTV Bosco							1377	0.00E+00	3.63E-04
PTV Clint		1					1377	7.26E-04	
PTV Jimmie				1			1377	7.26E-04	
PTV Koby							1377	0.00E+00	
PTE Akwaya-Jean							1377	0.00E+00	6.54E-04
PTE Banyo	1		2			1	1377	2.90E-03	
PTE Basho			1				1377	7.26E-04	
PTE Damian							1377	0.00E+00	
PTE Julie_LWC21			2				1377	1.45E-03	
PTE Kopongo							1377	0.00E+00	
PTE Koto							1377	0.00E+00	
PTE Paquita							1377	0.00E+00	
PTE Taweh							1377	0.00E+00	
PTE Tobi			1			1	1377	1.45E-03	
PTT Clara	1	2	1				1377	2.90E-03	2.18E-03
PTT Doris		1		1			1377	1.45E-03	
PTT Julie	1	1	1				1377	2.18E-03	
PTT Valliant		2	1				1377	2.18E-03	
PTS Andromeda							1377	0.00E+00	4.84E-04
PTS Bwambale		1					1377	7.26E-04	
PTS Harriet							1377	0.00E+00	
PTS Kindongo		1	1			1	1377	2.18E-03	
PTS Nakuu							1377	0.00E+00	
PTS Vincent							1377	0.00E+00	

CD4	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
Bonobo Bono							1377	0.00E+00	5.59E-04
Bonobo Catherine							1377	0.00E+00	
Bonobo Chipita	1						1377	7.26E-04	
Bonobo Desmond							1377	0.00E+00	
Bonobo Dzeeta							1377	0.00E+00	
Bonobo Hermien	2	1					1377	2.18E-03	
Bonobo Hortense	1	1					1377	1.45E-03	
Bonobo Kombote							1377	0.00E+00	
Bonobo Kosana							1377	0.00E+00	
Bonobo Kumbuka							1377	0.00E+00	
Bonobo LB502	1						1377	7.26E-04	
Bonobo Natalie	1	1					1377	1.45E-03	
Bonobo Salonga	1						1377	7.26E-04	
GBG Kaisi	1	1			1		1377	2.18E-03	2.18E-03
GBG Mkubwa	1	2	1		2		1377	4.36E-03	
GBG Victoria							1377	0.00E+00	

CD4	Ambiguity Code						Gene Size	Heterozygosity	Average Heterozygosity by Sub-Species
Species_Individual	Y	R	M	K	S	W			
GGG Abe			1	1			1377	1.45E-03	1.53E-03
GGG Akiba		1	1	1			1377	2.18E-03	
GGG Amani		1					1377	7.26E-04	
GGG Anthal		1	1	1			1377	2.18E-03	
GGG Azizi							1377	0.00E+00	
GGG Banjo	1	1	1				1377	2.18E-03	
GGG Bulera	1	1	1	1			1377	2.90E-03	
GGG Carolyn	1		1	1			1377	2.18E-03	
GGG Choomba							1377	0.00E+00	
GGG Coco		2					1377	1.45E-03	
GGG Delphi			1	1			1377	1.45E-03	
GGG Dian		1	1	1			1377	2.18E-03	
GGG Dolly							1377	0.00E+00	
GGG Helen		2					1377	1.45E-03	
GGG Katie B							1377	0.00E+00	
GGG Katie K							1377	0.00E+00	
GGG Kokomo		2	1	1			1377	2.90E-03	
GGG Kolo		1	1	1			1377	2.18E-03	
GGG Kowali		1	1	1			1377	2.18E-03	
GGG Mimi		1		1			1377	1.45E-03	
GGG Oko		2					1377	1.45E-03	
GGG Paki	1	1	1	1			1377	2.90E-03	
GGG Porta		1					1377	7.26E-04	
GGG Sandra		2					1377	1.45E-03	
GGG Suzie	1	1	1	1			1377	2.90E-03	
GGG Tzambo	2	1					1377	2.18E-03	
GGG Vila		1					1377	7.26E-04	
PA Babu							1377	0.00E+00	2.90E-04
PA Buschi		1					1377	7.26E-04	
PA Dunja							1377	0.00E+00	
PA Elsi		1					1377	7.26E-04	
PA Kiki							1377	0.00E+00	
PP Napoleon							1377	0.00E+00	1.45E-04
PP Nonja	1						1377	7.26E-04	
PP Sari							1377	0.00E+00	
PP Temmy							1377	0.00E+00	
PP Tila							1377	0.00E+00	