Microbiology and Infectious Disease

Adaptation of CD4 in gorillas and chimpanzees conveyed resistance to simian immunodeficiency viruses

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2
May 28, 2024 (this version)

Reviewed preprint version 1
January 10, 2024

Posted to preprint server
November 13, 2023

Sent for peer review
November 13, 2023

Cody J. Warren, Arturo Barbachano-Guerrero, Vanessa L. Bauer, Alex C. Stabell, Obaiah Dirasantha, Qing Yang, Sara L. Sawyer

BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA

Abstract

Simian immunodeficiency viruses (SIVs) comprise a large group of primate lentiviruses that endemically infect African monkeys. HIV-1 spilled over to humans from this viral reservoir, but the spillover did not occur directly from monkeys to humans. Instead, a key event was the introduction of SIVs into great apes, which then set the stage for infection of humans. Here, we investigate the role of the lentiviral entry receptor, CD4, in this key and fateful event in the history of SIV/HIV emergence. First, we reconstructed and tested ancient forms of CD4 at two important nodes in ape speciation, both prior to the infection of chimpanzees and gorillas with these viruses. These ancestral CD4s fully supported entry of diverse SIV isolates related to the viruses that made this initial jump to apes. In stark contrast, modern chimpanzee and gorilla CD4 orthologs are more resistant to these viruses. To investigate how this resistance in CD4 was gained, we acquired *CD4* gene sequences from 32 gorilla individuals of two species, and identified alleles that encode 8 unique CD4 protein variants. Functional testing of these identified variant-specific differences in susceptibility to virus entry. By engineering single point mutations from resistant gorilla CD4 variants into the permissive human CD4 receptor, we demonstrate that acquired substitutions in gorilla CD4 did convey resistance to virus entry. We provide a population genetic analysis to support the theory that selection is acting in favor of more and more resistant *CD4* alleles in ape species harboring SIV endemically (gorillas and chimpanzees), but not in other ape species that lack SIV infections (bonobos and orangutans). Taken together, our results show that SIV has placed intense selective pressure on ape *CD4*, acting to propagate SIV-resistant alleles in chimpanzee and gorilla populations.

eLife assessment

This study presents an **important** finding on how lentiviral infection has driven the diversification of the HIV/SIV entry receptor CD4. Using a combination of molecular evolution approaches coupled with functional testing of extant and ancestral reconstructions of great ape CD4, the authors provide **solid** evidence to support the idea that endemic simian immunodeficiency virus infection in gorillas have selected for gorilla CD4 alleles that are more resistant to SIV infection. Expanding the study to interrogate the evolution and function of additional primate CD4 sequences could yield more **convincing** evidence.

https://doi.org/10.7554/eLife.93316.2.sa2

Introduction

Simian immunodeficiency viruses (SIVs) cause life-long chronic infections in African monkeys and apes (Klatt et al., 2012; Sharp and Hahn, 2011). SIVs are classified in subfamily *Orthoretrovirinae* (genus *Lentivirus*). Currently, individuals from 45 different African primate species have demonstrated antibodies that react to SIVs in serological tests, and in 37 of these species, SIV exposure has been confirmed by obtaining at least partial sequence of the actual virus (Ayouba et al., 2014). Complete genome sequence is available for only 27 SIVs (Ayouba et al., 2014). Unique virus isolates are named "SIV" followed by a three-letter subscript that refers to the host primate species from which it was isolated (*e.g.*, SIVcpz was isolated from chimpanzees) (Ayouba et al., 2014). Remarkably, of the tested primate species in Africa, approximately 90% of them have been associated with at least one SIV strain, indicating nearly pervasive infection of African primates (Ayouba et al., 2014).

HIV-1 emerged into humans from this diverse viral reservoir, but was not a spillover of virus directly from monkeys to humans. Instead, a key transition was the spillover of SIVs into great apes, which then set the stage for infection of humans (**Figure 1**). First, SIV of chimpanzees (SIVcpz) arose following the cross-species transmission and recombination of multiple SIVs from infected monkeys upon which chimpanzees predate (Bailes et al., 2003; Sharp et al., 2005). It is unknown if this virus recombination event occurred in the monkey reservoir before the first chimpanzee was infected, or if it occurred within chimpanzee populations. Subsequently, SIVcpz transmitted to gorillas (giving rise to SIVgor) (Heuverswyn et al., 2006; Takehisa et al., 2009). Chimpanzees and gorillas have been endemically infected with SIVcpz and SIVgor since those spillover events (Sharp and Hahn, 2011). Spillover to humans from both chimpanzees and gorillas subsequently occurred on multiple occasions (Heuverswyn et al., 2006; Keele et al., 2006; Plantier et al., 2009). One of these spillovers yielded HIV-1 "group M" – the pandemic virus that has swept the globe, infecting over 80 million people. A third great ape species native to Africa, the bonobo, remains uninfected with SIV. Orangutans, the final great ape species, are native to Asia and also remain uninfected.

CD4 is the primary entry receptor for primate lentiviruses (SIV and HIV). CD4 is a surface protein expressed on T cells, where it is bound by the viral envelope (Env) glycoprotein to begin viral entry into the cell. To understand the role that CD4 plays in dictating the host tropism of SIVs, one must first appreciate the remarkable evolutionary signatures contained in the *CD4* gene. *CD4* has evolved under positive natural selection over the course of primate evolution (Meyerson et al., 2014; Zhang et al., 2008). This type of selection operates in favor of new alleles of *CD4* that have better resistance to virus entry (Meyerson and Sawyer, 2011). As such, it has been noted

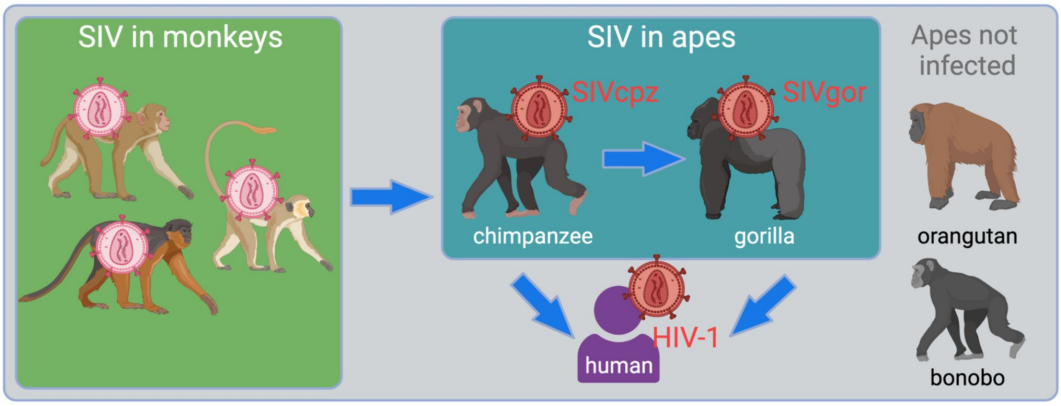

Figure 1.

Overview of the emergence of SIV into apes, ultimately giving rise to several groups of HIV-1.

The figure shows, in the green box, the SIV reservoir that exists in African monkeys. Chimpanzees became infected with viruses from this reservoir resulting in the emergence of a new virus lineage, SIVcpz (Bailes et al., 2003; Sharp et al., 2005). From there, chimpanzees infected both gorillas and humans. The final two great ape species, orangutans and bonobos, are not known to harbor an SIV.

that most of the sequence evolution in *CD4* has been concentrated to the region encoding the D1 domain that interacts with HIV and SIV Env (Meyerson et al., 2014; Zhang et al., 2008). Even though natural selection operates at the level of alleles circulating within primate populations (Ohainle and Malik, 2021; Russell et al., 2021), the ultimate outcome is fixed *CD4* sequence divergence between species. As a result, we have demonstrated that different primate orthologs of CD4 vary dramatically in the lentiviruses that they will engage (Warren et al., 2019a).

Here, we focus on a key event in the emergence of HIV-1 into humans – the spillover of SIVs from monkeys to apes. First, we reconstructed and tested ancestral forms of CD4 at two important nodes in ape speciation, prior to the infection of chimpanzees and gorillas with these viruses. These ancestral CD4s fully support entry of diverse SIV isolates representing the viruses that made this initial jump to apes. In stark contrast, modern chimpanzee and gorilla CD4 orthologs are less supportive of infection by these viruses, consistent with natural selection having shaped *CD4* to resist infection in these species. Second, we investigated the subsequent spillover of SIV from chimpanzees to gorillas. We gathered *CD4* sequences from 32 gorilla individuals of two species, and identified alleles that encode 8 unique CD4 protein variants. We then identified variant-specific differences in susceptibility to SIVcpz entry (the virus that spilled over to gorillas). By engineering single point mutations from resistant gorilla CD4 variants into a permissive human CD4 receptor, we demonstrate that these mutations are responsible for resistance to virus entry in gorillas. A population genetic analysis supports the theory the selection is acting in favor of more and more resistant *CD4* alleles in gorillas. Taken together with similar analyses in chimpanzees (Bibollet-Ruche et al., 2019; Warren et al., 2019b), our results show that SIV has placed intense selective pressure on ape *CD4*, retaining and propagating SIV-resistant *CD4* alleles.

Results

Receptor mediated resistance to SIV entry is a trait acquired during ape speciation

First, we wanted to know what ape CD4 was like before SIVs spilled over to apes and began to exert infection pressure on them. We used an alignment of *CD4* from diverse simian primates, and the program PAML (Yang, 2007a), to infer ancestral *CD4* sequences at the base of the hominin and hominid clades, at the evolutionary positions shown with red and blue nodes in **Fig. 2A**. The reconstructions of sequences at these ancestral nodes yielded *CD4* sequences that differ from human *CD4* by only 2 (hominin) and 5 (hominid) nonsynonymous substitutions. Only one of these changes mapped to the D1 domain of CD4 (N52S; **Fig. 2B**). We then synthesized these extinct *CD4* genes. We transduced Cf2Th (canine) cells with retroviral vectors that stably integrated each of these *CD4* genes (hominin ancestral *CD4*, hominid ancestral *CD4*, human *CD4*, gorilla *CD4*, chimpanzee *CD4*, or an empty vector; **Fig. S1**). All cell lines were also transduced to stably express human *CCR5*, a critical co-receptor for SIV and HIV entry.

We then tested these extinct and modern CD4 proteins for their ability to support viral entry mediated by SIVcpz Env. Since we don't know the actual genetic sequence of the first SIV(s) to infect chimpanzees, the best alternate strategy is to test a phylogenetically diverse set of extant SIVcpz strains (**Fig. 2C**). We also tested HIV-1 strains that are embedded within the SIVcpz clade, because these represent the virus spillovers from chimpanzees to humans. To generate pseudoviruses bearing SIVcpz and HIV-1 Env, different Env expression plasmids were co-transfected into 293T cells along with a plasmid encoding HIV-1ΔEnv-eGFP. The cell lines stably expressing various CD4 proteins and human CCR5 were then infected with each of these pseudoviruses. The percent of GFP+ (infected) cells was measured by flow cytometry and viral titers were calculated as transducing units per milliliter (TDU/mL). All tested pseudoviruses displayed similar levels of infection on cells expressing human or the ancestral CD4 proteins (**Fig. 2D**, 2E). This suggests that ancestral versions of CD4 in apes were susceptible to primate

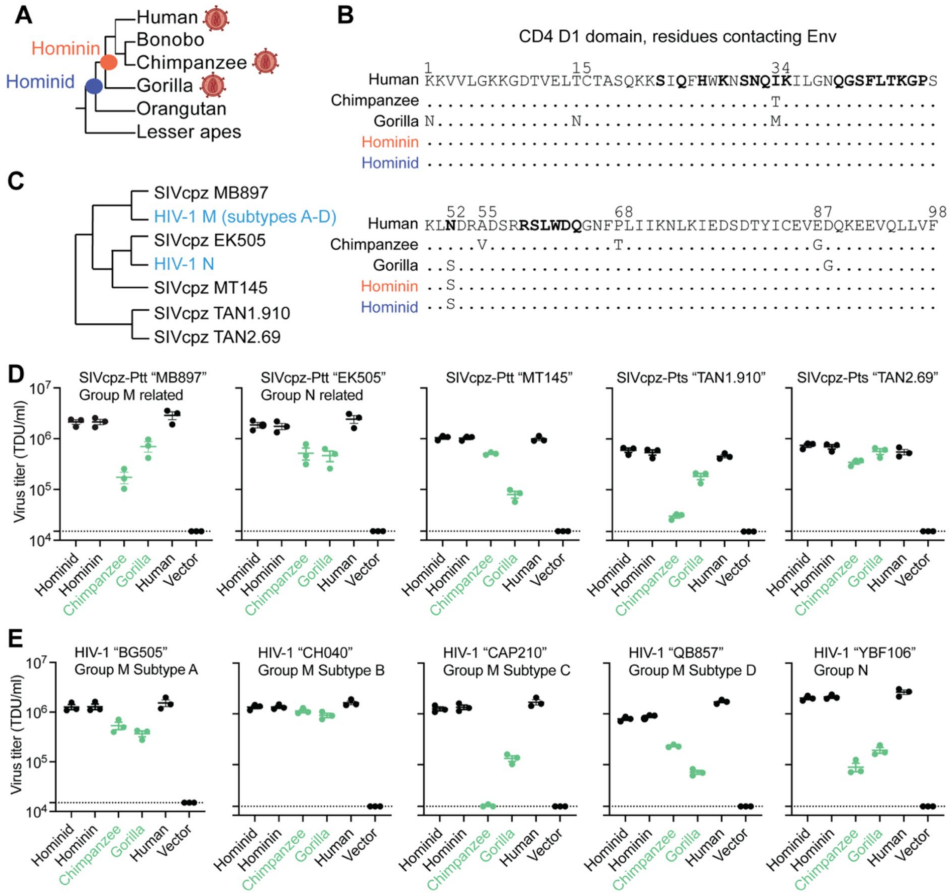

Figure 2.

Receptor mediated resistance to SIVcpz entry is a trait acquired in gorillas and chimpanzees.

(A) Cladogram of ape species, highlighting the nodes for which ancestral *CD4* sequences were reconstructed. The virion diagram next to some ape species represents apes that are infected by SIV/HIV. **(B)** An amino acid alignment of the CD4 D1 domain of human, chimpanzee, gorilla, and the inferred ancestral CD4 sequences. Dots represent identical residues compared to human and distinct amino acids and numerical positions are noted. Bolded residues on the human sequence represent sites known to directly interact with HIV-1 Envelope (Liu et al., 2017). **(C)** Cladogram of HIV-1 and SIVcpz based on previously published work (Takehisa et al., 2007), highlighting genetic relationships of the envelope (Env) clones used in this study. **(D, E)** HIV-1ΔEnv-GFP viruses were pseudotyped with Envs (top of graphs) from diverse **(D)** SIVcpz or **(E)** HIV-1 strains. Cf2Th cells stably expressing human CCR5 and various CD4s (X-axis) were infected with various volumes of these pseudoviruses and then analyzed by flow cytometry 48 hours post-infection. GFP positive cells were enumerated within the CD4/CCR5 positive gate and virus titers (transducing units per milliliter; TDU/mL) were determined for those samples falling within the linear infection range (n = 2 titration points). The mean virus titers obtained from each of three independent experiments were plotted (dots), with error bars representing the standard error of the mean (SEM). Dotted lines represent the lower limit of detection for this assay. SIVcpz-Ptt and SIVcpz-Pts refer to SIVs derived from the chimpanzee subspecies *Pan troglodytes troglodytes* or *Pan troglodytes schweinfurthii,* respectively.

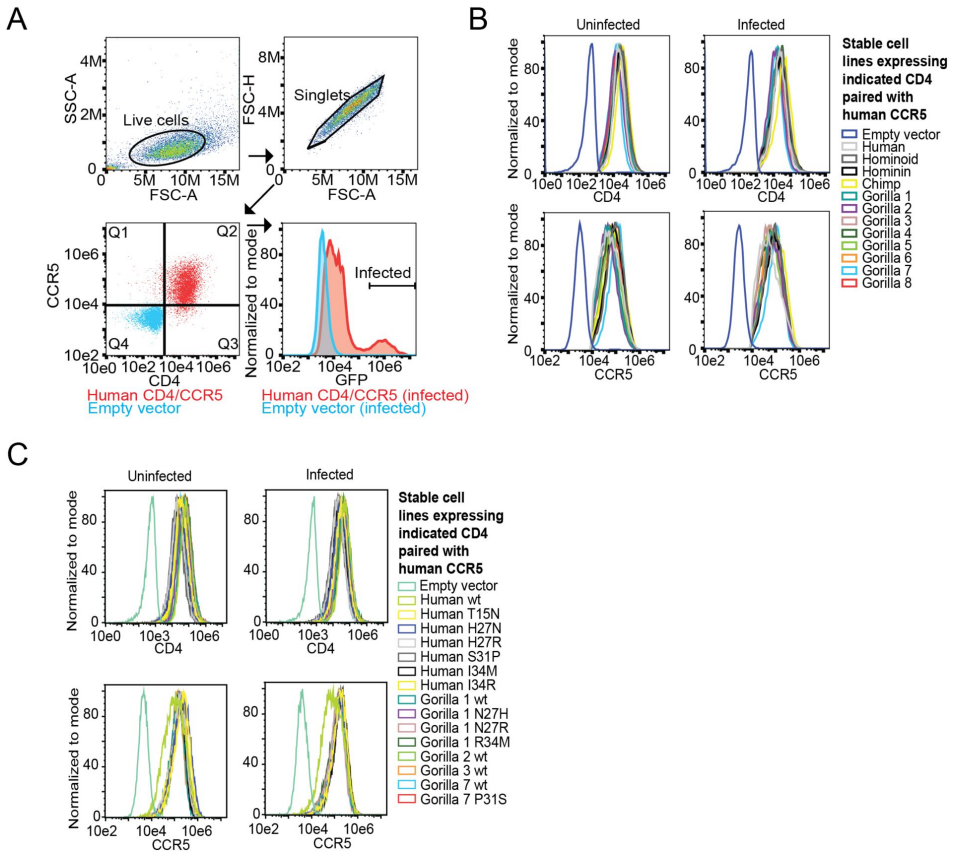

Figure S1.

Flow cytometry gating strategy.

(**A**) Collected events were selected for live cells and singlets based on forward and side scatter values. Singlets were gated for CD4 and CCR5 fluorescent signal and then the double-positive population (Q2) was further analyzed for viral infection based on a shift in GFP fluorescence compared to virus exposed cells lacking CD4/CCR5 receptors (empty vector transduced cells). (**B**) Expression levels for CD4 and CCR5 were compared amongst all stable cell lines under uninfected and infected conditions, demonstrating that viral infection does not impact receptor expression levels. For empty vector control cells, singlets were used for comparison. Data shown are representative of multiple independent experiments.

lentivirus entry, just as human CD4 is known to be today. On the other hand, cells bearing chimpanzee and gorilla CD4s were generally less permissive to virus entry (green in **Fig. 2D**, 2E). We conclude that CD4 was originally permissive to primate lentiviruses, but that selective pressures exerted by SIVs in the chimpanzee and gorilla lineages led to the retention of mutations that now confer resistance to primate lentivirus infection. This has not happened in humans where selective pressure by HIV-1 is relatively recent.

CD4 alleles circulating in gorilla populations differentially support entry of SIVcpz

Natural selection operates on individuals within populations, and only over time can the effects of this selection be seen in the divergence of gene orthologs between species. We and others have shown that multiple *CD4* alleles circulate in chimpanzee populations, which encode CD4 variants more resistant to SIVcpz entry than human CD4 (Bibollet-Ruche et al., 2019; Warren et al., 2019b). We next wanted to know if the same is true in gorilla populations. To similarly analyze gorilla *CD4*, we used the whole genome sequence data from the Great Ape Genome Project (Prado-Martinez et al., 2013) to identify extant *CD4* alleles. We analyzed genetic data from 32 gorillas (*Gorilla gorilla gorilla* [n = 28]; *Gorilla gorilla diehli* [n = 1]; *Gorilla beringei graueri* [n = 3]) and found six nonsynonymous and five synonymous SNPs separating the individual alleles encoded. Five out of six of the nonsynonymous polymorphisms are located within the domain 1-encoding region of *CD4* (two are in the same codon, codon 27) and one in domain 2 (**Fig. 3A**). A study of over 100 fecal samples from gorillas at field sites in Africa recently identified the same set of SNPs (Russell et al., 2021). In the definition of alleles considered here, we ignored synonymous polymorphisms. The nonsynonymous polymorphisms identified resulted in eight alleles that encode eight distinct CD4 protein variants. The frequencies of these eight alleles are heterogeneous, where allele 5 is the most common (**Fig. 3B**). (This, allele 5, was also the gorilla *CD4* that was tested in **Figure 2D**, E and shown in the alignment in **Figure 2B**.) From looking at the sequences of these different alleles, we noticed a predicted glycosylation site (N-glycosylation tripeptide NXT) at position 15 that is fixed in the gorilla population but absent in the other African apes (**Fig. 3A**). Interestingly, the gorilla *CD4* allele 2 codes for a proline at position 18, immediately after the tripeptide NCT, which strongly reduces the likelihood of glycosylation (Gavel and Heijne, 1990). Since five of the six protein-altering polymorphisms are in the region corresponding to domain 1 of the CD4 protein, which directly binds to the lentiviral Env (**Fig. 3C**), we next wanted to test their functional significance.

We made stable cell lines expressing each gorilla CD4 variant, along with human CCR5 (**Fig. S1**). We then infected each of these with GFP pseudoviruses displaying envelopes from different strains of SIVcpz, as described above. We quantified the number of GFP+ cells to measure viral entry through each CD4 variant. Again, we don't know the exact strain of SIVcpz that initially infected gorillas, so instead we have tested a phylogenetic diversity of SIVcpz strains. We found substantial differences in susceptibility to virus entry between the CD4 variants, varying by up to 2 orders of magnitude in some cases (**Figure 4**). All gorilla CD4 variants were equal to, or more resistant to infection than, the human CD4. We also tested pseudoviruses displaying a diverse set of Envs from HIV-1 groups M and N, and found similar patterns (**Figure S2**). These data are consistent with SIV exerting selective pressure on gorillas in favor of resistant alleles of *CD4*. However, as would be expected in a host-virus arms race, the viruses are evolving too. As such, we found considerable differences on the entry phenotype for each SIVcpz strain evaluated, where a single host CD4 variant can be highly restrictive to one strain, while being fully functional for entry of another. As an outlier, SIVcpz TAN2.69 showed a uniformly strong ability to use any of the gorilla CD4 variants.

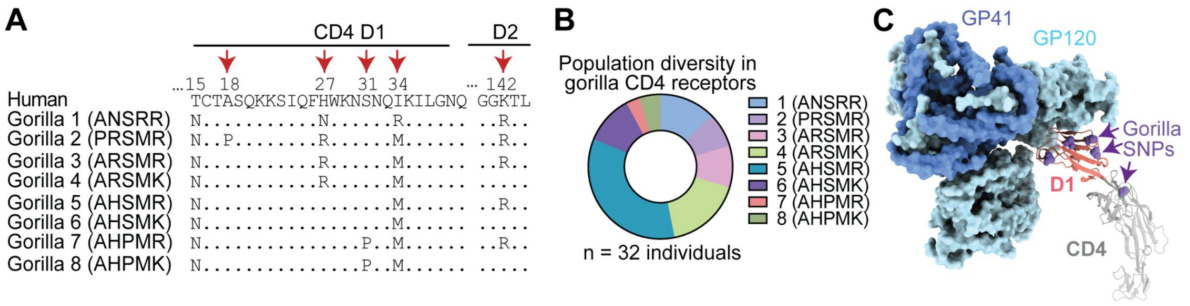

Figure 3.

Identification of diverse gorilla *CD4* alleles.

(A) Eight unique protein variants of gorilla CD4 were identified. The polymorphic sites (red arrows) are shown in the alignment, where dots indicate amino acid residues that are identical to human. **(B)** The frequencies of the eight alleles encoding unique CD4 protein variants are shown for three gorilla subspecies, *Gorilla gorilla gorilla* (n = 28), *Gorilla beringei graueri* (n = 3), and *Gorilla gorilla diehli* (n = 1). **(C)** Cryo-EM structure of an HIV-1 Env trimer in complex with human CD4 (PDB 5U1F) visualized in ChimeraX (Goddard et al., 2017). Individual gp120 and gp41 subunits are colored in light and dark blue, respectively. The CD4 D1 domain (red) and D2-D4 domains (gray) are shown, with gorilla nonsynonymous SNPs shown on the human sequence as purple spheres.

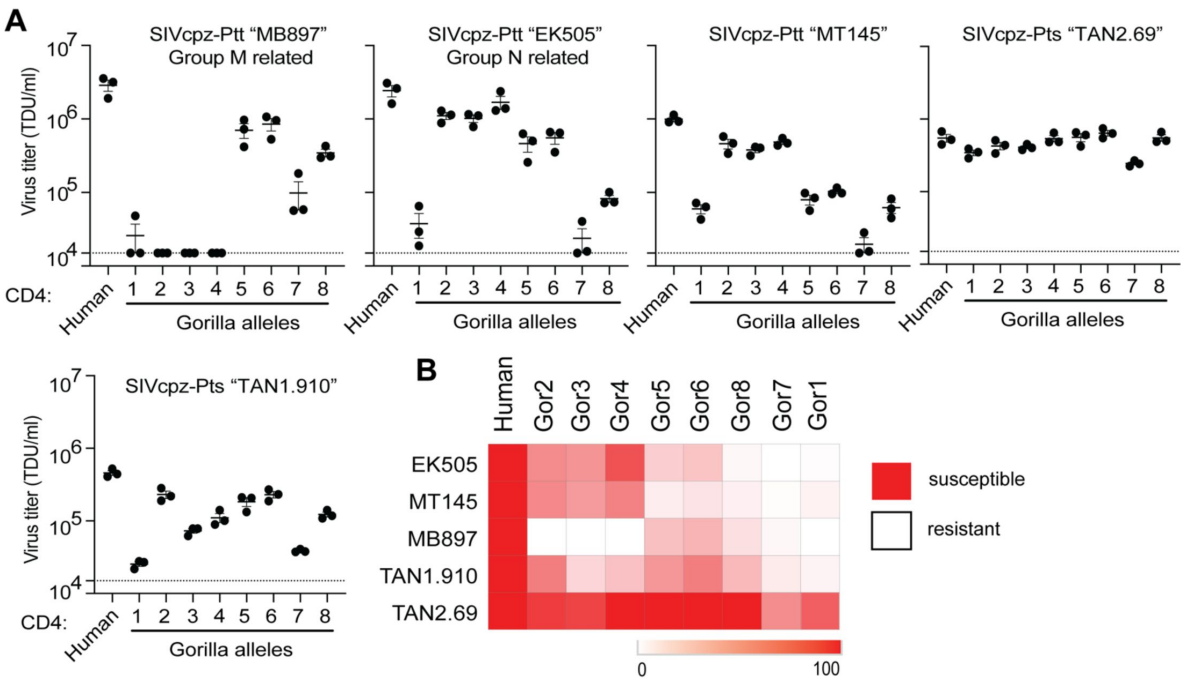

Figure 4.

Gorilla CD4 variants differentially support entry of SIVcpz.

(A) HIV-1ΔEnv-GFP viruses were pseudotyped with Envs (top of graphs) from diverse SIVcpz strains. Cf2Th cells stably expressing human CCR5 and various CD4s (X-axis) were infected with various volumes of these pseudoviruses and then analyzed by flow cytometry 48 hours post infection. GFP positive cells were enumerated and virus titers (transducing units per milliliter; TDU/mL) were determined for those samples falling within the linear infection range (n = 2 titration points). The mean virus titers obtained from each of three independent experiments were plotted (dots), with error bars representing the standard error of the mean (SEM). **(B)** Data from each pseudotyped Env in A were used to calculate virus titer means normalized to cells expressing human CD4 and were plotted as a heat map, where red and white represent susceptibility or resistance to viral entry, respectively.

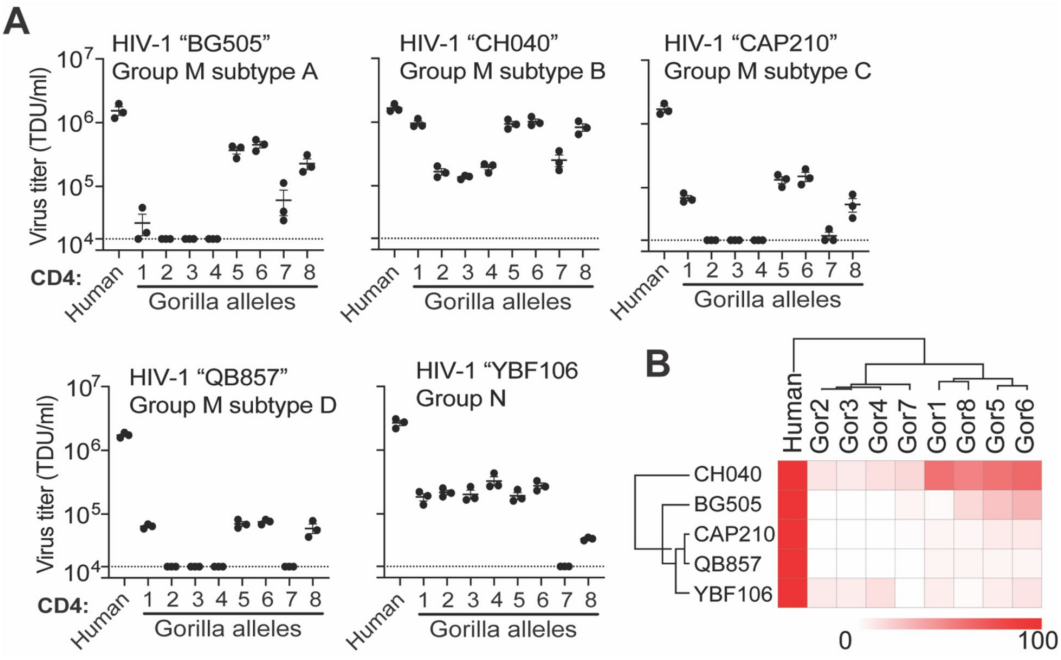

Figure S2.

Gorilla CD4 alleles differentially support entry of HIV-1.

(A) HIV-1ΔEnv-GFP viruses were pseudotyped with Envs (top of graphs) from globally diverse HIV strains. Cf2Th cells stably expressing human CCR5 and various CD4s (X-axis) were infected with various volumes of these pseudoviruses and then analyzed by flow cytometry 48 hours post infection. GFP positive cells were enumerated and virus titers (transducing units per milliliter; TDU/mL) were determined for those samples falling within the linear infection range (n = 2 titration points). The mean virus titers obtained from each of three independent experiments were plotted (dots), with error bars representing the standard error of the mean (SEM). **(B)** Data from each pseudotyped Env in "A" were used to calculate virus titer means normalized to human CD4 expressing cells and were plotted as a heat map. CD4 alleles and Envs were hierarchically clustered to depict similarities in phenotype.

Individual amino acid substitutions in gorilla CD4 protect against SIVcpz entry

We next sought to evaluate if CD4 polymorphisms found in gorilla individuals are protective when engineered into the human version of CD4, a widely susceptible receptor for primate lentiviruses. First, we investigated gorilla CD4 variant 2, which encodes a proline at position 18 that is predicted to prevent an otherwise fixed N-glycosylation at position 15 (**Fig. 3A**). We noticed that gorilla variant 2 CD4 is highly susceptible to most of the SIVcpz strains tested in this study (**Fig. 4**). Variant 3, which differs from variant 2 only by this proline, supported less entry by SIVcpz TAN1.910 due to this change in glycosylation status (**Fig. 5A**). To explore the effects of this gorilla specific glycan at residue 15, we generated cell lines stably expressing a mutated version of human *CD4* that encodes for the gorilla specific glycosylation motif. We then challenged these cells with pseudoviruses displaying the envelope of different SIVcpz strains and consistently found a decrease in susceptibility to entry compared to wild type human CD4 (**Fig. 5B**). To confirm the glycosylation status of CD4, we performed CD4 western blotting on lysates from cells stably expressing each of the different versions of CD4. As expected, human T15N CD4, as well as gorilla variant 3, migrated at a higher molecular weight compared to human wild type CD4 and gorilla variant 2, corresponding to the predicted number of glycans on CD4 domain 1 (**Fig. 5C**). We then treated the lysates with PNGase F, an N-linked glycosidase, and found that all CD4 variants migrated to the same molecular weight, confirming that the mobility shift is due to the glycosylation status of CD4. Thus, most gorilla CD4 variants have gained a glycan at position 15 that reduces entry of SIV as compared to human CD4. This phenotype is dependent on the lentiviral strain infecting. It seems that the gorilla *CD4* allele 2, which doesn't encode this glycan, would be at a fitness disadvantage. In support of this, allele 2 is one of the least frequent alleles in the gorilla population that we surveyed (**Figure 3B**).

We proceeded to test the amino acid residues at the other 3 polymorphic positions in domain 1 of gorilla CD4. We infected cells individually expressing mutant forms of human CD4 that coded for the gorilla specific residues at positions 27, 31, and 34. We found that in all cases, the mutant form of human CD4 encoding the gorilla specific amino acid was significantly more restrictive to at least two of the four SIVcpz strains when compared to human wild type CD4 (**Fig. 5D**). In several cases, such as H27R CD4 expressing cells infected with SIVcpz MB897, we found drastic effects where a fully supportive receptor was rendered highly refractory to infection by a single amino acid substitution. We found that the protective role of these gorilla specific substitutions was SIVcpz strain specific, demonstrating that collectively, the diversity found in gorilla individuals can confer relative protection to all the SIV strains we tested, but that SIVs are counter-evolving as well. These results suggest that single amino acid changes in domain 1 can drastically modify the interaction between CD4 and the lentivirus envelope, directly influencing virus entry.

To evaluate the reverse - if gorilla CD4 mutated to recapitulate the amino acids encoded in human CD4 may render the CD4 a better receptor for SIVcpz - we made and constructed cells expressing those CD4s and quantified the level of infection. We did not observe a full restoration of infection phenotype here and instead found only minimal increases in entry of SIVcpz through these receptors (**Fig. 5E**). These results imply that the resistance to SIVcpz found in gorilla individuals is not dependent on single amino acids, but rather the cumulative effect of multiple amino acid changes. Overall, our data suggest that population-level sequence diversity in *CD4* of gorillas confers some level of protection against multiple SIVcpz strains.

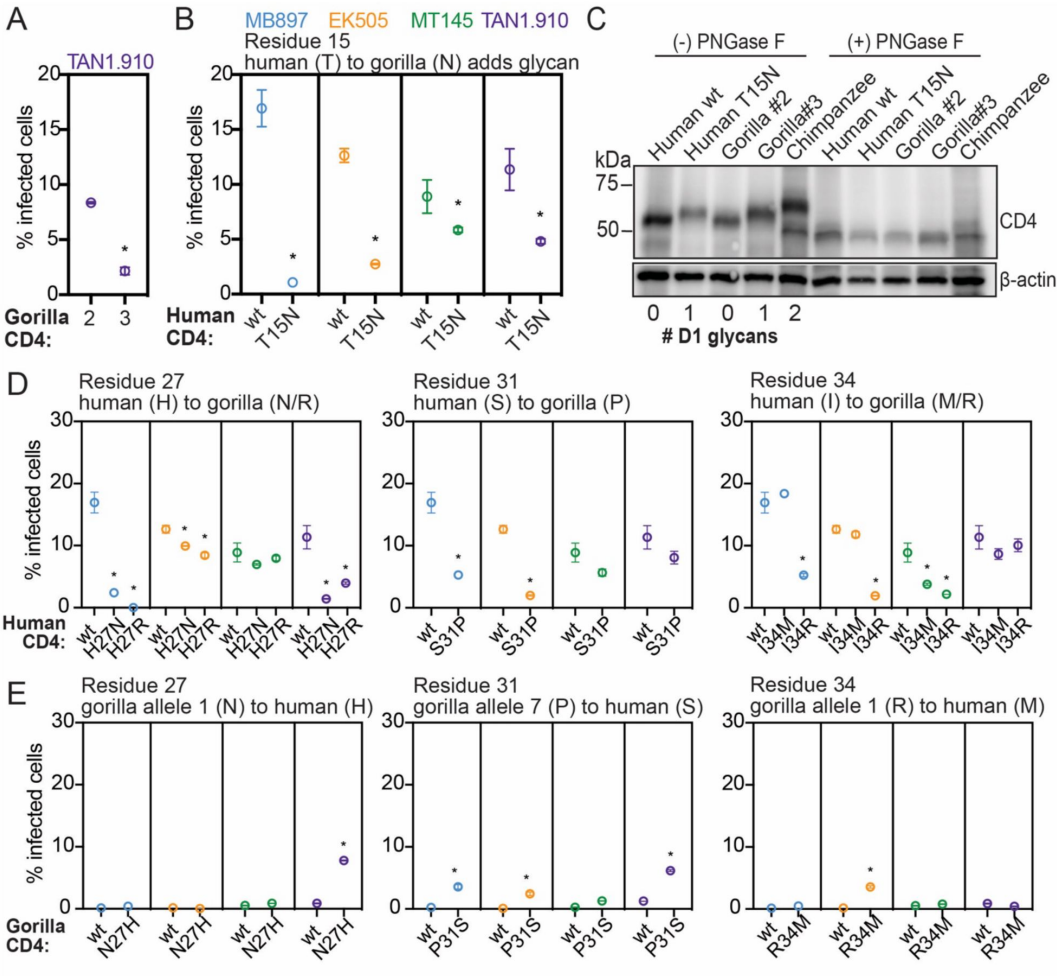

Figure 5.

Individual amino acid substitutions that have occurred in gorilla CD4 protect against SIVcpz entry.

(A,B,D,E) HIV-1ΔEnv-GFP viruses were pseudotyped with Envs from diverse SIVcpz isolates (MB897, blue; EK505, orange; MT145, green; TAN1.910, purple). Cf2Th cells stably expressing human CCR5 and wild-type (wt) or mutated human or gorilla CD4 (X-axis) were infected with these pseudoviruses and then the percent cells infected (GFP-positive cells) were enumerated by flow cytometry 48 hours post infection. Data represent the mean +/- SEM from two independent experiments, each with two technical replicates. Stars above data sets signify that both independent experiments showed significant statistical differences ($p < 0.05$) when compared to wild-type by one-way ANOVA. **(C)** Lysates of Cf2Th cells stably expressing the indicated CD4 receptors in "A" and "B" were treated with PNGase F (to remove N-specific glycans) or left untreated and then probed for CD4 expression by western blotting. The number of N-specific glycosylation sites within the D1-domain of CD4 was determined computationally (Gupta and Brunak, 2002) and is shown under the blot. β-Actin served as a loading control. Chimpanzee CD4, with two D1-domain glycans, has been described previously (Warren et al., 2019b).

Positive natural selection has shaped *CD4* polymorphism in SIV-endemic ape species

Natural selection influences the frequency of alleles within populations. Alleles with deleterious effects will be kept at low frequency by purifying selection. On the other hand, alleles that confer a selective advantage will reach higher frequencies and/or be maintained in a population longer than expected due to different forms of positive selection (i.e., selective sweeps or frequency dependent selection). We next tested how polymorphism in *CD4* has been shaped in ape species, with comparison made between apes that have been endemically infected with SIV (chimpanzees and gorillas) and those that have not (bonobos and orangutans).

Formal methods to detect the influence of selection on population-level nucleotide variation exist (Fay and Wu, 2000; Fu and Li, 1993; Tajima, 1989), but their statistical power is decreased in non-human ape species due to their small sample sizes and lower levels of variation (Prado-Martinez et al., 2013). Thus, we use a comparative approach to detect signatures of natural selection. To do this, we compared patterns of population level diversity in *CD4* versus its neighboring genes. We compared SIV endemic apes (chimpanzee and gorillas) to apes uninfected (bonobos and orangutans) or recently infected on an evolutionary timescale (humans). This was performed for both nonsynonymous (protein altering) and synonymous (not protein altering) polymorphisms. Sequences for chimpanzee, gorilla, bonobo, and two orangutan species were obtained from the Great Ape Genome Project (Prado-Martinez et al., 2013) for *CD4* and 11 neighboring genes spanning 250 kb of the X-chromosome. For these same genes, human variation was obtained from the 1000 Genomes project. We calculated nucleotide diversity either based on the number of single nucleotide polymorphisms (SNPs; Watterson's $\boldsymbol{\theta_\omega}$ (Watterson, 1975)) or mean pairwise difference between individuals ($\boldsymbol{\theta_\pi}$ (Tajima, 1983)). The mutation rate at *CD4* does not appear to be elevated given similar levels of variation between *CD4* and its neighboring loci when based on the number of SNPs ($\boldsymbol{\theta_\omega}$) (**Table S1** and **Fig. S3**).

However, within the endemically infected species, nonsynonymous SNPs in *CD4* are at a significantly higher frequency compared to neighboring loci, represented by $\boldsymbol{\theta_\pi}$ (**Fig. 6A**) (Tajima, 1983). This difference is not observed for synonymous variation. This discordance between nonsynonymous and synonymous variation suggests that the higher frequency of nonsynonymous variants at *CD4* in the endemically infected species is not explained by neutral or demographic evolutionary forces. In addition, the higher frequency of segregating nonsynonymous variation is restricted to the endemically infected species. Taken together, these patterns are consistent with positive selection increasing the frequency of and/or maintaining nonsynonymous SNPs in *CD4* within the endemically infected species only. Also, in support of this, we find that gorilla and chimpanzee nonsynonymous polymorphic sites are significantly concentrated on the domain 1 of CD4 when compared to the un/recently infected species (**Fig. 6B**). This difference is statistically significant (**Fig. 6C**). This data suggests that long-term endemic infection of SIV in ape populations may be driving nonsynonymous SNPs to higher frequency in *CD4*, particularly in the region corresponding to domain 1 that directly interacts with the virus Env glycoprotein.

Discussion

Pathogens are strong selective drivers of host gene evolution (A Demogines et al., 2012) (Meyerson and Sawyer, 2011; Warren and Sawyer, 2023, 2019). We and others have previously shown that the *CD4* gene has evolved under strong positive selection throughout the evolution and speciation of simian primates (Meyerson et al., 2014; Zhang et al., 2008). Selection on *CD4* is thought to be driven by the direct interaction between CD4 and the HIV/SIV envelope glycoprotein. Indeed, most of the sequence evolution in CD4 has occurred in the D1

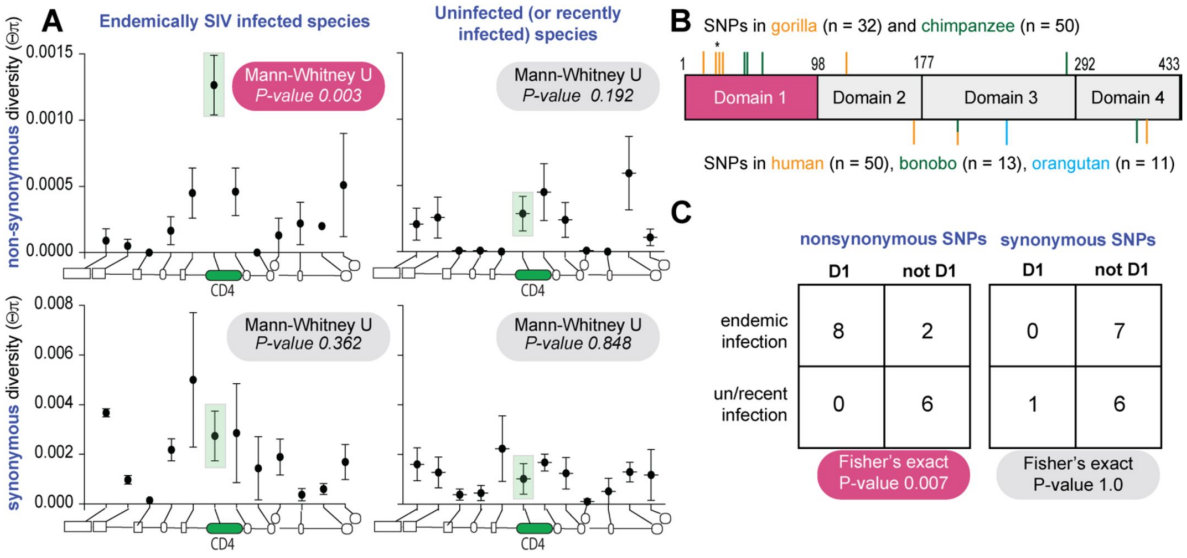

Figure 6.

Positive natural selection has shaped *CD4* polymorphism in SIV-endemic ape species.

(**A**) Mean and standard error of mean of synonymous and nonsynonymous nucleotide heterozygosity (Θπ) at *CD4* and neighboring loci across species endemically infected with SIV (chimpanzee and gorilla) or un/recently infected (human, bonobo and orangutans). Schematic along the bottom of each graph depicts the relative location of each locus as follows 5' to 3': *ZNF384, PIANP, COPS7A, MLF2, PTMS, CD4, GPR162, GNB3, CDCA3, TPI1, LRRC23,* and *ENO2*. Mann-Whitney test indicates whether heterozygosity at *CD4* is significantly different than neighboring loci. (**B**) Schematic of CD4 domain regions. Ticks above and below the CD4 box indicate the location of polymorphic sites for the infected and un/recently infected species groups, respectively. One of the polymorphic residues in gorilla contains two nonsynonymous changes in a single codon, marked by a star above the tick. (**C**) 2x2 contingency table and test results comparing synonymous and nonsynonymous polymorphism location relative to domain 1 between infected and un/recently infected species.

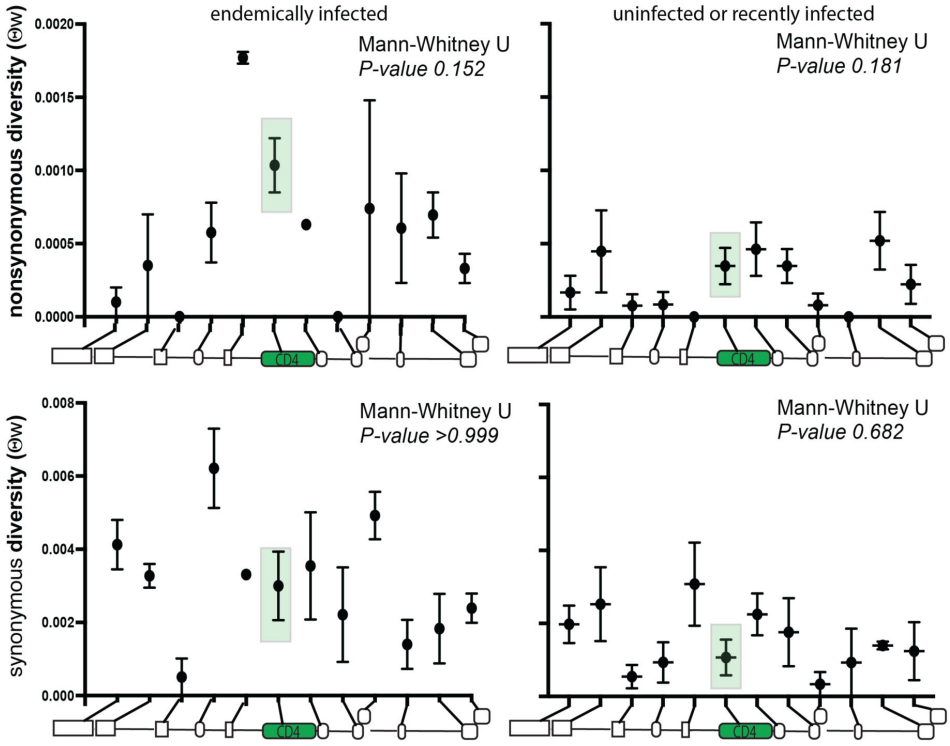

Figure S3.

Single nucleotide polymorphisms in ape species.

Population nucleotide diversity at a locus is estimated either based on the number of single nucleotide polymorphisms (SNPs; Watterson's $\boldsymbol{\theta}_{\omega}$ (Watterson, 1975)) or mean pairwise difference between individuals ($\boldsymbol{\theta}_{\pi}$ (Tajima, 1983)). The level of variability based on the number of single nucleotide polymorphisms at a locus ($\boldsymbol{\theta}_{\pi}$) is not significantly different between CD4 and neighboring loci. Y-axis shows the mean and standard error of $\boldsymbol{\theta}_{\pi}$ for synonymous and nonsynonymous nucleotide variants at CD4 and neighboring loci across species endemically infected with immunodeficiency viruses (chimpanzee and gorilla) or recently/uninfected (human, bonobo and orangutans). Schematic along bottom of each graph depicts the relative location of each locus and are as follows 5' to 3': *ZNF384*, *PIANP*, *COPS7A*, *MLF2*, *PTMS*, *CD4*, *GPR162*, *GNB3*, *CDCA3*, *TPI1*, *LRRC23,* and *ENO2*. Mann-Whitney test indicates whether heterozygosity at *CD4* is significantly different than neighboring loci. We observe no difference in the total number of nonsynonymous and synonymous SNPs, represented by $\boldsymbol{\theta}_{\omega}$ between *CD4* and its genomic neighbors in the endemic and un/recently infected species (see also **Table S1**).

domain that contacts Env (Meyerson et al., 2014; Zhang et al., 2008). We performed an updated analysis of positive selection in *CD4* including new *CD4* orthologs that have become available (**Figure S4**) (McBee et al., 2015). We found that removing the D1 sequence from the analysis renders the gene no longer under positive selection. This sets the stage for the current study, which focused on selection on *CD4* within ape populations.

Within populations of animals, when alleles of *CD4* arise that can resist SIV, they would be predicted to rise in frequency. We and others have demonstrated that many *CD4* alleles circulating in chimpanzees convey an increased ability to restrict viral entry by SIVs (Bibollet-Ruche et al., 2019; Warren et al., 2019b). This is also observed for many other African primate species, where amino acid polymorphisms in CD4 resist viral entry (Russell et al., 2021). It is important to note that most African primate species currently harbor lentiviruses endemically, despite *CD4* evolution, meaning that CD4 remains functional for viral entry of at least some viruses despite the selective pressure to resist it. Therefore, we understand *CD4* to be evolving to convey natural tolerance in primates. “Natural tolerance” refers to a species’ ability to resist or tolerate a virus to an acceptable level for peaceful co-existence of the virus and host. It is obtained by evolutionary adaptations that occur over time, allowing the species to develop mechanisms to reduce the negative effects of the virus (Pagán and García-Arenal, 2018). For example, some species may evolve barriers (like resistant forms of CD4) that reduce the titers that a virus can achieve in their body. Studying natural tolerance is key to understanding virus reservoirs in nature.

Herein, we strengthen the insight into lentiviral tolerance via CD4 evolution in three ways. First, we reconstructed extinct ancestral forms of ape CD4 that pre-date SIV, and find that they were highly vulnerable to SIV entry. We then show that CD4 became less permissive to SIV in species that experienced long-term endemic infection. This resistant phenotype is associated with the accumulation of specific amino acid substitutions in the D1 domain of CD4. Second, we show that gorillas harbor a diversity of *CD4* alleles, all of which encode protein variants that are more resistant to SIV entry than is human CD4. Again, we demonstrate that these protein variants are gaining resistance by accumulating amino acid substitutions in the D1 domain, one of which creates a new motif for post-translational addition of a glycan to the CD4 protein. Protective (to the host) glycosylation of CD4 has recently also been observed by us and others in chimpanzees (Bibollet-Ruche et al., 2019; Warren et al., 2019b), and in another population sample of gorillas (Russell et al., 2021). Indeed, the evolutionary acquisition of a glycan shield on CD4 may be a recurring theme in the evolution of primate species that are plagued with SIVs (Russell et al., 2021). Lastly, using population genetics analyses, we show that nonsynonymous SNPs are enriched within ape species that are endemically infected with SIV (chimpanzees and gorillas) relative to those that are not (bonobos and orangutans) or which have been infected for less than 100 years (humans). This increased population level diversity is observed only for *CD4*, and not shared by other genes neighboring the *CD4* loci.

Collectively, it is now clear that the sequence diversity (within species) and divergence (between species) of several primate genes have been driven by infection pressure from lentiviruses (Daugherty and Malik, 2012; Demogines et al., 2010; Han et al., 2011; Johnson and Sawyer, 2009; Judd et al., 2021; Lee et al., 2012; Malfavon-Borja et al., 2013; Meyerson et al., 2018, 2015, 2014; Meyerson and Sawyer, 2011; Sawyer et al., 2007, 2006, 2005, 2004; Sawyer and Elde, 2012; Stabell et al., 2016; Warren et al., 2024, 2019b, 2019a; Warren and Sawyer, 2019; White et al., 2014; Wilke and Sawyer, 2016). The outcome of this virus-driven host evolution is ultimately detrimental to the very viruses that drove this evolution. When host genes like *CD4* become highly diverse within species, a given virus strain may only be able to infect a small number of individuals within the population. For instance, gorilla CD4 variant 1 is highly resistant to most of the SIVs we tested (**Fig. 4**). In fact, we only found one SIV isolate, SIVcpz “TAN2.69,” that could enter cells through the receptor encoded by allele 1. That

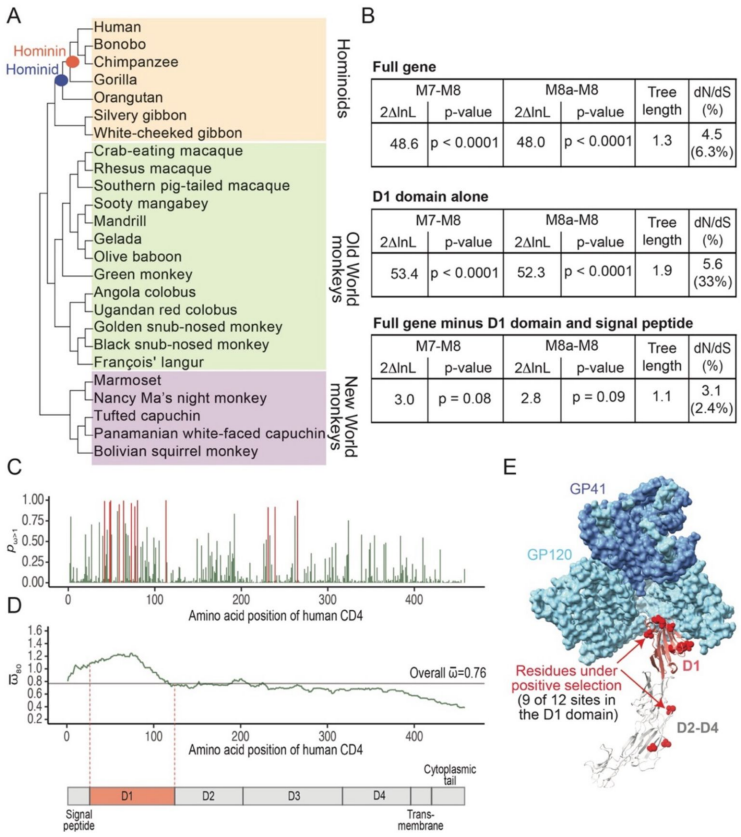

Figure S4.

CD4 is under positive selection in primates.

Previous studies have identified CD4 as an HIV-1 cofactor that is evolving under positive (diversifying) selection (Meyerson et al., 2014; Zhang et al., 2008). However, these studies were limited in that they analyzed *CD4* sequences from a narrow set of primate species (Zhang et al., 2008), or included a larger species panel but lacked the complete *CD4* coding sequence (Meyerson et al., 2014). To extend on these studies, we collected full-length *CD4* sequences from 25 primate species and tested for evidence of site-specific selective pressures using the *codeml* program on the Phylogenetic Analysis by Maximum Likelihood (PAML) package (Yang, 2007a). **(A)** Cladogram of the primate species (n = 25) analyzed in this study. **(B)** The most amino-terminal extracellular domain of CD4 (domain 1, D1) is bound by the primate lentivirus (HIV/SIV) envelope glycoprotein (Env) during entry (Bour et al., 1995). We next sought to assess whether D1 alone is evolving under positive selection (presumably due to selective pressures exerted by SIVs), or if other regions of CD4 are also experiencing selective pressures for diversification. Site-specific selective pressures in primate CD4 (full gene; top), the CD4 D1 domain alone (amino acids 26-123; middle), and CD4 minus the signal peptide and the D1 domain (amino acids 123-458; bottom) were detected using PAML (Yang, 2007a). Positive selection among amino acid sites was tested using two model comparisons, M7 vs. M8 and M8a vs. M8. In each of these comparisons, the null models (M7, M8a) do not allow for sites under positive selection, while the alternative model (M8) does. Tables summarize the likelihood ratio test between the M7-M8 and M8a-M8 models. The 2ΔlnL value (twice the difference in the natural log of the likelihoods) is shown, along with the p-value with which the neutral models (M7 or M8a) are rejected in favor of the model of positive selection (M8). **(C)** To further identify codon sites in CD4 under positive selection, we calculated the posterior probability of ω > 1 (where ω is the dN [nonsynonymous]/dS [synonymous] rate ratio, and values > 1 in the model M8 indicate sites under selection) using the Bayes empirical Bayes approach. Plot of posterior probabilities (ω>1 under maximum likelihood random-sites model M8) for all CD4 sites. Sites under positive selection ($p_\omega > 0.9$) are shown in red. (**D**) The posterior mean of ω over a sliding window of 80 amino acids is shown (green line), along with the overall mean of ω across the entire gene (grey line). In both panels C and D, the amino acid positions are shown in relationship to human CD4, and the D1 domain of CD4 is highlighted in orange. **(E)** Cryo-EM structure of an HIV-1 Env trimer in complex with human CD4 (PDB 5U1F) was visualized in ChimeraX (Goddard et al., 2017). Individual gp120 and gp41 subunits are colored in light and dark blue, respectively. The CD4 D1 domain (red) and D2-D4 domains (gray) are shown, with sites under positive selection ($P_\omega > 0.9$) shown on the human sequence as red spheres. 9 of the 12 sites passing this stringent cutoff map to the Env-CD4 D1 domain interface.

suggests that gorillas that are homozygous for allele 1 would largely be protected from most circulating SIV strains. Taking this example further, if allele 1 were to become fixed within gorilla populations, many strains of SIV in gorillas could go extinct.

In the long-term, the virus-driven evolution of genes like *CD4* also means that virus spillover between species – including the zoonotic spillovers that yield new human viruses - are less likely to happen. Indeed, a prevailing theme that has emerged in recent years is that receptor sequence divergence serves as a potent barrier to the movement of viruses between species (Barbachano-Guerrero et al., 2023; Ann Demogines et al., 2012; Demogines et al., 2013; Hu et al., 2022; Kaelber et al., 2012; Kerr et al., 2015; Makin, 2022; Meyerson et al., 2014; Nahabedian et al., 2017; Ng et al., 2015; Sawyer and Elde, 2012; Warren et al., 2022, 2019a; Warren and Sawyer, 2023, 2019). Specifically, CD4 and another host protein RanBP2 (Meyerson et al., 2018) played key roles in the transmission of SIVs to apes (see also (Sauter and Kirchhoff, 2019). Further, this study suggests that SIV entry is blocked by the CD4 receptor of some primate individuals that it might encounter. Therefore, spillover of lentiviruses between species will only happen when virus is transmitted between *key individuals* of two different species. The donor individual would need to have *CD4* alleles that yield high titers of SIV in its body, and the recipient individual would need to have *CD4* alleles that make it receptive to infection by this new virus.

Materials and Methods

Ancestral reconstruction of the CD4 sequence at the base of the hominin and hominid clades

The ancestral state of CD4 was determined using the PAML software package as previously described (Yang, 2007a, 2007b; Yang et al., 1995). As input, we used an alignment of CD4 sequences from the following species: human (*Homo* sapiens; NM_000616.4), common chimpanzee (*Pan troglodytes*; NM_001009043.1), western lowland gorilla (*Gorilla gorilla gorilla*; XM_004052582.2), bonobo (*Pan paniscus*; XM_008973678.1), northern white-cheeked gibbon (*Nomascus leucogenys*; XM_004092147.1), Sumatran orangutan (*Pongo abelii*; XM_024256502.1), rhesus monkey (*Macaca mulatta*; NM_001042662.1), grivet (*Chlorocebus sabaeus*; XM_007967413.1), sooty mangabey (*Cercocebus atys*; NM_001319342.1), pig-tailed macaque (NM_001305921.1), crab-eating macaque (*Macaca nemestrena*; XM_005569956.2), gelada (*Theropithecus gelada*; XM_025401282.1), black snub-nosed monkey (*Rhinopithecus bieti*; XM_017891844.1), drill (*Mandrillus leucophaeus*; XM_011982990.1), Angolan colobus (*Colobus angolensis palliatus*; XM_011952091.1), golden snub-nosed monkey (*Rhinopithecus roxellana*; XM_010385914.1), and olive baboon (*Papio anubis*; XM_003905871.3).

Genotype and allele determination of CD4 from gorillas

Short-read data available through the National Center for Biotechnology Information's (NCBI) Short Read Archive (BioProject PRJNA189439) were mapped onto the *G. gorilla* genome using BWA-MEM (Li, 2013). We applied GATK base quality score recalibration, indel realignment, duplicate removal, and SNP discovery and genotyping in each individual separately (McKenna et al., 2010). Joint genotyping and variant recalibration was performed in a species-specific manner and in accordance to the GATK best practices recommendations (Auwera et al., 2013; DePristo et al., 2011). Variant recalibration was performed using SNPs called by the neighbor quality score method of ssahaSNP on capillary sequencing runs from NCBI's Trace Read Archive (Ning et al., 2001), dbSNP (if available), and high-quality SNPs called on the hg18 genome lifted over to the assembly used for mapping (Prado-Martinez et al., 2013). Processing was performed using custom scrips written in Python. Nucleotide sequence data reported are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK063765-BK063795.

Receptor expression constructs and site directed mutagenesis

Human (Genbank ID# MK170450) and chimpanzee (Genbank ID# NM_001009043.1) CD4 expression plasmids were constructed in a previous study (Warren et al., 2019a). The chimpanzee CD4 allele tested here is "allele 6" as defined by us previously (Warren et al., 2019b), and has 2 glycans that impede virus binding to the receptor. Gorilla CD4 alleles and ancestral CD4s were commercially synthesized (IDT GeneBlocks) and gateway cloned into the pLPCX retroviral packaging vector (Clontech). Mutant versions of human and gorilla CD4 were constructed by standard site-directed mutagenesis methods using overlapping PCR primers encoding the modification. Both wild-type and mutant CD4 constructs were analyzed by Sanger sequencing prior to use.

Generation of stable cell lines expressing CD4

HEK293T cells (ATCC CRL-11268) were cultured in DMEM (Invitrogen) with 10% FBS, 2 mM L-glutamine, and 1X penicillin-streptomycin (complete medium) at 37 °C and 5% CO2. Cf2Th (ATCC CRL-1430) cells stably expressing human CCR5 (from (Warren et al., 2019a)) were cultured in complete medium supplemented with 250 µg/mL hygromycin. To produce retroviruses for transduction, HEK293T cells plated in antibiotic free media ($1x10^6$ cells per well in a six well plate) were transfected with 2 µg of pLPCX transfer vector containing the *CD4* gene of interest (or empty vector), 1 µg of pCS2-mGP (MLV gag/pol), and 0.2 µg of pC-VSV-G (VSV-G envelope) using a 3:1 ratio of TransIT-293 (Mirus) transfection reagent to DNA according to the manufacturer's instructions. Forty-eight hours post transfection, supernatant was collected, filtered through 0.22 µm cellulose acetate filters, and retrovirus stored at -80 °C in single-use aliquots. Cf2Th cells stably expressing human CCR5 were plated at $2x10^4$ cells per well of a 12-well dish (15% confluent) and 24-h later, transduced with 500 µL of retroviral supernatant by spinoculation at 1,200 xg for 75 min in the presence of 5 µg/mL polybrene. Forty-eight hours post transduction, the cells were placed in complete medium containing selection antibiotics (250 µg/mL hygromycin and 3 µg/mL puromycin) and cultured until stable outgrowth was noted (>1 week). Stable cell lines were maintained indefinitely in selection media. To confirm expression of CD4, cells were analyzed by flow cytometry (**Fig. S1**). Briefly, cells were harvested from culture plates, washed two times with PBS, fixed in 2% paraformaldehyde, and washed 2 times in flow buffer (1X PBS, 2% FBS, 1mM EDTA). Fixed cells were stained for 30 min at 4° C with PerCP-Cy5.5 mouse anti-human CD195 (CCR5, BD Biosciences 560635) and AlexaFluor647 mouse anti-human CD4 (BD Biosciences, 566681), and analyzed using a BD Accuri C6 Plus flow cytometer (BD Biosciences).

HIV/SIV Envelope clones used in this study

Envelope clones for HIV-1 and SIVcpz EK505 and MB897 were constructed in a previous study (Warren et al., 2019b). SIVcpz MT145, TAN1.910, and TAN2.69 molecular clones were a gift from Brandon Keele (Frederick National Laboratory for Cancer Research, Frederick, MD) and used as template for PCR amplification. The RevEnv cassettes of SIVcpz were amplified by PCR using the following primer pairs, where the lowercase sequence corresponds to an added Kozak sequence for enhanced translation: MT145 (JN835462) forward 5'-tcgccaccATGGCAGGAAGAAGCGAGGGAGACG-3', reverse 5'-TTAAAGCAAAGCTCTTTCTAAGCCTTGT-3'; TAN1.910 (AF447763.1) forward 5'-tcgccaccATGGCAGGAAGAGAAGAGGACGC-3', reverse 5'-TTAATTTAAGGCTAGTTCCAGACCC-3'; TAN2.69 (DQ374657.1) forward 5'-tcgccaccATGGCAGGAAGAGAAGAGGACGC-3', reverse 5'-TTAATTTAAGGCTATTTCTAGACCCTGT-3'. PCR products were cloned into the pCR8/GW/TOPO TA plasmid (Thermo Fisher) and then shuttled into a Gateway-converted pCDNA3.1 mammalian expression vector (Invitrogen).

Single-cycle HIV and SIV pseudovirus infections

To produce HIV-1ΔEnv-eGFP reporter viruses, $13x10^6$ HEK293T cells were seeded into a 15-cm dish in antibiotic free media and 24 h later transfected with 13.25 µg of Q23ΔEnv-GFP (group M backbone; (Humes and Overbaugh, 2011)) and 6.75 µg of envelope plasmid. Forty-eight hours post transfection, the cell supernatant was harvested, concentrated (~100-fold) using Amicon Ultracel 100K filters (Millipore), and stored at -80 °C in single use aliquots. Cf2Th cells stably expressing CD4 and CCR5 were plated at $3x10^4$ cells/well of a 48-well plate 24 h before infection. The cells (~80% confluent) were then infected with HIV-1 pseudoviruses in three different volumes (**Fig. 2** and **Fig. 4**), or a volume corresponding to 10-20% infection of cells expressing human CD4 (**Fig. 5**). Infections were carried out by spinoculation at 1,200 xg for 75 min in the presence of 5 µg/mL of polybrene. Forty-eight hours post infection, the cells were harvested from the plate and fixed in 2% paraformaldehyde. Fixed cells were washed three times with PBS and resuspended in 50 µL flow buffer (PBS buffer containing 2% FBS and 1 mM EDTA) and stained for 30 min at 4 °C with the following antibody mixture: PerCP-Cy5.5 mouse anti-human CD195 (CCR5, BD Biosciences 560635) and AlexaFluor647 mouse anti-human CD4 (BD Biosciences, 566681), and analyzed using a BD Accuri C6 Plus flow cytometer (BD Biosciences). Following singlet cell discrimination, gates were drawn to capture double-positive cells expressing CD4 and CCR5, and then the percent GFP+ cells was enumerated within that population. The data from ~$2x10^4$ cells per technical replicate were analyzed using FlowJo v10. To calculate virus titers (**Fig. 2** and **Fig. 4**), the linear range of the infectivity curve was determined, and two points within the linear range were selected to calculate the mean virus titer in TDU/mL. The limit of detection for the titer calculation corresponds to a value of 0.2 % GFP positive cells. TDU/mL mean values were normalized to the titer of infection in cells expressing human CD4, and data used to construct a heat map using the Morpheus server (*https://software.broadinstitute.org/morpheus*); rows and columns were hierarchically clustered by Euclidian distance.

Statistical comparisons were performed between percentages of infected cells in some cases. Values of technical replicates of each biological replicate were compared between mutant and wild type CD4 versions by one-way ANOVA. If a statistically significant difference was found ($p < 0.05$) in both independent biological replicates, an asterisk was added to the mutant column in the dot plot.

Glycosylation state of CD4 by western blotting

Cf2Th cells stably expressing CD4 cells were lysed in Nonidet P-40 buffer [150 mM NaCl, 50 mM Tris·HCl pH 7.4, 1% Nonidet P-40 substitute, 1 mM DTT, 1 µL/mL Benzonase (Sigma-Aldrich #E1014), and protease inhibitor mixture (Sigma-Aldrich, #11873580001)] by resuspending the cell pellet and rocking at 4° C for 30 min. Cell lysate was cleared by centrifugation at maximum speed for 15 min. Whole-cell extracts were quantified using the BCA assay and 10 µg was subjected to PNGase F (New England Biolabs, #P0705S) treatment according to the manufacturer's protocol, including a paired sample with no glycosidase as control. Treated whole cell extracts (5 µg per lane) were resolved on a 12% TGX Stain-free polyacrylamide gel (Bio-Rad, #1610185) by applying 180V until loading dye ran off the gel. Protein was transferred to a PVDF membrane (MilliporeSigma, #IPVH07850) using a wet transfer apparatus set at 100V for 60 min. The membrane was incubated with blocking buffer (tris-buffered saline 1X, Tween-20 0.1%, 5% nonfat dried milk) for 60 min at room temperature. Primary antibodies were diluted in blocking buffer and incubated with the membrane overnight at 4° C (1:1,000 anti-CD4, Abcam #ab133616). After primary antibody incubation, the membrane was washed 4 x 5 min in TBST (0.1% Tween-20). Secondary antibodies were diluted in blocking buffer and incubated with the membrane for 60 min at room temperature (1:10,000 anti-rabbit-HRP, Promega #W401B). After secondary antibody incubation, the membrane was washed 4 x 5 min in TBST (0.1% Tween-20), developed using ECL

reagent (Sigma-Aldrich, #GERPN2232), and imaged on a Bio-Rad ChemiDoc Imaging System. As a loading control, membranes were reblotted to detect β-actin expression using primary (Cell Signal #3700) and secondary-HRP (Promega #W402B) antibodies and developed as described.

Analysis of population-level selection acting on CD4

To compare the pattern of molecular evolution at CD4 relative to neighboring loci we pulled population level re-sequencing data for loci located within 100 kb downstream and upstream of CD4. Primate sequences were obtained from the Great Ape Genome project (Prado-Martinez et al., 2013) and size-matched (n=50, matched with the number of chimpanzee sequences) human sequences were randomly selected to represent diverse ethnic groups from the Human 1000 Genomes Project, selecting 10 individuals for each of the 5 superpopulations [Africans (AFR), Admixed Americans (AMR), East Asians (EAS), Europeans (EUR) and South Asians (SAS)].

To identify the individual-specific SNPs within the selected loci, genotype data in variant call format (VCF) was directly downloaded from International Genome Sample Resources (internationalgenome.org/) and the Great Ape Genome Project (biologiaevolutiva.org/greatape/). For human variants, the variant calls were made based on human reference genome annotation hg38, and individual-specific haplotypes were extracted by altering the reference sequence with the alternative SNPs annotated in the VCF files via a Perl script (github.com/santiagosnchez/vcf2fasta). For non-human primate variants, the short read genome sequences were mapped to human reference genome hg19 to generate the VCF files containing species- and population-level variants, as previously described(Prado-Martinez et al., 2013). The SNPs in the VCF files were further filtered by the variant call quality (GQ ≥ 15). Like the human sequences, the individual-specific haplotype sequences are re-constructed by correcting the reference sequence with VCF annotations. In total we obtained population level variation for CD4 plus 15 other loci (six upstream and nine downstream). Four loci were removed from analysis because they have previously been shown to directly interact with a viral protein (*USP5* and *SPSB2*; (Jia et al., 2020; Rathore et al., 2020; Wang et al., 2019; Zhang et al., 2021)) or non-human primate sequencing reads did not map well with the human reference due to repetitive sequence (*LAG3* and *P3H3*). Coding loci included in this study (in order 5' to 3') are: *ZNF384*, *PIANP*, *COPS7A*, *MLF2*, *PTMS*, *CD4*, *GPR162*, *GNB3*, *CDCA3*, *TPI1*, *LRRC23*, and *ENO2*. This was done for great ape species endemically infected with immunodeficiency viruses (chimpanzee and gorilla) and those newly or not infected (human, bonobo, Sumatran and Bornean orangutans).

Sequences were aligned for each species individually using the Muscle alignment program (Edgar, 2004). DnaSP (Rozas et al., 2017) was used to haplotype-phase the downloaded sequences and to calculate levels of nucleotide diversity for each locus. Rarely we would observe an internal stop codon within a locus' reading frame. In these cases, both haplotypes for that individual were removed from analysis. We analyzed the subspecies of gorilla and chimpanzee together. While there is evidence of genetic differentiation between these subspecies (Prado-Martinez et al., 2013), this should not affect our comparisons as the differentiation is expected to be similar across all loci.

Analysis of positive selection of CD4 in primates

Sequence alignments

CD4 sequences were aligned to the longest human isoform in MEGA X for macOS (Stecher et al., 2020) using the ClustalW alignment tool. Multiple sequence alignments were visually inspected, duplicate gene sequences were removed, and the gene isoform from each species that best aligned to the human reference was retained for further analysis. The terminal stop codon was removed and aligned DNA and protein sequences were exported as fasta files. Codon alignments were generated using PAL2NAL (Suyama et al., 2006). Species cladograms for use in PAML were

constructed following the species-level phylogenetic relatedness of primates (Perelman et al., 2011). Cladograms were generated using Newick formatted files and viewed with Njplot version 2.3.

Evolutionary analysis

Codon alignments and unrooted species cladograms were used as input files for analysis of positive selection using the PAML4.8 software package (Yang, 2007a). To detect selection, multiple sequence alignments were fit to the NSites models M7 (neutral model, codon values of dN/dS fit to a beta distribution bounded between 0 and 1), M8a (neutral model, similar to M7 but with an extra codon class fixed at dN/dS = 1) and M8 (positive selection model, similar to M8a but with the extra codon class allowed to have a dN/dS > 1). A likelihood ratio test was performed to assess whether the model of positive selection (M8) yielded a significantly better fit to the data compared to null models (model comparisons M7 vs. M8 and M8a vs M8). Posterior probabilities (Bayes Empirical Bayes analysis) were assigned to individual codons with dN/dS values > 1. To calculate the posterior mean of ω over a sliding window, the per-site ω value was extracted from the M8 model, and the average ω value within the designated window size (80 amino acids) was calculated across the open reading frame in a sliding manner. With the window slide 1 amino acid each time to calculate the smoothed mean ω values.

Acknowledgements

This work was funded by the NIH (DP1-DO-33547, DP1-DA-046108, R01-AI-137011, R01-OD-034046 to SLS; T32 A1007447-25, F32 GM125442, and K99 Al151256 to CJW). The flow cytometry work was performed at the BioFrontiers Institute Flow Cytometry Core supported by NIH Grant S10ODO21601.

References

1. Auwera GAV der *et al.* (2013) **From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline** *Current protocols in bioinformatics* **43**:11–33 https://doi.org/10.1002/0471250953.bi1110s43

2. Ayouba A, Heuverswyn FV, Peeters M (2014) **Human Immunodeficiency Viruses: OriginEncyclopedia of Virology (Second Edition) - SIMIAN IMMUNODEFICIENCY VIRUSES (RETROVIRIDAE)** *Elsevier* https://doi.org/10.1016/b978-0-12-801238-3.02597-6

3. Bailes E, Gao F, Bibollet-Ruche F, Courgnaud V, Peeters M, Marx PA, Hahn BH, Sharp PM (2003) **Hybrid Origin of SIV in Chimpanzees** *Science* **300**:1713–1713 https://doi.org/10.1126/science.1080657

4. Barbachano-Guerrero A, Perez D, Sawyer SL (2023) **How avian influenza viruses spill over to mammals** *Elife* **12**

5. Bibollet-Ruche F *et al.* (2019) **CD4 receptor diversity in chimpanzees protects against SIV infection** *P Natl Acad Sci Usa* **116**:3229–3238 https://doi.org/10.1073/pnas.1821197116

6. Bour S, Geleziunas R, Wainberg MA (1995) **The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection** *Microbiol Rev* **59**:63–93 https://doi.org/10.1128/mr.59.1.63-93.1995

7. Coffin J *et al.* (2021) **ICTV Virus Taxonomy Profile: Retroviridae 2021** *J Gen Virol* **102** https://doi.org/10.1099/jgv.0.001712

8. Daugherty MD, Malik HS (2012) **Rules of Engagement: Molecular Insights from Host-Virus Arms Races** *Annu Rev Genet* **46**:677–700 https://doi.org/10.1146/annurev-genet-110711-155522

9. Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL (2013) **Dual host-virus arms races shape an essential housekeeping protein** *PLoS Biol* **11** https://doi.org/10.1371/journal.pbio.1001571

10. Demogines A, East AM, Lee J-H, Grossman SR, Sabeti PC, Paull TT, Sawyer SL (2010) **Ancient and Recent Adaptive Evolution of Primate Non-Homologous End Joining Genes** *Plos Genet* **6** https://doi.org/10.1371/journal.pgen.1001169

11. Ann Demogines, Farzan M, Sawyer SL (2012) **Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats** *J Virol* **86**:6350–6353 https://doi.org/10.1128/jvi.00311-12

12. Demogines A, Truong KA, Sawyer SL (2012) **Species-Specific Features of DARC, the Primate Receptor for Plasmodium vivax and Plasmodium knowlesi** *Mol Biol Evol* **29**:445–449 https://doi.org/10.1093/molbev/msr204

13. DePristo MA *et al.* (2011) **A framework for variation discovery and genotyping using next-generation DNA sequencing data** *Nature genetics* **43**:491–498 https://doi.org/10.1038/ng.806

14. Edgar RC (2004) **MUSCLE: multiple sequence alignment with high accuracy and high throughput** *Nucleic Acids Res* **32**:1792–1797 https://doi.org/10.1093/nar/gkh340

15. Fay JC, Wu C-I (2000) **Hitchhiking Under Positive Darwinian Selection** *Genetics* **155**:1405–1413 https://doi.org/10.1093/genetics/155.3.1405

16. Fu YX, Li WH (1993) **Statistical tests of neutrality of mutations** *Genetics* **133**:693–709 https://doi.org/10.1093/genetics/133.3.693

17. Gavel Y, Heijne G von (1990) **Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering** *Protein Eng* **3**:433–442 https://doi.org/10.1093/protein/3.5.433

18. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE (2017) **UCSF ChimeraX: Meeting modern challenges in visualization and analysis** *Protein Sci Publ Protein Soc* **27**:14–25 https://doi.org/10.1002/pro.3235

19. Gupta R, Brunak S (2002) **Prediction of glycosylation across the human proteome and the correlation to protein function** *Pac Symposium Biocomput Pac Symposium Biocomput* :310–22

20. Han K, Lou DI, Sawyer SL (2011) **Identification of a Genomic Reservoir for New TRIM Genes in Primate Genomes** *Plos Genet* **7** https://doi.org/10.1371/journal.pgen.1002388

21. Heuverswyn FV *et al.* (2006) **SIV infection in wild gorillas** *Nature* **444**:164–164 https://doi.org/10.1038/444164a

22. Hu P, Bauer VL, Sawyer SL, Diaz-Griffero F (2022) **Human ACE2 Polymorphisms from Different Human Populations Modulate SARS-CoV-2 Infection** *Viruses* **14** https://doi.org/10.3390/v14071451

23. Humes D, Overbaugh J (2011) **Adaptation of Subtype A Human Immunodeficiency Virus Type 1 Envelope to Pig-Tailed Macaque Cells** *J Virol* **85**:4409–4420 https://doi.org/10.1128/jvi.02244-10

24. Jia P, Zhang W, Xiang Y, Lu X, Liu W, Jia K, Yi M (2020) **Ubiquitin-specific protease 5 was involved in the interferon response to RGNNV in sea perch (Lateolabrax japonicus)** *Fish Shellfish Immun* **103**:239–247 https://doi.org/10.1016/j.fsi.2020.04.065

25. Johnson WE, Sawyer SL (2009) **Molecular evolution of the antiretroviral TRIM5 gene** *Immunogenetics* **61**:163–176 https://doi.org/10.1007/s00251-009-0358-y

26. Judd EN, Gilchrist AR, Meyerson NR, Sawyer SL (2021) **Positive natural selection in primate genes of the type I interferon response** *Bmc Ecol Evol* **21** https://doi.org/10.1186/s12862-021-01783-z

27. Kaelber JT, Demogines A, Harbison CE, Allison AB, Goodman LB, Ortega AN, Sawyer SL, Parrish CR (2012) **Evolutionary reconstructions of the transferrin receptor of caniforms supports canine parvovirus being a re-emerged and not a novel pathogen in dogs** *PLoS Pathog* **8** https://doi.org/10.1371/journal.ppat.1002666

28. Keele BF *et al.* (2006) **Chimpanzee reservoirs of pandemic and nonpandemic HIV-1** *Science* **313**:523–526 https://doi.org/10.1126/science.1126531

29. Kerr SA, Jackson EL, Lungu OI, Meyer AG, Demogines A, Ellington AD, Georgiou G, Wilke CO, Sawyer SL (2015) **Computational and functional analysis of the virus-receptor interface reveals host range trade-offs in New World arenaviruses** *J Virol* **89**:11643–11653 https://doi.org/10.1128/jvi.01408-15

30. Klatt NR, Silvestri G, Hirsch V (2012) **Nonpathogenic Simian Immunodeficiency Virus Infections** *Csh Perspect Med* **2**:a007153–a007153 https://doi.org/10.1101/cshperspect.a007153

31. Lee K, Mulky A, Yuen W, Martin TD, Meyerson NR, Choi L, Yu H, Sawyer SL, KewalRamani VN (2012) **HIV-1 Capsid-Targeting Domain of Cleavage and Polyadenylation Specificity Factor 6** *J Virol* **86**:3851–3860 https://doi.org/10.1128/jvi.06607-11

32. Li H (2013) **Aligning sequences reads, clone sequences and assembly contigs with BWA-MEM** :1–3

33. Liu Q *et al.* (2017) **Quaternary contact in the initial interaction of CD4 with the HIV-1 envelope trimer** *Nat Struct Mol Biol* **24**:370–378 https://doi.org/10.1038/nsmb.3382

34. Makin S (2022) **On the lookout for viruses that could leap from animals into humans** *Nature* **610**:S41–S41 https://doi.org/10.1038/d41586-022-03357-5

35. Malfavon-Borja R, Sawyer SL, Wu LI, Emerman M, Malik HS (2013) **An Evolutionary Screen Highlights Canonical and Noncanonical Candidate Antiviral Genes within the Primate TRIM Gene Family** *Genome Biol Evol* **5**:2141–2154 https://doi.org/10.1093/gbe/evt163

36. McBee RM, Rozmiarek SA, Meyerson NR, Rowley PA, Sawyer SL (2015) **The Effect of Species Representation on the Detection of Positive Selection in Primate Gene Data Sets** *Mol Biol Evol* **32**:1091–1096 https://doi.org/10.1093/molbev/msu399

37. McKenna A *et al.* (2010) **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data** *Genome research* **20**:1297–1303 https://doi.org/10.1101/gr.107524.110

38. Meyerson NR, Rowley PA, Swan CH, Le DT, Wilkerson GK, Sawyer SL (2014) **Positive selection of primate genes that promote HIV-1 replication** *Virology* **454**:291–8 https://doi.org/10.1016/j.virol.2014.02.029

39. Meyerson NR, Sawyer SL (2011) **Two-stepping through time: mammals and viruses** *Trends Microbiol* **19**:286–294 https://doi.org/10.1016/j.tim.2011.03.006

40. Meyerson NR, Sharma A, Wilkerson GK, Overbaugh J, Sawyer SL (2015) **Identification of owl monkey CD4 receptors broadly compatible with early-stage HIV-1 Isolates** *J Virol* **89**:8611–8622 https://doi.org/10.1128/jvi.00890-15

41. Meyerson NR, Warren CJ, Vieira DASA, Diaz-Griffero F, Sawyer SL (2018) **Species-specific vulnerability of RanBP2 shaped the evolution of SIV as it transmitted in African apes** *PloS Pathog* **14** https://doi.org/10.1371/journal.ppat.1006906

42. Nahabedian J, Sharma A, Kaczmarek ME, Wilkerson GK, Sawyer SL, Overbaugh J (2017) **Owl monkey CCR5 reveals synergism between CD4 and CCR5 in HIV-1 entry** *Virology* **512**:180–186 https://doi.org/10.1016/j.virol.2017.09.018

43. Ng M *et al.* (2015) **Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats** *eLife* **4** https://doi.org/10.7554/elife.11785

44. Ning Z, Cox AJ, Mullikin JC (2001) **SSAHA: A Fast Search Method for Large DNA Databases** *Genome Res* **11**:1725–1729 https://doi.org/10.1101/gr.194201

45. Ohainle M, Malik HS (2021) **A balancing act between primate lentiviruses and their receptor** *Proc National Acad Sci* **118** https://doi.org/10.1073/pnas.2104741118

46. Pagán I, García-Arenal F (2018) **Tolerance to Plant Pathogens: Theory and Experimental Evidence** *Int J Mol Sci* **19** https://doi.org/10.3390/ijms19030810

47. Perelman P *et al.* (2011) **A molecular phylogeny of living primates** *PLoS Genet* **7** https://doi.org/10.1371/journal.pgen.1001342

48. Plantier J-C, Leoz M, Dickerson JE, Oliveira FD, Cordonnier F, Lemée V, Damond F, Robertson DL, Simon F (2009) **A new human immunodeficiency virus derived from gorillas** *Nat Med* **15**:871–872 https://doi.org/10.1038/nm.2016

49. Prado-Martinez J *et al.* (2013) **Great ape genetic diversity and population history** *Nature* **499**:471–475 https://doi.org/10.1038/nature12228

50. Rathore A, Iketani S, Wang P, Jia M, Sahi V, Ho DD (2020) **CRISPR-based gene knockout screens reveal deubiquitinases involved in HIV-1 latency in two Jurkat cell models** *Sci Rep-uk* **10** https://doi.org/10.1038/s41598-020-62375-3

51. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) **DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets** *Mol Biol Evol* **34**:3299–3302 https://doi.org/10.1093/molbev/msx248

52. Russell RM *et al.* (2021) **CD4 receptor diversity represents an ancient protection mechanism against primate lentiviruses** *Proc National Acad Sci* **118** https://doi.org/10.1073/pnas.2025914118

53. Sauter D, Kirchhoff F (2019) **Key viral adaptations preceding the AIDS pandemic** *Cell Host Microbe* **25**:27–38 https://doi.org/10.1016/j.chom.2018.12.002

54. Sawyer SL, Elde NC (2012) **A cross-species view on viruses** *Curr Opin Virol* **2**:561–568 https://doi.org/10.1016/j.coviro.2012.07.003

55. Sawyer SL, Emerman M, Malik HS (2007) **Discordant Evolution of the Adjacent Antiretroviral Genes TRIM22 and TRIM5 in Mammals** *Plos Pathog* **3** https://doi.org/10.1371/journal.ppat.0030197

56. Sawyer SL, Emerman M, Malik HS (2004) **Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G** *PLoS Biol* **2** https://doi.org/10.1371/journal.pbio.0020275

57. Sawyer SL, Wu LI, Akey JM, Emerman M, Malik HS (2006) **High-Frequency Persistence of an Impaired Allele of the Retroviral Defense Gene TRIM5α in Humans** *Curr Biol* **16**:95–100 https://doi.org/10.1016/j.cub.2005.11.045

58. Sawyer SL, Wu LI, Emerman M, Malik HS (2005) **Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain** *Proc National Acad Sci* **102**:2832–2837 https://doi.org/10.1073/pnas.0409853102

59. Sharp PM, Hahn BH (2011) **Origins of HIV and the AIDS pandemic** *CSH Perspect Med* **1** https://doi.org/10.1101/cshperspect.a006841

60. Sharp PM, Shaw GM, Hahn BH (2005) **Simian Immunodeficiency Virus Infection of Chimpanzees** *J Virol* **79**:3891–3902 https://doi.org/10.1128/jvi.79.7.3891-3902.2005

61. Stabell AC, Hawkins J, Li M, Gao X, David M, Press WH, Sawyer SL (2016) **Non-human primate schlafen11 inhibits production of both host and viral proteins** *Plos Pathog* **12** https://doi.org/10.1371/journal.ppat.1006066

62. Stecher G, Tamura K, Kumar S (2020) **Molecular evolutionary genetics analysis (MEGA) for macOS** *Mol Biol Evol* **37**:1237–1239 https://doi.org/10.1093/molbev/msz312

63. Suyama M, Torrents D, Bork P (2006) **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments** *Nucleic Acids Res* **34**:W609–W612 https://doi.org/10.1093/nar/gkl315

64. Tajima F (1989) **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism** *Genetics* **123**:585–595 https://doi.org/10.1093/genetics/123.3.585

65. Tajima F (1983) **EVOLUTIONARY RELATIONSHIP OF DNA SEQUENCES IN FINITE POPULATIONS** *Genetics* **105**:437–460 https://doi.org/10.1093/genetics/105.2.437

66. Takehisa J *et al.* (2009) **Origin and Biology of Simian Immunodeficiency Virus in Wild-Living Western Gorillas** *J Virol* **83**:1635–1648 https://doi.org/10.1128/jvi.02311-08

67. Takehisa J *et al.* (2007) **Generation of Infectious Molecular Clones of Simian Immunodeficiency Virus from Fecal Consensus Sequences of Wild Chimpanzees** *J Virol* **81**:7463–7475 https://doi.org/10.1128/jvi.00551-07

68. Wang M *et al.* (2019) **SPSB2 inhibits hepatitis C virus replication by targeting NS5A for ubiquitination and degradation** *Plos One* **14** https://doi.org/10.1371/journal.pone.0219989

69. Warren CJ, Barbachano-Guerrero A, Bauer VL, Stabell AC, Dirasantha O, Yang Q, Sawyer SL (2024) **Infection pressure in apes has driven selection for CD4 alleles that resist lentivirus (HIV/SIV) infection** *eLIFE* https://doi.org/10.7554/elife.93316

70. Warren CJ, Meyerson NR, Dirasantha O, Feldman ER, Wilkerson GK, Sawyer SL (2019) **Selective use of primate CD4 receptors by HIV-1** *PLoS Biol* **17** https://doi.org/10.1371/journal.pbio.3000304

71. Warren CJ, Meyerson NR, Stabell AC, Fattor WT, Wilkerson GK, Sawyer SL (2019) **A glycan shield on chimpanzee CD4 protects against infection by primate lentiviruses (HIV/SIV)** *Proc National Acad Sci* **116**:11460–11469 https://doi.org/10.1073/pnas.1813909116

72. Warren CJ, Sawyer SL (2023) **Identifying animal viruses in humans** *Sci New York N Y* **379**:982–983 https://doi.org/10.1126/science.ade6985

73. Warren CJ, Sawyer SL (2019) **How host genetics dictates successful viral zoonosis** *PLoS Biol* **17** https://doi.org/10.1371/journal.pbio.3000217

74. Warren CJ *et al.* (2022) **Primate hemorrhagic fever-causing arteriviruses are poised for spillover to humans** *Cell* **185**:3980–3991 https://doi.org/10.1016/j.cell.2022.09.022

75. Watterson GA (1975) **On the number of segregating sites in genetical models without recombination** *Theor Popul Biol* **7**:256–276 https://doi.org/10.1016/0040-5809(75)90020-9

76. White TE, Brandariz-Nuñez A, Valle-Casuso JC, Knowlton C, Kim B, Sawyer SL, Diaz-Griffero F (2014) **Effects of human SAMHD1 polymorphisms on HIV-1 susceptibility** *Virology* **460**:34–44 https://doi.org/10.1016/j.virol.2014.04.023

77. Wilke CO, Sawyer SL (2016) **At the mercy of viruses** *Elife* **5** https://doi.org/10.7554/elife.16758

78. Yang Z (2007) **PAML 4: phylogenetic analysis by maximum likelihood** *Mol Biol Evol* **24**:1586–1591 https://doi.org/10.1093/molbev/msm088

79. Yang Z (2007) **PAML User Manual**

80. Yang Z, Kumar S, Nei M (1995) **A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequence** *Genetics* **141**:1641–1650

81. Zhang H *et al.* (2021) **Ubiquitin-Modified Proteome of SARS-CoV-2-Infected Host Cells Reveals Insights into Virus–Host Interaction and Pathogenesis** *J Proteome Res* **20**:2224–2239 https://doi.org/10.1021/acs.jproteome.0c00758

82. Zhang ZD, Weinstock G, Gerstein M (2008) **Rapid Evolution by Positive Darwinian Selection in T-Cell Antigen CD4 in Primates** *J Mol Evol* **66**:446–456 https://doi.org/10.1007/s00239-008-9097-1

Article and author information

Cody J. Warren
BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA

Arturo Barbachano-Guerrero
BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA
ORCID iD: 0000-0001-7483-839X

Vanessa L. Bauer
BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA

Alex C. Stabell
BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA

Obaiah Dirasantha
BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA

Qing Yang
BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA

Sara L. Sawyer
BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA
For correspondence: ssawyer@colorado.edu
ORCID iD: 0000-0002-6965-1085

Editors

Reviewing Editor
John Schoggins
The University of Texas Southwestern Medical Center, Dallas, United States of America

Senior Editor
John Schoggins
The University of Texas Southwestern Medical Center, Dallas, United States of America

Reviewer #1 (Public Review):

Summary:

Authors previously demonstrated that species-specific variation in primate CD4 impacts its ability to serve as a functional receptor for diverse SIVs. Here, Warren and Barbachano-Guerrero et al. perform population genetics analyses and functional characterization of great ape CD4 with a particular focus on gorillas, which are natural hosts of SIVgor. They first used ancestral reconstruction to derive the ancestral hominin and hominid CD4. Using pseudotyped viruses representing a panel of envelopes from SIVcpz and HIV strains, they find that these ancestral reconstructions of CD4 are more similar to human CD4 in terms of being a broadly susceptible entry receptor (in the context of mediating entry into Cf2Th cells stably expressing human CCR5). In contrast, extant gorilla and chimpanzee CD4 are functional entry receptors for a narrower range of HIV and SIVcpz isolates. Based on these differences, authors next surveyed gorilla sequences and identified several CD4 haplotypes, specifically in the region encoding the CD4 D1 domain, which directly contacts the viral glycoprotein and thus may impact the interaction. Consistent with this possibility, authors demonstrated that gorilla CD4 haplotypes are, on average, less capable of supporting entry than human CD4, and that some are largely unable to function as SIV entry receptors. Interestingly, individual residues found at key positions in the gorilla CD4 D1 when tested in the context of human CD4 reduce entry of some virions pseudotyped with diverse SIVcpz envelopes, suggesting that individual amino acids can in part explain the observed differences across gorilla CD4 haplotypes. Finally, the authors perform statistical tests to infer that CD4 from great apes with endemic SIV (i.e., chimpanzees and gorillas) but not non-reservoirs (i.e., orangutans, bonobos) or recent spillover hosts (i.e., humans), have been subject to selection as a result of pressure from endemic SIV.

The conclusions of this paper are mostly well supported by data.

Strengths:

(1) The functional assays are appropriate to test the stated hypothesis, and the authors use a broad diversity of envelopes from HIV and SIVcpz strains. Authors also partially characterize one potential mechanism of gorilla CD4 resistance - receptor glycosylation at the derived N15 found in 5/6 gorilla haplotypes.

(2) Ancestral reconstruction provides a particularly interesting aspect of the study, allowing authors to infer the ancestral state of hominid CD4 relative to modern CD4 from gorillas and chimpanzees. This, coupled with evidence supporting SIV-driven selection of gorilla CD4 diversity and the characterization of functional diversity of extant haplotypes provides several interesting findings.

Weaknesses:

(3). The major inference of the work is that SIV infection of gorillas drove the observed diversity in gorilla CD4. This is supported by the majority of SNPs being localized to the CD4 D1, which directly interacts with envelope, and the demonstrated functional consequences of that diversity for viral entry. However, SIVgor (to the best of my knowledge) only infects Western lowland gorillas (Gorilla gorilla gorilla), and one Gorilla gorilla diehli and three Gorilla beringei graueri individuals were included in the haplotype and allele frequency analyses. The presence of these haplotypes or the presence of similar allele frequencies in Eastern lowland and mountain gorillas would impact this conclusion. It would be helpful for the authors to clarify this point.

(4) The authors appear to use a somewhat atypical approach to assess intra-population selection to compensate for relatively small numbers of NHP sequences (Fig. 6). However, they do not cite precedence for the robustness of the approach or the practice of grouping sequences from multiple species for the endemic vs other comparison. They also state in the methods that some genes encoded in the locus were removed from the analysis "because they have previously been shown to directly interact with a viral protein." This seems to undercut the analysis, and prevents alternative explanations for the observed diversity in CD4 (e.g., passenger mutations from selection at a neighboring locus).

(5) Data in Figure 5 is graphed as % infected cells instead of virus titer (TDU/mL). It's unclear why this is the case, and prevents a comparison to data in Figure 2 and Figure 4.

(6) The lack of pseudotyping with SIVgor envelope is a surprising omission from this study, that would help to contextualize the findings. Similarly, building gorilla CD4 haplotype SNPs onto the hominin ancestor (as opposed to extant human CD4) may provide additional insights that are meaningful towards understanding the evolutionary trajectory of gorilla CD4.

Comments on revised version:

In the revised manuscript, the authors more appropriately contextualize conclusions that can be made based on their data versus inferences, which are now much more clearly described in the discussion. The authors also included more references to substantiate claims, additional description of methodology, and provided well-reasoned responses to the weaknesses described in my primary review.

Re: #3. As the authors point out, we do not know if eastern gorillas were at one time exposed to SIV. The authors use a variety of phylogenetic and functional approaches to infer that SIVcpz is the selective pressure-shaping gorilla CD4. While I agree this is a highly likely scenario, the allelic diversity of CD4 across gorilla subpopulations raises multiple evolutionary scenarios consistent with the data.

Re: #4. The explanation provided by the authors is reasonable. However, a demonstration that this approach is robust to potential factors that might skew the data (e.g., recombination) is argued but not tested. Part of the concern here is that the study is limited by very small sample sizes, and to the best of my knowledge, grouping sequences from multiple species to make claims about selection is not an established practice. The authors note in their response that they confirmed the existence of CD4 alleles in this study with those identified in 100 gorilla individuals from Russell et al. 2021 (unavailable to the authors at the time of

submission) - a re-analysis that includes that data from Russell et al. 2021 would have strengthened the analyses.

https://doi.org/10.7554/eLife.93316.2.sa1

Reviewer #2 (Public Review):

Lentiviral infection of primate species has been linked to the rapid mutational evolution of numerous primate genes that interact with these viruses, including genes that inhibit lentiviruses as well as genes required for viral infection. In this manuscript, Warren et al. provide further support for the diversification of CD4, the lentiviral entry receptor, to resist lentiviral infection in great ape populations. This work builds on their prior publication (Warren et al. 2019, PMCID: PMC6561292) and that of other groups (e.g., Russell et al. 2021, PMCID: PMC8020793; Bibollet-Ruche et al. 2019, PMCID: PMC6386711) documenting both sequence and functional diversity in CD4, specifically within (1) the CD4 domain that binds to the lentiviral envelope and (2) great ape populations with endemic lentiviruses. Thus, the paper's finding that gorilla populations exhibit diverse CD4 alleles that differ in their susceptibility to lentiviral infection is well demonstrated both here and in a prior publication.

Strengths:

By reconstructing the CD4 sequence from the ancestor of gorillas and chimpanzees, the authors document that modern species have evolved more resistance to (admittedly modern) lentiviruses. They also deconstruct the molecular basis of this resistance by showing that one mutation, which adds a glycosylation site to CD4, is sufficient to confer lentiviral resistance to the susceptible human allele.

Weaknesses:

Warren et al. also pursue two novel lines of evidence to suggest that lentiviruses are the causative driver of great ape CD4 diversification, which seems likely from a logical perspective but is difficult to prove. First, they demonstrate that resistance to lentiviral infection is a derived trait in chimpanzees and gorillas, which have been co-evolving with endemic lentiviruses, but not in humans, which only recently acquired HIV. Nevertheless, these three examples are insufficient to prove that derived resistance is not stochastic or due to drift. The argument would be strengthened by demonstrating that bonobo and orangutan CD4, which also do not have endemic lentiviruses, resemble the ancestral and human susceptibility to great-ape-infecting lentiviruses.

Second, Warren et al. provide a population genetic argument that only endemically infected primates exhibit diversifying selection, again arguing for endemic lentiviruses being the evolutionary driver. The authors compare SNP occurrence in CD4 to neighboring genes, demonstrating that non-synonymous SNP frequency is only elevated in endemically infected species. Moreover, these amino-acid-coding changes are significantly concentrated in the CD4 domain that binds the lentiviral envelope. This is a creative analysis to overcome the problem of very small sample sizes, with very few great ape individuals sequenced. However, the small number of species compared (2-4 in each group) also limits the power of the analysis. Expanding the analysis to Old World Monkey species that do or do not have endemic lentiviruses, as well as great apes, would strengthen this argument.

Overall, this manuscript lends additional support to a well-documented example of a host-virus arms race: that of lentiviruses and the viral entry receptor.

https://doi.org/10.7554/eLife.93316.2.sa0

Author response:

The following is the authors' response to the original reviews.

We are thankful for the comments and suggestions from the Editor and Reviewers about our manuscript submitted to the eLife Journal. We have addressed all the comments, and we think these modifications will help bring clarity to our message and be helpful to your readership. Here we include an outline of the corrections performed, as well as a detailed response to each of the reviewer's comments.

As per the Editor and Reviewers suggestions, outline of corrections:

· The title of the manuscript has been changed to reflect a more conservative conclusion.

· Changes in the main manuscript text were made to enhance clarity, including the use genetic terminology and naming.

· Specific responses to some comments from the reviewers are included in this document. We combined some comments that would be better addressed together.

Accompanied to this letter is an updated version of our manuscript with the track changes feature enabled. Again, we are thankful of the comments and suggestions we received, and we hope this revised version of our manuscript will be accompanied by an updated assessment and public reviews and a final eLife Version of Record.

Response to the public review and minor recommendations.

> ***From Reviewer #1:***
>
> *The major inference of the work is that SIV infection of gorillas drove the observed diversity in gorilla CD4. This is supported by the majority of SNPs being localized to the CD4 D1, which directly interacts with the envelope, and the demonstrated functional consequences of that diversity for viral entry. However, SIVgor (to the best of my knowledge) only infects Western lowland gorillas (Gorilla gorilla gorilla), and one Gorilla gorilla diehli and three Gorilla beringei graueri individuals were included in the haplotype and allele frequency analyses. The presence of these haplotypes or the presence of similar allele frequencies in Eastern lowland and mountain gorillas would impact this conclusion. It would be helpful for the authors to clarify this point.*
>
> ***From Reviewer #1 (minor comment):***
>
> *Which subspecies of gorilla are the nsSNPs coming from? Gorilla gorilla diehli [n =1]; Gorilla beringei graueri [n = 3]) are not extant reservoirs of SIV and to my knowledge are not thought to have been, and so it's important to point out where the diversity is coming from if the authors are asserting that SIVgor drove this population-level diversity in gorilla CD4.*

We initially included genomic data from all the gorilla individuals available to maximize sensitivity to identify allelic variants. Although evidence points to eastern gorillas not being currently infected with SIV, our results show that all allelic variants identified have differential susceptibility to the HIV-1 and SIVcpz strains tested. The allelic variants we identified with this genomic data set match the variants identified by Russell et al (doi.org/10.1073/pnas.2025914118), including the ones found in eastern gorillas, and recapitulate that those variants have differential susceptibility to lentiviral entry, similar to the variants of western populations. Whether eastern gorillas have been exposed to lentiviruses in the past remains unknown.

> ***From Reviewer #1:***
>
> *The authors appear to use a somewhat atypical approach to assess intra-population selection to compensate for relatively small numbers of NHP sequences (Fig. 6). However, they do not cite precedence for the robustness of the approach or the practice of grouping sequences from multiple species for the endemic vs other comparison. They also state in the methods that some genes encoded in the locus were removed from the analysis "because they have previously been shown to directly interact with a viral protein." This seems to undercut the analysis and prevents alternative explanations for the observed diversity in CD4 (e.g., passenger mutations from selection at a neighboring locus).*

Given the nature of our samples, to detect any influence of natural selection acting on CD4, we chose to compare patterns of molecular evolution of CD4 to its neighboring loci. Comparisons of molecular evolution signatures across genomic regions are the basis of methods to detect positive selection (e.g., Sabeti DOI: 10.1038/nature01140). For our comparison, the neighboring loci represent our neutral standard for the genomic region CD4 resides. Our rationale is that demographic and neutral influences on the number and frequency of polymorphic sites in a region would equally affect all loci in a genomic region. Because these neighboring loci are our neutral benchmark, we excluded before analysis other genes in this genomic region that interact with viruses. The logic is that these loci may be evolving under the influence of positive selection and would decrease the power of our comparison. None of the excluded loci are direct neighbors to CD4. This, and given that the CD4 genomic region in humans is of average recombination rate, dampens the possibility that what we are observing at CD4 is due to selection acting at a neighboring locus. In addition, the classic population genetic method to detect positive selection, the McDonald-Kreitman test (McDonald DOI: 10.1038/351652a0), was originally presented combining polymorphism data across species. We assume that any effect on levels of diversity created by combining variability between species would equally affect all loci included in the study, not just CD4.

> ***From Reviewer #1:***
>
> *Data in Figure 5 is graphed as % infected cells instead of virus titer (TDU/mL). It's unclear why this is the case, and prevents a comparison to data in Figure 2 and Figure 4.*
>
> ***From Reviewer #1 (minor comment):***
>
> *Figure 5: the data presentation is now shown as % infected cells instead of viral titer. This makes it difficult to compare data from Figure 5 to other figures. Can the authors please either justify this change, display data consistently or provide matched data displays as a Supplemental Figure?*

For the experiments presented in figures 2 and 4 we used different volumes of infecting pseudoviruses, which allowed us to identify the linear range of infection. Then, based on the number of cells plated per experimental replicate, we calculated a virus titer. In follow-up experiments (Fig. 5), we used fixed volumes of virus that would infect ~10-20% of control (wild-type; wt) CD4-expressing cells. Comparisons were then made between wt and mutated CD4s, and these data are best presented in their raw forms as percent cells infected. Although this change in method prevents direct comparison between the figures, we focused on the differences observed between the experimental conditions per experimental panel.

> ***From Reviewer #1:***
>
> *The lack of pseudotyping with SIVgor envelope is a surprising omission from this study, that would help to contextualize the findings.*

> ***From Reviewer #2 (minor comment):***
>
> *The inclusion of HIV-1 but not SIVgor strains in Figures 2D/E is somewhat conspicuous since chimpanzee alleles certainly differ in susceptibility to SIVcpz (and SIVgor) strains per Russell et al. 2021. The authors should either test some SIVgor infections, cite published data on at least extant human/chimpanzee/gorilla CD4 susceptibility to SIVgor, or address why they did not include it.*

We agree the data of host susceptibility to SIVgor strains would have been an interesting question to explore. However, we opted to focus on the transmission of SIVcpz strains into gorilla populations for this study. It is worth mentioning that we have cloned SIVgor envelope genes from some strains into our expression system, but we were unable to recover infectious pseudoviruses using an HIV-1DEnv-GFP backbone. This suggests that HIV-1 may be incompatible with incorporating SIVgor Env into virus particles. Recently, Russell et al (DOI: 10.1073/pnas.2025914118) managed to generate SIVgor Env pseudotyped virions using a different backbone (SIVcpzDEnv-GFP) that was unavailable to us at the time of this study.

> ***From Reviewer #1:***
>
> *Similarly, building gorilla CD4 haplotype SNPs onto the hominin ancestor (as opposed to extant human CD4) may provide additional insights that are meaningful toward understanding the evolutionary trajectory of gorilla CD4.*

We decided to use the extant human CD4 as a backbone to test the effects on the individual amino acid variants found in the allelic diversity of the gorilla population since the human protein is highly susceptible to all the HIV-1 and SIV strains tested, and the expected phenotype is a loss-of-function. Since the D1 of the human and ancestral sequences for CD4 are almost identical (except for a change that is fixed in gorillas), and they showed similar levels of susceptibility to lentivirus entry, we expect that the phenotypes found would be the same if the gorilla SNPs were built into the ancestral CD4 backbone.

> ***From Reviewer #2:***
>
> *To bolster the argument that lentiviruses are indeed the causative driver of this diversification, which seems likely from a logical perspective but is difficult to prove, Warren et al. pursue two novel lines of evidence. First, the authors reconstruct ancestral CD4 genes that predate lentiviral infection of hominid populations. They then demonstrate that resistance to lentiviral infection is a derived trait in chimpanzees and gorillas, which have been co-evolving with endemic lentiviruses, but not in humans, which only recently acquired HIV. Nevertheless, the derived resistance could be stochastic or due to drift. This argument would be strengthened by demonstrating that bonobo and orangutan CD4, which also do not have endemic lentiviruses, resemble the ancestral and human susceptibility to great-ape-infecting lentiviruses.*
>
> ***From Reviewer #2 (minor comment):***
>
> *The data presented in Figure 2, showing that chimp and gorilla (but not human) CD4 resistance to lentiviral infection is a derived trait, is very intriguing for suggesting that endemic lentiviruses are the causative driver of CD4 evolution. Nevertheless, this could be stochastic or due to genetic drift. Given the later emphasis on several other non-endemically infected species, the authors should at the very least include the sequences for bonobo and orangutan CD4 in the presented alignment (Fig 2B). Ideally, they would also test these orthologs to demonstrate that they are not resistant to lentiviruses infecting great apes (SIVcpz / HIV-1 / SIVgor). If they have also derived resistance, this would suggest a possible other evolutionary driver or genetic drift.*

Based on our analysis on polymorphic sites using available data from populations of apes, we strongly believe the accumulation of resistant polymorphisms in CD4 did not arise in a stochastic manner. The frequency and accumulation of these changes strongly correlate with the function of CD4 as a receptor for lentivirus entry. We agree that experimentally testing the CD4 protein from bonobo and orangutan would strengthen our conclusions; however, based on our genomic analyses, we decided to focus on the species that would present a higher level of variability of susceptibility to the lentivirus tested, namely gorillas and chimpanzees.

> ***From Reviewer #2:***
>
> *Warren et al. provide a population genetic argument that only endemically infected primates exhibit diversifying selection, again arguing for endemic lentiviruses being the evolutionary driver. The authors compare SNP occurrence in CD4 to neighboring genes, demonstrating that non-synonymous SNP frequency is only elevated in endemically infected species. Moreover, these amino-acid-coding changes are significantly concentrated in the CD4 domain that binds the lentiviral envelope. This is a creative analysis to overcome the problem of very small sample sizes, with very few great ape individuals sequenced. The additional small number of species compared (2-3 in each group) also limits the power of the analysis; the authors could consider expanding their analysis to Old World Monkey species that do or do not have endemic lentiviruses, as well as great apes.*

The scope of this project was to evaluate the differential phenotype of the accumulated polymorphisms found in the ape branch of the primates. Although evaluating the accumulation of polymorphisms in a broader range of primates would generate interesting observations, this would likely require increasing the total number of primate species to include sampling along the speciation tree, many of which lack population level data.

> ***From Reviewer #1 (minor comment):***
>
> *Ancestral reconstruction methods and associated data tables should be included to indicate statistical support for assigned codons. A comment on ambiguity at relevant positions is needed. Similarly, given the polymorphic nature of gorilla and chimpanzee CD4, how confident are the authors in their ancestral reconstructions based on a single representative genome per species? Does this change when you include the broader panel of gorilla sequences? Is the ancestral reconstruction robust to other methods besides PAML?*

We used the PAML software package to reconstruct the ancestral hominin and hominid sequence of CD4 because it is a standard and well recognized method for this purpose. For this analysis, we used the set of primate sequences selected for positive selection analyses (see methods), namely the longest isoform sequences for each of the available species that best aligned with human CD4. We feel that the best way to perform to the ancestral state reconstruction was to use only these curated sequences instead of the population level sequences, removing potential biases introduced by having different numbers of variants per species.

> ***From Reviewer #1 (minor comment):***
>
> *Page 10: "It seems that allele 2, which doesn't have this glycan, would be at a fitness disadvantage. In support of this, allele 2 is one of the least frequent alleles in the gorilla population that we surveyed (Figure 3B)." - this inference depends on the gorilla species*

> *that encode allele 2 and allele frequencies. There are statistical tests to address this inference.*

Population genetic statistics that test for skews in sample allele frequencies are not appropriate here due to the nature of the samples in this study. However, the reviewer is correct that our inference in allele frequency is dependent on the gorilla species that we find this allele in. Allele 2 is found in the Gorilla beringei graueri subspecies of gorilla included in this study. We only have data for three individuals (six alleles) from this subspecies compared to 51 individual (102 alleles) from Gorilla gorilla gorilla. As such, genetic subdivision between the gorilla subspecies could also produce the low frequency of allele 2 observed in our sample.

> ***From Reviewer #1 (minor comment):***
>
> *Page 11: "These results imply that the resistance to SIVcpz found in gorilla individuals is not dependent on single amino acids, but rather the cumulative effect of multiple SNPs." Would it be more relevant (or relevant in other ways) to test this statement by putting those mutations into the hominid ancestor? Testing individual residues in the context of human CD4 may be subject to epistasis or several other factors.*

We agree that constructing multiple of the resistant SNPs in the susceptible human background would have strengthened our hypothesis, as all these amino acid changes are associated with increased resistance to at least one of the lentiviruses tested. However, the number of CD4 variants to test would increase significantly and we feel that this approach was out of the scope of this manuscript.

> ***From Reviewer #1 (minor comment):***
>
> *Figure 6: If you perform this analysis on chimpanzee CD4 alone do you get the same result? Just gorillas? If you remove eastern/mountain gorillas? The very small numbers of non-human non-SIV-reservoir great apes may preclude a strong conclusion.*

We agree that our study is limited by the small number of available sequences from individuals of the studied species. If we remove a whole species or subspecies the statistical power would be greatly reduced. Removing all chimpanzees or gorillas (or a subspecies) would still show that only each of those species accumulate SNPs in the D1 region of CD4, although with less statistical significance.

> ***From Reviewer #2 (minor comment):***
>
> *Related to Figure 2: It would strengthen the argument that resistance is a derived trait if the authors mapped the causative mutations from gorilla CD4 onto the ancestral hominin CD4. However, this experiment is not particularly critical, merely a suggestion.*

We appreciate this suggestion. We decided to use the human CD4 backbone as it is widely susceptible to lentiviral entry. The hominid and hominin ancestral sequences are almost identical to the human sequence in domain 1, except for a fixed mutation shared with the gorilla CD4. We expect that the SNPs observed in the gorilla population would also reduce susceptibility to lentivirus entry in the ancestral CD4 reconstructions.

> ***From Reviewer #2 (minor comment):***
>
> *Related to Figure 3B: It is difficult to make much of the allele frequency for 8 alleles in 32 individuals. Can the authors collate this with allele frequency for the referenced 100 individuals from Russell et al. 2021, to give a better sense of population frequency? This*

> *may allow the authors to better correlate allele frequency with SIVcpz resistance patterns in Figure 4, strengthening their argument that more resistant alleles should be over-represented in the population.*

At the time of our analysis the data from Russell (DOI: 10.1073/pnas.2025914118) was not available to collate or compare. When that data became available, we immediately compared the existence of the alleles found and confirmed that the ones we found were also detected in the samples used in that study.

> ***From Reviewer #2 (minor comment):***
>
> *Related to Figure 6: As written, several methodological details should be clarified. How were human genomes selected to limit the sample size to 50?*

We selected a total of 50 human individuals in order to size-match the sample size of the largest group in Fig 6B (chimpanzee, n=50). We randomly selected 10 individuals for each of the 5 superpopulations [Africans (AFR), Admixed Americans (AMR), East Asians (EAS), Europeans (EUR) and South Asians (SAS)] defined by the 1000 Genome Project.

> ***From Reviewer #2 (minor comment):***
>
> *Related to Figure 6: What comparison is being reported for the Mann-Whitney U test (CD4 vs. which gene)? Are the means shown in A an average of 2 (endemic) or 3 (non-endemic) species - if so, the authors should show the individual data points to give a clearer depiction of the data spread. In addition, it is not clear that a statistical test with sample sizes of 2 is meaningful, since Mann Whitney typically assumes n > 5. To strengthen this statistical argument, it may be necessary to include additional species that have (a) multiple genomes (or at least this locus) sequenced, and (b) have or lack lentiviral sequences. This may necessitate expanding the analysis to include Old World Monkeys (e.g. Rhesus Macaque Genome Project).*

In the Figure 6 we use the Mann-Whitney U test to compare variation between CD4 and the neighboring loci. The average and SEM are for two endemic and four non-endemic species (two orangutan datasets are from two distinct species vs the gorilla subspecies). It is true our sample size is small for any statistical testing. For the Mann-Whitney U-test it is generally preferred to have n > 5 in each group. So, we do run into problems with the endemically infected comparisons as we only have two data points (chimpanzee and gorilla) for the CD4 group. For the uninfected species, CD4 has four data points.

> ***From Reviewer #1 (minor comment):***
>
> *Page 6. "This suggests that the ancestral versions of CD4 in apes were susceptible to primate lentivirus entry" - The data show that tested virus pseudotyped with SIV/HIV envs can engage ancestral CD4 in the context of a canine cell line expressing human CCR5, but not necessarily that this interaction was sufficient for the process of entry per se, especially in the context of a gorilla (or hominid) cell. Some additional context would be useful for a broad readership.*
>
> ***From Reviewer #1 (minor comment):***
>
> *Page 6: "but that selective pressures exerted by SIVs in the chimpanzee and gorilla lineages have led to the retention of mutations that confer resistance to primate lentivirus infection. This has not happened in humans where selective pressure by HIV-1 is too new" - this cannot be concluded from the data in Figure 1. It would be more appropriate as a Discussion point.*

From Reviewer #1 (minor comment):

Page 14: "Natural tolerance is often required before a virus can establish itself long term in a host reservoir, and thus understanding it is key to understanding virus reservoirs in nature" - please provide a reference. This is one among several theories of long-term host-virus evolution dynamics/outcomes, and further discussion may benefit the broad readership of eLife.

From Reviewer #1 (minor comment):

Page 15: "There is a surprising outcome of virus-driven host evolution in that the divergence and diversity of these host genes ultimately comes at a detriment to the very viruses that drove this evolution." - it is not clear to this reviewer why this is surprising.

From Reviewer #2 (minor comment):

Related to Figure 5A: The authors suggest that the gorilla glycosylation site provides resistance to SIVcpz, based on TAN1.910, but in fact the glycosylated allele is no more resistant than the un-glycosylated allele to most SIVcpz strains (in Figure 4). The authors should acknowledge this more clearly in the text.

From Reviewer #2 (minor comment):

The title of this article (that infection "has driven selection") is somewhat overstated - though it seems very likely that lentiviruses are driving CD4 diversification, this is difficult to prove. The arguments presented here rely on very few data points: modern chimp and gorilla compared to ancestral CD4, and a population genetic analysis relying on 2 or 3 species with 10-50 individuals each. The authors should either bolster these arguments (see the above suggestions) and/or soften the claim in the title.

Modifications to the main text of the manuscript have been made to enhance clarity on the subjects stated above.

https://doi.org/10.7554/eLife.93316.2.sa3