Environmentally-mediated selection parallels population divergence across a chimpanzee subspecies contact zone

Matthew W. Mitchell[1,2,*], Walker Alexander[3], Dana V. Mitchell[1], Adam H. Freedman[4], Janina Dordel[1], Ryan J. Harrigan[5], Ahmet Sacan[3], Fabrice Kentatchime[1,6], Bryan S. Featherstone[1], Ekwoge E. Abwe[1,7,8], Paul R. Sesink Clee[1], Abwe E. Abwe[8], Sabrina Locatelli[9], Bethan J. Morgan[7,8,10], Bernard Fosso[11], Roger Fotso[11], Sarah A. Tishkoff[12], Evan E. Eichler[13,14], Nicola M. Anthony[15], Thomas B. Smith[5,16], Mary Katherine Gonder[1,6,*]

[1]Department of Biology, Drexel University, Philadelphia, PA, USA

[2]Coriell Institute for Medical Research, Camden, NJ, USA

[3]School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA, USA

[4]Faculty of Arts and Sciences Informatics Group, Harvard University, Cambridge, MA, USA

[5]Center for Tropical Research, Institute of Environment and Sustainability, University of California, Los Angeles, CA, USA

[6]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA

[7]San Diego Zoo Wildlife Alliance, Escondido, CA, USA

[8]Cameroon Biodiversity Association, Douala, Cameroon

[9]Institut de Recherche pour le Développement (IRD), Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution et Contrôle (MIVEGEC) (IRD 224-CNRS 5290-Université de Montpellier), Montpellier, France

[10]School of Natural Sciences, University of Stirling, Stirling, UK

[11]Wildlife Conservation Society Cameroon, Yaoundé, Cameroon

[12]Departments of Genetics and Biology, University of Pennsylvania, Philadelphia, PA, USA

[13]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA

[14]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

[15]Department of Biological Sciences, University of New Orleans, New Orleans, LA, USA

[16]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA

*Matthew W. Mitchell
Email: mmitchell@coriell.org

*Mary Katherine Gonder
katy.gonder@ag.tamu.edu

Abstract

Species evolve from populations with ancestor-descendant relationships in a bifurcating process shaped by geography, gene flow, genetic drift, and natural selection leading to local adaptation to prevailing environmental and ecological conditions. Building on this foundational understanding, we explored local adaptation in chimpanzees (*Pan troglodytes*) at a key geographical intersection in Cameroon where the two main chimpanzee phylogenetic lineages converge. The Nigeria-Cameroon chimpanzee (*P. t. ellioti*) and central chimpanzee (*P. t. troglodytes*) last shared a common ancestor about 500 thousand years ago, with occasional gene flow between them. The evolutionary processes driving their prolonged separation are not fully understood, but neutral evolutionary mechanisms alone cannot account for the observed divergence pattern. Cameroon is often referred to as 'Africa in miniature' because the Gulf of Guinea Forest, Congo Basin Forest, and savanna converge there, forming an ecotone. Thus, this contact zone between subspecies in Cameroon provides a unique natural laboratory that enabled us to investigate how environmental variation and natural selection shape divergence in chimpanzees. We developed a genome-wide panel of single-nucleotide polymorphisms (SNPs) in 112 wild chimpanzees sampled in multiple habitats across this contact zone. We augmented SNP discovery by sequencing eight new chimpanzee genomes from Cameroon and analyzing them with previously published chimpanzee genomes. We found that *P. t. ellioti* and *P. t. troglodytes* diverged from one another around 478,000 years ago and occasionally exchange migrants. We identified 1,690 unique SNPs across 905 genes associated with 31 environmental variables that describe the habitat. These genes are involved in essential biological processes, including immune response, neurological development, behavior, and dietary adaptations. This study highlights the importance of understanding the geographical context of natural selection, paving the way for future studies to interpret evidence for genetic variation with phenotypic traits and deepening our understanding of how populations diverge in response to environmental pressures.

Author Summary

We investigated how local adaptation contributes to shaping the diversification of chimpanzee subspecies at the geographical convergence point for the two major branches of the chimpanzee phylogenetic tree. We analyzed genome-wide SNP genotypes of 112 chimpanzees sampled from natural communities located in this understudied area. We used tiered methods that identified 905 genes subject to selection, each associated with one or more of 31 environmental predictors describing the habitat. We found strong signals of selection in immune response genes that separate *P. t. troglodytes* from *P. t. ellioti*, highlighting the important role of different pathogen histories in their evolution. We also found evidence of selection in genes associated with neurological development, behavior, and diet, that separate both the subspecies and populations of *P. t. ellioti* that occupy different niches. These findings suggest that ecological and cultural factors may also contribute to shaping the diversification of chimpanzees across the contact zone.

Introduction

Species consist of populations of reproductively compatible individuals with ancestor–descendant relationships that evolve through time [1]. Speciation may result from various factors. It may have a geographical dimension ranging from allopatry to sympatry with varying degrees of gene flow among populations, genetic drift, and natural selection [2]. Ecological factors often play a decisive role in this process through the local adaptation of populations to prevailing environmental conditions [3]. The fusion of genomics with ecological modeling has advanced the ability to identify loci under environmental selection. It contributes to understanding how species adapt to specific habitats and its impact on speciation [4, 5]. While this link has been studied in many taxa [6], it has been an especially strong focus in studies of human evolution. Human populations have adapted to a multitude of environments [5], disease landscapes [4, 7, 8], and diets including the ability to digest milk into adulthood [9], fatty acid digestion [10], foraging practices in tropical African rainforests [11], cereal-rich diets [12], and persistence in high-altitude environments [13-15].

By comparison, the factors that contribute to shaping the evolution of non-human great ape species are poorly understood. Genomic tools have contributed substantially to resolving the evolutionary relationships and histories of great ape species, subspecies, and some populations [16-20]. However, these studies generally assumed that neutral evolutionary processes (i.e., genetic drift) largely explain the partitioning of genetic variation in great apes. In particular, population genetic structure has been presented as evidence for allopatric speciation in ‘Pleistocene Refugia,' among gorillas [21], isolation across conspicuous geographic boundaries like rivers [21-23], and separation on different islands [20]. However, a growing body of evidence supports the hypothesis that local adaptation due to natural selection occupies an essential role in shaping the patterning of genetic variation and speciation in great apes [19, 24-26].

Among the great apes, chimpanzees (*Pan troglodytes*) have been particularly well-studied, including analysis of genomes from a representative sample of captive individuals [16, 18, 19] and population genetic studies of natural populations [22-24, 27]. The overall picture from these studies is that the species originated in western equatorial Africa about 1mya. By 500kya in the Middle Pleistocene, two lineages began to diverge from the ancestral *Pan* population: a western lineage composed of the subspecies *P. t. verus* and *P. t. ellioti*, and a central/eastern lineage comprising *P. t. troglodytes* and *P. t. schweinfurthii* (**Fig. 1a**). Major rivers, lakes, and the Dahomey Gap are thought to have acted as dispersal barriers that separate the subspecies to different degrees and timescales, potentially leading to allopatric speciation among chimpanzee subspecies. Among these dispersal barriers, the Sanaga River in Cameroon stands out (**Fig. 1b**). It separates the chimpanzee phylogenetic tree into its two main branches yet remains permeable to occasional gene flow between *P. t. ellioti* and *P. t. troglodytes* [18, 24, 27]. The Sanaga River has likely enabled some degree of allopatric divergence due to genetic drift but the role that natural selection may have played in separating these chimpanzee subspecies remains unknown.

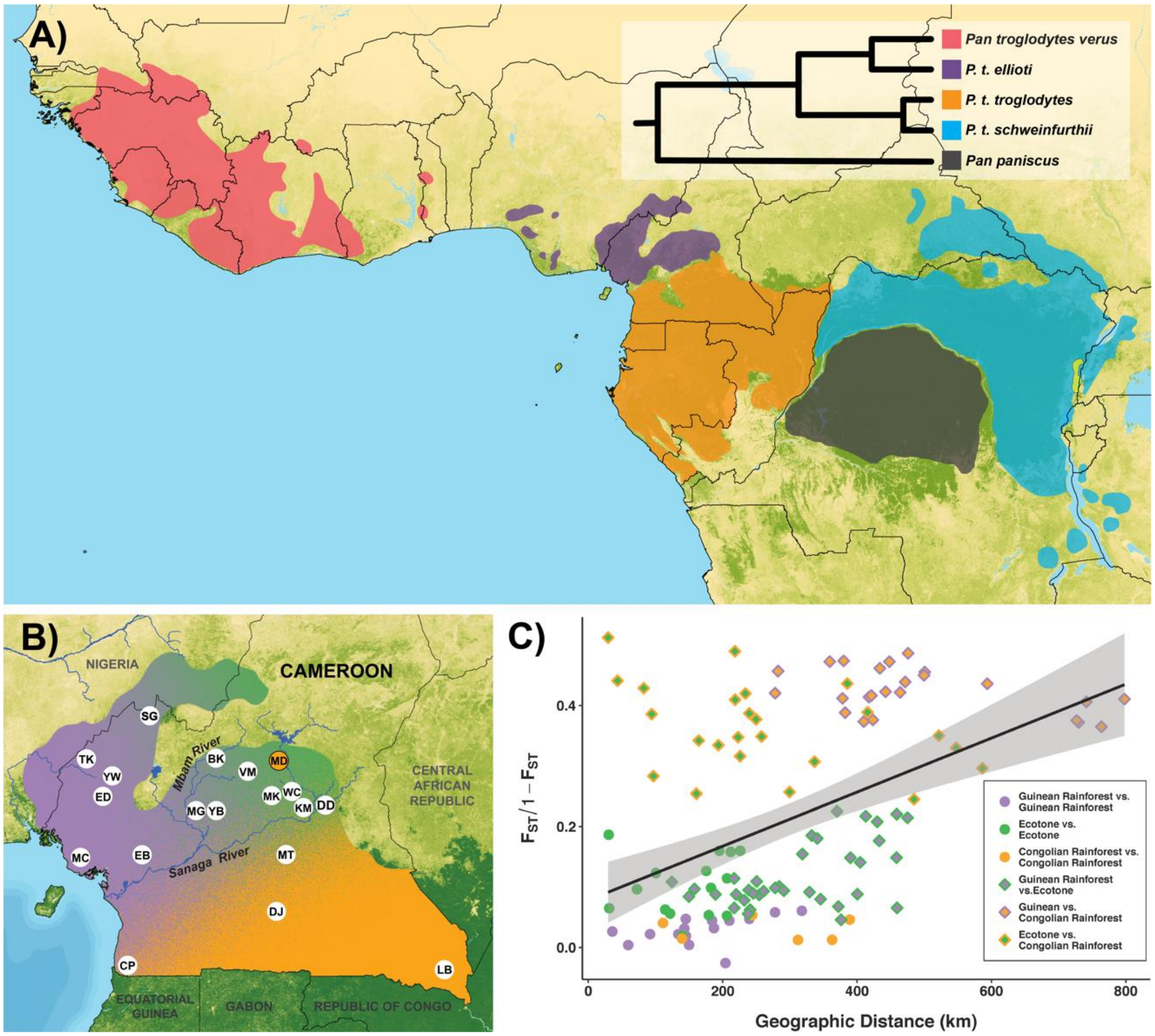

Fig 1. Chimpanzee evolutionary history across Africa and population structure in Cameroon.
(A) Distribution and phylogeny of the genus *Pan*.
(B) Sampling locations of wild chimpanzee populations in Cameroon overlaid on spatial interpolation of population structure using SNPs from wild chimpanzees. The 'MD' sampling location is shaded orange to signify the presence of a *P. t. ellioti*/*P. t. troglodytes* F1 hybrid (CMMD06).
(C) Isolation-by-environment in wild chimpanzees in Cameroon. Correlation between 'linearized F_{ST}' and geographic distance (km) generated using SNPs from wild chimpanzees. Solid circles represent pairs of sampling locations from the same habitat. Dual-colored diamonds represent pairs of sampling locations from different habitats. Colors correspond to chimpanzee population origin: purple – *P. t. ellioti* (Rainforest), green – *P. t. ellioti* (Ecotone), and orange – *P. t. troglodytes*.

Natural selection has numerous opportunities to contribute to genetic divergence that may vary between subspecies or populations in different habitats. Life history traits and pathogen defense stand out as likely candidates for establishing among-population divergence due to local adaptation Among these, the role of pathogens is best understood. Differences in pathogen presence and prevalence have long been associated with genotypic differences among great apes, especially chimpanzees. For instance, wild chimpanzee populations are infected to different degrees with several disease-causing pathogens, including malaria [28], Ebola [29], and viruses like simian immunodeficiency virus (SIV) [30]. In the case of SIV and similar viruses, it is relatively well established that these pathogens have exerted selective pressure on chimpanzees, particularly the central and eastern subspecies [31-33]. Interestingly, Cameroon is a unique disease landscape for chimpanzees, especially concerning the puzzling distribution of SIVcpz. Unlike *P. t. troglodytes* and *P. t. schweinfurthii*, SIVcpz has not been found in *P. t. ellioti* or *P. t. verus*, despite extensive sampling [34-36] (**Fig. 1a**).

Secondly, each chimpanzee subspecies occupies a distinct set of environmental niches [37], creating opportunities for adaptation to local environmental conditions. Although little is known about the links among genotypes, phenotypes, and environmental conditions, chimpanzees in arid environments are more efficient in salt removal than their counterparts in more humid forested environments [38]. However, the role of local adaptation to specific environments remains largely unexplored yet is perhaps the most intriguing avenue of investigation in their evolution. Chimpanzees, like humans, have complex social systems and behaviors and maintain diverse cultural traditions [39]. Similarly, cultural variation among chimpanzee communities may lead to localized gene-culture co-evolution, potentially facilitating adaptation to diverse habitats [37] that are vulnerable to human encroachment [40]. Habitat variation and resource availability, specifically food types, are also known to affect chimpanzee socioecological patterns directly [41], yet whether this variation translates into heritable genetic differences remains speculative.

We investigated how local adaptation has influenced the evolution of chimpanzees in Cameroon, a key region where the western and central/eastern lineages of chimpanzees converge. Despite the wealth of research on the contributions of neutral evolutionary processes to the genetic variation found in wild chimpanzees, the contribution of natural selection remains a significant knowledge gap that our study aimed to fill. We employed a two-tier approach to identify genic regions under selection from a comprehensive analysis of natural chimpanzee communities sampled intensively across Cameroon. First, we used whole-genome sequencing (WGS) data from 24 previously published chimpanzee genomes [16], along with eight newly sequenced genomes of individuals from Cameroon to create and annotate a map of genomic regions under natural selection from this expanded sample of complete genomes of chimpanzees originating from Cameroon. Second, we used the analysis of this expanded sample of genomes to create a genome-wide panel of ancestry-informative putatively neutral SNPs, as well as SNPs that fell within signals of positive selection (inferred with the WGS data) and, thus, were good candidates for performing tests to assess local adaptation. We genotyped these SNPs in 112 wild chimpanzees sampled across multiple habitats in Cameroon, encircling the contact zone between *P. t. ellioti* and *P. t. troglodytes*, and that represent the diversity of habitats occupied by chimpanzees across the contact zone [42], including the northern extent of the Congo Basin Forest, the lowland and montane Gulf of Guinea Forest, and the forest/savanna ecotone that bridges these two forest ecosystems. Finally, we used these SNP panels to investigate the relationship between individual SNPs and a suite of SNPs representing the genome to understand the relationship between allele frequencies and environmental variability. Our objective was to assess whether environmental pressures from differing ecologies have influenced allele frequency variation across these wild populations.

Results

We used WGS data from 24 previously published chimpanzee genomes [16], along with eight newly sequenced genomes of individuals from Cameroon from the Limbe Wildlife Center referred to hereafter as ‘captive chimpanzees’ (**Fig. S1** and **Table S1**). We used the captive chimpanzee dataset and previously published data [16] to create an annotated map of genomic regions under natural selection. Second, we used a genome-wide panel of SNPs in 112 wild chimpanzees sampled across multiple habitats across Cameroon (**Fig. 1b**) to develop a high-resolution, spatially explicit map of allele frequencies to understand the link between habitat variation and loci under selection.

Captive chimpanzee genome analysis and SNP discovery

Developing SNP datasets

We identified SNPs from 32 chimpanzee genomes across all four subspecies, which included eight newly sequenced genomes from the contact zone between the western and central/eastern chimpanzee lineages. After quality filtering, we retained

12,754,225 high-quality SNPs. Based on this initial whole-genome SNP set, two datasets were created. The first dataset was thinned for linkage disequilibrium (LD), retaining only SNPs with $r^2 \leq 0.1$, which resulted in 1,113,142 SNPs retained. The second dataset was thinned to include only SNPs that followed our neutrality criteria (**Methods**), resulting in 147,700 SNPs. **S1 Text** provides additional details on heterozygosity (**Fig. S2**) and population cluster analyses (**Figs. S3 and S4**).

Genome scans for signals of selection and defining genomic 'outlier' regions

To calculate a test statistic for cross-population extended haplotype homozygosity (XP-EHH) and integrated haplotype score (iHS), SNP-based results were summarized into windows following Pickrell et al. [43], but chromosomes were split into 100kb windows and SNPs were binned in 100 SNP increments. We merged windows indicating positive selection for each method and population. The analysis identified regions specific to the two lineages, and those shared among the Western and Central/Eastern lineages were analyzed separately. **Table 1** summarizes outlier regions for the XP-EHH and iHS and combined outlier tests. In the Western lineage, we found 335 outlier windows stretching 83.5 Mb with 695 candidate genes. The Central/Eastern lineage had 318 windows stretching 81 Mb with 682 genes. We found 25 windows over 13.6 Mb with 80 candidate genes shared between lineages. We plotted the distribution of the outlier regions on individual chromosomes (**Fig. 2a**).

Table 1. Summary of captive chimpanzee whole genome "outlier" regions.

	XP-EHH		iHS		Combined		
	Western lineage	Central/ Eastern lineage	Western lineage	Central/ Eastern lineage	Western lineage	Central/ Eastern lineage	Shared
Windows found	257	270	118	95	335	318	25
Base pairs in windows	53,753,675	55,149,110	34,000,000	32,200,000	83,453,675	81,049,110	13,600,085
Protein coding genes used in enrichment analysis	563	563	152	144	695/610	682/593	80/70

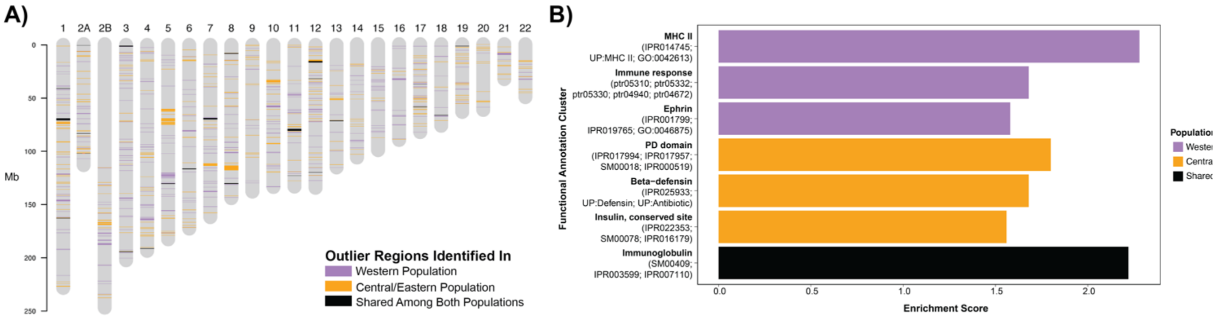

Fig 2. Natural selection in chimpanzees.
(A) Regions under selection found using captive chimpanzee genomes plotted on individual chromosomes.
(B) Functional enrichment clustering of genes and pathways under selection in chimpanzees. Enriched functional annotation clusters (based on genes in outlier regions) including their respective enrichment score. The name of one functional annotation of each cluster was taken to represent the complete cluster.

While all chromosomes are affected by selective sweeps some chromosomes show more regions under selection in one lineage or the other. The most extreme example in chimpanzees is chromosome 20, showing 6 times as many genetic regions under selection in the Central/Eastern lineage than in the Western lineage. Less extreme examples are found on chromosomes 8, 9, 13, and 19 with 2-fold more genome space showing evidence of selective sweeps in the Central/Eastern than the

Western lineage. In the Western lineage chromosomes 16, 15, 18, 3, and 11 show 5-, 4-, 3-, 2-, and 2-fold more genome space to be under selection than the Central/Eastern lineage, respectively.

While the number of regions under selection in the Western as well as the Central/Eastern lineage was equal on chromosomes 1, 2A, 2B, 4-8, 10-14, and 17, there were some differences in the remaining chromosomes. Chromosomes 20 and 21 in the Central/Eastern lineage had five and four times more regions affected by selective sweeps than the Western lineage. Chromosomes 9, 19, and 22 showed two times more regions. In the Western lineage chromosomes 16 and 15 exhibited four and three times more regions under selection than chromosomes in the Central/Eastern lineage. Chromosomes 3 and 18 showed two times more regions under selection in the Western lineage compared to the Central/Eastern lineage. There was no evidence for selection shared between both lineages on chromosomes 2B, 9, 10, 14-16, and 20, 21-22.

Functional annotation, enrichment, and cluster analysis of outlier regions under selection

We analyzed annotated outlier regions with complete or partially overlapping genes and other genetic content (*e.g.*, non-coding genes, pseudogenes). Both lineages had significantly more protein-coding and non-coding genes than randomly sampled genome regions (one sample t-test, *p*=0.0001 & *p*=0.0001) (**Table 2**). Additionally, the number of non-coding genes (ncRNA) was also significantly higher (one sample t-test, *p*=0.0001 & *p*=0.0016), while the number of pseudogenes showed no significant differences (one sample t-test, *p*=0.3023 & *p*=0.6518) (**Table 2**). Closer inspection of the most enriched regions (**Table S2**) revealed these contained mostly protein-coding genes and ncRNAs. The region with the highest significance value in the Western lineage carries exclusively ncRNAs and one window did not contain any annotated genetic features at all.

Table 2. Genetic content of 200 kb windows under selection and ten randomly sampled genome regions.

POPULATION	Real Regions	Random Regions (n=10)		*p*-value[a]
WEST		**Average**	**StDev[b]**	
Protein coding genes	500	400.8	31.84	0.0001
ncRNAs	178	141.1	18.60	0.0001
Pseudogenes	5	6.2	2.90	0.3023
CENTRAL/EAST				
Protein coding genes	536	401.1	30.65	0.0001
ncRNAs	155	136.7	24.00	0.0016
Pseudogenes	5	6.4	2.72	0.6518

[a] *p*-values were obtained using the “one sample t-test”.
[b] StDev: Standard Deviation

We examined enriched gene ontology (GO) terms in the ‘Biological Processes’ category (**Table S3**) and enriched KEGG pathways (**Table S4**) for genes under selection in the Western lineage, the Central/Eastern lineage, or shared between the two populations. Genes significantly enriched in the Western lineage are involved in developmental processes (hair follicle development, embryonic development, pattern specification, melanocyte differentiation), cellular and metabolic processes, and protein localization and degradation. Enriched KEGG pathways in the Western lineage were mainly related to diseases caused by pathogens or internal dysfunctions, branched-chain amino acids (BCAAs) degradation, and neurological development. The

Central/Eastern lineage genes are enriched for innate immune system response, cellular processes, and wound healing. Enriched KEGG pathways in the Central/Eastern lineage are involved in several diseases affecting the heart muscle and Amoebiasis. The shared dataset showed enrichment only in bone mineralization without any KEGG pathway.

To minimize annotation redundancy and clarify the biological functions in each lineage, we grouped genes into functional clusters based on similar biological meaning, not physical distance [44]. **Fig 2b** and **Table S5** show functional enrichment clusters of genes that were unique to the western group (purple), unique to the central/eastern group (orange) and shared between the western and central/eastern lineage.

We grouped 610 candidate genes from the Western lineage into three clusters. The cluster with the highest enrichment score (ES = 2.3) included four genes (PATR-DOB, PATR-DMB, MAMU-DMA, HLA-DOA) functionally associated with the Major Histocompatibility Complex (MHC) II (**Fig. 3a**). MHC II genes, located on chromosome 6, play a crucial role in the adaptive immune response by activating CD4 T cells to respond to extracellular pathogens [45]. The second cluster (ES = 1.8) contains the same four genes as the first cluster, plus gene HLA-DQA1. This cluster is defined by additional gene functions and displays enrichment in additional disease pathways active in diseases like Asthma, Graft-versus-host disease, Allograft rejection, type I diabetes mellitus, and the intestinal immune network for IgA production [45]. The third cluster (ES 1.58) contains three genes containing the Ephrin receptor-binding domain. These three genes (EFNA4, EFNA3, EFNA1) form a gene cluster on chromosome 1 from position 133,320,040 to 133,391,332. Depending on the context, Eph signaling pathways are key determinants of neurological development, cell morphogenesis, tissue patterning, angiogenesis, and neural plasticity [46, 47].

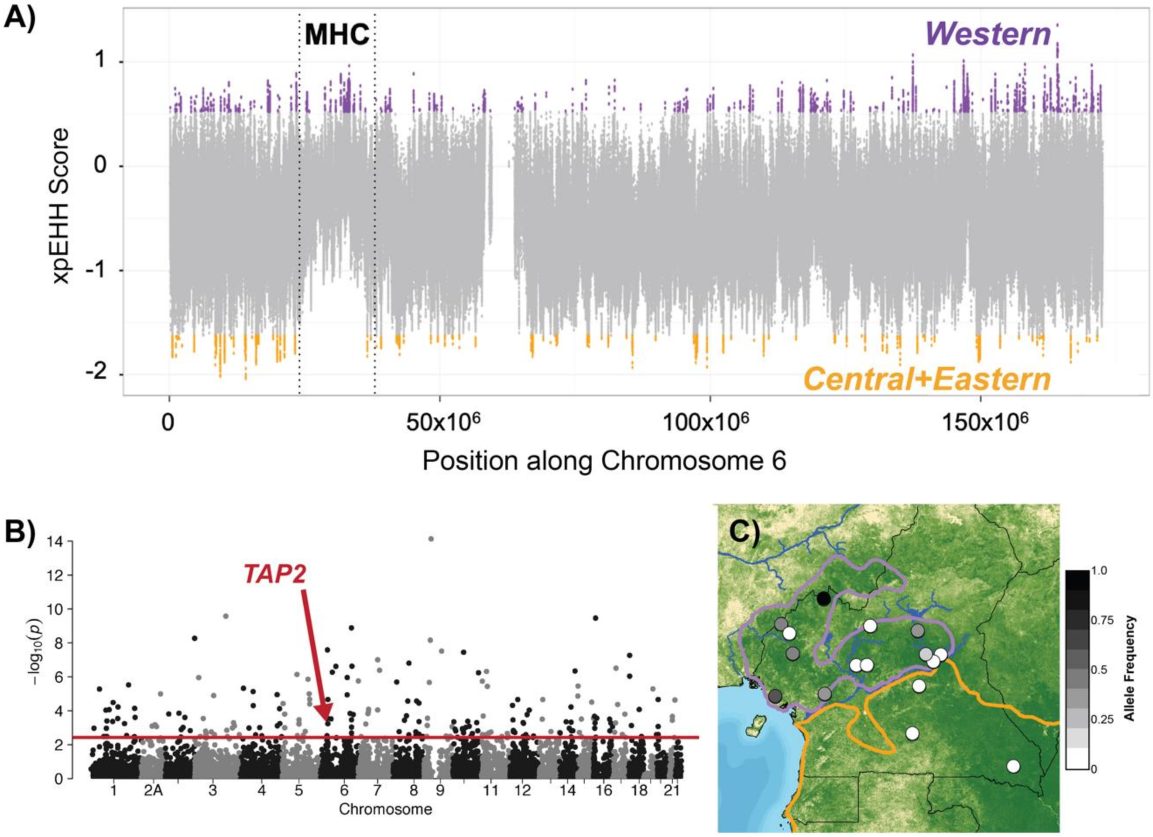

Fig 3. Genome-wide variation of immune response genes under selection in chimpanzees.
(A) XP-EHH analysis of SNPs on chromosome 6 from whole genome sequences of captive chimpanzees. Colored points represent SNPs within the 1% tail of the XP-EHH scores across the genome. The entire MHC region is notated, showing SNPs in MHC genes under selection in the Western lineage (*P. t. verus* and *P. t. ellioti*).
(B) Manhattan plot shows the genome-wide significance level (solid red line) for SNPs associated with Normalized Difference Vegetation Index (NDVI) - Brown with the *TAP2* SNP noted.
(C) Map of allele frequencies for the *TAP2* SNP superimposed onto NDVI and chimpanzee subspecies ranges in Cameroon.

In the Central/Eastern lineage, 593 genes were analyzed, forming three functional clusters. The first (ES=1.8) showed enrichment of three genes (TFF3, TFF2, TFF1) with a PD (or trefoil) domain. These three genes form a cluster on chromosome 21, but their functions are not understood: The peptides coded for in these segments are in several tissues but are most abundant in the GI tract where they may stabilize the mucosa and promote healing [48]. The second cluster (ES=1.7) contained five genes belonging to the beta-defensin gene including DEFB126, DEFB127, DEFB129, and DEFB132 are located on chromosome 20, and DEFB125 on chromosome 8. As antimicrobial peptides they are important in the innate response, including resistance of epithelial surfaces to microbial colonization and encapsulating viruses [49]. The third cluster (ES=1.6) comprises genes INS, RLN3, and INSL6, all sharing an Insulin-like domain.

Functional enrichment analysis of genes shared between both lineages revealed only one cluster of six genes with an enrichment score of 2.2: IL1RL2 and IL18RAP form a gene cluster on chr2A, CNTN6 is located on chr3, and ROBO3, ROBO4, and HEPACAM form a gene cluster on chr11. These genes are all annotated with an Immunoglobulin-like domain.

Wild chimpanzee SNP genotyping, population structure, and selection analysis

Sequence analysis, filtering, SNP calling, and on-target read assessment

We isolated DNA from fecal samples collected non-invasively from unhabituated natural communities of chimpanzees sampled across Cameroon. For 192 of these samples, we obtained 412,081,940 raw reads from single Illumina HiSeq PE125 lane – an average of ~2.15 million reads per sample. In total, 275,443,720 of these reads mapped to the chimpanzee reference genome; from these, we removed approximately

75 million reads and were left with 38,657,083 reads that mapped to our target sites (**Fig. S5a**) – an average of 201,339 on-target reads per sample. The 9,986 targeted sites had a mean read depth of 20x with one site showing as much as 166x coverage (**Fig. S6**). After removing samples for missing data and relatedness, we were left with two datasets ('10k' and '1k'). The '10k dataset' samples had significantly more on-target reads per sample than the total dataset; an average of 328,863 on-target reads per sample, representing ~16% of the total reads from these samples (**Figs. S5b** and **S5d**). The '10k dataset' filtering process resulted in 7,878 SNPs and 112 samples, and all samples from Boumba Bek (BB) and Campo Ma'an (CP) were removed. To retain more geographic representation of samples from at least one of these sites, we created another dataset ('1k dataset') by applying a more stringent site filter and the same individual missingness filter above which resulted in 994 SNPs and 142 samples (including two individuals from CP, but none from BB). The '1k dataset' samples had an average of 268,773 on-target reads per sample, representing ~12% of the total reads from these samples (**Figs. S5b** and **S5f**).

Testing for isolation-by-distance and isolation-by-environment

We found that pairwise F_{ST} values between sampling locations from different habitats were significantly higher than pairwise F_{ST} between sampling locations within the same habitat for both the '10k' (one-tailed *t*-test, *p*-value = 2.2e-16; **Figs. 1c** and **S7a**) and the '1k' dataset (one-tailed *t*-test, *p*-value = 2.1e-16; **Fig. S7b** and **S9a**). Additionally, the geographic distance between sampling locations from different habitats was significantly higher than between locations within the same habitat for all 19 sampling locations included in this study (one-tailed *t*-test, *p*-value = 6.324e-11).

We also performed a permutation test to account for the fact that population structure across habitats can confound the detection of isolation-by-distance (IBD). This categorized population pairs by geographic distance and randomized their habitat origins, forming a null distribution of t-statistics. Using this distribution, we assessed if F_{ST} differed more between than within habitats/populations. For the '10k' and '1k' datasets, we found that F_{ST} was significantly higher between populations/habitats than within them (*p*-value < 0.0001; **Figs. S8b** and **S9b**). This suggests that IBD alone cannot fully explain the high F_{ST} values between populations/habitats. We also ran the permutation test for *P. t. ellioti* sampling locations alone and found that F_{ST} is significantly higher between *P. t. ellioti* (Rainforest) and *P. t. ellioti* (Ecotone) than within them compared to the null distribution (*p*-value = 0.0002; **Fig. S10**). Taken together, the results of the permutation tests suggest that habitat differences play a much stronger role than geographic distance alone, although the signal is slightly stronger within *P. t. ellioti* than in *P. t. troglodytes*. This may be attributed to the fact that *P. t. troglodytes* in Cameroon occupies more uniform Congo Basin forested habitat south and east of the Sanaga River. In contrast, *P. t. ellioti* occupies the comparatively diverse Gulf of Guinea forest comprising lowland forest, montane forest, and the forest-savanna gradient north of the Sanaga River.

Population structure, hybridization, and demographic history

We next investigated population structure, hybridization, and demographic history. Principal Components Analysis (PCA) (**Figs. S11, S12, S13, S14** and **Table S10**), population clustering analysis results (**Figs. S15** and **S16**), and Analysis of Molecular Variance (AMOVA) (**Table S11**) consistently distinguished between *P. t. ellioti* and *P. t. troglodytes*. The results from wild chimpanzee samples were consistent with results from the genome analysis of captive individuals (**Figs. S17**, **S18**, **S19**, **S20** and **S21**), indicating that our SNP discovery approach from captive individuals is likely capturing pockets of genetic differentiation present in wild individuals. In addition, certain individuals showed hybrid ancestry, notably an F1 hybrid in the *P. t. ellioti* population and a potential backcrossed hybrid in *P. t. troglodytes*. The demographic history model indicates that *P. t. ellioti* and *P. t. troglodytes* split from one another around 478,000 years ago, with continuous but rare gene flow between them since splitting, underlining a complex demographic history characterized by significant admixture and evolutionary divergence within the region. **S1 Text** provides more detailed results from these analyses. Based on these results, we concluded that neither the IBD model nor simple allopatric divergence along the banks of the Sanaga River fully explains the separation of *P. t. ellioti* from *P. t. troglodytes*.

Mapping wild chimpanzee genomic variation across habitats

These findings drew our attention to investigating how habitat variation corresponds with neutral and adaptive genetic differentiation among chimpanzees in Cameroon. Using a gradient forest model [50] and 31 environmental predictor variables sourced from publicly available databases (See **Methods**, **SI Text,** and **Table S7**), we quantified environmental associations with genomic loci, pinpointing key environmental drivers and projecting genomic diversity spatially. We identified 581 SNPs with significant environmental associations, representing 6% of all SNPs from wild chimpanzees genotyped in this study. From these, 346 unique candidate genes within 10kb windows of these SNPs, matched outliers from captive chimpanzee genome scans. When mapped to the study, these showed clear signals driven by a phylogenetic split between *P. t. ellioti* and *P. t. troglodytes* across the Sanaga River *and* habitat variation across Cameroon (**Fig. 4c**). Latitude (a proxy for geographic distance) had a pronounced effect along PC1 (**Fig. 4d**). Isothermality and surface moisture also contributed heavily to the model in differentiating between coastal and interior rainforest habitats, as well as rainforest versus ecotone habitats (**Fig. 4c** and **4d**).

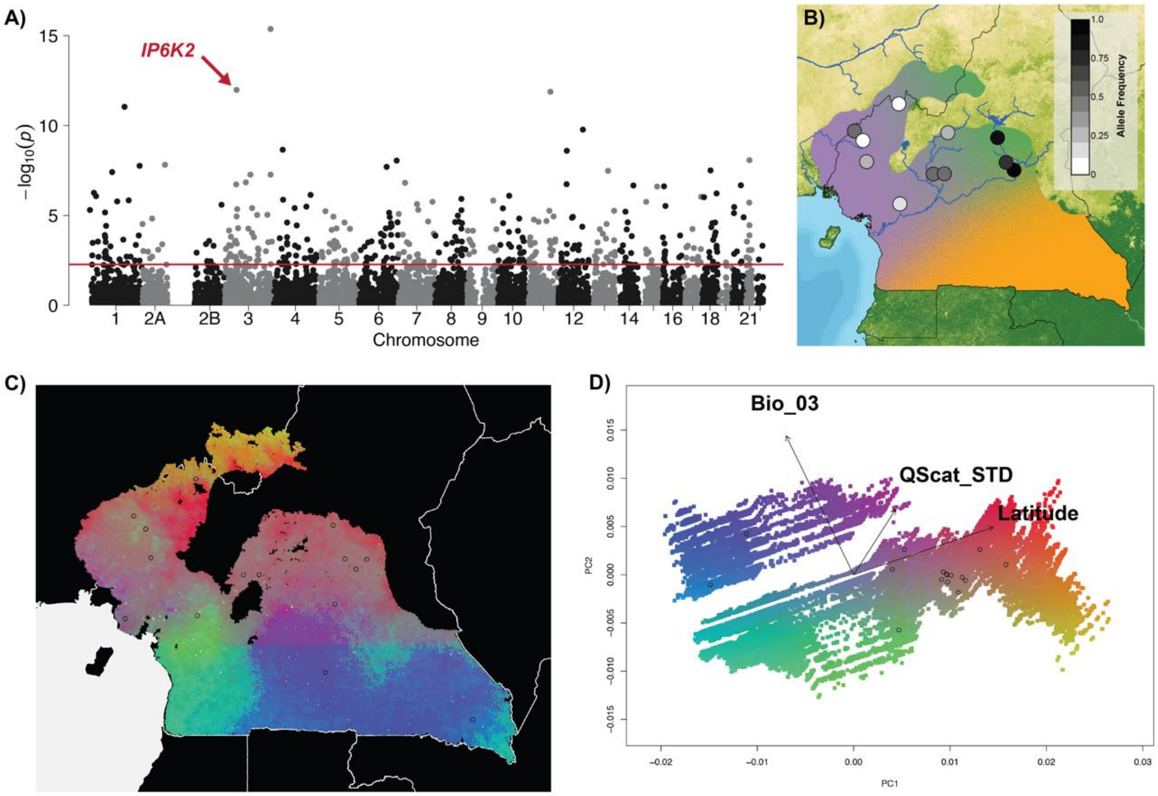

Fig 4. Dietary gene under selection and gene-environment relationships.
(A) Manhattan plot shows the genome-wide significance level (solid red line) for SNPs associated with Annual Mean Temperature (BIO1) with the *IP6K2* SNP noted.
(B) Spatialized allele frequencies for the *IP6K2* SNP showing differentiation between *P. t. ellioti* populations.
(C) Gradient forest-transformed climate variables show climate adaptation across the study area.
(D) Colors are based on a PCA of transformed climate variables.

Among all predictors tested in the model, latitude had the highest R^2 weighted importance, likely reflecting the deep split between chimpanzee subspecies and/or bioclimate turnover across the rainforest-savanna gradient. Precipitation during the dry season and vegetation density were also important for predicting chimpanzee genomic diversity (**Fig. S24**). The second most important axis of variation in the gradient forest model primarily contributed to isothermality (bio3) and surface moisture (QScat_STD). Thus, the variables contributing the most align with a rainforest/savanna ecotone split

(**Figs. 4c** and **4d**), consistent with previous studies of niche modeling [42].

Detecting environmentally associated loci under selection in wild chimpanzee populations

We used Latent Factor Mixed Models (LFMM) to test for signals of selection on individual SNPs in a manner that controls for confounding effects of population structure. We identified 1,690 SNPs significantly associated with one of 31 environmental predictors (**Table S7**) after accounting for population structure (K=3). We then identified 905 unique candidate genes within 10kb windows of the environmentally associated outlier SNPs, all of which were outliers in the captive chimpanzee genomes selection scan. Of the population groupings, we identified 695 associated with General Temperature variables, 388 associated with Temperature Range, 66 associated with Temperature Seasonality, 160 associated with Precipitation (Wet/Cold), 305 associated with Precipitation (Dry/Warm), 456 associated with Surface Moisture, 448 associated with Tree Cover, 355 associated with Vegetation Greenness, 325 associated with Vegetation Brownness, and 428 associated with Topography. A simple Mantel test revealed a significant correlation between pairs of environmental predictor variables and shared outlier SNPs returned by LFMM between variable pairs (Mantel r = 0.425, p = 1.00e-6) demonstrating that independent LFMM performed as expected.

Quantifying environmental relationships with candidate genes

We searched for functional enrichment signals in environment-associated genes in two complementary ways. First, we compared the functions of genes associated with environmental variation with genes that show no signs of positive selection. This comparison helped us determine whether genes influenced by environmental factors and potentially under selection differ functionally from those evolving under neutral conditions. We used genes outside outlier regions from the captive chimpanzee whole-genome analysis as a reference. We identified 47 biological processes enriched in 1,018 unique environmentally associated outlier genes from both the gradient forest and LFMM models (**Table S12**). There were several enrichment clusters, notably two processes functionally associated with immune response, one was related to the Major Histocompatibility Complex (MHC) II – an important part of the adaptive immune system - and eight processes associated with neurological development, including 60 unique genes. We also found 48 enriched KEGG pathways in this subset of outliers (**Table S13**). Key clusters included pathways in neurological development (56 genes), digestion and metabolism (40 genes), and immune response (40 genes).

Our second analysis examined if genes influenced by environmental variation showed functional enrichments compared to those under positive selection without a clear environmental impact. This test uses a much smaller set of background genes composed only of those assayed in wild chimpanzee SNP scan but were not environmentally associated outliers in the LFMM and gradient forest models. Unsurprisingly, the enrichment analysis using this more limited background set of genes resulted in one significantly enriched biological process and KEGG pathway each, both relating to neurological development, specifically axon guidance (**Table S14).**

Of the genes linked with immune response and MHC II, Transporter 2, ATP binding cassette subfamily B member (*TAP2*) stands out. It contains a SNP that significantly associated with Vegetation Brownness (NDVI BRN) ($-\log_{10} = 3.231762344$, $p < 0.001$) (**Fig. 3b**) and is associated with *General Temperature* variables in the LFMM analysis of wild chimpanzees. The *TAP2* SNP in wild chimpanzees is nearly fixed in *P. t. troglodytes* and is variable across *P. t. ellioti* habitats (**Fig. 3c**). Additionally, *TAP2* was found to be under natural selection in the analysis of captive chimpanzee whole genomes, and it is part of the enriched KEGG pathways under selection in *P. t. verus* and *P. t. ellioti* (**Table S4**). *TAP2* is a component of the transporter associated with antigen processing (TAP) complex, which plays a role in ensuring that MHC class I (MHC-I) molecules are expressed on the cell surface [51]. TAP complex proteins, including TAP2, are essential for viral peptide transport from the cytoplasm onto MHC-I receptors within the endoplasmic reticulum [52]. In humans, several *TAP2* gene variants are linked to an increased HIV-1 infection risk [53].

We identified the gene, Leucine rich repeat and Ig domain containing 2 (*LINGO2*), as having one of the strongest associations with the environmental predictor variable, mean annual normalized vegetation index (NDVI), in the LFMM analysis ($-\log_{10} = 8.32330639$, $p = 0.00000000475$) (**Fig. 5a**). A linear regression also revealed a significant association between the allele frequencies of the *LINGO2* SNP and mean annual NDVI ($R^2 = 0.4826$, $p = 0.003495$) (**Fig. 5b**). Low allele frequencies were found in *P. t. ellioti* (Rainforest), with variable frequencies found in *P. t. ellioti* (Ecotone) (**Fig. 5c**). *LINGO2* is highly expressed in human brain tissue [54] and affects synapse development and function [55]. *LINGO2* is also under positive selection in Lidia cattle breed subpopulations and partially drives neurobehavioral phenotype variation among them [56].

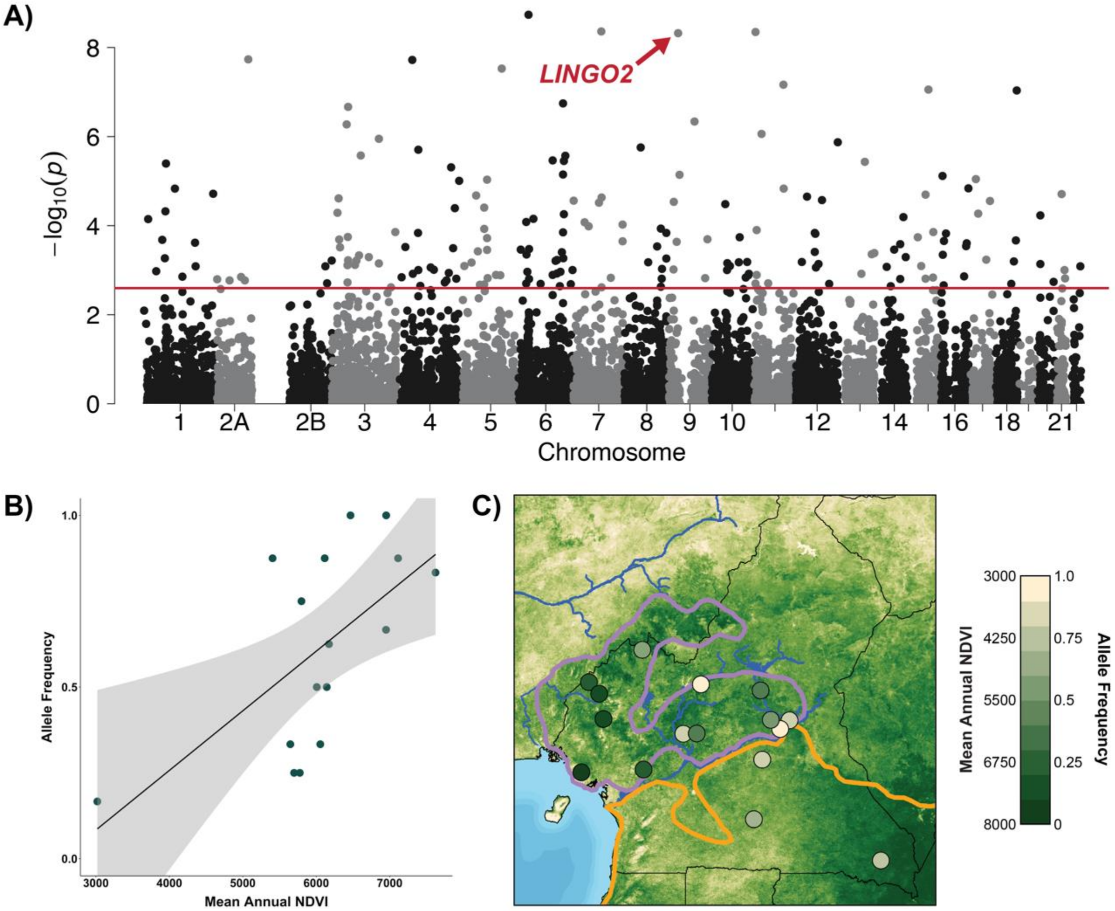

Fig 5. Genome-wide variation of neurological development genes under selection.
(A) Manhattan plot shows the genome-wide significance level (solid red line) for SNPs association with mean annual normalized vegetation index (NDVI) with the *LINGO2* SNP noted.
(B) Correlation between *LINGO2* allele frequency and mean annual NDVI.
(C) Spatialization of allele frequencies for this SNP superimposed onto mean annual NDVI.

Of the 24 digestion and metabolism-related genes identified in the biological processes and KEGG pathways, we further narrowed down our search by using additional measures to quantify relationships of each of the genes with environmental variables and were able to identify two genes with associated SNPs exhibiting significant linear relationships directly with their associated environmental variables across space, suggesting a potential role for diversifying selection across the forest/savanna ecotone gradient. Acetyl-CoA acetyltransferase 2 (*ACAT2*) contains the

SNP at position 161,530,902 on chromosome 6 (**Fig. S26a**), which had the strongest association of all SNPs with temperature seasonality according to results of the LFMM analysis ($-\log_{10} = 3.594887672$, $p = 0.000254163$) (**Fig. S26b**). Linear regression revealed a strong and significant relationship between *ACAT2*'s outlier SNP and temperature seasonality ($R^2 = 0.5651$, $p = 0.0005$) **(Fig. S26c**). When plotting allele frequencies of *ACAT2*'s outlier SNP, higher frequencies were observed in the ecotone's northern sampling sites. Sampling sites within the range of the *P. t. troglodytes* population had lower frequencies of the allele (**Fig. S26d**). The product of the *ACAT2* gene is known to be involved in cholesterol and beta-oxidation lipid metabolism [57].

Another gene identified to have an environmentally associated outlier SNP was Phospholipase C like 2 (*PLCL2*). *PLCL2* contains the SNP at position 17,268,745 on chromosome 3 (**Fig. S27a**). We identified a strong relationship between this SNP and the environmental variable *precipitation of the wettest month* through LFMM analysis ($-\log_{10} = 3.36472742$, $p = 0.00043179$) (**Fig. S27b**). Linear regression revealed a strong and significant relationship between *PLCL2*'s outlier SNP and the environmental variable *precipitation of the wettest month* ($R^2 = 0.3422$, $p = 0.0102$) **(Fig. S27c**). We observed higher allele frequencies of the *PLCL2* SNP in the *P. t. ellioti* ecotone population, with the *P. t. ellioti* rainforest population having the lowest frequencies across Cameroon (**Fig. S27d**). *PLCL2* is associated with obesity in mouse models. Individuals lacking the allele were shown to have a leaner phenotype; were able to resist induced obesity due to increased protection from glucose metabolism disorders and insulin resistance; and exhibited higher energy expenditure [58].

Finally, a SNP in the Inositol hexakisphosphate kinase 2 (*IP6K2*) gene was identified as a significant outlier differentiating the *P. t. ellioti* ecotone and rainforest populations and associated with the environmental variable *Mean Annual Temperature* through LFMM analysis ($-\log_{10}$ = 11.98296666, $p < 0.0001$) (**Fig. 4a**). *IP6K2*'s SNP was significantly more divergent than neutral SNPs between the two *P. t. ellioti* populations (F_{ST} = 0.49, $p < 0.0001$) (**Fig. 4b**). In humans, the *IP6K2* gene is linked with inflammatory bowel disease and cellular response to flavonoids, plant metabolites found in fruits and vegetables [59]. The human KEGG pathway containing *IP6K2* is associated with VACTERL/VATER syndrome, often associated with congenital heart disease and chondrodysplasia [60, 61].

Discussion

We presented genome-wide SNP genotyping from a representative sample of 112 wild chimpanzees from across Cameroon, along with eight newly sequenced genomes of captive chimpanzees to enhance SNP discovery. We supplemented these data by combining genetic analyses with environmental association scans to search for evidence of environmentally-mediated selection. While prior studies have largely concentrated on neutral evolution mechanisms across this contact zone between chimpanzee lineages, our findings support a role for diversifying selection in the divergence of chimpanzee subspecies across different environments. The proposed population history of chimpanzees across this contact zone is well supported in this study. *P. t. ellioti* and *P. t. troglodytes* last shared a common ancestor around 478,000 years ago, with occasional gene flow between them evidenced by an F1 hybrid in *P. t.*

ellioti and a potential backcrossed hybrid in *P. t. troglodytes*. These findings support prior studies suggesting that this contact zone between subspecies best fits an isolation-with-migration population model in which allopatric divergence and positive selection contribute to the partitioning of genetic variation [62].

The evidence supporting a role for environmentally-mediated selection across this contact zone is also compelling. We found 1,690 unique SNPs were associated with at least one of 31 environmental predictors, indicating that prevailing environmental conditions contribute to local adaptation in *P. t. elloti* and *P. t. troglodytes*, and to a lesser extent, among populations within *P. t. ellioti*. These SNPs are distributed among 905 outlier genes enriched for 48 biological processes. Overall, the sets of genes with highly divergent allele frequencies that separate *P. t. ellioti* from *P. t. troglodytes* suggest a role for selection in pathways important in two main categories: immune response and life history traits (neurological development, behavior, and dietary function).

It is important to reiterate that all outliers identified in wild chimpanzees using LFMM-based approaches were also identified as outliers in the haplotype homozygosity selection scans of captive chimpanzee genomes. This two-tiered approach offers heightened reliability of selection scans in wild populations while mitigating the incidence of false positives in our final dataset. Moreover, the congruence of these identified genomic regions between the two methods and two complementary datasets suggests that these outliers are subject to positive selection and not merely an artifact of demographic history or neutral population genetic structure.

Signatures of selection from variable pathogen histories

The lack of natural SIVcpz infection in *P. t. ellioti* sparks interest because it persists despite opportunities for transmission. SIVcpz*Ptt* virus infects *P. t. troglodytes*, crossed the species barrier on at least four occasions: from chimpanzees to humans in southern Cameroon, giving rise to the HIV-1 group M pandemic and to HIV-1 group N [63, 64]. HIV-1 group O and P also arose from transmission from chimpanzees to gorillas before subsequent transmission to humans [65, 66]. Thus, SIVcpz can cross genus boundaries which makes its absence in *P. t. ellioti* particularly striking since this subspecies still exchanges occasional migrants with *P. t. troglodytes*. Finally, the presence of prey primate species that harbor endemic SIV strains also creates multiple pathways for cross-species transmissions [34, 36, 67, 68]. Thus, we speculate that the absence of an SIVcpz in *P. t. ellioti* must be at least partially explained by adaptations that interrupt SIVcpz cell entry and/or boost immune response to clear SIVcpz infection.

Four processes functionally associated with the Major Histocompatibility Complex (MHC) II on chromosome 6 play a crucial role in the adaptive immune response. MHC II peptides stimulate CD4+ T cells that activate downstream immune responses to intracellular pathogens, including viruses. In particular, the Th1/Th2 cell differentiation pathway determines the type of helper cell a CD4+ T cell will become. Naïve CD4+ T cells recognize an MHC class II molecule, activate, and divide to produce clone effector CD4+ T cells specific for a particular antigen. CD4+ T cells can differentiate into T helper type-1 (Th1), T helper type-2 (Th2), or other T helper types, each with distinct cytokine-secretion phenotypes, production of distinct interferons, and different downstream immune responses. This finding corresponds well with a growing body of evidence that positive selection associated with pathogen defenses has contributed to the genetic and phenotypic differentiation of chimpanzee subspecies, especially *P. t. troglodytes* and *P. t schweinfurthii*, which is perhaps due to exposure to different viruses [32].

This finding naturally called our attention to the absence of SIVcpz in *P. t. ellioti* attributed to a lack of gene flow between *P. t. ellioti* and *P. t. troglodytes* [30]. Given that gene flow occurs between *P. t. ellioti* and *P. t. troglodytes*, and that SIVcpz*ptt P. t. troglodytes* is the source of multiple cross-species infections in both humans and gorillas, it is logical to assume that SIVcpz should naturally infect *P. t. ellioti*. We observed evidence of positive selection in *P. t. ellioti* due to highly differentiated SNPs enriched in genic sites. Among these, *TAP2* (**Fig. 3c**) variants increase the risk of HIV infection in humans [53], and may have a similar function in chimpanzees. Given the low level of gene flow, and the absence of sequence data upstream or downstream of the *TAP2* in our data, we cannot conclude whether this is evidence for recent adaptation to SIVcpz*ptt* or evidence of ancient selection resulting from exposure to SIV-like viruses. Evidence is mounting that chimpanzees have had a long and continuing relationship with SIV-like viruses such that differences in viral exposures and immune responses have likely been a central feature of the evolution of chimpanzees [32, 33, 69].

For instance, *P. t. troglodytes* and *P. t. schweinfurthii* also show evidence of recent positive selection in genes involved in SIV/HIV cell entry and immune response to SIV, biological pathways responsible for T-helper cell differentiation, including CD4 [33], and multiple genes that SIV/HIV use to infect and control host cells including CCR3, CCR9 and CXCR6 [32]. There is also compelling evidence of past selective sweeps leading to reduced diversity in the MHC II repertoire of *P. t. verus* that has been attributed to past infections with SIV or SIV-like viruses [70]. Although we cannot speculate further given the nature of the data in this study, our findings add to the mounting evidence that chimpanzees have experienced long-lasting host-virus relationships with SIV-like viruses and that these relationships have been a critical process underpinning their evolution. More detailed investigations are needed on whether the positive selection in *P. t. ellioti* is due to past, recent, or ongoing infection with SIVcpz and/or related viruses.

Signatures of selection across variable habitats

Cameroon is also a uniquely positioned 'natural laboratory' to examine the relative contributions of neutral evolutionary forces versus natural selection in the evolution of many animals, including chimpanzees. In addition to being home to the Sanaga River, a well-known biogeographic boundary for many species, the country is exceptionally ecologically diverse. We speculate that the area is important for understanding how habitat variation and behavioral diversity may impact chimpanzee evolution. The Congo Basin Forest in the south, the Gulf of Guinea Forest in the west, and the Sahelian habitats in the north of Cameroon all converge and interdigitate to form a unique ecotone habitat composed of open woodland, savanna, and riparian forest [71]. This ecotone is a known engine of diversification for many species [72-78]. Differences between *P. t. ellioti* and *P. t. troglodytes* have been linked with habitat variation across Cameroon [62], which suggests a possible role of local adaptation in their genetic differentiation. Finally, there is a further genetic distinction within *P. t. ellioti* itself, with one gene pool associated with the mountainous rainforest in western

Cameroon and the other with the ecotone in central Cameroon [24] (**Fig. 1b**). Each gene pool has a unique ecological niche [42, 79] with marked differences in key socioecological variables, including sex-specific differences in community structure [80] and dietary preferences [81].

We observed compelling evidence for positive selection that distinguished *P. t. troglodtyes* from *P. t. ellioti* across this contact zone. We also found evidence of diversifying selection that distinguished *P. t. ellioti* populations that occupy different niches [42], which adds strength to our previous findings that both allopatric divergence due to genetic drift and environmentally-mediated local adaptation contribute to sustaining the prolonged separation of these two subspecies across this narrow contact zone between them. In particular, we found 246 genes involved in cellular, metabolic, and developmental processes were associated with one or more of the 31 environmental predictor variables. Genes with the most divergent allele frequencies were associated with latitudinal variation and separate *P. t. ellioti* from *P. t. troglodytes*. We detected an additional more subtle signal of positive selection among *P. t. ellioti* chimpanzees located in western Cameroon's mountainous, forested regions compared to the population inhabiting central Cameroon's drier ecotone forests. Environmentally driven pressures between habitats shape adaptive variation, especially between rainforest and ecotone habitats. Chimpanzees in these two different habitat types were previously identified as distinctive ecological populations occupying unique niches [42, 79]. The two ecological populations also exhibit distinctive differences in diet and nesting preferences [81, 82], key elements of chimpanzee cultural diversity.

Thus, multiple phenotypic axes appear to be under environmentally mediated selection that can be linked to habitat variation and variation in chimpanzee socioecology. Genes under environmentally mediated selection associated with neurological development (e.g., *LINGO2*) could be shaped by selective pressures linked to the development of cultural traits in diverse habitats [37], while those associated with diet and metabolism are likely shaped by pressures related to fruit availability and seasonality [81]. One of the most compelling findings of our study is the identification of 24 genes related to digestion and metabolism with the strongest signals of environmentally-mediated selection, including *ACAT2*, *PLCL2*, and *IP6K2*, which present promising avenues for future research.

This study adds to the emerging evidence that neutral evolutionary forces alone cannot explain the prolonged persistence of the separation of *P. t. ellioti* from *P. t. troglodytes* across the narrow contact zone between them. Local adaptation to prevailing conditions has led to divergence in sets of genes important in immune response, neurological development, behavior, and dietary function. Together, these findings suggest that local adaptation, notably to varying pathogen pressure and different habitat types, has shaped chimpanzee subspecies differentiation in Cameroon, and likely across their broad range. Future studies exploring how these genetic differences map to phenotypic differences in wild populations are needed to better understand precisely which traits — particularly those associated with pathogen defense, diet, social organization, and other aspects of chimpanzee cultural diversity — provide for local adaptation and divergence among chimpanzee populations across this region.

Materials and Methods

Captive chimpanzee genomes

Sequencing and read mapping

Our captive chimpanzee genomic dataset includes 24 previously sequenced [16] and 8 new chimpanzee genomes from Cameroon, representing all four subspecies: 4 *P. t. verus*, 10 *P. t. ellioti*, 12 *P. t. troglodytes*, and 6 *P. t. schweinfurthii* (**Table S1**). These eight genomes were sequenced using established methods [16] and deposited in GenBank. Details on the samples, the estimated origins of the captive chimpanzees [83], and GenBank accession numbers are in **Fig S1** and **Table S1**. We mapped raw sequencing reads against the chimpanzee reference genome Pan_troglodytes-2.1.4 (*panTro4*; https://www.ncbi.nlm.nih.gov/assembly/GCF_000001515.6/) using BWA-MEM v0.7.12 [84] with default parameters. After removing PCR duplicates using PICARD v1.119 (https://broadinstitute.github.io/picard/index.html), we called variants using FREEBAYES v0.9.20 [85]. After filtering, 12,754,225 high-quality bi-allelic SNPs on the autosomes were retained.

Genome scans for signals of selection

We divided SNP datasets into a ‘Western lineage’ (*P. t. verus* & *P. t. ellioti*; *n*=15) and ‘Central/Eastern lineage’ (*P. t. troglodytes* & *P. t. schweinfurthii*; *n*=17). We applied two selection scan methods, cross-population extended haplotype homozygosity (XP-EHH) [86] and integrated haplotype score (iHS) [87] to detect sweeps using hapbin v1.2.0 [88]. Since both tests require haplotypes, we phased the whole-genome SNP datasets (12,754,225 SNPs) with SHAPEIT v2.r837 [89] following established methods [18]. Genetic maps for *panTro4* were provided by de Manuel & Kuhlwilm *et al.* [18] and Auton & Feldel-Alon *et al.* [90]. iHS calculations used SNPs with a minor allele frequency (MAF) over 5%. We determined the ancestral state of each allele using the 6-primate EPO alignment (ftp://ftp.ensembl.org/pub/release-80/fasta/ancestral_alleles/) [91, 92]. After phasing, ancestral allele assignment, and MAF filtering we used 4,577,055 SNPs in the Western lineage and 6,475,338 SNPs in the Central/Eastern lineage for iHS. XP-EHH scores compared both lineages using 12,450,633 SNPs, and the results were normalized across the genome.

Defining population-informative neutral SNPs

Using normalized XP-EHH and iHS values, we identified SNPs expected to follow neutral evolution that met the following criteria: (*i*) a *p*-value of > 0.05, (*ii*) absent from the top 1% genomic regions under selection (see *Defining genomic 'outlier' regions in captive chimpanzees*), (*iii*) be located >10kb from a gene, and (*iv*) be in linkage equilibrium. Using these parameters, we defined 147,700 neutral SNPs reflecting chimpanzee population structure. We annotated these using the Variant Effect Predictor (VEP) v82 and the UpDownDistance plugin [93].

Defining genomic 'outlier' regions

To understand the amount of selection on the genome, we considered numbers of base pairs under selection (magnitude) and the size of regions affected (genome space). We employed iHS [87] and XP-EHH [86] to detect signatures of local positive selection. Both assume that long-range haplotypes remain unaffected by recombination, signifying natural selection even with small datasets [43, 94]. They are also complementary: while iHS detects partial sweeps, XP-EHH identifies near-fixation events. Following Pickrell *et al.* [43], chromosomes were split into 100kb non-overlapping windows, and the fraction of SNPs with |iHS| > 2 and the maximum XP-EHH was used as a test statistic. We analyzed the fraction of SNPs with |iHS| > 2 and the maximum XP-EHH per window. We turned these into empirical *p*-values by binning windows by SNP count, with iHS dropping windows with < 100 SNPs. Each window's statistic value was compared against others in its bin to determine an empirical p-value. All bins were then sorted by this *p*-value. The top 1% of each test statistic was noted. ‘Outlier regions’ were windows in this 1% (*p*-value < 0.01). Adjacent windows were merged, retaining the smallest *p*-value.

Characterizing genomic regions under selection

As XP-EHH and iHS are complementary, we analyzed the 1% tail of each test merging adjacent windows. Windows were extended by 50kb on either side and annotated for gene content using Ensembl’s BioMart [95], including protein-coding genes, pseudogenes, and RNA coding genes. Genes within outlier regions were considered candidate genes. We tested whether these genomic regions carry certain types of gene content more often than expected by chance by randomly selecting regions equivalent in length and annotating them as described above. We repeated this process 10 times for the Western and Central/Eastern populations, respectively. We counted the number of protein coding genes, non-coding RNAs, and pseudogenes in the real and randomized datasets. We then calculated mean and standard deviation for each and performed a one-sample t-test to determine significance.

Functional annotation and enrichment analysis of whole-genome datasets

We used DAVID Bioinformatics Resources v6.8 [44] to annotate candidate genes and perform an enrichment analysis with default functional annotations (GO terms, KEGG pathways, protein domains). We concentrated on the 'functional annotation clustering' function using the highest classification stringency and adjusted the enrichment thresholds for EASE to 0.05, reducing non-significant term inclusion. This clustering reduces redundancy by grouping similar annotations. Clusters received a Group Enrichment Score based on their *p*-value, ranking their biological importance. High scores likely mean lower *p*-values for annotation members [44]. We omitted windows found in both Western and Central/Eastern lineages, analyzing them separately. We set a background gene population as the entire chimpanzee genome, as recommended for genome-wide studies [44].

Wild chimpanzee SNP genotyping, population history, and selection analysis

Fecal sample collection, DNA extraction, and quantification

We sampled wild, non-habituated chimpanzee populations using non-invasive methods during a series of field studies from 2003 to 2015 spanning remote forested regions of Cameroon (**Fig. 1b**). Sampling occurred in protected and unprotected areas, as detailed in **Table S6.** Chimpanzee fecal samples were collected and stored following established protocols [24]. All samples were transported from Cameroon to the United States in full compliance with the Convention of International Trade in Endangered

Species of Wild Fauna and Flora (CITES), the Centers for Disease Control (CDC) export and import regulations, and with approval from the Government of Cameroon.

Following established protocols [24], DNA was extracted from fecal samples with the QIAamp DNA Stool Mini Kit (Qiagen). Due to the low proportion of endogenous DNA in fecal gDNA extracts [96, 97], samples were sometimes extracted up to six times to ensure enough chimpanzee DNA for sequencing processes. The concentration of endogenous DNA was measured via quantitative real-time PCR using the Quantifiler™ Human DNA Quantification Kit (Applied Biosystems) and methods from prior studies [24].

SNP ascertainment, library preparation, DNA capture enrichment, and sequencing

We genotyped 9,986 SNPs of wild chimpanzees from Cameroon, chosen from the larger set of 12,450,633 SNPs identified in the captive chimpanzee genome dataset. This selection comprised: (i) population informative neutral SNPs (n=3,492) randomly selected 147,700 neutral SNPs defined in the whole-genome dataset from above; (ii) ‘outlier’ SNPs (n=6,494) identified through iHS or xpEHH tests as being in the top 1% for selection signals and within or 10k bp up- or down- stream of a known gene; and, (iii) SNPs in genes involved in immune response, disease resistance, and dietary adaptation in humans (n=20) [98, 99]. For each targeted SNP, we designed two 80 nucleotide biotinylated RNA probes, overlapping by 20bp, to create 100bp windows around each SNP using the panTro4 chimpanzee reference genome. After rigorous filtration using the Arbor Biosciences BLAST pipeline, we finalized a bait-set of 19,974 probes, assigning one or two probes to each SNP based on the outcome of the stringent filtering process.

gDNA samples were prepared in clean facilities at Arbor Biosciences to prevent contamination. DNA was quantified, sonicated, and size-selected for around 300nt fragment lengths. Samples were converted to sequencing libraries via adapter ligation. They were index-amplified based on the DNA input amount. Up to 2µg of each library was then enriched using the myBaits system v3. Different enrichment and amplification protocols were applied depending on the DNA quantity in the starting extract. Libraries were combined for equal representation, sequenced on an Illumina HiSeq PE125 lane at HudsonAlpha, and protocols were consistent with studies on degraded or low endogenous DNA samples [96, 97].

SNP calling and on-target read assessment

We filtered sequence reads with the Illumina CASAVA-1.8 FASTQ Filter (http://cancan.cshl.edu/labmembers/gordon/fastq_illumina_filter/) and mapped them to the chimpanzee panTro4 genome using BWA-MEM. After removing PCR duplicates with PICARD, we called variants using FREEBAYES. We evaluated DNA capture enrichment and sequencing for the raw and filtered sequence reads using SAMtools v1.3.1 (101) and VCFtools v0.1.15 [100] following established methods (97) for all 192 individuals. Using VCFtools, we filtered variant calls based on quality and coverage. SNPs with <5x coverage or quality score <30 were recorded as missing data [97]. Positions with >60% missing data, a minor allele frequency below 5%, or individuals with >70% missing data were removed. This resulted in 7,878 SNPs from 112 samples, termed the ‘10k dataset’. Due to removing all Boumba Bek (BB) and Campo Ma’an (CP) samples, a second ‘1k dataset’ was made with stricter site filtering, yielding 994 SNPs and 142 samples, which included two from CP but none from BB. We also removed closely related and duplicate samples using the R package related [101] using the triadic likelihood method [102], resulting in 85 individuals in the '10k dataset' and 108 individuals in the '1k dataset.'

Testing for isolation-by-environment and inferring population structure

S1 Text provides full details on methods to test IDB, IBE and to infer population structure, hybridization, and demographic history. In brief, we examined IDB versus IBE using pairwise F_{ST} values between sampling locations using Arlequin v3.5 [103], while geographic distances were determined with the geosphere package in R [104], focusing on areas with more than one individual. Population structure was inferred by PCA and ADMIXTURE analysis [105], DISTRUCT v1.1 [106], CLUMPP v1.1.2 [107], with geographic without pre-assigned population labels using the SNPrelate package in R, focusing on 'neutral' SNPs identified from captive chimpanzee genomes. We mapped genetic clusters using TESS [108] and Ad-Mixer v1.0 [109], accounting for IDB. We calculated observed and expected heterozygosity using the adegenet package in R [110], identified potential hybrids using NEWHYBRIDS v1.0 [111] as implemented in the R packages *hybriddetective* [112] and *parallelnewhybrid* [113], and investigated demographic history using δaδi [114] to model asymmetric migration patterns between *P. t. ellioti* and *P. t. troglodytes.*

Environmental data layers (**Table S7**) were compiled and analyzed to assess habitat suitability and IBE for chimpanzees in Cameroon and Nigeria. These layers, sourced from publicly available databases, included diverse variables such as topography, hydrography, climate, vegetation, moisture content, and tree cover. After standardizing these layers to a 30-arcsecond resolution and converting them to the WGS84 coordinate system, the dataset underwent cross-correlation analysis to pinpoint environmental factors significantly influencing chimpanzee distribution (**Table S8**).

Mapping genomic variation across habitats

Using the R package *gradientForest* [50], we calculated associations between allele frequencies and environmental variation across suitable habitat. This extended random forest model identifies links between response variables (e.g., SNP allele frequencies) and spatial environmental factors [115] by iteratively processing datasets, assessing outliers and predictor significance. Gradient forests further apply regression to multiple responses, revealing genomic variation from environmental shifts. This can pinpoint areas of high intraspecific variation, subspecies transitions, or barriers separating genomic variation related to the environment [116]. Following established methods, we refined the environmental dataset (**Table S7**) to reduce noise and applied the gradient forest model [76].

We ran gradient forests on 7,878 SNP allele frequencies used as a response dataset, and 17 environmental variables as the predictors in our final model, including measures of temperature, precipitation, vegetation, surface moisture, and geographic features at the sampling locations. We ran 100 trees in our model, noting SNPs significantly associated with any environmental variable ($R^2 > 0$) and the average regression of all associated SNPs. To assess model performance, we randomized the environmental data and ran 200 permutations of the model, creating a distribution of R^2

and significant SNP associations. We then ran 200 permutations of the actual model, comparing these distributions. (**Fig. S25**).

Detecting environmentally associated loci under selection

In order to understand the degree to which environmentally driven natural selection may cause chimpanzees to be locally adapted to different habitats, we used latent factor mixed models implemented in the program LFMM v1.5 [117]. LFMM quantifies statistical associations between allele frequencies and environmental variables, accounts for underlying population genetic structure, and detects loci with stronger environmental correlations than population structure. We ran five MCMC replicates for all environmental variable with 25,000 burn-in steps, 100,000 iterations, and a latent factor of *K*=3 from *a priori* knowledge of wild chimpanzee population structure in Cameroon (**Figs. S11, S12, S13, S14 S15, S16, S20** and **S21**). We calculated median z-scores across runs and used them to calculate the genomic inflation factor (λ) and adjusted *p*-values. To correct for multiple testing, we applied a conservative false discovery rate (FDR) of 0.1 using the Benjamini-Hochberg algorithm. We identified unique candidate SNPs linked to at least one environmental variable, presented via Manhattan plots using the *qqman* package [118].

Using outlier analysis with LFMM, we grouped highly correlated environmental variables together to create ‘environmental groupings’ (**Table S9**). These included: General Temperature (n=7), Temperature Range (n=2), Temperature Seasonality (n=2), Precipitation – Wet/Cold (n=4), Precipitation – Dry/Warm (n=4), Tree Cover (n=2), Vegetation Brownness (n=2), Vegetation Greenness (n=3), Surface Moisture Content (n=2), and Topography (n=3).To determine the degree to which types of environmental variation may drive selection of different genomic regions, we identified panels of unique candidates from each ‘environmental grouping.’ We also analyzed the impact of multicollinearity of our environmental predictors on the LFMM results by correlating the degree of association between pairs of environmental predictors and the number of shared outlier SNPs, using a Mantel test in R. (**Table S8**).

Enriched gene ontologies and KEGG pathways

We identified candidate genes near candidate SNPs positions with Ensembl’s BioMart tool [119, 120], including both complete and partial genes within these windows. We used the DAVID database [44] for annotation and enrichment analysis of candidate gene lists focusing on the ‘Biological Processes’ category of the ‘Gene Ontology’ database [121] and KEGG pathways [60] using a *p*-value threshold of 0.05 and two different background populations of genes to control for potential bias since the SNPs assayed in wild chimpanzees were selected from a subset of those identified using whole-genome data from captive individuals. The first background we used was composed of the population of genes found outside regions under selection identified in the whole-genome sequencing data (**Fig 2a**). This resulted in a broad view of environmentally mediated selection in wild chimpanzees by including only genes in putatively neutral regions of the genome. The second background population of genes consisted of all genes assayed in wild chimpanzees, excluding environmental outliers.

Acknowledgments

We thank the government of Cameroon for permission to conduct this research. We thank the Congo Basin Institute, the International Institute for Tropical Agriculture, the Cameroon Biodiversity Association, the San Diego Zoo Wildlife Alliance, the Wildlife Conservation Society, and the World Wildlife Fund for their support in Cameroon. We thank Kevin Njabo, Beatrice Hahn, Martine Peeters, Louis Nkembi, and Amy Pokempner for their assistance in collecting fecal samples in Cameroon. We thank Felice Elefant for the use of her lab's 7500 Real-Time PCR System. We thank Alison Devault and Jacob Enk of Arbor Biosciences for assistance in the experimental design and execution of the wild chimpanzee DNA capture enrichment and sequencing as part of Arbor Biosciences' myReads services.

References

1. Coyne JA, Orr HA. Speciation. Sunderland, Mass.: Sinauer Associates; 2004. xiii, 545, 2 p. of plates p.
2. Nielsen R. Molecular Signatures of Natural Selection. Annu Rev Genet. 2005;39(1):197-218. doi: 10.1146/annurev.genet.39.073003.112420.
3. Moritz C, Patton JL, Schneider CJ, Smith TB. Diversification of rainforest faunas: An integrated molecular approach. Annu Rev Ecol Syst. 2000;31(1):533-63. doi: doi:10.1146/annurev.ecolsys.31.1.533.
4. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. Science. 2006;312(5780):1614-20. doi: 10.1126/science.1124309
5. Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet. 2008;4(2):e32. doi: 10.1371/journal.pgen.0040032.
6. Manel S, Holderegger R. Ten years of landscape genetics. Trends Ecol Evol. 2013;28(10):614-21. doi: 10.1016/j.tree.2013.05.012.
7. Novembre J, Galvani AP, Slatkin M. The Geographic Spread of the CCR5 Δ32 HIV-Resistance Allele. PLoS Biol. 2005;3(11):1954-62. doi: 10.1371/journal.pbio.0030339.
8. Hedrick PW. Population genetics of malaria resistance in humans. Heredity. 2011;107(4):283-304. doi: 10.1038/hdy.2011.16.
9. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genet. 2007;39(1):31-40. doi: 10.1038/ng1946.
10. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science. 2015;349(6254):1343-7. doi: doi:10.1126/science.aab2319.
11. Jarvis JP, Scheinfeldt LB, Soi S, Lambert C, Omberg L, Ferwerda B, et al. Patterns of Ancestry, Signatures of natural selection, and genetic association with stature in Western African Pygmies. PLoS Genet. 2012;8(4):e1002641. doi: 10.1371/journal.pgen.1002641.
12. Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, et al. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci USA. 2010;107(supplement_2):8924-30. doi: doi:10.1073/pnas.0914625107.
13. Bhandari S, Zhang X, Cui C, Yangla, Liu L, Ouzhuluobu, et al. Sherpas share genetic variations with Tibetans for high-altitude adaptation. Mol Genet Genomic Med. 2017;5(1):76-84. doi: 10.1002/mgg3.264.
14. Hanaoka M, Droma Y, Basnyat B, Ito M, Kobayashi N, Katsuyama Y, et al. Genetic variants in *EPAS1* contribute to adaptation to high-altitude hypoxia in Sherpas. PLoS ONE. 2012;7(12):e50566. doi: 10.1371/journal.pone.0050566.
15. Scheinfeldt LB, Soi S, Thompson S, Ranciaro A, Woldemeskel D, Beggs W, et al. Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol. 2012;13(1):R1. doi: 10.1186/gb-2012-13-1-r1.
16. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. Nature. 2013;499(7459):471-5. doi: 10.1038/nature12228
17. Xue YL, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science. 2015;348(6231):242-5. doi: 10.1126/science.aaa3952. PubMed PMID: WOS:000352613700048.
18. de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science. 2016;354(6311):477-81. doi: 10.1126/science.aag2602.

19. Cagan A, Theunert C, Laayouni H, Santpere G, Pybus M, Casals F, et al. Natural selection in the great apes. Mol Biol Evol. 2016;33(12):3268-83. doi: 10.1093/molbev/msw215. PubMed PMID: WOS:000387925300021.
20. Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M, Desai T, et al. Morphometric, behavioral, and genomic evidence for a new orangutan species. Current Biol. 2017;27(22):3487-98.e10. doi: 10.1016/j.cub.2017.09.047.
21. Anthony NM, Johnson-Bawe M, Jeffery K, Clifford SL, Abernethy KA, Tutin CE, et al. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. Proc Natl Acad Sci USA. 2007;104(51):20432-6. Epub 2007/12/14. doi: 10.1073/pnas.0704816105.
22. Gonder MK, Oates JF, Disotell TR, Forstner MR, Morales JC, Melnick DJ. A new west African chimpanzee subspecies? Nature. 1997;388(6640):337. doi: 10.1038/41005.
23. Gonder MK, Disotell TR, Oates JF. New genetic evidence on the evolution of chimpanzee populations, and implications for taxonomy. Int J Primatol. 2006;27(4):1103-27. doi: 10.1007/s10764-006-9063-y.
24. Mitchell MW, Locatelli S, Ghobrial L, Pokempner AA, Sesink Clee PR, Abwe EE, et al. The population genetics of wild chimpanzees in Cameroon and Nigeria suggests a positive role for selection in the evolution of chimpanzee subspecies. BMC Evol Biol. 2015;15:3. doi: 10.1186/s12862-014-0276-y.
25. Mattle-Greminger MP, Bilgin Sonay T, Nater A, Pybus M, Desai T, de Valles G, et al. Genomes reveal marked differences in the adaptive evolution between orangutan species. Genome Biol. 2018;19(1):193. doi: 10.1186/s13059-018-1562-6.
26. Rodrigues MF, Kern AD, Ralph PL. Shared evolutionary processes shape landscapes of genomic variation in the great apes. Genet. 2024;226(4):iyae006. doi: 10.1093/genetics/iyae006.
27. Fontsere C, Kuhlwilm M, Morcillo-Suarez C, Alvarez-Estape M, Lester JD, Gratton P, et al. Population dynamics and genetic connectivity in recent chimpanzee history. Cell Genom. 2022;2(6). doi: 10.1016/j.xgen.2022.100133.
28. Sharp PM, Plenderleith LJ, Hahn BH. Ape Origins of Human Malaria. Annu Rev Microbiol. 2020;74(1):39-63. doi: 10.1146/annurev-micro-020518-115628.
29. Leendertz SAJ, Wich SA, Ancrenaz M, Bergl RA, Gonder MK, Humle T, Leendertz FH. Ebola in great apes – current knowledge, possibilities for vaccination, and implications for conservation and human health. Mammal Rev. 2016:n/a-n/a. doi: 10.1111/mam.12082.
30. Locatelli S, McKean KA, Sesink Clee PR, Gonder MK. The Evolution of Resistance to Simian Immunodeficiency Virus (SIV): A Review. Int J Primatol. 2014;35(2):349-75. doi: 10.1007/s10764-014-9763-7.
31. Bibollet-Ruche F, Russell RM, Liu W, Stewart-Jones GBE, Sherrill-Mix S, Li Y, et al. CD4 receptor diversity in chimpanzees protects against SIV infection. Proc Natl Acad Sci USA. 2019;116(8):3229-38. doi: 10.1073/pnas.1821197116.
32. Schmidt JM, de Manuel M, Marques-Bonet T, Castellano S, Andrés AM. The impact of genetic adaptation on chimpanzee subspecies differentiation. PLoS Genet. 2019;15(11):e1008485. doi: 10.1371/journal.pgen.1008485.
33. Pawar H, Ostridge HJ, Schmidt JM, Andrés AM. Genetic adaptations to SIV across chimpanzee populations. PLoS Genet. 2022;18(8):e1010337. doi: 10.1371/journal.pgen.1010337.
34. Locatelli S, Harrigan RJ, Sesink Clee PR, Mitchell MW, McKean KA, Smith TB, Gonder MK. Why are Nigeria-Cameroon Chimpanzees (*Pan troglodytes ellioti*) free of SIVcpz infection? PLoS ONE. 2016;11(8):e0160788. doi: 10.1371/journal.pone.0160788.
35. Prince AM, Brotman B, Lee D-H, Andrus L, Valinsky J, Marx P. Lack of evidence for HIV Type 1-related SIVcpz infection in captive and wild chimpanzees (*Pan troglodytes verus*) in West Africa. AIDS Res Hum Retroviruses. 2002;18(9):657-60. doi: 10.1089/088922202760019356. PubMed PMID: 12079561.
36. Leendertz SAJ, Locatelli S, Boesch C, Kücherer C, Formenty P, Liegeois F, et al. No evidence for transmission of SIVwrc from western red colobus monkeys (*Piliocolobus badius badius*) to wild west african chimpanzees (*Pan troglodytes verus*) despite high exposure through hunting. BMC Microbiol. 2011;11(1):24. doi: 10.1186/1471-2180-11-24.
37. Kalan AK, Kulik L, Arandjelovic M, Boesch C, Haas F, Dieguez P, et al. Environmental variability supports chimpanzee behavioural diversity. Nat Commun. 2020;11(1):4451. doi: 10.1038/s41467-020-18176-3.

38. Wessling EG, Deschner T, Mundry R, Pruetz JD, Wittig RM, Kühl HS. Seasonal Variation in Physiology Challenges the Notion of Chimpanzees (*Pan troglodytes verus*) as a Forest-Adapted Species. Front Ecol Evol. 2018;6. doi: 10.3389/fevo.2018.00060.

39. Boesch C, Kalan AK, Mundry R, Arandjelovic M, Pika S, Dieguez P, et al. Chimpanzee ethnography reveals unexpected cultural diversity. Nat Hum Behav. 2020;4(9):910-6. doi: 10.1038/s41562-020-0890-1.

40. Kühl HS, Boesch C, Kulik L, Haas F, Arandjelovic M, Dieguez P, et al. Human impact erodes chimpanzee behavioral diversity. Science. 2019;363(6434):1453-5. doi: doi:10.1126/science.aau4532.

41. Stumpf R. Chimpanzees and Bonobos, Diveristy within and between species. In: Campbell CJ, Fuentes A, MacKinnon KC, Panger M, Bearder SK, editors. Primates in Perspective. New York: Oxford University Press; 2007. p. 321-44.

42. Sesink Clee PR, Abwe EE, Ambahe RD, Anthony NM, Fotso R, Locatelli S, et al. Chimpanzee genetic structure in Cameroon and Nigeria is associated with habitat variation that may be lost under climate change. BMC Evolutionary Biology. 2015;15:2. doi: 10.1186/s12862-014-0275-z.

43. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, et al. Signals of recent positive selection in a worldwide sample of human populations. Genome Res. 2009;19(5):826-37. doi: 10.1101/gr.087577.108.

44. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44. doi: 10.1038/nprot.2008.211.

45. Villadangos JA. Presentation of antigens by MHC class II molecules: getting the most out of them. Mol Immunol. 2001;38(5):329-46. doi: 10.1016/s0161-5890(01)00069-4.

46. Pasquale EB. EPH receptor signalling casts a wide net on cell behaviour. Nat Rev Mol Cell Biol. 2005;6(6):462-75. doi: 10.1038/nrm1662. PubMed PMID: WOS:000229629100013.

47. Kania A, Klein R. Mechanisms of ephrin-Eph signalling in development, physiology and disease. Nat Rev Mol Cell Biol. 2016;17(4):240-56. doi: 10.1038/nrm.2015.16.

48. Aihara E, Engevik KA, Montrose MH. Trefoil Factor Peptides and Gastrointestinal Function. Annu Rev Physiol. 2017; 79: 357–380.

49. Weinberg A, Jin G, Sieg S, McCormick T. The Yin and Yang of Human Beta-Defensins in Health and Disease. Front Immunol. 2012;3(294). doi: 10.3389/fimmu.2012.00294.

50. Ellis N, Smith SJ, Pitcher CR. Gradient forests: calculating importance gradients on physical predictors. Ecol. 2012;93(1):156-68. doi: 10.1890/11-0252.1.

51. Hansen TH, Bouvier M. MHC class I antigen presentation: learning from viral evasion strategies. Nat Rev Immunol. 2009;9(7):503-13. doi: 10.1038/nri2575.

52. Luo Y, Jacobs EY, Greco TM, Mohammed KD, Tong T, Keegan S, et al. HIV–host interactome revealed directly from infected cells. Nat Microbiol. 2016;1(7):16068. doi: 10.1038/nmicrobiol.2016.68.

53. Abitew AM, Sobti RC, Sharma VL, Wanchu A. Analysis of transporter associated with antigen presentation (TAP) genes polymorphisms with HIV-1 infection. Mol Cel Biochem. 2020;464(1):65-71. doi: 10.1007/s11010-019-03649-x.

54. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics *. Mol Cell Proteomics. 2014;13(2):397-406. doi: 10.1074/mcp.M113.035600.

55. Williams SM, An JY, Edson J, Watts M, Murigneux V, Whitehouse AJO, et al. An integrative analysis of non-coding regulatory DNA variations associated with autism spectrum disorder. Mol Psychiatry. 2019;24(11):1707-19. doi: 10.1038/s41380-018-0049-x.

56. Eusebi PG, Cortés O, Carleos C, Dunner S, Cañon J. Detection of selection signatures for agonistic behaviour in cattle. J Anim Breed Genet. 2018;135(3):170-7. doi: 10.1111/jbg.12325.

57. Parini P, Davis M, Lada AT, Erickson SK, Wright TL, Gustafsson U, et al. ACAT2 Is Localized to Hepatocytes and Is the Major Cholesterol-Esterifying Enzyme in Human Liver. Circulation. 2004;110(14):2017-23. doi: 10.1161/01.CIR.0000143163.76212.0B.

58. Oue K, Zhang J, Harada-Hada K, Asano S, Yamawaki Y, Hayashiuchi M, et al. Phospholipase C-related Catalytically Inactive Protein Is a New Modulator of Thermogenesis Promoted by β-Adrenergic Receptors in Brown Adipocytes. J Biol Chem. 2016;291(8):4185-96. doi: 10.1074/jbc.M115.705723.

59. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24. doi: 10.1038/nature11582.

60. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27-30. doi: 10.1093/nar/28.1.27.

61. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947-51. doi: 10.1002/pro.3715.

62. Mitchell MW, Locatelli S, Sesink Clee PR, Thomassen HA, Gonder MK. Environmental variation *and* rivers govern the structure of chimpanzee genetic diversity in a biodiversity hotspot. BMC Evol Biol. 2015;15:1. doi: 10.1186/s12862-014-0274-0.

63. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature. 1999;397(6718):436-41. doi: 10.1038/17130.

64. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science. 2006;313(5786):523-6. doi: 10.1126/science.1126531

65. Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, et al. SIV infection in wild gorillas. Nature. 2006;444(7116):164-. doi: 10.1038/444164a.

66. D'arc M, Ayouba A, Esteban A, Learn GH, Boué V, Liegeois F, et al. Origin of the HIV-1 group O epidemic in western lowland gorillas. Proc Natl Acad Sci USA. 2015;112(11):E1343-E52. doi: doi:10.1073/pnas.1502022112.

67. Bailes E, Gao F, Bibollet-Ruche F, Courgnaud V, Peeters M, Marx PA, et al. Hybrid origin of SIV in chimpanzees. Science. 2003;300(5626):1713-. doi: doi:10.1126/science.1080657.

68. Sharp PM, Shaw GM, Hahn BH. Simian Immunodeficiency virus infection of chimpanzees. J Virol. 2005;79(7):3891-902. doi: 10.1128/JVI.79.7.3891-3902.2005.

69. Rudicell RS, Holland Jones J, Wroblewski EE, Learn GH, Li Y, Robertson JD, et al. Impact of Simian Immunodeficiency Virus infection on chimpanzee population dynamics. PLoS Pathog. 2010;6(9):e1001116. doi: 10.1371/journal.ppat.1001116

70. de Groot NG, Otting N, Doxiadis GGM, Balla-Jhagjhoorsingh SS, Heeney JL, van Rood JJ, et al. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci USA. 2002;99(18):11748-53. doi: 10.1073/pnas.182420799.

71. Maisels F. Mbam Djerem National Park, Cameroon: at the forest's edge. Canopee. 2005;27:2-6.

72. Smith TB, Wayne RK, Girman DJ, Bruford MW. A role for ecotones in generating rainforest biodiversity. Science. 1997;276(5320):1855-7. doi: 10.1126/science.276.5320.1855.

73. Simard F, Ayala D, Kamdem G, Pombi M, Etouna J, Ose K, et al. Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. BMC Ecol. 2009;9(1):17. doi:10.1186/1472-6785-9-17.

74. Smith TB, Thomassen HA, Freedman AH, Sehgal RNM, Buermann W, Saatchi S, et al. Patterns of divergence in the olive sunbird *Cyanomitra olivacea* (Aves: Nectariniidae) across the African rainforest-savanna ecotone. Biol J Linnean Soc. 2011;103(4):821-35. doi: 10.1111/j.1095-8312.2011.01674.x.

75. Freedman AH, Thomassen HA, Buermann W, Smith TB. Genomic signals of diversification along ecological gradients in a tropical lizard. Mol Ecol. 2010;19(17):3773-88. doi: 10.1111/j.1365-294X.2010.04684.x.

76. Zhen Y, Harrigan RJ, Ruegg KC, Anderson EC, Ng TC, Lao SRN, et al. Genomic divergence across ecological gradients in the Central African rainforest songbird (*Andropadus virens*). Mol Ecol. 2017;26(19):4966-77. doi: 10.1111/mec.14270.

77. Morgan K, Mboumba J-F, Ntie S, Mickala P, Miller CA, Zhen Y, et al. Precipitation and vegetation shape patterns of genomic and craniometric variation in the central African rodent *Praomys misonnei*. Proc Biol Sci B. 2020;287(1930):20200449. doi: doi:10.1098/rspb.2020.0449.

78. Freedman AH, Harrigan RJ, Zhen Y, Hamilton AM, Smith TB. Evidence for ecotone speciation across an African rainforest-savanna gradient. Mol Ecol. 2023;32(9):2287–300. doi: 10.1111/mec.16867.

79. Abwe EE, Morgan BJ, Tchiengue B, Kentatchime F, Doudja R, Ketchen ME, et al. Habitat differentiation among three Nigeria–Cameroon chimpanzee (*Pan troglodytes ellioti*) populations. Ecol Evol. 2019;9(3):1489-500. doi: 10.1002/ece3.4871.

80. Mitchell MW, Locatelli S, Abwe EE, Ghobrial L, Gonder MK. Male-driven differences in chimpanzee (*Pan troglodytes*) population genetic structure across three habitats in Cameroon and Nigeria. Int J Primatol. 2018;39(4):581-601. doi: 10.1007/s10764-018-0053-7.

81. Abwe EE, Morgan BJ, Doudja R, Kentatchime F, Mba F, Dadjo A, et al. Dietary ecology of the Nigeria–Cameroon chimpanzee (*Pan troglodytes ellioti*). Int J Primatol. 2020;41(1):81-104. doi: 10.1007/s10764-020-00138-7.

82. Abwe EE. Linking behavioral diversity with genetic and ecological variation in the Nigeria-Cameroon chimpanzee (*Pan troglodytes ellioti*): Drexel University; 2018.

83. Ghobrial L, Lankester F, Kiyang JA, Akih AE, de Vries S, Fotso R, et al. Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. BMC Ecol. 2010;10(1):2. doi: 10.1186/1472-6785-10-2.

84. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]. 2013. doi: 10.48550/arXiv.1303.3997

85. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. ArXiv e-prints [Internet]. 2012 July 01, 2012. Available from: https://ui.adsabs.harvard.edu/#abs/2012arXiv1207.3907G.

86. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449:913. doi: 10.1038/nature06250

87. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A Map of Recent Positive Selection in the Human Genome. PLoS Biol. 2006;4(3):e72. doi: 10.1371/journal.pbio.0040072.

88. Maclean CA, Hong NPC, Prendergast JGD. hapbin: An Efficient Program for Performing Haplotype-Based Scans for Positive Selection in Large Genomic Datasets. Mol Biol Evol. 2015;32(11):3027-9. doi: 10.1093/molbev/msv172.

89. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9:179. doi: 10.1038/nmeth.1785

90. Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, et al. A fine-scale chimpanzee genetic map from population sequencing. Science. 2012;336(6078):193-8. doi: 10.1126/science.1216872.

91. Paten B, Herrero J, Beal K, Fitzgerald S, Birney E. Enredo and Pecan: Genome-wide mammalian consistency-based multiple alignment with paralogs. Genome Res. 2008;18(11):1814-28. doi: 10.1101/gr.076554.108.

92. Paten B, Herrero J, Fitzgerald S, Beal K, Flicek P, Holmes I, Birney E. Genome-wide nucleotide-level mammalian ancestor reconstruction. Genome Res. 2008;18(11):1829-43. doi: 10.1101/gr.076521.108.

93. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:14. doi: 10.1186/s13059-016-0974-4.

94. Huff CD, Harpending HC, Rogers AR. Detecting positive selection from genome scans of linkage disequilibrium. BMC Genomics. 2010;11:9. doi: 10.1186/1471-2164-11-8. PubMed PMID: WOS:000274642200001.

95. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015;43(W1):W589-W98. doi: 10.1093/nar/gkv350.

96. Perry GH, Marioni JC, Melsted P, Gilad Y. Genomic-scale capture and sequencing of endogenous DNA from feces. Mol Ecol. 2010;19(24):5332-44. doi: 10.1111/j.1365-294X.2010.04888.x.

97. Hernandez-Rodriguez J, Arandjelovic M, Lester J, Filippo C, Weihmann A, Meyer M, et al. The impact of endogenous content, replicates and pooling on genome capture from faecal samples. Mol Ecol Res. 2018;18(2):319-33. doi: doi:10.1111/1755-0998.12728.

98. Fan S, Hansen MEB, Lo Y, Tishkoff SA. Going global by adapting local: A review of recent human adaptation. Science. 2016;354(6308):54-9. doi: 10.1126/science.aaf5098.

99. Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, et al. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci USA. 2010;107(Supplement 2):8924-30. doi: 10.1073/pnas.0914625107.

100. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8. doi: 10.1093/bioinformatics/btr330.

101. Pew J, Muir PH, Wang J, Frasier TR. related: an R package for analysing pairwise relatedness from codominant molecular markers. Mol Ecol Res. 2015;15(3):557-61. doi: 10.1111/1755-0998.12323.

102. Wang J. Triadic IBD coefficients and applications to estimating pairwise relatedness. Genetical Res. 2007;89(3):135-53. doi: 10.1017/S0016672307008798.

103. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resources. 2010;10(3):564-7. doi: 10.1111/j.1755-0998.2010.02847.x.

104. Hijmans RJ. geosphere: Spherical Trigonometry. R package version 15-5. 2016. Available from: https://CRAN.R-project.org/package=geosphere.

105. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655-64. doi: 10.1101/gr.094052.109.

106. Rosenberg NA. DISTRUCT: A program for the graphical display of population structure. Mol Ecol Notes. 2004;4(1):2. doi: 10.1046/j.1471-8286.2003.00566.x.

107. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label and switching and multimodality in analysis of population structure. Bioinformatics. 2007;23(14):6. doi: 10.1093/bioinformatics/btm233.

108. Chen C, Durand E, Forbes F, François O. Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. Mol Ecol Notes. 2007;7(5):747-56. doi: 10.1111/j.1471-8286.2007.01769.x.

109. Mitchell MW, Rowe B, Sesink Clee PR, Gonder MK. TESS Ad-Mixer: A novel program for visualizing TESS Q matrices. Conserv Genet Res. 2013;5(4):1075-8. doi: 10.1007/s12686-013-9987-4.

110. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27(21):3070-1. doi: 10.1093/bioinformatics/btr521.

111. Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. Genetics. 2002;160(3):1217-29.

112. Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. hybriddetective: A workflow and package to facilitate the detection of hybridization using genomic data in R. Mol Ecol Res. 2017;17(6):e275-e84. doi: doi:10.1111/1755-0998.12704.

113. Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. parallelnewhybrid: an R package for the parallelization of hybrid detection using newhybrids. Mol Ecol Res. 2017;17(1):91-5. doi: doi:10.1111/1755-0998.12597.

114. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLoS Genet. 2009;5(10):e1000695. doi: 10.1371/journal.pgen.1000695.

115. Fitzpatrick MC, Keller SR. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol Let. 2015;18(1):1-16. doi: https://doi.org/10.1111/ele.12376.

116. Láruson ÁJ, Fitzpatrick MC, Keller SR, Haller BC, Lotterhos KE. Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. Evol Appl. 2022;15(3):403-16. doi: 10.1111/eva.13354.

117. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Bio Evol. 2013;30(7):1687-99. doi: 10.1093/molbev/mst063.

118. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J Open Source Software. 2018;3(25):731. doi: https://doi.org/10.21105/joss.00731.

119. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database. 2011;2011:bar030. doi: 10.1093/database/bar030.

120. Zerbino DR, Achuthan P, Akanni W, Amode M R, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46(D1):D754-D61. doi: 10.1093/nar/gkx1098.

121. Consortium GO. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research. 2004;32(suppl_1):D258-D61. doi: 10.1093/nar/gkh036.

Supporting Information

S1 File. Supplemental Results.

S2. File Extended Methods and Materials.

S1 Fig. Origins of chimpanzees of Cameroon included in this study. Sample locations and proportions of estimated ancestry were estimated in previous studies [83, 122].

Table S1. Captive chimpanzee genomes included in this study

S2 Fig. Heterozygosity estimates of captive chimpanzee genomes.

(A) Individual heterozygosity.

(B) Subspecies heterozygosity.

S3 Fig. Population structure of captive chimpanzee genomes.

(A) PCA of LD pruned SNP data set consisting of 1,113,142 SNPs.

(B) sNMF individual ancestry analysis of the LD pruned data set in a range of K values.

(C) PCA of Neutral SNP data set consisting of 147,000 SNPs.

(D) sNMF individual ancestry analysis of the neutral SNP data set in a range of K values.

S4 Fig. Cross Entropy Results.
Value of the cross-entropy criterion as a function of the number of ancestral populations in sNMF for (A) the LD pruned SNP panel and (B) the neutral SNP panel.

Table S2. Top 10 regions under selection including their genetic content.

Table S3. Enriched GO terms in the “Biological Processes” domain.

Table S4. Enriched KEGG pathways.

Table S5. Functional enrichment clustering.

Table S6. Number of wild chimpanzee samples collected and used in this study.

Table S7. Environmental predictor variables used characterize chimpanzee habitats.

Table S8. Pearson correlation table of environmental variables.

Table S9. Environmental variable groupings.

S5 Fig. Illumina reads of wild chimpanzee samples evaluated into 4 categories.
Raw reads – yellow, mapped reads (sequence reads that mapped to the panTro4

reference genome) – green, mapped and deduplicated reads (PCR duplicates were removed) – light blue, and on-target reads mapped to our sites – dark blue.

(A) Proportion of read types for all 192 sequenced chimpanzee samples.

(B) Number of read types for 85 samples included in the '10k' dataset.

(C) Number of read types for samples removed due to missingness (>30% missing sites).

(D) Proportion of read types for 85 samples included in the 'complete' dataset.

(E) Number of read types for duplicate samples removed following relatedness analysis.

(F) Number of read types for 23 individuals included in the '1k' dataset, but not the '10k' dataset.

S6 Fig. Frequency of mean sequencing coverage of wild chimpanzee samples for each site.

Mean coverage across all sites was 20.2 reads/site. The red vertical line represents the minimum coverage needed to accurately call SNPs (5x coverage).

S7 Fig. Pairwise F_{ST} between sites shows population structure. (A) '10k' dataset. (B) '1k' dataset.

S8 Fig. Isolation-by-distance for '10k' dataset.

(A) Correlation between 'linearized F_{ST}' and geographic distance (km) generated using the '10k' dataset. Solid circles represent pairs of sampling locations from the same habitat. Dual-colored diamonds represent pairs of sampling locations from different habitats.

(B) Null distribution of t-statistics from 10,000 permutations same- or different habitat/population pairs in four bins of geographic distance. The red dotted line shows the t-statistic value for actual data.

S9 Fig. Isolation-by-distance for '1k' dataset.

(A) Correlation between 'linearized F_{ST}' and geographic distance (km) generated using the '1k' dataset. Solid circles represent pairs of sampling locations from the same habitat. Dual-colored diamonds represent pairs of sampling locations from different habitats.

(B) Null distribution of *t*-statistics from 10,000 permutations same- or different habitat/population pairs in four bins of geographic distance. The red dotted line shows the *t*-statistic value for actual data.

S10 Fig. Isolation-by-distance for only *P. t. ellioti* populations.

(A) Correlation between 'linearized F_{ST}' and geographic distance (km) generated using the '10k' dataset. Solid circles represent pairs of sampling locations from the same habitat. Dual-colored diamonds represent pairs of sampling locations from different habitats.

(B) Null distribution of *t*-statistics from 10,000 permutations same- or different habitat/population pairs in four bins of geographic distance. The red dotted line shows the *t*-statistic value for actual data.

S11 Fig. PCA of all SNPs from the '10k' dataset.

The first two principal components recapitulate known population structure of chimpanzees in Cameroon. They show 3 clear populations, and one *P. t. ellioti* (Ecotone) individual (CMMD06) clustering with *P. t. troglodytes*, as well as multiple *P. t. ellioti* (Rainforest) clustering together with *P. t. ellioti* (Ecotone) and vice versa.

S12 Fig. PCA of only neutral SNPs from the '10k' dataset.

S13 Fig. PCA of all SNPs from the '1k' dataset

S14 Fig. PCA of only neutral SNPs from the '1k' dataset

Table S10. The results of the Tracy-Widom test for all SNPs from the '10k' SNP dataset.

S15 Fig. ADMIXTURE bar plots for *K*=2-3.

These three populations correspond to known population structure. However, at *K*=2, there is a signal of possible historic gene flow of *P. t. troglodytes* into *P. t. ellioti* (Ecotone). Moreover, the is one individual (CMMD06 – also identified in the PCA) as being a potential *ellioti/troglodytes* hybrid. At *K*=3, we see evidence of three additional individuals that may be Rainforest/Ecotone hybrids, as well as evidence of mixing between the populations.

S16 Fig. The cross-validation error results of ADMIXTURE analysis of wild chimpanzees ('10k' dataset).

S17 Fig. PCA results of the merged captive and wild datasets.

S18 Fig. ADMIXTURE bar plots for *K*=2-5 for merged captive and wild datasets.

S19 Fig. The cross-validation error results of ADMIXTURE analysis of merged captive and wild chimpanzees ('10k' dataset).

S20 Fig. Cluster analysis and spatial interpolation of population structure.

(A) TESS bar plots showing individual proportions of ancestry of wild chimpanzees.

(B) Spatial interpolation of the Q matrix for *K*=3 generated using TESS and Ad-Mixer.

S21 Fig. Estimating K_{MAX} from TESS analysis.

D*K* values estimated for *K*=1-5 across 10 replicate runs.

Table S11. Analysis of Molecular Variance (AMOVA).

S22 Fig. Mean observed heterozygosity.

(A) Heterozygosity of all loci for all individuals grouped by population.
(B) Heterozygosity for all individuals.
There were no significant differences between heterozygosity for each population.

S23 Fig. Posterior plots of model performance.

(A) The observed Joint SFS for *P. t. troglodytes* and *P. t. ellioti*.

(B) the simulated Joint SFS for *P. t. troglodytes* and *P. t. ellioti* under the most likely asymmetric migration scenario obtained from *δaδi.*

(C) The residuals between the modeled and observed Joint SFS.

(D) A 1D histogram of the residual values between the model and the observed data

S24 Fig. *R*^{2} weighted importance of the environmental predictor variables to the Gradient Forest model of gene-environment relationships.

S25 Fig. Results of randomized gradient forest models (n=200), as compared to results from the observed data (n=200).

An average 588 of 7,878 SNPs demonstrated a positive R^2 with at least one environmental variable, with an average R^2 = 0.155 in the Observed data distribution (n=200, represented by the red histograms in A) and B) above). A significantly different average was obtained when randomizing the associations between the genomic data (SNPs) and environmental predictors for both total SNPs with a positive R^2 (average total = 504, t = 5.011(unequal variances df = 202.28), p < 0.0001), as well as for the average R^2 (average = 0.152, t = 2.806 (unequal variance df = 261.97), p = 0.0054) of the randomized gradient forests runs (n=200).

Table S12. Enriched GO terms in the 'Biological Processes' domain for environmentally associated outliers (LFMM and gradient forest) in wild chimpanzees.

Table S13. Enriched KEGG pathways for environmentally associated outliers (LFMM and gradient forest) in wild chimpanzees.

Table S14. Enriched GO terms in the 'Biological Processes' domain for environmentally associated outliers (LFMM and gradient forest) in wild chimpanzees.

S26 Fig. Evidence of selective pressures on acetyl-CoA acetyltransferase 2 (*ACAT2*).

(A) Map of the *ACAT2* gene on chromosome 6 with brown star representing the SNP identified through outlier analysis between *P. t. troglodytes* and *P. t. ellioti*.

(B) Manhattan plot showing the significance (as the negative $\log_{10}$ p-value) of SNP associations with the environmental variable temperature seasonality. Grey colors distinguish different chromosomes. The red line represents the threshold for significant association ($p = 0.05$). The SNP contained in the ACAT2 gene is highlighted by the red arrow.

(C) Correlation between the allele frequency of the SNP contained in the ACAT2 gene and temperature seasonality values at each corresponding sampling location ($R^2 = 0.5615$, $p = 0.0005$).

(D) Allele frequencies of the SNP contained in the ACAT2 gene across Cameroon. Sampling sites are represented by circles that are shaded according to the frequency of the allele within the population. SNP frequencies are plotted against temperature seasonality across the region.

S27 Fig. Evidence of selective pressures on phospholipase C like 2 (*PLCL2*).

(A) Map of the *PLCL2* gene on chromosome 3 with brown star representing the SNP identified through outlier analysis between *P. t. troglodytes* and *P. t. ellioti*.

(B) Manhattan plot showing the significance (as the negative $\log_{10}$ p-value) of SNP associations with the environmental variable precipitation of wettest month. Grey colors distinguish different chromosomes. The red line represents the threshold for significant association ($p = 0.05$). The SNP contained in the PLCL2 gene is highlighted by the red arrow.

(C) Correlation between the allele frequency of the SNP contained in the PLCL2 gene and precipitation of wettest month values at each corresponding sampling location ($R^2 = 0.3422$, $p = 0.0102$).

(D). Allele frequencies of the SNP contained in the PLCL2 gene across Cameroon. Sampling sites are represented by circles that are shaded according to the frequency of the allele within the population. SNP frequencies are plotted against precipitation of wettest month across the region.