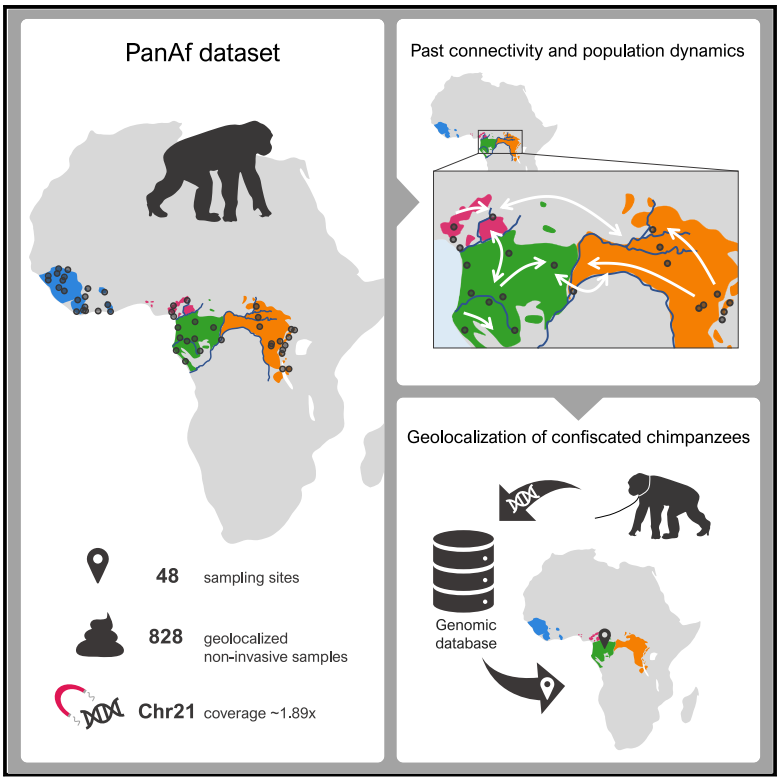
Cell Genomics

Article

Population dynamics and genetic connectivity in recent chimpanzee history

Graphical abstract



Highlights

- Chromosome 21 capture of geolocalized non-invasive chimpanzee samples
- Support for four differentiated subspecies with local population structure
- Fine-scale description of gene flow barriers and corridors since Late Pleistocene
- Power to infer geographical origin of confiscated chimpanzees

Authors

Claudia Fontsere, Martin Kuhlwilm, Carlos Morcillo-Suarez, ..., Esther Lizano, Mimi Arandjelovic, Tomas Marques-Bonet

Correspondence

claudia.fontsere@upf.edu (C.F.), arandjel@eva.mpg.de (M.A.), tomas.marques@upf.edu (T.M.-B.)

In brief

Fontsere et al. captured and sequenced chromosome 21 from 828 non-invasively collected chimpanzee samples, providing an extensive catalog of genomic diversity for wild chimpanzee populations. The authors describe patterns of isolation and connectivity between localities and implement a fine-grained geolocalization approach to infer the origin of confiscated chimpanzees.

Fontsere et al., 2022, Cell Genomics 2, 100133
June 8, 2022
https://doi.org/10.1016/j.xgen.2022.100133

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Population dynamics and genetic connectivity in recent chimpanzee history

Claudia Fontsere,[1,*] Martin Kuhlwilm,[1,2,3] Carlos Morcillo-Suarez,[1] Marina Alvarez-Estape,[1] Jack D. Lester,[4] Paolo Gratton,[4,5] Joshua M. Schmidt,[6,7] Paula Dieguez,[4,8] Thierry Aebischer,[9] Paula Álvarez-Varona,[10] Anthony Agbor,[4] Samuel Angedakin,[4] Alfred K. Assumang,[11] Emmanuel A. Ayimisin,[4] Emma Bailey,[4] Donatienne Barubiyo,[4] Mattia Bessone,[4] Andrea Carretero-Alonso,[12] Rebecca Chancellor,[13] Heather Cohen,[4,14] Emmanuel Danquah,[11] Tobias Deschner,[4] Andrew Dunn,[15] Jef Dupain,[16] Villard E. Egbe,[4] Olga Feliu,[17] Annemarie Goedmakers,[18] Anne-Céline Granjon,[4] Josephine Head,[4] Daniela Hedwig,[19] Veerle Hermans,[20] R. Adriana Hernandez-Aguilar,[10,21] Inaoyom Imong,[15] Sorrel Jones,[4,22] Jessica Junker,[8] Parag Kadam,[23] Mike Kaiser,[4] Mbangi Kambere,[4]

(Author list continued on next page)

[1]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain
[2]Department of Evolutionary Anthropology, University of Vienna, 1030 Vienna, Austria
[3]Human Evolution and Archaeological Sciences (HEAS), University of Vienna, 1030 Vienna, Austria
[4]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[5]Dipartimento di Biologia, Università di Roma "Tor Vergata", Rome, Italy
[6]UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London, UK
[7]School of Biological Sciences, The University of Adelaide, Adelaide, Australia
[8]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Puschstrasse 4, 04103 Leipzig, Germany
[9]University of Fribourg, Fribourg, Switzerland
[10]Jane Goodall Institute Spain and Senegal, Dindefelo Biological Station, Dindefelo, Kedougou, Senegal
[11]Department of Wildlife and Range Management, Faculty of Renewable Natural Resources, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[12]Fundación Rainfer, Fuente el Saz de Jarama, Madrid, Spain
[13]Depts of Anthropology & Sociology and Psychology, West Chester University, West Chester, PA, USA
[14]Wild Chimpanzee Foundation (WCF), Deutscher Platz 6, 04103 Leipzig, Germany
[15]Wildlife Conservation Society (WCS), 2300 Southern Boulevard, Bronx, NY 10460, USA
[16]Antwerp Zoo Foundation-KMDA, Royal Zoological Society of Antwerp, Antwerp, Belgium
[17]Research Department, Fundació MONA, Riudellots de la Selva, Girona, Spain
[18]Chimbo Foundation, Huningspaed 6. 8567 LL Oudemirdum, the Netherlands

(Affiliations continued on next page)

SUMMARY

Knowledge on the population history of endangered species is critical for conservation, but whole-genome data on chimpanzees (*Pan troglodytes*) is geographically sparse. Here, we produced the first non-invasive geolocalized catalog of genomic diversity by capturing chromosome 21 from 828 non-invasive samples collected at 48 sampling sites across Africa. The four recognized subspecies show clear genetic differentiation correlating with known barriers, while previously undescribed genetic exchange suggests that these have been permeable on a local scale. We obtained a detailed reconstruction of population stratification and fine-scale patterns of isolation, migration, and connectivity, including a comprehensive picture of admixture with bonobos (*Pan paniscus*). Unlike humans, chimpanzees did not experience extended episodes of long-distance migrations, which might have limited cultural transmission. Finally, based on local rare variation, we implement a fine-grained geolocalization approach demonstrating improved precision in determining the origin of confiscated chimpanzees.

INTRODUCTION

Genetic data on chimpanzee (*Pan troglodytes*) populations have been used to study the species' diversity and population structure, as well as to characterize their demographic history and patterns of admixture at a broad subspecies level[1–7] and with their sister species, bonobos (*Pan paniscus*).[6,8] Due to a limited fossil record and absence of ancient DNA record, chimpanzee population genetics is inherently restricted to modern-day individuals.[9]

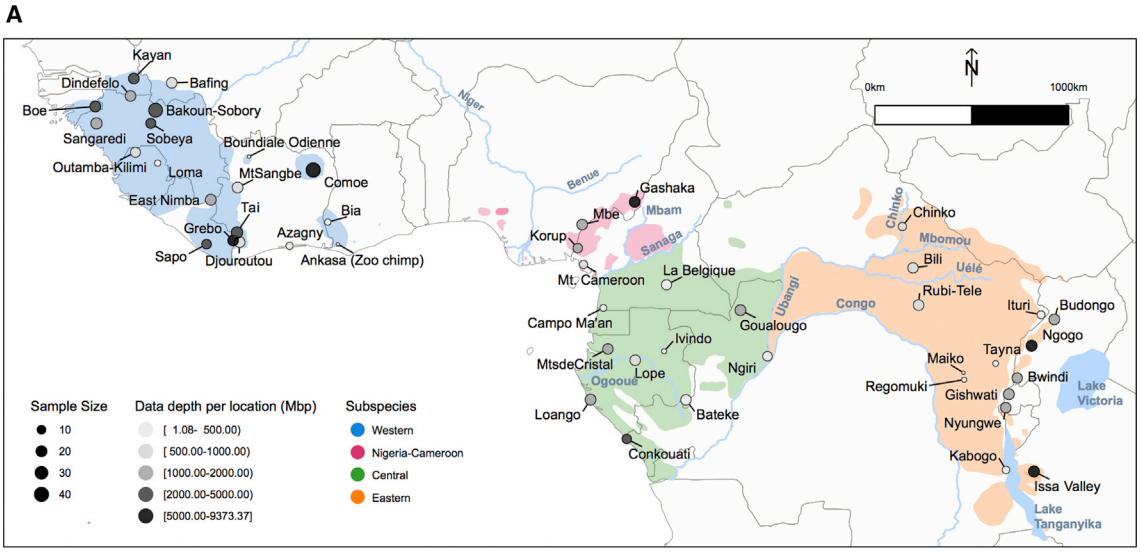
Four chimpanzee subspecies are currently recognized (western -*P. t. verus*-, Nigeria-Cameroon -*P. t. ellioti*-, central -*P. t. troglodytes*-, and eastern -*P. t. schweinfurthii*-, Figure 1A) but conflicting hypotheses still exist about whether genetic diversity

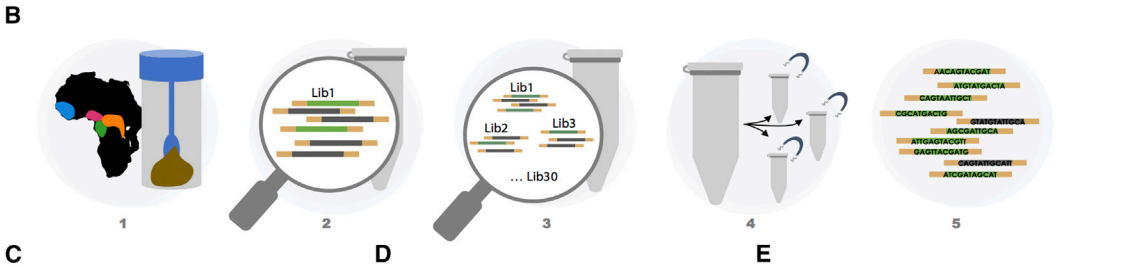
Magloire V. Kambale,[4] Ammie K. Kalan,[23,24] Ivonne Kienast,[4] Deo Kujirakwinja,[15] Kevin Langergraber,[25,26,27] Juan Lapuente,[4] Bradley Larson,[4] Anne Laudisoit,[28,29] Kevin Lee,[4,25] Manuel Llana,[10] Miquel Llorente,[17,30] Sergio Marrocoli,[4] David Morgan,[31] Felix Mulindahabi,[15] Mizuki Murai,[4] Emily Neil,[4] Sonia Nicholl,[4] Stuart Nixon,[32] Emma Normand,[33] Chris Orbell,[34] Lucy J. Ormsby,[4] Liliana Pacheco,[35] Alex Piel,[23,36] Laura Riera,[17] Martha M. Robbins,[4] Aaron Rundus,[37] Crickette Sanz,[38,39] Lilah Sciaky,[4] Volker Sommer,[36,40] Fiona A. Stewart,[23,36,41] Nikki Tagg,[20] Luc Roscelin Tédonzong,[20] Els Ton,[18] Joost van Schijndel,[4,18] Virginie Vergnes,[33] Erin G. Wessling,[43] Jacob Willie,[20,42] Roman M. Wittig,[4,44] Yisa G. Yuh,[4,45] Kyle Yurkiw,[4] Klaus Zuberbuehler,[46,47] Jochen Hecht,[48] Linda Vigilant,[4] Christophe Boesch,[4] Aida M. Andrés,[6] David A. Hughes,[49,50] Hjalmar S. Kühl,[4,8] Esther Lizano,[1,51] Mimi Arandjelovic,[4,8,53,54,*] and Tomas Marques-Bonet[1,48,51,52,53,54,*]

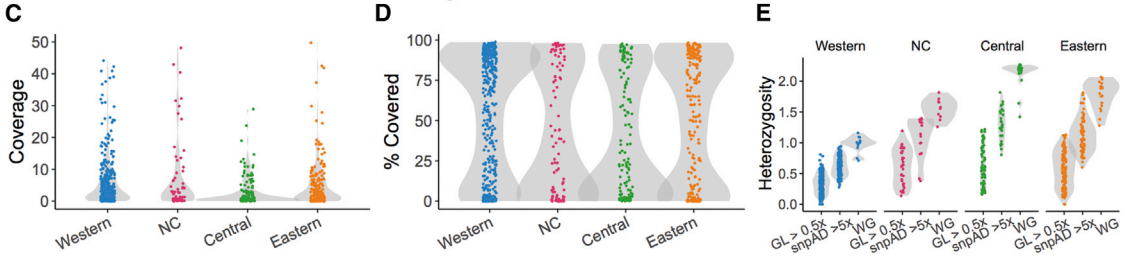
[19]The Elephant Listening Project, K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA
[20]KMDA, Centre for Research and Conservation, Royal Zoological Society of Antwerp, Koningin Astridplein 20-26, B-2018 Antwerp, Belgium
[21]Serra Húnter Professor, Department of Social Psychology and Quantitative Psychology, University of Barcelona, Barcelona, Spain
[22]RSPB Centre of Conservation Science, David Attenborough Building, Pembroke Street, Cambridge, CB2 3QZ, UK
[23]Greater Mahale Ecosystem Research and Conservation Project (GMERC, LTD), Mpanda, Tanzania
[24]Department of Anthropology, University of Victoria, BC V8P 5C2, Canada
[25]School of Human Evolution and Social Change, Arizona State University, 900 Cady Mall, Tempe, AZ 85287, USA
[26]Institute of Human Origins, Arizona State University, 900 Cady Mall, Tempe, AZ 85287, USA
[27]Arizona State University, PO Box 872402, Tempe, AZ 85287-2402, USA
[28]EcoHealth Alliance, 520 Eighth Ave, Suite 1200, New York, NY 10018, USA
[29]Evolutionary Ecology group (EVECO), Campus Drie Eiken, Universiteitsplein 1, 2610 Wilrijk, Belgium
[30]Serra Húnter Fellow, Departament de Psicologia, Facultat d'Educació i Psicologia, Universitat de Girona, Girona, Spain
[31]Lester E. Fisher Center for the Study and Conservation of Apes, Lincoln Park Zoo, 2001 North Clark Street, Chicago, IL 60614, USA
[32]Chester Zoo, Field Programmes, Cedar House, Caughall Road, Upton-by Chester, CH2 1LH, UK
[33]Wild Chimpanzee Foundation (WCF), 23BP238 Abidjan, Côte d'Ivoire
[34]Panthera, 8 W 40th St., New York, NY 10018, USA
[35]Centres Molina, Carrer de Lorena, 61, 08042 Barcelona, Spain
[36]University College London, Department of Anthropology, 14 Taviton Street, London, WC1H 0BW, UK
[37]Department of Psychology, West Chester University, West Chester, PA, USA
[38]Washington University in Saint Louis, Department of Anthropology, One Brookings Drive, St. Louis, MO 63130, USA
[39]Wildlife Conservation Society, Congo Program, Brazzaville, Republic of Congo
[40]Primate Project, Serti, Taraba State 663001, Nigeria
[41]School of Biological & Environmental Sciences, Liverpool John Moores University, James Parsons Building, Byrom Street, Liverpool L3 3AF, UK
[42]Terrestrial Ecology Unit (TEREC), Department of Biology, Ghent University (UGent), K.L. Ledeganckstraat 35, 9000 Ghent, Belgium
[43]Department of Human Evolutionary Biology, Harvard University, 11 Divinity Avenue, Cambridge, MA 02138, USA
[44]Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, BP 1301, Abidjan 01, Côte d'Ivoire
[45]Department of Geography, Planning and Environmental Studies, University of Concordia, 1455 Boulevard de Maisonneuve Ouest, Montréal, QC H3G 1M8, Canada
[46]Université de Neuchâtel, Institut de Biologie, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland
[47]School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, UK
[48]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain
[49]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[50]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK
[51]Institut Catala de Paleontologia Miquel Crusafont, Universitat Autonoma de Barcelona, Cerdanyola del Vallès, Spain
[52]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain
[53]These authors contributed equally
[54]Lead contact
*Correspondence: claudia.fontsere@upf.edu (C.F.), arandjel@eva.mpg.de (M.A.), tomas.marques@upf.edu (T.M.-B.)
https://doi.org/10.1016/j.xgen.2022.100133

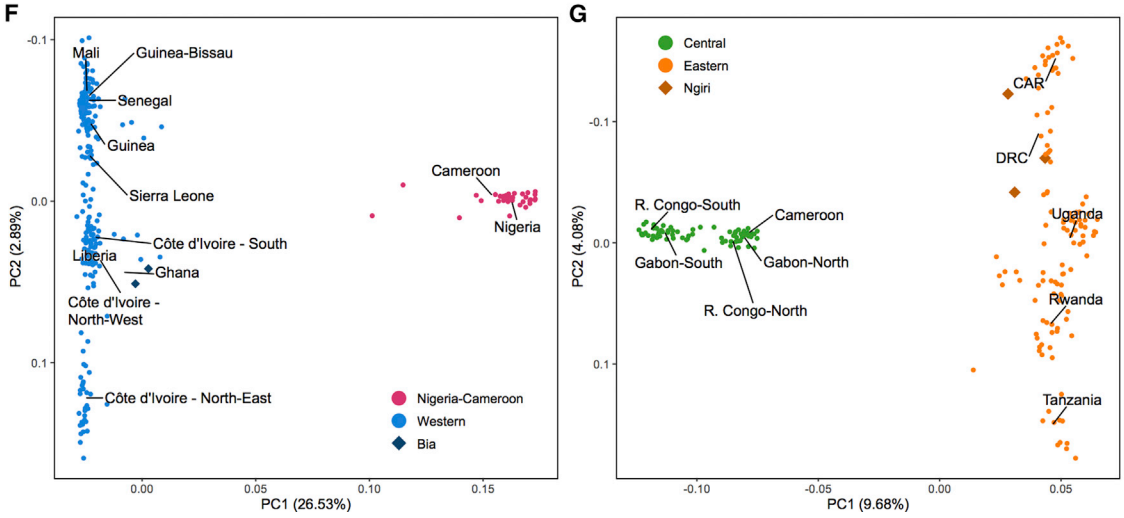
in central and eastern chimpanzee populations reflects two distinctly separated subspecies,[6] or a cline of variation under isolation-by-distance.[1,10,11] This long-standing question also relates to the degree of connectivity among subspecies over time, which requires a fine-scaled reconstruction of the demographic history of chimpanzee populations after their split more than ~100 thousand years ago (kya)[6] and their inter-connectivity since the Last Glacial Maximum (LGM). Identifying genetic connections between present-day chimpanzee communities and the role of past environmental change in shaping these[12,13] may be linked to behavioral variation in chimpanzee communities,[14] similar to what has been explored extensively in humans as a strongly migratory species.[15] Also, it will provide crucial tools for the development of conservation strategies for an endangered species that has suffered a dramatic decline in the last decades.[16,17] A comprehensive genomic knowledge of a threatened species[18] can guide conservation plans both *in situ* and *ex situ*.[19] Furthermore, genetic information has proven useful to infer the populations of origin of confiscated individuals from illegal trade, detect poaching hotspots,[20,21] and guide repatriation planning.[22,23]

For a detailed reconstruction of chimpanzee population structure and demographic history, it is crucial to gather data from a large number of individuals covering the current range of the species and of sufficient data depth. Since practical and ethical









(legend on next page)

concerns impede the collection of blood samples from wild ape populations, non-invasive samples, such as feces,[26,27] are a promising alternative, although low quality and quantities of host DNA (hDNA)[27] have typically precluded population data analysis using single-nucleotide polymorphisms (SNPs). However, in the last years, several technical advances in target capture methods have allowed the use of non-invasive samples in large-scale genomic studies.[25,27–29] Here, we take advantage of these advances to generate an extensive dataset on genomic variation in georeferenced chimpanzees to infer their demographic history and develop a tool for the geolocation of chimpanzee samples.

RESULTS

Capturing the diversity of wild chimpanzees

A total of 828 unique individuals were identified from non-invasive samples collected from 48 sampling sites across the chimpanzee range (Figure 1A) as part of the PanAfrican Program: The Cultured Chimpanzee.[11] Using previously developed methods, we captured chromosome 21 from chimpanzee fecal DNA[11,25,28,29] (Figure 1B) and generated sequencing data to a median coverage of 1.89-fold (0- to 90.14-fold) in the target space (Figure 1C), covering on average 12.9 million positions per sample (STAR Methods; Notes S1 and S2; Table S1).

Numerous samples have high levels of sequencing reads mapping to other primate species than chimpanzee (n = 100, Figures S9, S10, S14, and S15; Note S3), likely due to the inclusion of sympatric primate species in the diet, a well-known phenomenon,[30,31] or sample misidentification during collection of feces.[11,32] We also assessed human contamination among the remaining 728 samples using an approach very sensitive in low-coverage data,[33] finding 36 of those with more than 1% of such contamination (Figure S16). This is similar to patterns of contamination observed in ancient DNA studies on humans.[34] There is also large variation in coverage and hDNA content according to the sampling site, suggesting that environmental and/or dietary factors influence DNA quantity and preservation (Figures 1C, 1D, S4, and S7). Heterozygosity estimates, after careful quality assessment (Figures S21–S29; Note S3), are consistent with known patterns from high-coverage samples:[3,6] highest in central chimpanzees, followed by eastern, Nigeria-Cameroon, and western chimpanzee subspecies (Figure 1E).

The history of *Pan* populations during the Middle Pleistocene

We deemed samples with more than 0.5-fold average coverage (on the target regions of chr21) and low levels of contamination to be of sufficient data depth and quality (n = 555) (Figures S9–S13; Note S3) for a PCA from genotype likelihoods,[35,36] and found that these cluster according to the four described subspecies (Figures 1F, 1G, and S13) that were previously estimated to have diverged during the Middle Pleistocene (139–633 kya), after the split from bonobos (<2 million years ago [mya]) (Figure 2A).[6] Low levels of ancient introgression from bonobos into the non-western chimpanzee subspecies (<1%) had previously been identified, most likely as the result of bonobo admixture into the ancestral population of eastern and central chimpanzees more than 200 kya[6,37] (Figure 2A), possibly associated with a reduction of the Congo River discharge, the natural barrier separating both species.[38] On chromosome 21, we did not observe a significant enrichment of allele sharing (F statistics)[39] between bonobos and chimpanzees, likely due to limitations in the data, a small extent of admixture, and the small number of independent loci (Note S8). However, given the information from previous models based on whole genomes, we sought to determine introgressed fragments on chromosome 21 with the larger number of individuals used in this study. To this end, we inferred bonobo introgression using admixfrog,[40] a method developed to reliably detect introgressed fragments even in low-coverage ancient genomes. With this hidden Markov model we inferred local ancestry for each sample (target) using different sources, which represent the admixing population (bonobo and all chimpanzee subspecies) from the reference panel[6] (STAR Methods; Note S8). We found that all central chimpanzee communities sampled south of the Ogooué River (Figure 2A) (Loango, Lopé, Conkouati, and Batéké) harbor significantly more bonobo-like genomic fragments than those north of the river (two-sided Wilcoxon rank-sum test, Benjamini-Hochberg adjusted p value = 5.735×10^{-8}), or any other chimpanzee population (adjusted p value < 0.01; Figure S64; Table S5; Note S8), with some of the

Figure 1. Overview on sampling, capture, and chimpanzee population history

(A) Geographic distribution of chimpanzee subspecies and PanAf sampling locations. The western chimpanzee range is shown in blue, Nigeria-Cameroon in pink, central in green, and eastern in orange. The size of the dots represents the number of sequenced samples (n = 828) and color intensity represents the amount of chimpanzee genetic data generated (mega-base pairs of mapped sequence) from each sampling site.

(B) Experimental pipeline. (1) Samples were collected from 48 sampling sites, DNA extracted and screened for amplification success, uniqueness, and relatedness using microsatellites;[11] (2) one library per individual[24] was prepared; (3) between 10 and 30 libraries were pooled equi-endogenously;[25] (4) enrichment for chromosome 21 with target capture methods, between three and five times per library;[25] (5) sequencing data were generated with Illumina.

(C) Average coverage on the target region of chromosome 21 for each sample.

(D) Percentage of the target space covered by at least one read.

(E) Heterozygosity estimates per subspecies derived from ANGSD genotype likelihood on PanAf samples with more than 0.5-fold coverage (GL > 0.5×), from snpAD genotype calls on PanAf samples with more than 5-fold coverage, and from GATK genotype calls on previously published whole-genome (WG) chimpanzee samples.[6]

(F) PCA of western (blue) and Nigeria-Cameroon (pink) chimpanzee subspecies. Dark blue diamonds, Bia sampling site in Ghana at the eastern fringe of the extant western chimpanzee range.

(G) PCA of central (green) and eastern (orange) chimpanzee subspecies. Dark orange diamonds, Ngiri sampling site at the western fringe of the eastern chimpanzee distribution. CAR, Central African Republic; DRC, Democratic Republic of Congo; R. Congo, Republic of Congo. See also Figures S3–S13, S29–S39, and Table S4

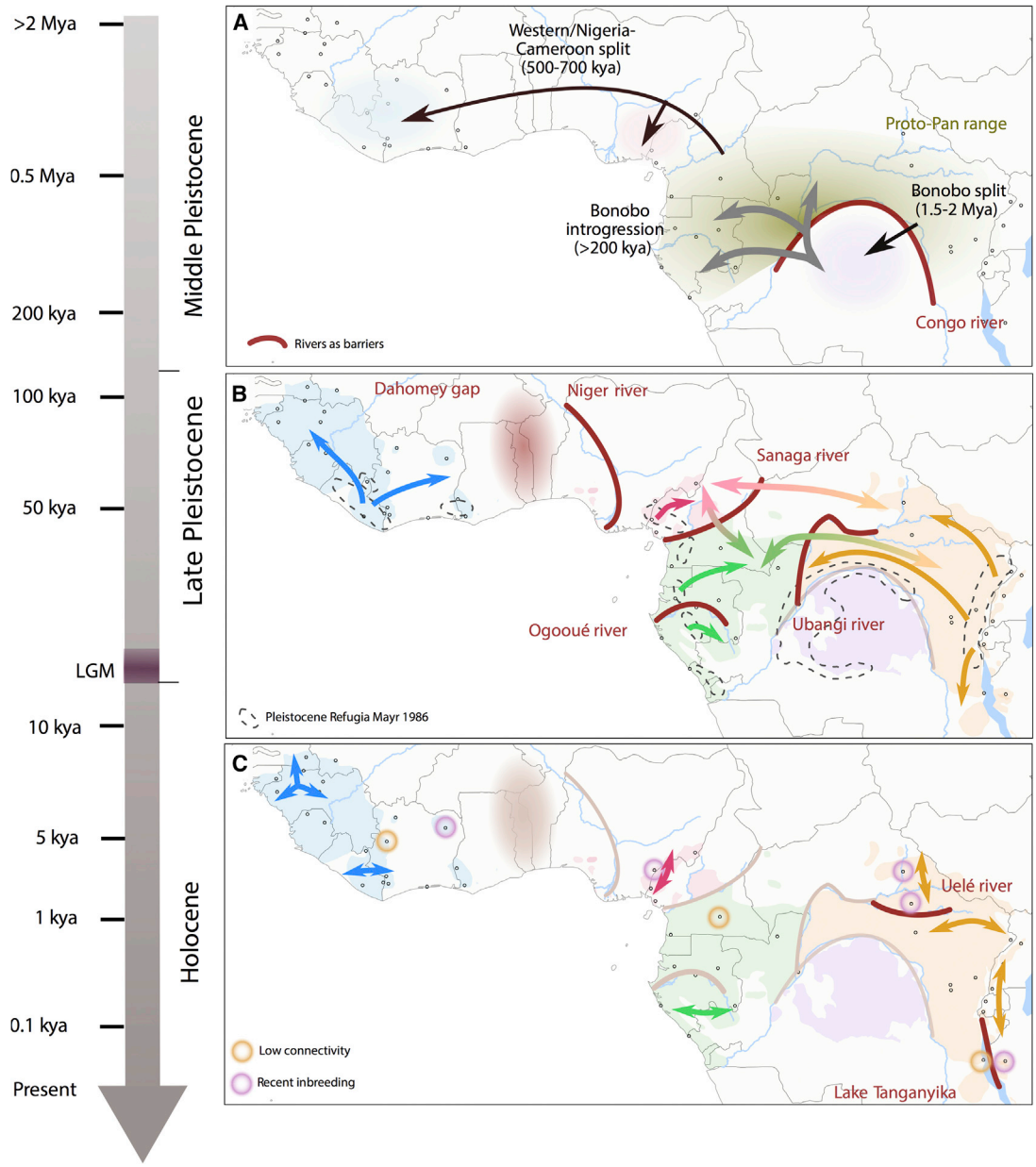


Figure 2. Reconstruction of chimpanzee genetic history

Major rivers and lakes (red lines) and the Dahomey gap (red shading) represent geographical barriers separating populations at different timescales.

(A) Formation of and migration between *Pan* species (chimpanzees and bonobos) and subspecies formation during the Middle Pleistocene; separation and migration events inferred in previous studies,[6,8,38] additional gene flow into southern central populations was inferred here using admixfrog.

(B) Corridors of gene flow (arrows) during the Late Pleistocene and after the Last Glacial Maximum (LGM), when chimpanzee populations expanded from refugia;[12,14] within subspecies based on migration surfaces obtained with EEMS and shared rare variation, between subspecies based on short IBD-like tracts (<0.5 Mbp) and shared fragments of ancestry inferred with admixfrog.

(C) Population connectivity and isolation during the Holocene; connectivity was determined by long (>0.5 Mbp) IBD-like fragments between sampling locations within subspecies and supported by presence or absence of shared rare variation; signatures of recent inbreeding are represented by long regions of homozygosity in individuals from a given sampling location. See also Figures S40–S53, S57–S60, S64, S66, S76–S79, S82–S85, and S92.

individuals from the Lopé and Loango sampling sites showing the highest bonobo ancestry (Figure S64). These fragments are also longer than in other chimpanzee populations (Table S5; Figure S66J), which may hint at a separate, more recent admixture event, although this observation was not significant (two-sided Wilcoxon rank-sum test).

Within chimpanzees, our dense sampling approach, including communities at the border between subspecies (eastern Ngiri in

DRC, on the eastern bank of the Ubangi River) and thousands of markers (Table S1), allows us to assess the relationship between central and eastern chimpanzee subspecies. Despite Ngiri being geographically closer to Goualougo (a central chimpanzee sampling site, ~280 km) than to any eastern chimpanzee location in our dataset (Rubi-Télé, ~845 km, and Bili, ~900 km) (Figure 1A), individuals from Ngiri clearly fall within the genetic diversity of eastern chimpanzees in the PCA (Figures 1G and S33), pointing to a clear long-term separation of these subspecies. These findings support an unequivocal separation of central and eastern chimpanzee subspecies over a large evolutionary time. However, subsequent recent interbreeding has been suggested by other studies.[6,11]

Long-term subspecies differentiation and genetic exchange during the Late Pleistocene

The sustained genetic differentiation of chimpanzee subspecies can be interpreted in the context of geographical barriers impeding gene flow, especially the major rivers in tropical Africa.[41] We applied the EEMS method[42] to analyze long-term migration landscapes during the Late Pleistocene and Early Holocene[43] (Figures 2B and S82). We found evidence for regions of reduced effective migration that overlap with geographic barriers, such as the Sanaga River (separating Nigeria-Cameroon and central chimpanzees) and the Ubangi River (separating central and eastern chimpanzees) (Figures 2B and S82; Note S10). These patterns of stratification and shared drift were also supported by F_{ST} and f_3 statistics (Figures S45, and S54; Notes S6 and S7).

Previous evidence suggested that some chimpanzee subspecies have not been fully isolated since their separation, but rather experienced migration events.[3,6,11,44] To analyze the permeability of subspecies barriers to gene flow, we used two methods designed to capture signatures of gene flow at different timescales. First, we used identical-by-descent-like (IBD-like) segments detected between individuals from different subspecies using IBDseq,[45] i.e., regions of the chromosome where two individuals share variation. Since the detected segments are smaller than 0.5 mega-base pairs (Mbp) between subspecies, they represent genetic exchange that happened more than approximately 5 kya, assuming an exponential decay of fragment length due to recombination (STAR Methods; Note S10; Figure S89; Table S7). Second, we inferred shorter introgressed fragments between chimpanzee subspecies with the aforementioned method admixfrog,[40] using four genomes of each chimpanzee subspecies from the reference panel as sources[6] to partition genomic regions into the subspecies state they resemble most (STAR Methods; Note S8). We found evidence of gene flow between the central, eastern, and Nigeria-Cameroon subspecies with both methods (Figures 2B, 3, and S66; Table S7), indicating low levels of genetic exchange at different timescales despite their long-term separation. We observed that Nigeria-Cameroon Gashaka individuals carry more fragments of central and eastern chimpanzee ancestry than other Nigeria-Cameroon chimpanzees, while the central Goualougo individuals carry more eastern and Nigeria-Cameroon chimpanzee fragments than the other central communities (Figure S66). This indicates gene flow between these local populations, which is also supported by an analysis of shared rare alleles, which are likely to have emerged more recently[46] and whose sharing patterns are informative on recent admixture (Figures S76, and S77; Note S9). The observation of a northern area of past genetic exchange between the three subspecies is broadly consistent with conclusions from microsatellite data,[11] and with previous studies suggesting a hybrid zone between central and Nigeria-Cameroon chimpanzees in central Cameroon.[44,47]

Recent history between communities since the LGM

Local population stratification within subspecies, probably arising during the Late Pleistocene, has been partially explored previously for eastern and central chimpanzees using whole genomes, but with a much smaller sample size and sampling density.[6] Here, for the first time, we can explore the fine-scale population structure and recent connectivity across the whole geographic range since the LGM and into the Holocene for all subspecies, partially down to the specific site level (Figure S31). To do this, we combine information from different methods that can specifically identify connectivity and isolation at different timescales, specifically EEMS[42] (more than 6 kya), shared rare alleles (~1.5–15 kya, Note S9),[48] long (>0.5 Mbp) IBD-like tracts shared between communities of the same subspecies (less than 5 kya; please see more on possible caveats to this approach in the Limitations of the study and Note S10),[49] as well as recent inbreeding with regions of homozygosity (RoH) (Figures S40–S42; Note S6). This yields a comprehensive and detailed picture of genetic connectivity across the chimpanzee range and within subspecies, beyond the broad genetic clines in eastern and western chimpanzees (Figures 1F, 1G, and S31–S34; Note S5).

Overall, western chimpanzees exhibit higher levels of connectivity across their range and across timescales than the other subspecies, as detected with IBD-like shared fragments, rare variation, and EEMS (Figures 2B, 2C, 3C, S79, S82, and S84; Note S10). Remarkably, for the same geographic distances, western chimpanzee sampling sites share more and longer IBD-like tracts than the other subspecies (Figures 3C and S90), especially within their northern range (Senegal, Mali, northern Guinea, and Guinea-Bissau). Also, shared rare variation resembles the results from IBD-like shared fragments (Figure S79). It is important to note that western chimpanzees have the lowest diversity and likely suffered a strong bottleneck,[6] so our results could support two different scenarios: either high levels of recent connectivity between persisting populations during the past ~780 years (according to the IBD-like tract length; range 117–2,200 years) (Table S9), or a range expansion into the fringe areas of the chimpanzee habitat within the same time frame, resulting in a very recent separation of these populations[50] (Figure 2C). However, at this stage we cannot distinguish these scenarios based on genetic data only. All four sampling sites of Nigeria-Cameroon chimpanzees seem to have been connected within the past 2,500 years (mean 1,600 until 1,000 years ago), indicated by both IBD-like segments and rare allele connectivity (Figures 3D and S77; Note S9). Furthermore, a signature of recent inbreeding in Mbe (i.e., long RoH) suggests that this population was strongly isolated only very recently[51] (Figure S40).

Eastern chimpanzee sampling sites largely follow a pattern of isolation-by-distance, shown as an exponential decay of IBD-

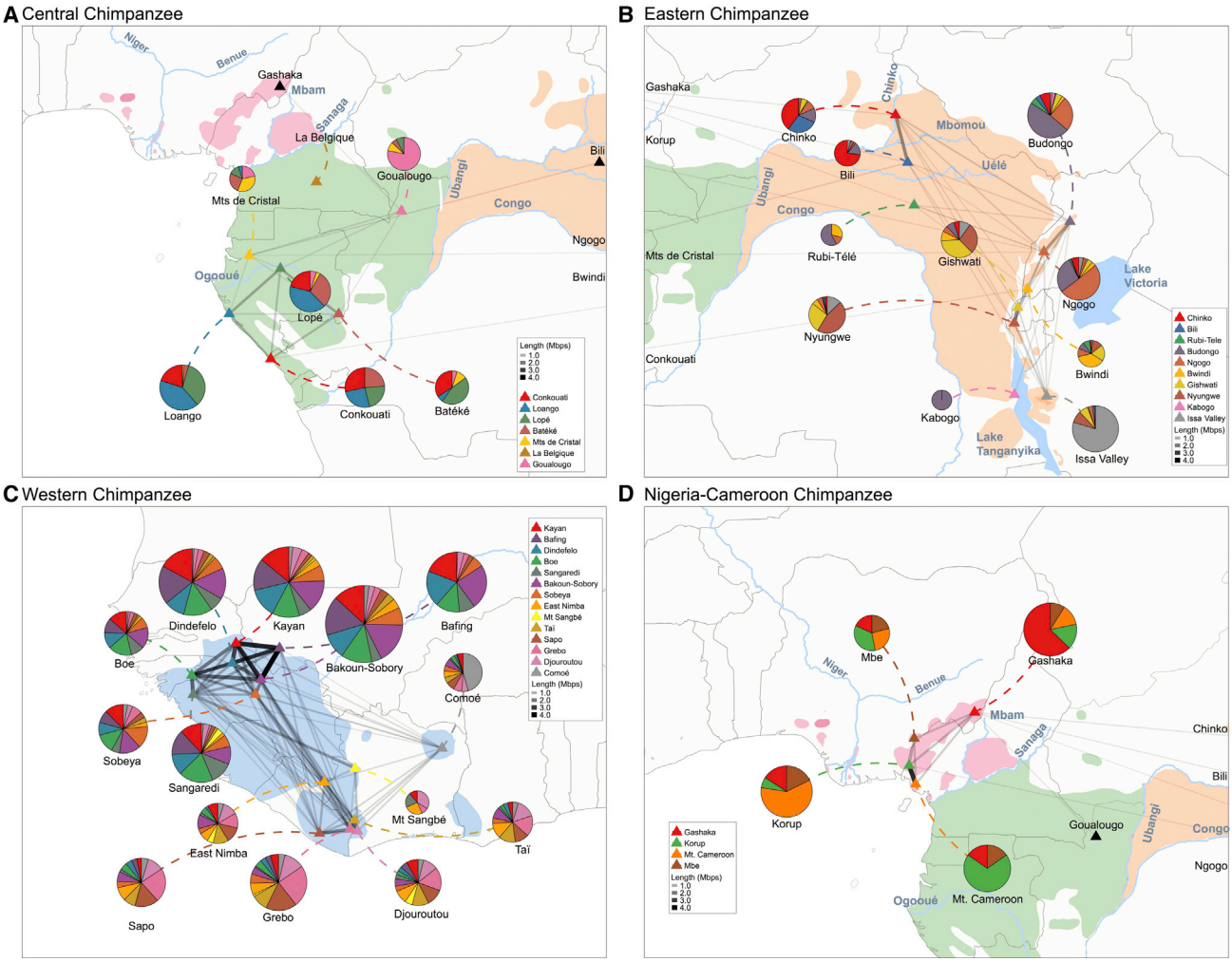


Figure 3. Recent connectivity between chimpanzee populations

(A–D) The size of the pie charts represents the pairwise number of shared fragments, normalized by the number of pairs. Thickness of lines indicates the average length of IBD-like tracts (in Mbp). Triangles show the location of sites. Colors in pies indicate the origin of IBD-like tracts, including comparison between samples from the same site. (A) Central chimpanzees, (B) eastern chimpanzees, (C) western chimpanzees, and (D) Nigeria-Cameroon chimpanzees. Note also few and short IBD-like fragment connections between central, eastern, and Nigeria-Cameroon subspecies. See also Figures S86–S94.

like fragment length (Figure S90) along a genetic North-South cline also found in the PCA (Figure 1G). However, we observed three clusters of recent connectivity reflected in a higher number and longer IBD-like segments (Figure 3B, thicker lines between Chinko-Bili, Budongo-Ngogo, and Gishwati-Nyungwe). Also, the Uéle River and Lake Tanganyika likely acted as isolation barriers in eastern chimpanzee populations in recent times, which is supported by IBD-like segments and shared rare variation (Figures 2C, 3B, S76–S79, and S92; Table S3; Notes S9 and S10). Dispersal corridors suggested for populations in western Uganda[52] and between western Uganda and the eastern DRC[53] (Figure 2B) are supported by these types of analyses (Figures 3B and S78). Finally, all eastern chimpanzee populations share rare variation with those communities living in the area of previously proposed Pleistocene refugia[12,14] (Budongo, Bwindi, Gishwati, Ngogo, and Nyungwe), suggesting an expansion into the southeast (Issa Valley[54]), central and southwest (Regomuki), and northwest (Rubi-Télé, Bili, Chinko, and Ngiri) after the LGM (Figures 2B and S78; Note S9). In central chimpanzees, we detected two strongly differentiated population clusters rather than a cline (Figures 2B,S76, S82, and S84; Notes S6, S7, and S10), separated by the Ogooué River in Gabon, which appears to have been a barrier reducing migration between these regions at least since the LGM, and maintained through the Holocene. Meanwhile, connectivity was higher within each central chimpanzee cluster, indicated by IBD-like tracts, rare allele sharing, and the EEMS surface (Figures 3A, S76, and S84; Notes S9 and S10). The southern cluster also matches with those populations that show a larger amount of bonobo-like introgressed fragments (Figure S64).

Geolocalization of chimpanzees using rare alleles

Our unique sampling breadth allowed the discovery of ~50% more new genetic variants on chromosome 21 (Figure S29) in

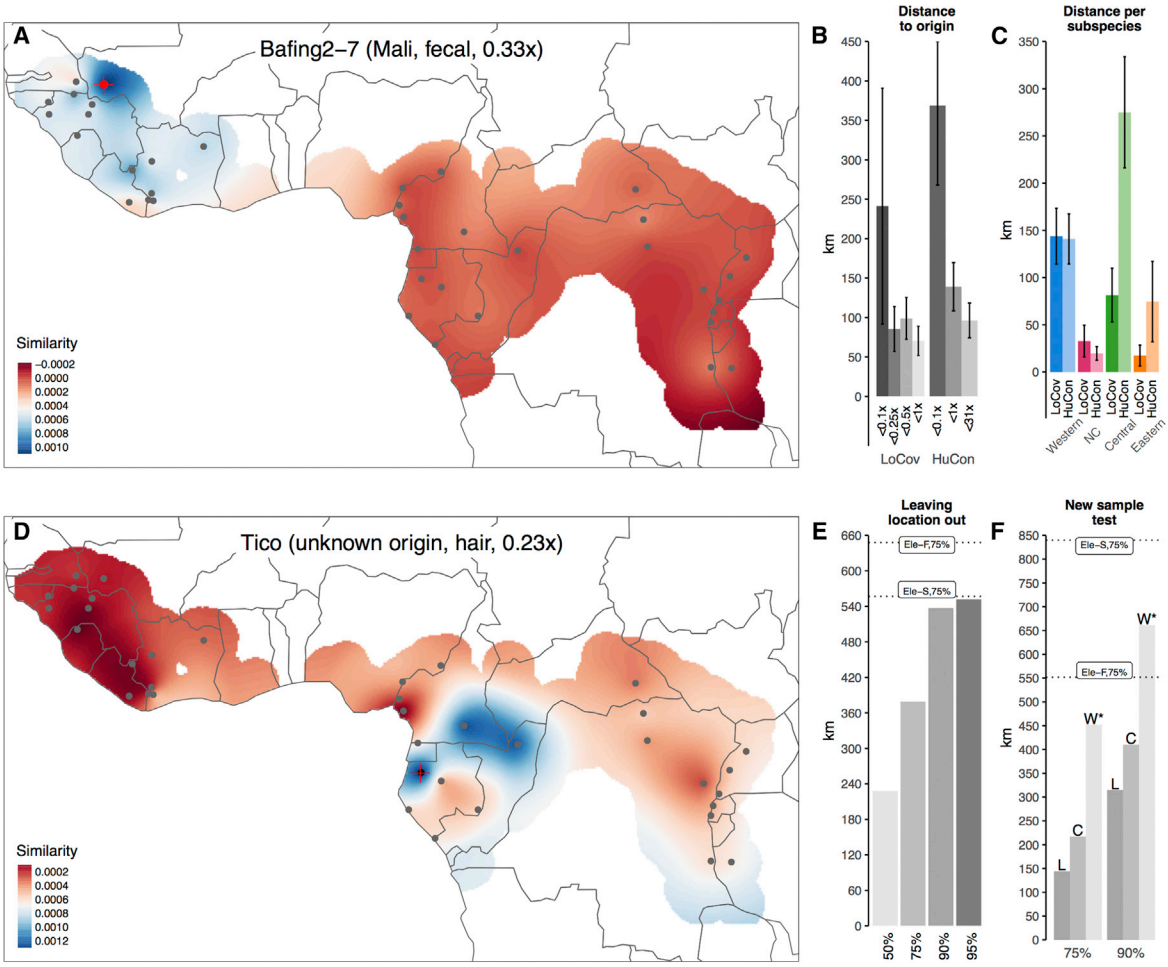


Figure 4. Chimpanzee geolocalization based on rare variation

(A) Spatial model of shared rare alleles with 38 sampling locations. Red indicates lower amounts, while blue indicates larger amounts of shared rare alleles. Black dots, locations included in the reference panel; red dot, known place of origin (low-coverage sample Baf2-7 from Bafing, Mali); red cross, inferred origin. Red cross and dot overlap in this correctly assigned sample.

(B) Average distance (km) of best matching to true location in bins of coverage for samples with low coverage and contamination (LoCov) (n = 99) and samples with human contamination of more than 0.5% (HuCon) (n = 139). Error bars represent the SEM.

(C) Average distance of best matching to true location per subspecies, stratified by low-coverage and human-contaminated samples; note that the Nigeria-Cameroon chimpanzee range is smaller than that of other subspecies, thus resulting in smaller distances. Error bars represent the SEM.

(D) Geolocalization of the chimpanzee Tico from a rescue center in Spain, here assigned to Gabon or Equatorial Guinea.

(E) Assignment accuracy when leaving full locations out (n = 434), with 50th, 75th, 90th, and 95th percentiles for the distance of inferred to true origin; for comparison, best 75th percentiles for geolocalization of elephants[21] are shown as dotted lines (Ele-S, Savanna elephant; Ele-F, forest elephant).

(F) Assignment accuracy for samples not included in the reference panel (L, low coverage; C, contaminated; W, whole-genome data[6]); for comparison, the 75th percentiles of the single sample test in elephants are shown as dotted lines; the asterisk marks that the origin for whole genomes may be different from the place of confiscation reported for these individuals. See also Figures S67–S75, S80, and S81.

comparison with previously published chimpanzee whole genomes.[6] In particular, rare variation likely emerged recently (during few hundreds to thousands of years[46]), and will be geographically structured. Hence, rare alleles are particularly useful for geolocalization because the chimpanzee groups studied here do show local stratification in the sharing of these alleles (Figures S76–S79; Note S9). Here, we developed a strategy to use rare variants (STAR Methods; Figures S67–S75; Note S9) to infer the geographic origin of samples. In brief, we used a reference panel of 434 samples of sufficient quality (Note S3) across 38 sampling locations, obtained the derived frequency of each SNP within each population, and retained SNPs that were observed at one given sampling location but at low cumulative frequency (lower than 1) across all other locations (STAR Methods; Note S9). We then tested samples by calculating their proportion of matching genotypes, across all such positions, to each reference population, and applied a spatial interpolation (kriging) across the chimpanzee range, allowing the visualization of regions of putative origin (e.g., Figure 4A; Data S1, Figure S96).

First, we applied this strategy to 99 samples excluded from the reference panel due to low coverage (<1-fold), as well as 139 samples with human contamination (>0.5%) (Data S1, Figures S96 and S97). At a coverage of more than 0.1-fold, samples are, on average, located 81 km (0–502 km) from their true origin (Figure 4B). In the presence of human contamination (>0.5%, coverage >0.1-fold), this average increases to 139 km, on average, mostly due to central chimpanzee samples (Figure 4C). Samples from locations not included in the reference panel are assigned to nearby regions of the corresponding subspecies (Data S1, Figure S98).

We assessed the accuracy of our method using an approach from a previous study in elephants[21] by inferring the origin of samples when leaving their sampling location out of the reference panel. We find that 75% of the samples are inferred to originate from within 379 km of their sampling location (Figure 4E), considerably closer than the closest 75th percentile in elephants (557 km for sample groups of savannah elephants[21]). Remarkably, when comparing the closest 75th percentile of testing single samples where the sampling location was included in elephants (552 km for forest elephants), in chimpanzees we find that this distance from the true location is less than half for the low-coverage (144 km) and contaminated (217 km) samples (Figure 4E). Our geographically dense reference panel with thousands of markers, likely enhanced by a lower overall mobility of chimpanzees compared with elephants, makes our methodology outperform the elephant one, even though genotype data are extremely patchy and incomplete. Also, the approximate origin of previously published chimpanzee whole genomes[6] (Note S9; Data S1, Figure S102) is closer to the known place of origin or confiscation (75th percentile: 452 km; Figure 4F) than what has been found for elephants of known origin. Finally, we used this strategy to estimate the most probable origin of 20 chimpanzees from two Spanish rescue centers (Fundació Mona and Fundación Rainfer), which were sequenced at low coverage from hair and blood samples (median 0.35-fold coverage, ranging from 0.15- to 4.3-fold) (Figures S80 and S81; Note S9). Hence, with our method even shallow sequencing (without target capture) provides enough information for the approximate geolocalization of chimpanzees with unknown or low confidence origin information (e.g., Figures 4D, S80, and S81; Note S9).

DISCUSSION

Our study shows how non-invasive samples can be used as a source of genomic DNA for population and conservation genomic purposes. Here, we have implemented target capture on chimpanzee fecal samples, although it is worth noting that the same approach could be applied to other great ape and primate species, broadening their application from a few autosomal, sex-linked, or mtDNA markers to an entire chromosome. Precisely, by target capturing a complete chromosome we have the power to discover variation previously unreported and detect contiguous segments of DNA that are inherited together. We found evidence supporting the genetic differentiation of the four recognized subspecies of chimpanzee populations,[3,6] whose differentiation could be linked to historical geographical barriers, in particular the Sanaga River and Ubangi River. Such barriers of gene flow[41,55] have been proposed before, particularly the Congo river separating bonobos from chimpanzees. However, rivers have not been immutable throughout history, and a reduction of river discharge during glaciation periods likely opened corridors for migration;[38] for example, allowing ancient introgression from bonobos into non-western chimpanzees[6] and also between chimpanzee subspecies.[44] Here, we detected differential amounts of ancient introgression from bonobos to central chimpanzee populations north and south of the Ogooué River. This could be explained either by multiple phases of genetic exchange between chimpanzees and bonobos, as has been suggested previously,[6] or by a dilution of bonobo ancestry due to admixture with other chimpanzee populations, as supported by a higher Nigeria-Cameroon and eastern chimpanzee ancestry in the central chimpanzee populations north of the Ogooué River. However, these scenarios are not mutually exclusive, and need to be further investigated using multiple whole genomes from these different regions.

Importantly, this dataset is useful to study the population history and connectivity of wild chimpanzee communities in more recent times. Population stratification in chimpanzee populations can be explained by isolation-by-distance to some degree,[11] but known ecological or geographical barriers have also reduced gene flow between certain populations for extended periods of time, leading to substantial substructure in chimpanzees. This is the case for the Ogooué River acting as a barrier between northern and southern central chimpanzee populations, or Lake Tanganyika separating eastern chimpanzee populations in the south.[53] The Uélé River, isolating eastern chimpanzees since the LGM in the north, is concordant with observed behavioral differences to its north and south.[56,57] Corridors of gene flow between non-western chimpanzee subspecies have been suggested previously,[3,6,11] and we restrict these events mainly to specific areas between central, Nigeria-Cameroon, and eastern chimpanzee populations in the north of their range, particularly between Goualougo and Gashaka, located at the northern fringe of the distribution of these subspecies. However, due to the lack of sampling in eastern Cameroon, we propose that a historical corridor may have reached from the northern range of central chimpanzees to Gashaka through central Cameroon, in concordance with previous results on mtDNA.[44]

These patterns of isolation-by-distance over tens of thousands of years, with genetic interactions occurring on a local scale, stand in apparent contrast to the demographic history of most human populations during the same time frame, which is characterized by high levels of migration.[46] We speculate that chimpanzee's comparably lower migration pattern might be related to a lower extent of information transmission, which is a fundamental difference between them and humans.[58] We speculate that limited genetic and cultural exchange in chimpanzees compared with humans might be a consequence of the social structure of chimpanzees.[59] The higher inter-connectivity of western chimpanzees may also help to explain their larger behavioral diversity compared with non-western chimpanzee populations. A large degree of sharing of IBD-like fragments in the northwestern range of western chimpanzees, resulting from

either recent expansion or high recent connectivity, might reflect population movements from Pleistocene refugia in the south (Liberia, Côte d'Ivoire) after the LGM (Figure 2B),[12,13] possibly related to the proposed cultural expansion in western chimpanzees.[14] However, the Comoé sites in the east of Côte d'Ivoire are genetically closer to forest populations in the south (Figures S45 and S54), despite seemingly being behaviorally similar to the north-eastern mosaic woodland habitat populations.[60] We also find genomic support for an expansion from Pleistocene refugia in eastern chimpanzee populations to the south, west, and northwest after the LGM (Figure 2B).

Using our knowledge of genetic diversity linked to geographical locations, we present a strategy for geolocalization with improved accuracy and precision, even when using low-coverage or contaminated samples (Figure 4). Geolocalization of chimpanzees has direct conservation applications: first, it can help ensure that confiscated chimpanzees from illegal pet trade[61] are placed into sanctuaries in their countries of origin as mandated by the international standards.[62] Second, when sequencing confiscated individuals or wildlife products (e.g., bushmeat), it can allow for the detection of poaching hotspots, so relevant authorities can enforce national and international laws enacted for protected species.[21,63] Successful methods have been developed for African elephants,[21,64] but past attempts in chimpanzees did not provide sufficient spatial resolution;[19,20] while, unfortunately, microsatellite data do not yield a sufficient degree of genetic structure in chimpanzees.[11] However, some geographic regions are not well resolved, resulting in different possible countries of origin, as is the case for other species.[21] Considering that samples are assigned to nearby locations when their sampling site is not covered (Data S1; Figure S99), this is likely to be improved with yet better sampling. Our strategy is based on low-coverage shotgun sequencing, with lower costs but requiring state-of-the-art laboratory facilities and bioinformatic know-how to process and identify the origin of a confiscated individual, which is not accessible for *in situ* genotyping.[65,66] However, new optimizations on sequencing technologies, such as Oxford Nanopore Technologies,[67] might be helpful to obtain genotype information on-site, to ascertain the origins of confiscated wildlife and products.

In conclusion, using the capture of chromosome 21 on hundreds of chimpanzee fecal samples, we presented the first geographically linked catalog of genomic diversity in extant wild chimpanzee populations. This resource allows for the determination of fine-scale population structure, past and recent gene flow, and migration events, and the construction of a geo-genetic map for the geolocalization of orphaned chimpanzees and confiscated bushmeat.

Limitations of the study

The use of non-invasive samples for population genomics is still limited by their low quality and low proportions of hDNA. Under these circumstances, whole-genome sequencing, which would provide stronger support in many analyses, is prohibited by both low library complexity and economical constraints. Since sequencing was limited to a portion of the genome, we could not reach enough confidence to resolve the origin of the differential amount of bonobo introgression in central chimpanzees, and we cannot apply the standard methods to study gene flow. The nature of our dataset also impedes the reconstruction of recent connectivity using IBD-like segments since the accuracy to detect those segments is directly limited by the missingness inherent in low-coverage datasets. Therefore, the timing of the events using the length of the IBD-like fragments can encompass large confidence intervals since the low coverage and high missingness in the data could result in underestimating their length, leading to inaccurate timings.

Fecal samples may be subject to contamination from mammalian or primate DNA from species included in the diet of the chimpanzee. Even though we used a very thorough quality control, due to our limited coverage we cannot discard small remnants of contamination in our dataset.

Finally, the geolocalization approach is based on rare variation, and relies on having a dense georeferenced panel of samples; even after our extensive sampling effort there are some under-represented areas where future studies should focus on gathering samples to fill in the current gaps.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Sample selection and sequencing
  - Data processing and filtering
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Population genetics
  - Rare alleles

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2022.100133.

ACKNOWLEDGMENTS

We would like to thank Marc de Manuel, Juan Antonio Rodriguez, Sojung Han, Christopher Barratt, Richard McElreath, and Lauren White for helpful discussion, and Alan Riedel, Katharina Madl, Veronika Staedele, and Amy Heilman for assistance in the laboratory. We thank Martijn Ter Heegde, Nadege Wangue Njomen, Joshua M. Linder, John Hart, Thurston Cleveland Hicks, Arcel Bamba, Richard Tshombe, Bruno Perodeau, Gita Chelluri, Theophile Desarmeaux, Vianet Mihindou, David Fine, Laura Kehoe, Lucy D'Auvergne, Nuria Maldonado, Luz Calia Miramontes Sequeiros, Theo Freeman, Hilde Vanleeuwe, Jean Claude Dengui, Paul Telfer, Michael Masozera, Nicolas Ntare, Alhaji Malikie Siaka, Henk Eshuis, Jill Pruetz, Geoffrey Muhanguzi, Karsten Dierks, Marcel Ketchan Eyong, Manasseh Eno-Nku, Abel Nzeheke, Yasmin Moebius, Floris Aubert, Matthieu Bonnet, Gregory Brazzola, Chloe Cipoletta, Katherine Corogenes, Charlotte Coupland, Bryan Curran, Emmanuel Dilambaka, Dervla Dowd, Kathryn J. Jeffery, Mohamed Kambi, Vincent Lapeyre, Vera Leinert, Giovanna Maretti, Rumen Martin, Amelia Meier, Protais Niyigabae, Robinson Orume, Jodie Preece, Sebastien Regnaut, Emilien Terrade,

and Alexander Tickle for assistance in field site coordination and sample collection. We thank the field assistants and volunteers from the Jane Goodall Institute Spain and Senegal for their help with sample collection in Dindefelo. We thank the team of Rescue & Rehabilitation of the Fundació Mona. We thank Agilent for their collaboration in the project.

We thank the following government agencies for their support in conducting field research in their countries: Ministère de la Recherche Scientifique et de l'Innovation, Cameroon; Ministère des Forêts et de la Faune, Cameroon; Ministère des Eaux et Forêts, Cote d'Ivoire; Ministère de l'Enseignement Supérieur et de la Recherche Scientifique in Côte d'Ivoire; Institut Congolais pour la Conservation de la Nature, DR-Congo; Ministère de la Recherche Scientifique, DR-Congo; Agence Nationale des Parcs Nationaux, Gabon; Center National de la Recherche Scientifique (CENAREST), Gabon; Société Equatoriale d'Exploitation Forestière (SEEF), Gabon; Department of Wildlife and Range Management, Ghana; Forestry Commission, Ghana; Ministère de l'Agriculture de l'Elevage et des Eaux et Forets, Guinea; Instituto da Biodiversidade e das Áreas Protegidas (IBAP), Guinea-Bissau; Ministro da Agricultura e Desenvolvimento Rural, Guinea-Bissau; Forestry Development Authority, Liberia; Eaux et Forets, Mali; Ministre de l'Environnement et de l'Assainissement et du Developpement Durable du Mali; Conservation Society of Mbe Mountains (CAMM), Nigeria; National Park Service, Nigeria; Ministère de l'Economie Forestière, R-Congo; Ministère de le Recherche Scientifique et Technologique, R-Congo; Ministry of Education, Rwanda; Rwanda Development Board, Rwanda; Direction des Eaux, Forêts, et Chasses, Senegal; Reserve Naturelle Communautaire de Dindefelo, Senegal; Ministry of Agriculture, Forestry, and Food Security, Sierra Leone; National Protected Area Authority, Sierra Leone; Tanzania Commission for Science and Technology, Tanzania; Tanzania Wildlife Research Institute, Tanzania; Makerere University Biological Field Station (MUBFS), Uganda; Uganda National Council for Science and Technology (UNCST), Uganda; Uganda Wildlife Authority, Uganda; National Forestry Authority, Uganda; Agence Congolaise de la Faune et des Aires Protégées; and The Wild Chimpanzee Foundation (WCF).

Funding: "la Caixa" Foundation doctoral fellowship program LCF/BQ/DE15/10360006 (to C.F.). FPI (Formación de Personal Investigador) PRE2018-083966 from Ministerio de Ciencia, Universidades e Investigación (to M.A.-E.) and CGL2017-82654-P (MINECO/FEDER, UE) (to E.L.). "la Caixa" Foundation (ID 100010434), fellowship code LCF/BQ/PR19/11700002 (to M.K.). Vienna Science and Technology Fund (WWTF) and the City of Vienna project VRG20-001 (to M.K.). The European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 864203) (to T.M.-B.). BFU2017-86471-P (MINECO/FEDER, UE) (to T.M.-B.). "Unidad de Excelencia María de Maeztu", funded by the AEI (CEX2018-000792-M) (to T.M.-B.). Howard Hughes International Early Career (to T.M.-B.). NIH 1R01HG010898-01A1 (to T.M.-B.). Secretaria d'Universitats i Recerca and CERCA Program del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880) (to T.M.-B.). UCL's Wellcome Trust ISSF3 award 204841/Z/16/Z (to A.M.A. and J.M.S.). Generalitat de Catalunya (2017 SGR-1040) (to M. Llorente). Wellcome Trust Investigator Award 202802/Z/16/Z (to D.A.H.). The Pan African Program: The Cultured Chimpanzee (PanAf) is generously funded by the Max Planck Society, the Max Planck Society Innovation Fund, and the Heinz L. Krekeler Foundation.

AUTHOR CONTRIBUTIONS

T.M.-B. and M.A. conceived and supervised the study. M.A., C.B., and H.S.K. direct the Pan African Program: The Cultured Chimpanzee. P.D., T.A., P.A.-V., A.A., S.A., A.K.A., E.A.A., E.B., D.B., M.B., A.C.-A., R.C., H.C., E.D., T.D., A.D., J.D., V.E.E., O.F., A.G., A.C.G., J.H., D.H., V.H., R.A.H.A., I.I., SJ., J.J., P.K., M.K., M.V.K., A.K.K., I.K., D.K., K.L., J.L., B.L., A.L., K.L., M. Llana, M. Llorente, S.M., D.M., F.M., M.M., E.N., S. Nicholl, S. Nixon, E.N., C.O., L.J.O., L.P., A.P., L.R., M.M.R., A.R., C.S., L.S., V.S., F.A.S., N.T., L.R.T., E.T., J.v.S., V.V., E.G.W., J.W., R.M.W., Y.G.Y., K.Y., and K.Z. supervised, conducted fieldwork, and collected samples. C.F., M.A.-E., J.D.L., J.H., M.A., and E.L. performed experimental laboratory work. C.F. and M.K. performed the analysis. C.M.-S., P.G., J.M.S., L.V., A.M.A., D.A.H., H.S.H., E.L., M.A., and T.M.-B. provided analytical support. C.F. and M.K. wrote the manuscript with input from all coauthors.

DECLARATION OF INTEREST

The authors declare no competing interests.

Received: July 2, 2021
Revised: December 29, 2021
Accepted: April 15, 2022
Published: June 1, 2022

REFERENCES

1. Fischer, A., Pollack, J., Thalmann, O., Nickel, B., and Pääbo, S. (2006). Demographic history and genetic differentiation in apes. Curr. Biol. *16*, 1133–1138. https://doi.org/10.1016/j.cub.2006.04.033.
2. Becquet, C., and Przeworski, M. (2007). A new approach to estimate parameters of speciation models with application to apes. Genome Res. *17*, 1505–1519. https://doi.org/10.1101/gr.6409707.
3. Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature *499*, 471–475. https://doi.org/10.1038/nature12228.
4. Lobon, I., Tucci, S., De Manuel, M., Ghirotto, S., Benazzo, A., Prado-Martinez, J., Lorente-Galdos, B., Nam, K., Dabad, M., Hernandez-Rodriguez, J., et al. (2016). Demographic history of the genus Pan inferred from whole mitochondrial genome reconstructions. Genome Biol. Evol. *8*, 2020–2030. https://doi.org/10.1093/gbe/evw124.
5. Hallast, P., Maisano Delser, P., Batini, C., Zadik, D., Rocchi, M., Schempp, W., Tyler-Smith, C., and Jobling, M.A. (2016). Great ape Y Chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. Genome Res. *26*, 427–439. https://doi.org/10.1101/gr.198754.115.
6. De Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V.C., Desai, T., Prado-Martinez, J., Hernandez-Rodriguez, J., Dupanloup, I., Lao, O., Hallast, P., et al. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science *354*, 477–481. https://doi.org/10.1126/science.aag2602.
7. Kuhlwilm, M., de Manuel, M., Nater, A., Greminger, M.P., Krützen, M., and Marques-Bonet, T. (2016). Evolution and demography of the great apes. Curr. Opin. Genet. Dev. *41*, 124–129. https://doi.org/10.1016/j.gde.2016.09.005.
8. Kuhlwilm, M., Han, S., Sousa, V.C., Excoffier, L., and Marques-Bonet, T. (2019). Ancient admixture from an extinct ape lineage into bonobos. Nat. Ecol. Evol. *3*, 957–965. https://doi.org/10.1038/s41559-019-0881-7.
9. McBrearty, S., and Jablonski, N.G. (2005). First fossil chimpanzee. Nature *437*, 105–108. https://doi.org/10.1038/nature04008.
10. Fünfstück, T., Arandjelovic, M., Morgan, D.B., Sanz, C., Reed, P., Olson, S.H., Cameron, K., Ondzie, A., Peeters, M., and Vigilant, L. (2015). The sampling scheme matters: Pan troglodytes troglodytes and P. t schweinfurthii are characterized by clinal genetic variation rather than a strong subspecies break. Am. J. Phys. Anthropol. *156*, 181–191. https://doi.org/10.1002/ajpa.22638.
11. Lester, J.D., Vigilant, L., Gratton, P., McCarthy, M.S., Barratt, C.D., Dieguez, P., Agbor, A., Álvarez-Varona, P., Angedakin, S., Ayimisin, E.A., et al. (2021). Recent genetic connectivity and clinal variation in chimpanzees. Commun. Biol. *4*, 283. https://doi.org/10.1038/s42003-021-01806-x.
12. Mayr, E., and O'Hara, R.J. (1986). The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. Evolution (N. Y). *40*, 55–67. https://doi.org/10.1111/j.1558-5646.1986.tb05717.x.
13. Barratt, C., Lester, J., Gratton, P., Onstein, R., Kalan, A., McCarthy, M., Bocksberger, G., White, L., Vigilant, L., Dieguez, P., et al. (2020). Late

Quaternary habitat suitability models for chimpanzees (Pan troglodytes) since the Last Interglacial (120,000 BP). Preprint at bioRxiv. https://doi.org/10.1101/2020.05.15.066662.

14. Kalan, A.K., Kulik, L., Arandjelovic, M., Boesch, C., Haas, F., Dieguez, P., Barratt, C.D., Abwe, E.E., Agbor, A., Angedakin, S., et al. (2020). Environmental variability supports chimpanzee behavioural diversity. Nat. Commun. *11*, 4451. https://doi.org/10.1038/s41467-020-18176-3.

15. Roberts, P., and Stewart, B.A. (2018). Defining the 'generalist specialist' niche for Pleistocene Homo sapiens. Nat. Hum. Behav. *2*, 542–550. https://doi.org/10.1038/s41562-018-0394-4.

16. Humle, T., Maisels, F., Oates, J.F., Plumptre, A., and Williamson, E.A. (2016). Pan troglodytes (Errata Version Published in 2018). IUCN Red List Threat. Species, e.T15933A129038584. 10.2305/IUCN.UK.2016-2RLTS.T15933A17964454.en.

17. Kühl, H.S., Sop, T., Williamson, E.A., Mundry, R., Brugière, D., Campbell, G., Cohen, H., Danquah, E., Ginn, L., Herbinger, I., et al. (2017). The Critically Endangered western chimpanzee declines by 80%. Am. J. Primatol. *79*, e22681. https://doi.org/10.1002/ajp.22681.

18. Supple, M.A., and Shapiro, B. (2018). Conservation of biodiversity in the genomics era. Genome Biol. *19*, 131. https://doi.org/10.1186/s13059-018-1520-3.

19. Frandsen, P., Fontsere, C., Nielsen, S.V., Hanghøj, K., Castejon-Fernandez, N., Lizano, E., Hughes, D., Hernandez-Rodriguez, J., Korneliussen, T.S., Carlsen, F., et al. (2020). Targeted conservation genetics of the endangered chimpanzee. Heredity *125*, 15–27. https://doi.org/10.1038/s41437-020-0313-0.

20. Ghobrial, L., Lankester, F., Kiyang, J.A., Akih, A.E., de Vries, S., Fotso, R., Gadsby, E.L., Jenkins, P.D., and Gonder, M.K. (2010). Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. BMC Ecol. *10*, 2. https://doi.org/10.1186/1472-6785-10-2.

21. Wasser, S.K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C., and Weir, B.S. (2015). Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. Science *349*, 84–87. https://doi.org/10.1126/science.aaa2457.

22. Banes, G.L., Galdikas, B.M.F., and Vigilant, L. (2016). Reintroduction of confiscated and displaced mammals risks outbreeding and introgression in natural populations, as evidenced by orang-utans of divergent subspecies. Sci. Rep. *6*, 22026. https://doi.org/10.1038/srep22026.

23. Oklander, L.I., Caputo, M., Solari, A., and Corach, D. (2020). Genetic assignment of illegally trafficked neotropical primates and implications for reintroduction programs. Sci. Rep. *10*, 3676. https://doi.org/10.1038/s41598-020-60569-3.

24. Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. Methods Ecol. Evol. *9*, 410–419. https://doi.org/10.1111/2041-210x.12871.

25. Fontsere, C., Alvarez-Estape, M., Lester, J., Arandjelovic, M., Kuhlwilm, M., Dieguez, P., Agbor, A., Angedakin, S., Ayuk Ayimisin, E., Bessone, M., et al. (2021). Maximizing the acquisition of unique reads in noninvasive capture sequencing experiments. Mol. Ecol. Resour. *21*, 745–761. https://doi.org/10.1111/1755-0998.13300.

26. Vigilant, L., and Guschanski, K. (2009). Using genetics to understand the dynamics of wild primate populations. Primates *50*, 105–120. https://doi.org/10.1007/s10329-008-0124-z.

27. Perry, G.H., Marioni, J.C., Melsted, P., and Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. Mol. Ecol. *19*, 5332–5344. https://doi.org/10.1111/j.1365-294x.2010.04888.x.

28. Hernandez-Rodriguez, J., Arandjelovic, M., Lester, J., de Filippo, C., Weihmann, A., Meyer, M., Angedakin, S., Casals, F., Navarro, A., Vigilant, L., et al. (2018). The impact of endogenous content, replicates and pooling on genome capture from faecal samples. Mol. Ecol. Resour. *18*, 319–333. https://doi.org/10.1111/1755-0998.12728.

29. White, L.C., Fontsere, C., Lizano, E., Hughes, D.A., Angedakin, S., Arandjelovic, M., Granjon, A., Hans, J.B., Lester, J.D., Rabanus-Wallace, M.T., et al. (2019). A roadmap for high-throughput sequencing studies of wild animal populations using noninvasive samples and hybridization capture. Mol. Ecol. Resour. *19*, 609–622. https://doi.org/10.1111/1755-0998.12993.

30. Boesch, C., and Boesch, H. (1989). Hunting behavior of wild chimpanzees in the taï national park. Am. J. Phys. Anthropol. *78*, 547–573. https://doi.org/10.1002/ajpa.1330780410.

31. Stanford, C.B., Wallis, J., Matama, H., and Goodall, J. (1994). Patterns of predation by chimpanzees on red colobus monkeys in gombe national park, 1982–1991. Am. J. Phys. Anthropol. *94*, 213–228. https://doi.org/10.1002/ajpa.1330940206.

32. Arandjelovic, M., Head, J., Kühl, H., Boesch, C., Robbins, M.M., Maisels, F., and Vigilant, L. (2010). Effective non-invasive genetic monitoring of multiple wild western gorilla groups. Biol. Conserv. *143*, 1780–1791. https://doi.org/10.1016/j.biocon.2010.04.030.

33. Kuhlwilm, M., Fontsere, C., Han, S., Alvarez-Estape, M., and Marques-Bonet, T. (2021). HuConTest: testing human contamination in great ape samples. Genome Biol. Evol. *13*, evab117. https://doi.org/10.1093/gbe/evab117.

34. Peyrégne, S., and Peter, B.M. (2020). AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. Genome Biol. *21*, 246. https://doi.org/10.1186/s13059-020-02123-y.

35. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinf. *15*, 356. https://doi.org/10.1186/s12859-014-0356-4.

36. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. Genetics *210*, 719–731. https://doi.org/10.1534/genetics.118.301336.

37. Nye, J., Laayouni, H., Kuhlwilm, M., Mondal, M., Marques-Bonet, T., and Bertranpetit, J. (2018). Selection in the introgressed regions of the chimpanzee genome. Genome Biol. Evol. *10*, 1132–1138. https://doi.org/10.1093/gbe/evy077.

38. Takemoto, H., Kawamoto, Y., and Furuichi, T. (2015). How did bonobos come to range south of the Congo river? Reconsideration of the divergence of *Pan paniscus* from other *Pan* populations. Evol. Anthropol. Issues News Rev. *24*, 170–184. https://doi.org/10.1002/evan.21456.

39. Peter, B.M. (2016). Admixture, population structure, and f-statistics. Genetics *202*, 1485–1501. https://doi.org/10.1534/genetics.115.183913.

40. Peter, B. (2020). 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. Preprint at bioRxiv. https://doi.org/10.1101/2020.03.13.990523.

41. Mitchell, M.W., Locatelli, S., Sesink Clee, P.R., Thomassen, H.A., and Gonder, M.K. (2015). Environmental variation and rivers govern the structure of chimpanzee genetic diversity in a biodiversity hotspot. BMC Evol. Biol. *15*, 1. https://doi.org/10.1186/s12862-014-0274-0.

42. Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. *48*, 94–100. https://doi.org/10.1038/ng.3464.

43. Al-Asadi, H., Petkova, D., Stephens, M., and Novembre, J. (2019). Estimating recent migration and population-size surfaces. PLoS Genet. *15*, e1007908. https://doi.org/10.1371/journal.pgen.1007908.

44. Mitchell, M.W., Locatelli, S., Ghobrial, L., Pokempner, A.A., Sesink Clee, P.R., Abwe, E.E., Nicholas, A., Nkembi, L., Anthony, N.M., Morgan, B.J., et al. (2015). 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. *15*, 3. https://doi.org/10.1186/s12862-014-0276-y.

45. Browning, B.L., and Browning, S.R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. Am. J. Hum. Genet. *93*, 840–851. https://doi.org/10.1016/j.ajhg.2013.09.014.

46. Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., and Willerslev, E. (2017). Tracing the peopling of the world through genomics. Nature *541*, 302–310. https://doi.org/10.1038/nature21347.

47. Gagneux, P., Gonder, M.K., Goldberg, T.L., and Morin, P.A. (2001). Gene Flow in Wild Chimpanzee Populations: What Genetic Data Tell Us about Chimpanzee Movement over Space and Time (Philosophical Transactions of the Royal Society B: Biological Sciences), pp. 889–897.

48. Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., et al. (2016). Iron age and Anglo-saxon genomes from east england reveal British migration history. Nat. Commun. *7*, 10408. https://doi.org/10.1038/ncomms10408.

49. Thompson, E.A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. Genetics *194*, 301–326. https://doi.org/10.1534/genetics.112.148825.

50. Wessling, E.G., Kühl, H.S., Mundry, R., Deschner, T., and Pruetz, J.D. (2018). The costs of living at the edge: seasonal stress in wild savanna-dwelling chimpanzees. J. Hum. Evol. *121*, 1–11. https://doi.org/10.1016/j.jhevol.2018.03.001.

51. Kuhlwilm, M., Gronau, I., Hubisz, M.J., De Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H.A., Lalueza-Fox, C., De La Rasilla, M., et al. (2016). Ancient gene flow from early modern humans into Eastern Neanderthals. Nature *530*, 429–433. https://doi.org/10.1038/nature16544.

52. McCarthy, M.S., Lester, J.D., Howe, E.J., Arandjelovic, M., Stanford, C.B., and Vigilant, L. (2015). Genetic censusing identifies an unexpectedly sizeable population of an endangered large mammal in a fragmented forest landscape. BMC Ecol. *15*, 21. https://doi.org/10.1186/s12898-015-0052-x.

53. Rich, A.M., Wasserman, M.D., Hunt, K.D., and Kaestle, F.A. (2020). Chimpanzee (Pan troglodytes schweinfurthii) population spans multiple protected areas in the albertine rift. Folia Primatol *91*, 595–609. https://doi.org/10.1159/000508073.

54. Moore, D.L., and Vigilant, L. (2014). Genetic diversity at the edge: comparative assessment of Y-chromosome and autosomal diversity in eastern chimpanzees (Pan troglodytes schweinfurthii) of Ugalla, Tanzania. Conserv. Genet. *15*, 495–507. https://doi.org/10.1007/s10592-013-0556-x.

55. Brunke, J., Radespiel, U., Russo, I.R., Bruford, M.W., and Goossens, B. (2019). Messing about on the river: the role of geographic barriers in shaping the genetic structure of Bornean small mammals in a fragmented landscape. Conserv. Genet. *20*, 691–704. https://doi.org/10.1007/s10592-019-01159-3.

56. Hicks, T.C., Tranquilli, S., Kuehl, H., Campbell, G., Swinkels, J., Darby, L., Boesch, C., Hart, J., and Menken, S.B.J. (2014). Absence of evidence is not evidence of absence: discovery of a large, continuous population of Pan troglodytes schweinfurthii in the Central Uele region of northern DRC. Biol. Conserv. *171*, 107–113. https://doi.org/10.1016/j.biocon.2014.01.002.

57. Hicks, T.C., Kühl, H.S., Boesch, C., Dieguez, P., Ayimisin, A.E., Fernandez, R.M., Zungawa, D.B., Kambere, M., Swinkels, J., Menken, S.B.J., et al. (2019). Bili-uéré: a chimpanzee behavioural realm in northern democratic republic of Congo. Folia Primatol *90*, 3–64. https://doi.org/10.1159/000492998.

58. Boesch, C., and Tomasello, M. (1998). Chimpanzee and human cultures. Curr. Anthropol. *39*, 591–614. https://doi.org/10.1086/204785.

59. Langergraber, K.E., Siedel, H., Mitani, J.C., Wrangham, R.W., Reynolds, V., Hunt, K., and Vigilant, L. (2007). The genetic signature of sex-biased migration in patrilocal chimpanzees and humans. PLoS One *2*, e973. https://doi.org/10.1371/journal.pone.0000973.

60. Kühl, H.S., Kalan, A.K., Arandjelovic, M., Aubert, F., D'Auvergne, L., Goedmakers, A., Jones, S., Kehoe, L., Regnaut, S., Tickle, A., et al. (2016). Chimpanzee accumulative stone throwing. Sci. Rep. *6*, 22219. https://doi.org/10.1038/srep22219.

61. Stiles, D., Redmond, I., Cress, D., Nellemann, C., and Formo, R.K. (2013). In Stolen Apes - The Illicit Trade in Chimpanzees, Gorillas, Bonobos and Orangutans, R.K. Stiles, I. Redmond, D. Cress, C. Nellemann, and Formo., eds. (United Nations Environment Programme, GRID-Arendal), pp. 1–56.

62. PASA (2016). Pan African Sanctuary Alliance. Operation Manual.

63. Gouda, S., Kerry, R.G., Das, A., and Chauhan, N.S. (2020). Wildlife forensics: a boon for species identification and conservation implications. Forensic Sci. Int. *317*, 110530. https://doi.org/10.1016/j.forsciint.2020.110530.

64. Wasser, S.K., Torkelson, A., Winters, M., Horeaux, Y., Tucker, S., Otiende, M.Y., Sitam, F.A.T., Buckleton, J., and Weir, B.S. (2018). Combating transnational organized crime by linking multiple large ivory seizures to the same dealer. Sci. Adv. *4*, eaat0625, eaat0625. https://doi.org/10.1126/sciadv.aat0625.

65. Mcmahon, B.J., Teeling, E.C., and Höglund, J. (2014). How and why should we implement genomics into conservation? Evol. Appl. *7*, 999–1007. https://doi.org/10.1111/eva.12193.

66. R Taylor, H., Dussex, N., and van Heezik, Y. (2017). Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. Glob. Ecol. Conserv. *10*, 231–242. https://doi.org/10.1016/j.gecco.2017.04.001.

67. Krehenwinkel, H., Pomerantz, A., Henderson, J.B., Kennedy, S.R., Lim, J.Y., Swamy, V., Shoobridge, J.D., Graham, N., Patel, N.H., Gillespie, R.G., and Prost, S. (2019). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. Gigascience *8*, giz006. https://doi.org/10.1093/gigascience/giz006.

68. Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.C., Agarwala, R., McLaren, W.M., Ritchie, G.R.S., et al. (2011). Modernizing reference genome assemblies. Plos Biol. *9*, e1001091. https://doi.org/10.1371/journal.pbio.1001091.

69. Kronenberg, Z.N., Fiddes, I.T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson, O.S., Underwood, J.G., Nelson, B.J., Chaisson, M.J.P., Dougherty, M.L., et al. (2018). High-resolution comparative analysis of great ape genomes. Science *360*, eaar6343. https://doi.org/10.1126/science.aar6343.

70. Rogers, J., Raveendran, M., Harris, R.A., Mailund, T., Leppälä, K., Athanasiadis, G., Schierup, M.H., Cheng, J., Munch, K., Walker, J.A., et al.; Baboon Genome Analysis Consortium (2019). The comparative genomics and complex population history of Papio baboons. Sci. Adv. *5*, eaau6947. https://doi.org/10.1126/sciadv.aau6947.

71. Warren, W.C., Jasinska, A.J., García-Pérez, R., Svardal, H., Tomlinson, C., Rocchi, M., Archidiacono, N., Capozzi, O., Minx, P., Montague, M.J., et al. (2015). The genome of the vervet (Chlorocebus æthiops sabæus). Genome Res. *25*, 1921–1933. https://doi.org/10.1101/gr.192922.115.

72. Genereux, D.P., Serres, A., Armstrong, J., Johnson, J., Marinescu, V.D., Murén, E., Juan, D., Bejerano, G., Casewell, N.R., Chemnick, L.G., et al. (2020). A comparative genomics multitool for scientific discovery and conservation. Nature *587*, 240–245. https://doi.org/10.1038/s41586-020-2876-6.

73. Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al. (2012). Insights into hominid evolution from the gorilla genome sequence. Nature *483*, 169–175. https://doi.org/10.1038/nature10842.

74. Yin, Y., Yang, T., Liu, H., Huang, Z., Zhang, Y., Song, Y., Wang, W., Guang, X., Sahu, S.K., and Kristiansen, K. (2020). The draft genome of mandrill (Mandrillus sphinx): an Old World monkey. Sci. Rep. *10*, 2431. https://doi.org/10.1038/s41598-020-59110-3.

75. Rohland, N., and Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. *22*, 939–946. https://doi.org/10.1101/gr.128124.111.

76. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*. https://doi.org/10.1101/pdb.prot5448.

77. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093. https://doi.org/10.1534/genetics.112.145037.

78. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

79. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

80. Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program: table 1. Mol. Biol. Evol. *32*, 2798–2800. https://doi.org/10.1093/molbev/msv150.

81. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

82. Jónsson, H., Ginolhac, A.A., Schubert, M., Johnson, P.L.F.F., Orlando, L., Jonsson, H., Ginolhac, A.A., Schubert, M., Johnson, P.L.F.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193.

83. Skotte, L., Korneliussen, T.S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. Genetics *195*, 693–702. https://doi.org/10.1534/genetics.113.154138.

84. Vieira, F.G., Lassalle, F., Korneliussen, T.S., and Fumagalli, M. (2016). Improving the estimation of genetic distances from Next-Generation Sequencing data. Biol. J. Linn. Soc. *117*, 139–149. https://doi.org/10.1111/bij.12511.

85. Korneliussen, T.S., Moltke, I., Albrechtsen, A., and Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinf. *14*, 289. https://doi.org/10.1186/1471-2105-14-289.

86. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.

87. Bivand, R., and Lewin-Koh, N. (2013). Maptools: Tools for Handling Spatial Objects. R Packag. version 0.8.

88. Petr, M., Vernot, B., and Kelso, J. (2019). Admixr-R package for reproducible analyses using ADMIXTOOLS. Bioinformatics *35*, 3194–3195. https://doi.org/10.1093/bioinformatics/btz030.

89. Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

90. Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics *27*, 592–593. https://doi.org/10.1093/bioinformatics/btq706.

91. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223. https://doi.org/10.1111/j.2041-210x.2011.00169.x.

92. Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. R. J. *10*, 439–446. https://doi.org/10.32614/rj-2018-009.

93. Pebesma, E.J., and Bivand, R.S. (2005). Classes and methods for spatial data in {R. R. News *5*, 9–13.

94. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2020). Vegan: Community Ecology Package.

95. R Core Team (2020). A Language and Environment for Statistical Computing. R. Found. Stat. Comput. *2*. https://www.R-project.org.

96. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

97. Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., and Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS One *7*, e37558. https://doi.org/10.1371/journal.pone.0037558.

98. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

99. Prüfer, K. (2018). SNPAD: an ancient DNA genotype caller. Bioinformatics *34*, 4165–4171. https://doi.org/10.1093/bioinformatics/bty507.

100. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967.

101. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

102. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

103. Arandjelovic, M., Guschanski, K., Schubert, G., Harris, T.R., Thalmann, O., Siedel, H., and Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. Mol. Ecol. Resour. *9*, 28–36. https://doi.org/10.1111/j.1755-0998.2008.02387.x.

104. Arandjelovic, M., Head, J., Rabanal, L.I., Schubert, G., Mettke, E., Boesch, C., Robbins, M.M., and Vigilant, L. (2011). Non-invasive genetic monitoring of wild central chimpanzees. PLoS One *6*, e14761. https://doi.org/10.1371/journal.pone.0014761.

105. Korneliussen, T.S., and Moltke, I. (2015). NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics *31*, btv509–4011. https://doi.org/10.1093/bioinformatics/btv509.

106. Nater, A., Mattle-Greminger, M.P., Nurcahyo, A., Nowak, M.G., de Manuel, M., Desai, T., Groves, C., Pybus, M., Sonay, T.B., Roos, C., et al. (2017). Morphometric, behavioral, and genomic evidence for a new orangutan species. Curr. Biol. *27*, 3487–3498.e10. https://doi.org/10.1016/j.cub.2017.09.047.e10.

107. Nait Saada, J., Kalantzis, G., Shyr, D., Cooper, F., Robinson, M., Gusev, A., and Palamara, P.F. (2020). Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. Nat. Commun. *11*, 6130. https://doi.org/10.1038/s41467-020-19588-x.

108. Auton, A., Fledel-Alon, A., Pfeifer, S., Venn, O., Ségurel, L., Street, T., Leffler, E.M., Bowden, R., Aneas, I., Broxholme, J., et al. (2012). A fine-scale chimpanzee genetic map from population sequencing. Science *336*, 193–198. https://doi.org/10.1126/science.1216872.

109. Stevison, L.S., Woerner, A.E., Kidd, J.M., Kelley, J.L., Veeramah, K.R., McManus, K.F., Bustamante, C.D., Hammer, M.F., Wall, J.D., Lorente-Galdos, B., et al. (2016). The time scale of recombination rate evolution in great apes. Mol. Biol. Evol. *33*, 928–945. https://doi.org/10.1093/molbev/msv331.

110. Langergraber, K.E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J.C., Muller, M.N., et al. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc. Natl. Acad. Sci. U S A *109*, 15716–15721. https://doi.org/10.1073/pnas.1211740109.

111. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The simons genome diversity project: 300 genomes from 142 diverse populations. Nature *538*, 201–206. https://doi.org/10.1038/nature18964.

112. Pebesma, E., and Gräler, B. (2014). Spatio-temporal geostatistics using gstat. R. J. *8*, 204–218.

113. Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., et al. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science *348*, 242–245. https://doi.org/10.1126/science.aaa3952.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Chimpanzee hair sample	This Study	AZA-01-01
Chimpanzee hair sample	This Study	AZA-01-02
Chimpanzee hair sample	This Study	AZA-01-03
Chimpanzee hair sample	This Study	AZA-01-06
Chimpanzee hair sample	This Study	AZA-01-08
Chimpanzee hair sample	This Study	AZA-01-09
Chimpanzee hair sample	This Study	AZA-01-10
Chimpanzee hair sample	This Study	AZA-01-11
Chimpanzee hair sample	This Study	AZA-01-12
Chimpanzee fecal sample	This Study	Baf1-12
Chimpanzee fecal sample	This Study	Baf1-13
Chimpanzee fecal sample	This Study	Baf1-16
Chimpanzee fecal sample	This Study	Baf1-17
Chimpanzee fecal sample	This Study	Baf1-19
Chimpanzee fecal sample	This Study	Baf1-2
Chimpanzee fecal sample	This Study	Baf1-5
Chimpanzee fecal sample	This Study	Baf2-1
Chimpanzee fecal sample	This Study	Baf2-26
Chimpanzee fecal sample	This Study	Baf2-27
Chimpanzee fecal sample	This Study	Baf2-4
Chimpanzee fecal sample	This Study	Baf2-42
Chimpanzee fecal sample	This Study	Baf2-43
Chimpanzee fecal sample	This Study	Baf2-44
Chimpanzee fecal sample	This Study	Baf2-46
Chimpanzee fecal sample	This Study	Baf2-68
Chimpanzee fecal sample	This Study	Baf2-7
Chimpanzee fecal sample	This Study	Baf2-73
Chimpanzee fecal sample	This Study	Baf2-74
Chimpanzee fecal sample	This Study	Baf2-75
Chimpanzee fecal sample	This Study	Bat1-1
Chimpanzee fecal sample	This Study	Bat1-10
Chimpanzee fecal sample	This Study	Bat1-11
Chimpanzee fecal sample	This Study	Bat1-14
Chimpanzee fecal sample	This Study	Bat1-15
Chimpanzee fecal sample	This Study	Bat1-16
Chimpanzee fecal sample	This Study	Bat1-18
Chimpanzee fecal sample	This Study	Bat1-19
Chimpanzee fecal sample	This Study	Bat1-20
Chimpanzee fecal sample	This Study	Bat1-22
Chimpanzee fecal sample	This Study	Bat1-24
Chimpanzee fecal sample	This Study	Bat1-3
Chimpanzee fecal sample	This Study	Bat1-37
Chimpanzee fecal sample	This Study	Bat1-39

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Bat1-4
Chimpanzee fecal sample	This Study	Bat1-42
Chimpanzee fecal sample	This Study	Bat1-48
Chimpanzee fecal sample	This Study	Bat1-5
Chimpanzee fecal sample	This Study	Bat1-50
Chimpanzee fecal sample	This Study	Bat1-6
Chimpanzee fecal sample	This Study	Bil1-1
Chimpanzee fecal sample	This Study	Bil1-10
Chimpanzee fecal sample	This Study	Bil1-15
Chimpanzee fecal sample	This Study	Bil1-20
Chimpanzee fecal sample	This Study	Bil1-24
Chimpanzee fecal sample	This Study	Bil1-25
Chimpanzee fecal sample	This Study	Bil1-28
Chimpanzee fecal sample	This Study	Bil1-3
Chimpanzee fecal sample	This Study	Bil1-30
Chimpanzee fecal sample	This Study	Bil1-37
Chimpanzee fecal sample	This Study	Bil1-38
Chimpanzee fecal sample	This Study	Bil1-41
Chimpanzee fecal sample	This Study	Bil1-46
Chimpanzee fecal sample	This Study	Bil1-47
Chimpanzee fecal sample	This Study	Bil1-50
Chimpanzee fecal sample	This Study	Bil1-51
Chimpanzee fecal sample	This Study	Bil1-7
Chimpanzee fecal sample	This Study	Bil1-8
Chimpanzee fecal sample	This Study	Bili2-5
Chimpanzee fecal sample	This Study	Bili2-7
Chimpanzee fecal sample	This Study	Boe1-30
Chimpanzee fecal sample	This Study	Boe1-32
Chimpanzee fecal sample	This Study	Boe1-33
Chimpanzee fecal sample	This Study	Boe1-36
Chimpanzee fecal sample	This Study	Boe1-43
Chimpanzee fecal sample	This Study	Boe1-44
Chimpanzee fecal sample	This Study	Boe1-57
Chimpanzee fecal sample	This Study	Boe1-60
Chimpanzee fecal sample	This Study	Boe1-71
Chimpanzee fecal sample	This Study	Boe1-75
Chimpanzee fecal sample	This Study	Boe1-77
Chimpanzee fecal sample	This Study	Boe2-10
Chimpanzee fecal sample	This Study	Boe2-11
Chimpanzee fecal sample	This Study	Boe2-18
Chimpanzee fecal sample	This Study	Boe2-22
Chimpanzee fecal sample	This Study	Boe2-26
Chimpanzee fecal sample	This Study	Boe2-30
Chimpanzee fecal sample	This Study	Boe2-33
Chimpanzee fecal sample	This Study	Boe2-39
Chimpanzee fecal sample	This Study	Boe2-6
Chimpanzee fecal sample	This Study	Bou1-1
Chimpanzee fecal sample	This Study	Bud1-1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Bud1-10
Chimpanzee fecal sample	This Study	Bud1-12
Chimpanzee fecal sample	This Study	Bud1-15
Chimpanzee fecal sample	This Study	Bud1-17
Chimpanzee fecal sample	This Study	Bud1-2
Chimpanzee fecal sample	This Study	Bud1-21
Chimpanzee fecal sample	This Study	Bud1-6
Chimpanzee fecal sample	This Study	Bud1-7
Chimpanzee fecal sample	This Study	Bud1-9
Chimpanzee fecal sample	This Study	Bud2-29
Chimpanzee fecal sample	This Study	Bud2-45
Chimpanzee fecal sample	This Study	Bud2-48
Chimpanzee fecal sample	This Study	Bud2-52
Chimpanzee fecal sample	This Study	Bud2-54
Chimpanzee fecal sample	This Study	Bud2-73
Chimpanzee fecal sample	This Study	Bud2-81
Chimpanzee fecal sample	This Study	Bud2-86
Chimpanzee fecal sample	This Study	Bud3-13
Chimpanzee fecal sample	This Study	Bud3-30
Chimpanzee fecal sample	This Study	Bwi-2-26
Chimpanzee fecal sample	This Study	Bwi-2-39
Chimpanzee fecal sample	This Study	Bwi1-10
Chimpanzee fecal sample	This Study	Bwi1-16
Chimpanzee fecal sample	This Study	Bwi1-29
Chimpanzee fecal sample	This Study	Bwi1-3
Chimpanzee fecal sample	This Study	Bwi1-36
Chimpanzee fecal sample	This Study	Bwi1-4
Chimpanzee fecal sample	This Study	Bwi1-55
Chimpanzee fecal sample	This Study	Bwi1-58
Chimpanzee fecal sample	This Study	Bwi1-6
Chimpanzee fecal sample	This Study	Bwi1-61
Chimpanzee fecal sample	This Study	Bwi1-64
Chimpanzee fecal sample	This Study	Bwi1-66
Chimpanzee fecal sample	This Study	Bwi1-69
Chimpanzee fecal sample	This Study	Bwi1-7
Chimpanzee fecal sample	This Study	Bwi1-71
Chimpanzee fecal sample	This Study	Bwi1-74
Chimpanzee fecal sample	This Study	Bwi1-78
Chimpanzee fecal sample	This Study	Bwi1-90
Chimpanzee fecal sample	This Study	Cam1-13
Chimpanzee fecal sample	This Study	Cam1-14
Chimpanzee fecal sample	This Study	Cam1-18
Chimpanzee fecal sample	This Study	Cam1-2
Chimpanzee fecal sample	This Study	Cam1-21
Chimpanzee fecal sample	This Study	Cam1-26
Chimpanzee fecal sample	This Study	Cam1-27
Chimpanzee fecal sample	This Study	Cam1-29
Chimpanzee fecal sample	This Study	Cam1-44

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Cam1-49
Chimpanzee fecal sample	This Study	Cam1-50
Chimpanzee fecal sample	This Study	Cam1-7
Chimpanzee fecal sample	This Study	Cam1-71
Chimpanzee fecal sample	This Study	Cam1-74
Chimpanzee fecal sample	This Study	Cam1-78
Chimpanzee fecal sample	This Study	Cam2-77
Chimpanzee fecal sample	This Study	Cam3-29
Chimpanzee fecal sample	This Study	Cam3-40
Chimpanzee fecal sample	This Study	Cam3-41
Chimpanzee fecal sample	This Study	Cam3-45
Chimpanzee fecal sample	This Study	Chinko-1
Chimpanzee fecal sample	This Study	Chinko-10
Chimpanzee fecal sample	This Study	Chinko-12
Chimpanzee fecal sample	This Study	Chinko-13
Chimpanzee fecal sample	This Study	Chinko-14
Chimpanzee fecal sample	This Study	Chinko-16
Chimpanzee fecal sample	This Study	Chinko-3
Chimpanzee fecal sample	This Study	Chinko-5
Chimpanzee fecal sample	This Study	Chinko-8
Chimpanzee fecal sample	This Study	CMNP1-19
Chimpanzee fecal sample	This Study	CMNP1-24
Chimpanzee fecal sample	This Study	CMNP1-43
Chimpanzee fecal sample	This Study	CMNP1-8
Chimpanzee fecal sample	This Study	CMNP2-1_B
Chimpanzee fecal sample	This Study	CMNP2-5
Chimpanzee fecal sample	This Study	CMNP2-6
Chimpanzee fecal sample	This Study	Cnp1-1
Chimpanzee fecal sample	This Study	Cnp1-14
Chimpanzee fecal sample	This Study	Cnp1-2
Chimpanzee fecal sample	This Study	Cnp1-36
Chimpanzee fecal sample	This Study	Cnp1-37
Chimpanzee fecal sample	This Study	Cnp1-47
Chimpanzee fecal sample	This Study	Cnp1-63
Chimpanzee fecal sample	This Study	Cnp1-70
Chimpanzee fecal sample	This Study	Cnp1-75
Chimpanzee fecal sample	This Study	CNPE1-1
Chimpanzee fecal sample	This Study	CNPE1-12
Chimpanzee fecal sample	This Study	CNPE1-2
Chimpanzee fecal sample	This Study	CNPE1-22
Chimpanzee fecal sample	This Study	CNPE1-26
Chimpanzee fecal sample	This Study	CNPE1-3
Chimpanzee fecal sample	This Study	CNPE1-31
Chimpanzee fecal sample	This Study	CNPE1-36
Chimpanzee fecal sample	This Study	CNPE1-6
Chimpanzee fecal sample	This Study	CNPE1-7
Chimpanzee fecal sample	This Study	CNPN1-20
Chimpanzee fecal sample	This Study	CNPN1-35

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	CNPN1-63
Chimpanzee fecal sample	This Study	CNPW1-13
Chimpanzee fecal sample	This Study	CNPW1-16_2
Chimpanzee fecal sample	This Study	CNPW1-17
Chimpanzee fecal sample	This Study	CNPW1-2
Chimpanzee fecal sample	This Study	CNPW1-40
Chimpanzee fecal sample	This Study	CNPW1-7
Chimpanzee fecal sample	This Study	CNPW2-29
Chimpanzee fecal sample	This Study	CNPW2-43
Chimpanzee fecal sample	This Study	Con1-12
Chimpanzee fecal sample	This Study	Con2-23
Chimpanzee fecal sample	This Study	Con2-25
Chimpanzee fecal sample	This Study	Con2-27
Chimpanzee fecal sample	This Study	Con2-38
Chimpanzee fecal sample	This Study	Con2-48
Chimpanzee fecal sample	This Study	Con2-49
Chimpanzee fecal sample	This Study	Con2-50
Chimpanzee fecal sample	This Study	Con2-53
Chimpanzee fecal sample	This Study	Con2-56
Chimpanzee fecal sample	This Study	Con2-57
Chimpanzee fecal sample	This Study	Con2-64
Chimpanzee fecal sample	This Study	Con2-66
Chimpanzee fecal sample	This Study	Con2-67
Chimpanzee fecal sample	This Study	Con2-71
Chimpanzee fecal sample	This Study	Con2-80
Chimpanzee fecal sample	This Study	Con3-10
Chimpanzee fecal sample	This Study	Con3-8
Chimpanzee fecal sample	This Study	Din1-10
Chimpanzee fecal sample	This Study	Din1-22
Chimpanzee fecal sample	This Study	Din1-26
Chimpanzee fecal sample	This Study	Din1-3
Chimpanzee fecal sample	This Study	Din1-4
Chimpanzee fecal sample	This Study	Din1-53
Chimpanzee fecal sample	This Study	Din1-6
Chimpanzee fecal sample	This Study	Din1-68
Chimpanzee fecal sample	This Study	Din1-7
Chimpanzee fecal sample	This Study	Din2-22
Chimpanzee fecal sample	This Study	Din2-29
Chimpanzee fecal sample	This Study	Din2-3
Chimpanzee fecal sample	This Study	Din2-38
Chimpanzee fecal sample	This Study	Din2-43
Chimpanzee fecal sample	This Study	Din2-79
Chimpanzee fecal sample	This Study	Din2-83
Chimpanzee fecal sample	This Study	Din3-5
Chimpanzee fecal sample	This Study	Din3-7
Chimpanzee fecal sample	This Study	Din3-8
Chimpanzee fecal sample	This Study	Din3-9
Chimpanzee fecal sample	This Study	Dja1-16

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Dja1-17
Chimpanzee fecal sample	This Study	Dja1-23
Chimpanzee fecal sample	This Study	Dja1-8
Chimpanzee fecal sample	This Study	Dja2-20
Chimpanzee fecal sample	This Study	Dja2-21
Chimpanzee fecal sample	This Study	Dja2-22
Chimpanzee fecal sample	This Study	Dja2-23
Chimpanzee fecal sample	This Study	Dja2-25
Chimpanzee fecal sample	This Study	Dja2-27
Chimpanzee fecal sample	This Study	Dja2-30
Chimpanzee fecal sample	This Study	Dja2-36
Chimpanzee fecal sample	This Study	Dja2-39
Chimpanzee fecal sample	This Study	Dja2-42
Chimpanzee fecal sample	This Study	Dja2-57
Chimpanzee fecal sample	This Study	Dja3-19
Chimpanzee fecal sample	This Study	Dja3-20
Chimpanzee fecal sample	This Study	Dja3-21
Chimpanzee fecal sample	This Study	Dja3-6
Chimpanzee fecal sample	This Study	Dja3-7
Chimpanzee fecal sample	This Study	Djo1-13
Chimpanzee fecal sample	This Study	Djo1-14
Chimpanzee fecal sample	This Study	Djo1-2
Chimpanzee fecal sample	This Study	Djo1-20
Chimpanzee fecal sample	This Study	Djo1-22
Chimpanzee fecal sample	This Study	Djo1-37
Chimpanzee fecal sample	This Study	Djo1-5
Chimpanzee fecal sample	This Study	Djo1-50
Chimpanzee fecal sample	This Study	Djo1-54
Chimpanzee fecal sample	This Study	Djo1-6
Chimpanzee fecal sample	This Study	Djo1-60
Chimpanzee fecal sample	This Study	Djo1-66
Chimpanzee fecal sample	This Study	Djo2-29
Chimpanzee fecal sample	This Study	Djo2-4
Chimpanzee fecal sample	This Study	Djo2-5
Chimpanzee fecal sample	This Study	Djo2-50
Chimpanzee fecal sample	This Study	Djo2-68
Chimpanzee fecal sample	This Study	Djo2-8
Chimpanzee fecal sample	This Study	Djo3-1
Chimpanzee fecal sample	This Study	Djo3-2
Chimpanzee fecal sample	This Study	El3-16
Chimpanzee fecal sample	This Study	El3-17
Chimpanzee fecal sample	This Study	El3-18
Chimpanzee fecal sample	This Study	El3-4
Chimpanzee fecal sample	This Study	El3-5
Chimpanzee fecal sample	This Study	El3-6
Chimpanzee fecal sample	This Study	Fjn1-10
Chimpanzee fecal sample	This Study	Fjn1-20
Chimpanzee fecal sample	This Study	Fjn1-21

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Fjn1-22
Chimpanzee fecal sample	This Study	Fjn1-42
Chimpanzee fecal sample	This Study	Fjn2-13
Chimpanzee fecal sample	This Study	Fjn2-50
Chimpanzee fecal sample	This Study	Fjn2-52
Chimpanzee fecal sample	This Study	Fjn2-62
Chimpanzee fecal sample	This Study	Fjn2-7
Chimpanzee fecal sample	This Study	Fjn2-9
Chimpanzee fecal sample	This Study	Fjn3-24
Chimpanzee fecal sample	This Study	Fjn3-43
Chimpanzee fecal sample	This Study	Fjn3-53
Chimpanzee fecal sample	This Study	Fjn3-54
Chimpanzee fecal sample	This Study	Fjn3-56
Chimpanzee fecal sample	This Study	Fjn3-68
Chimpanzee fecal sample	This Study	Fjn3-84
Chimpanzee fecal sample	This Study	Fouta1-10
Chimpanzee fecal sample	This Study	Fouta1-17
Chimpanzee fecal sample	This Study	Fouta1-6
Chimpanzee fecal sample	This Study	Fouta2-8
Chimpanzee fecal sample	This Study	Fouta3-1
Chimpanzee fecal sample	This Study	Fouta3-15
Chimpanzee fecal sample	This Study	Fouta3-25
Chimpanzee fecal sample	This Study	Fouta3-29
Chimpanzee fecal sample	This Study	Fouta3-30
Chimpanzee fecal sample	This Study	Fouta3-32
Chimpanzee fecal sample	This Study	Fouta3-34
Chimpanzee fecal sample	This Study	Fouta3-35
Chimpanzee fecal sample	This Study	Fouta3-37
Chimpanzee fecal sample	This Study	Fouta3-38
Chimpanzee fecal sample	This Study	Fouta3-40
Chimpanzee fecal sample	This Study	Fouta3-51
Chimpanzee fecal sample	This Study	Fouta3-55
Chimpanzee fecal sample	This Study	Fouta3-80
Chimpanzee fecal sample	This Study	Fouta3-82
Chimpanzee fecal sample	This Study	Fouta3-87
Chimpanzee fecal sample	This Study	Gas1-10
Chimpanzee fecal sample	This Study	Gas1-17
Chimpanzee fecal sample	This Study	Gas1-22
Chimpanzee fecal sample	This Study	Gas1-23
Chimpanzee fecal sample	This Study	Gas1-26
Chimpanzee fecal sample	This Study	Gas1-27
Chimpanzee fecal sample	This Study	Gas1-36
Chimpanzee fecal sample	This Study	Gas1-5
Chimpanzee fecal sample	This Study	Gas2-14
Chimpanzee fecal sample	This Study	Gas2-19
Chimpanzee fecal sample	This Study	Gas2-23
Chimpanzee fecal sample	This Study	Gas2-28
Chimpanzee fecal sample	This Study	Gas2-29

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Gas2-34
Chimpanzee fecal sample	This Study	Gas2-37
Chimpanzee fecal sample	This Study	Gas2-4
Chimpanzee fecal sample	This Study	Gas2-40
Chimpanzee fecal sample	This Study	Gas2-55
Chimpanzee fecal sample	This Study	Gas2-67
Chimpanzee fecal sample	This Study	Gas2-7
Chimpanzee fecal sample	This Study	GB-10-03
Chimpanzee fecal sample	This Study	GB-11-10
Chimpanzee fecal sample	This Study	GB-11-11
Chimpanzee fecal sample	This Study	GB-13-13
Chimpanzee fecal sample	This Study	GB-13-21
Chimpanzee fecal sample	This Study	GB-14-05
Chimpanzee fecal sample	This Study	GB-22-06
Chimpanzee fecal sample	This Study	GB-25-02
Chimpanzee fecal sample	This Study	GB-25-05
Chimpanzee fecal sample	This Study	GB-28-02
Chimpanzee fecal sample	This Study	GB-29-06
Chimpanzee fecal sample	This Study	GB-30-11
Chimpanzee fecal sample	This Study	GB-34-16
Chimpanzee fecal sample	This Study	GB-34-22
Chimpanzee fecal sample	This Study	GB-36-07
Chimpanzee fecal sample	This Study	GB-36-16
Chimpanzee fecal sample	This Study	GB-37-04
Chimpanzee fecal sample	This Study	GB-37-09
Chimpanzee fecal sample	This Study	Gbo1-10
Chimpanzee fecal sample	This Study	Gbo1-13
Chimpanzee fecal sample	This Study	Gbo1-15
Chimpanzee fecal sample	This Study	Gbo1-27
Chimpanzee fecal sample	This Study	Gbo1-41
Chimpanzee fecal sample	This Study	Gbo1-53
Chimpanzee fecal sample	This Study	Gbo1-86
Chimpanzee fecal sample	This Study	Gbo2-10
Chimpanzee fecal sample	This Study	Gbo2-2
Chimpanzee fecal sample	This Study	Gbo2-25
Chimpanzee fecal sample	This Study	Gbo2-43
Chimpanzee fecal sample	This Study	Gbo2-48
Chimpanzee fecal sample	This Study	Gbo2-57
Chimpanzee fecal sample	This Study	Gbo2-59
Chimpanzee fecal sample	This Study	Gbo2-63
Chimpanzee fecal sample	This Study	Gbo2-66
Chimpanzee fecal sample	This Study	Gbo2-85
Chimpanzee fecal sample	This Study	Gbo3-17
Chimpanzee fecal sample	This Study	Gbo3-2
Chimpanzee fecal sample	This Study	Gco1-25
Chimpanzee fecal sample	This Study	Gco1-32
Chimpanzee fecal sample	This Study	Gco1-33_2
Chimpanzee fecal sample	This Study	Gco1-37

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Gco1-39
Chimpanzee fecal sample	This Study	Gco1-42_2
Chimpanzee fecal sample	This Study	Gco1-43
Chimpanzee fecal sample	This Study	Gco1-44
Chimpanzee fecal sample	This Study	Gco1-48
Chimpanzee fecal sample	This Study	Gco1-5
Chimpanzee fecal sample	This Study	Gco1-50
Chimpanzee fecal sample	This Study	Gco1-51
Chimpanzee fecal sample	This Study	Gco1-55
Chimpanzee fecal sample	This Study	Gco1-56
Chimpanzee fecal sample	This Study	Gco1-60
Chimpanzee fecal sample	This Study	Gco1-61
Chimpanzee fecal sample	This Study	Gco1-8
Chimpanzee fecal sample	This Study	Gco2-5
Chimpanzee fecal sample	This Study	Gco2-7
Chimpanzee fecal sample	This Study	Gco2-8
Chimpanzee fecal sample	This Study	Gco2-9
Chimpanzee fecal sample	This Study	Gco4-2
Chimpanzee fecal sample	This Study	Gep1-21
Chimpanzee fecal sample	This Study	Gep1-23
Chimpanzee fecal sample	This Study	Gep1-25
Chimpanzee fecal sample	This Study	Gep1-26
Chimpanzee fecal sample	This Study	Gep1-62
Chimpanzee fecal sample	This Study	Gep1-65
Chimpanzee fecal sample	This Study	Gep2-10
Chimpanzee fecal sample	This Study	Gep2-20
Chimpanzee fecal sample	This Study	Gep2-28
Chimpanzee fecal sample	This Study	Gep2-29
Chimpanzee fecal sample	This Study	Gep2-30
Chimpanzee fecal sample	This Study	Gep2-37
Chimpanzee fecal sample	This Study	Gep2-40
Chimpanzee fecal sample	This Study	Gep2-41
Chimpanzee fecal sample	This Study	Gep2-45
Chimpanzee fecal sample	This Study	Gep2-48
Chimpanzee fecal sample	This Study	Gep2-52
Chimpanzee fecal sample	This Study	Gep2-53
Chimpanzee fecal sample	This Study	Gep2-61
Chimpanzee fecal sample	This Study	Gha-01-01
Chimpanzee fecal sample	This Study	Gha-01-04
Chimpanzee fecal sample	This Study	Gha-01-05
Chimpanzee fecal sample	This Study	Gha-01-06
Chimpanzee fecal sample	This Study	Gha-01-07
Chimpanzee fecal sample	This Study	Gha-01-08
Chimpanzee fecal sample	This Study	Gha-01-11
Chimpanzee fecal sample	This Study	Gis1-1
Chimpanzee fecal sample	This Study	Gis1-10
Chimpanzee fecal sample	This Study	Gis1-11
Chimpanzee fecal sample	This Study	Gis1-13

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Gis1-17
Chimpanzee fecal sample	This Study	Gis1-20
Chimpanzee fecal sample	This Study	Gis1-21
Chimpanzee fecal sample	This Study	Gis1-23
Chimpanzee fecal sample	This Study	Gis1-24
Chimpanzee fecal sample	This Study	Gis1-25
Chimpanzee fecal sample	This Study	Gis1-4
Chimpanzee fecal sample	This Study	Gis1-47
Chimpanzee fecal sample	This Study	Gis1-5
Chimpanzee fecal sample	This Study	Gis1-59
Chimpanzee fecal sample	This Study	Gis1-6
Chimpanzee fecal sample	This Study	Gis1-70
Chimpanzee fecal sample	This Study	Gis1-8
Chimpanzee fecal sample	This Study	Gis2-2
Chimpanzee fecal sample	This Study	Gis2-50
Chimpanzee fecal sample	This Study	Gis2-7
Chimpanzee fecal sample	This Study	Gou1-14
Chimpanzee fecal sample	This Study	Gou1-15
Chimpanzee fecal sample	This Study	Gou1-18
Chimpanzee fecal sample	This Study	Gou1-20
Chimpanzee fecal sample	This Study	Gou1-21
Chimpanzee fecal sample	This Study	Gou1-23
Chimpanzee fecal sample	This Study	Gou1-24
Chimpanzee fecal sample	This Study	Gou1-27
Chimpanzee fecal sample	This Study	Gou1-38
Chimpanzee fecal sample	This Study	Gou1-4
Chimpanzee fecal sample	This Study	Gou1-40
Chimpanzee fecal sample	This Study	Gou1-51
Chimpanzee fecal sample	This Study	Gou1-58
Chimpanzee fecal sample	This Study	Gou1-61
Chimpanzee fecal sample	This Study	Gou1-66
Chimpanzee fecal sample	This Study	Gou1-7
Chimpanzee fecal sample	This Study	Gou1-70
Chimpanzee fecal sample	This Study	Gou1-75
Chimpanzee fecal sample	This Study	Gou1-8
Chimpanzee fecal sample	This Study	Gou1-9
Chimpanzee fecal sample	This Study	Itu-01-01
Chimpanzee fecal sample	This Study	Itu-01-02
Chimpanzee fecal sample	This Study	Itu-01-03
Chimpanzee fecal sample	This Study	Itu-01-04
Chimpanzee fecal sample	This Study	Itu-01-05
Chimpanzee fecal sample	This Study	Itu-01-06
Chimpanzee fecal sample	This Study	Itu-01-07
Chimpanzee fecal sample	This Study	Itu-01-08
Chimpanzee fecal sample	This Study	Itu-01-09
Chimpanzee fecal sample	This Study	Itu-01-10
Chimpanzee fecal sample	This Study	Itu-01-11
Chimpanzee fecal sample	This Study	Itu-01-12

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Ivi1-1
Chimpanzee fecal sample	This Study	Ivi1-2
Chimpanzee fecal sample	This Study	Kab1-1
Chimpanzee fecal sample	This Study	Kab1-2
Chimpanzee fecal sample	This Study	Kab1-3
Chimpanzee fecal sample	This Study	Kab1-4
Chimpanzee fecal sample	This Study	Kab1-5
Chimpanzee fecal sample	This Study	Kab2-1
Chimpanzee fecal sample	This Study	Kab2-4
Chimpanzee fecal sample	This Study	Kab2-5
Chimpanzee fecal sample	This Study	Kay1-12
Chimpanzee fecal sample	This Study	Kay1-13
Chimpanzee fecal sample	This Study	Kay1-15
Chimpanzee fecal sample	This Study	Kay1-16
Chimpanzee fecal sample	This Study	Kay1-17
Chimpanzee fecal sample	This Study	Kay1-20
Chimpanzee fecal sample	This Study	Kay1-23
Chimpanzee fecal sample	This Study	Kay1-4
Chimpanzee fecal sample	This Study	Kay2-20
Chimpanzee fecal sample	This Study	Kay2-24
Chimpanzee fecal sample	This Study	Kay2-25
Chimpanzee fecal sample	This Study	Kay2-26
Chimpanzee fecal sample	This Study	Kay2-29
Chimpanzee fecal sample	This Study	Kay2-3
Chimpanzee fecal sample	This Study	Kay2-32
Chimpanzee fecal sample	This Study	Kay2-4
Chimpanzee fecal sample	This Study	Kay2-41
Chimpanzee fecal sample	This Study	Kay2-49
Chimpanzee fecal sample	This Study	Kay2-52
Chimpanzee fecal sample	This Study	Kay2-54
Chimpanzee fecal sample	This Study	Kor1-12
Chimpanzee fecal sample	This Study	Kor1-14
Chimpanzee fecal sample	This Study	Kor1-15
Chimpanzee fecal sample	This Study	Kor1-24
Chimpanzee fecal sample	This Study	Kor1-25
Chimpanzee fecal sample	This Study	Kor1-27
Chimpanzee fecal sample	This Study	Kor1-34
Chimpanzee fecal sample	This Study	Kor1-35
Chimpanzee fecal sample	This Study	Kor1-65
Chimpanzee fecal sample	This Study	Kor1-79
Chimpanzee fecal sample	This Study	Kor1-8
Chimpanzee fecal sample	This Study	Kor1-84
Chimpanzee fecal sample	This Study	Kor2-1
Chimpanzee fecal sample	This Study	Kor2-14
Chimpanzee fecal sample	This Study	Kor2-17
Chimpanzee fecal sample	This Study	Kor2-26
Chimpanzee fecal sample	This Study	Kor2-35
Chimpanzee fecal sample	This Study	Kor2-5

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Kor2-8
Chimpanzee fecal sample	This Study	LCA-3-10
Chimpanzee fecal sample	This Study	LCA-3-12
Chimpanzee fecal sample	This Study	Lib1-25D
Chimpanzee fecal sample	This Study	Lib1-6-D
Chimpanzee fecal sample	This Study	Lib2-10
Chimpanzee fecal sample	This Study	Lib2-14
Chimpanzee fecal sample	This Study	Lib2-15
Chimpanzee fecal sample	This Study	Lib2-17
Chimpanzee fecal sample	This Study	Lib2-2
Chimpanzee fecal sample	This Study	Lib2-23
Chimpanzee fecal sample	This Study	Lib2-26
Chimpanzee fecal sample	This Study	Lib2-27
Chimpanzee fecal sample	This Study	Lib2-28
Chimpanzee fecal sample	This Study	Lib2-48
Chimpanzee fecal sample	This Study	Lib2-62
Chimpanzee fecal sample	This Study	Lib2-66
Chimpanzee fecal sample	This Study	Lib3-34
Chimpanzee fecal sample	This Study	Loma2-1
Chimpanzee fecal sample	This Study	Loma2-2
Chimpanzee fecal sample	This Study	Loma2-3
Chimpanzee fecal sample	This Study	Loma2-4
Chimpanzee fecal sample	This Study	Loma2-5
Chimpanzee fecal sample	This Study	Loma2-6
Chimpanzee fecal sample	This Study	Loma2-7
Chimpanzee fecal sample	This Study	Lop1-13
Chimpanzee fecal sample	This Study	Lop1-14
Chimpanzee fecal sample	This Study	Lop1-23
Chimpanzee fecal sample	This Study	Lop1-24
Chimpanzee fecal sample	This Study	Lop1-25
Chimpanzee fecal sample	This Study	Lop2-11
Chimpanzee fecal sample	This Study	Lop2-16
Chimpanzee fecal sample	This Study	Lop2-3
Chimpanzee fecal sample	This Study	Lop2-34
Chimpanzee fecal sample	This Study	Lop2-35
Chimpanzee fecal sample	This Study	Lop2-43
Chimpanzee fecal sample	This Study	Lop2-45
Chimpanzee fecal sample	This Study	Lop2-76
Chimpanzee fecal sample	This Study	Lop2-77
Chimpanzee fecal sample	This Study	Lop2-80
Chimpanzee fecal sample	This Study	Lop2-82
Chimpanzee fecal sample	This Study	Lop2-88
Chimpanzee fecal sample	This Study	Lop3-11
Chimpanzee fecal sample	This Study	Lop3-14
Chimpanzee fecal sample	This Study	Lop3-20
Chimpanzee fecal sample	This Study	Mbe-02-01
Chimpanzee fecal sample	This Study	Mbe-02-04
Chimpanzee fecal sample	This Study	Mbe-02-05

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Mbe-02-07
Chimpanzee fecal sample	This Study	Mbe-02-09
Chimpanzee fecal sample	This Study	Mbe-02-12
Chimpanzee fecal sample	This Study	Mbe-02-13
Chimpanzee fecal sample	This Study	Mbe1-10
Chimpanzee fecal sample	This Study	Mbe1-12
Chimpanzee fecal sample	This Study	Mbe1-13
Chimpanzee fecal sample	This Study	Mbe1-15
Chimpanzee fecal sample	This Study	Mbe1-16
Chimpanzee fecal sample	This Study	Mbe1-17
Chimpanzee fecal sample	This Study	Mbe1-18
Chimpanzee fecal sample	This Study	Mbe1-19
Chimpanzee fecal sample	This Study	Mbe1-2
Chimpanzee fecal sample	This Study	Mbe1-20
Chimpanzee fecal sample	This Study	Mbe1-21
Chimpanzee fecal sample	This Study	Mbe1-22_2
Chimpanzee fecal sample	This Study	Mbe1-23
Chimpanzee fecal sample	This Study	Mbe1-24
Chimpanzee fecal sample	This Study	Mbe1-25
Chimpanzee fecal sample	This Study	Mbe1-26
Chimpanzee fecal sample	This Study	Mbe1-4
Chimpanzee fecal sample	This Study	Mbe1-5
Chimpanzee fecal sample	This Study	Mbe1-7
Chimpanzee fecal sample	This Study	Mbe1-9
Chimpanzee fecal sample	This Study	Mtc1-26
Chimpanzee fecal sample	This Study	Mtc1-40
Chimpanzee fecal sample	This Study	Mtc1-43
Chimpanzee fecal sample	This Study	Mtc1-54
Chimpanzee fecal sample	This Study	Mtc1-55
Chimpanzee fecal sample	This Study	Mtc1-56
Chimpanzee fecal sample	This Study	Mtc1-58
Chimpanzee fecal sample	This Study	Mtc1-63
Chimpanzee fecal sample	This Study	Mtc1-66
Chimpanzee fecal sample	This Study	Mtc1-67
Chimpanzee fecal sample	This Study	Mtc1-71
Chimpanzee fecal sample	This Study	Mtc1-72
Chimpanzee fecal sample	This Study	MTC2-24
Chimpanzee fecal sample	This Study	MTC2-31
Chimpanzee fecal sample	This Study	MTC2-33
Chimpanzee fecal sample	This Study	MTC2-40
Chimpanzee fecal sample	This Study	MTC2-42
Chimpanzee fecal sample	This Study	MTC2-5
Chimpanzee fecal sample	This Study	MTC2-6
Chimpanzee fecal sample	This Study	MTC2-7
Chimpanzee fecal sample	This Study	N173-11
Chimpanzee fecal sample	This Study	N173-14
Chimpanzee fecal sample	This Study	N173-17
Chimpanzee fecal sample	This Study	N181-11

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	N181-14
Chimpanzee fecal sample	This Study	N182-2
Chimpanzee fecal sample	This Study	N183-5
Chimpanzee fecal sample	This Study	N183-6
Chimpanzee fecal sample	This Study	N186-8
Chimpanzee fecal sample	This Study	N186-9
Chimpanzee fecal sample	This Study	N190-3
Chimpanzee fecal sample	This Study	N259-5
Chimpanzee fecal sample	This Study	N259-6
Chimpanzee fecal sample	This Study	N259-8
Chimpanzee fecal sample	This Study	N260-10
Chimpanzee fecal sample	This Study	N260-6
Chimpanzee fecal sample	This Study	N260-8
Chimpanzee fecal sample	This Study	N261-3
Chimpanzee fecal sample	This Study	N261-5
Chimpanzee fecal sample	This Study	N262-4
Chimpanzee fecal sample	This Study	Ngi1-1
Chimpanzee fecal sample	This Study	Ngi1-2
Chimpanzee fecal sample	This Study	Ngi1-3
Chimpanzee fecal sample	This Study	Ngi1-4
Chimpanzee fecal sample	This Study	Ngi1-5
Chimpanzee fecal sample	This Study	Ngi1-7
Chimpanzee fecal sample	This Study	Ngi1-8
Chimpanzee fecal sample	This Study	Ngi2-1
Chimpanzee fecal sample	This Study	Ngi2-3
Chimpanzee fecal sample	This Study	Ngi2-4
Chimpanzee fecal sample	This Study	Ngi2-5
Chimpanzee fecal sample	This Study	Ngi2-6
Chimpanzee fecal sample	This Study	Ngi2-7
Chimpanzee fecal sample	This Study	Ngi2-8
Chimpanzee fecal sample	This Study	Nim1-10
Chimpanzee fecal sample	This Study	Nim1-2
Chimpanzee fecal sample	This Study	Nim1-3
Chimpanzee fecal sample	This Study	Nim1-47
Chimpanzee fecal sample	This Study	Nim1-49
Chimpanzee fecal sample	This Study	Nim1-5
Chimpanzee fecal sample	This Study	Nim1-51
Chimpanzee fecal sample	This Study	Nim1-52
Chimpanzee fecal sample	This Study	Nim1-7
Chimpanzee fecal sample	This Study	Nim1-77
Chimpanzee fecal sample	This Study	Nim1-78
Chimpanzee fecal sample	This Study	Nim1-79
Chimpanzee fecal sample	This Study	Nim2-12
Chimpanzee fecal sample	This Study	Nim2-17
Chimpanzee fecal sample	This Study	Nim2-3
Chimpanzee fecal sample	This Study	Nim2-33
Chimpanzee fecal sample	This Study	Nim2-34
Chimpanzee fecal sample	This Study	Nim2-35

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Nim2-44
Chimpanzee fecal sample	This Study	Nim2-58
Chimpanzee fecal sample	This Study	NNP1-11
Chimpanzee fecal sample	This Study	NNP1-15
Chimpanzee fecal sample	This Study	NNP1-26
Chimpanzee fecal sample	This Study	NNP1-34
Chimpanzee fecal sample	This Study	NNP1-40
Chimpanzee fecal sample	This Study	NNP1-44
Chimpanzee fecal sample	This Study	NNP1-54
Chimpanzee fecal sample	This Study	NNP1-57
Chimpanzee fecal sample	This Study	NNP1-77
Chimpanzee fecal sample	This Study	NNP1-86
Chimpanzee fecal sample	This Study	NNP2-2
Chimpanzee fecal sample	This Study	NNP2-35
Chimpanzee fecal sample	This Study	NNP2-4
Chimpanzee fecal sample	This Study	NNP2-54
Chimpanzee fecal sample	This Study	NNP2-55
Chimpanzee fecal sample	This Study	NNP2-67
Chimpanzee fecal sample	This Study	NNP2-68
Chimpanzee fecal sample	This Study	NNP2-74
Chimpanzee fecal sample	This Study	NNP2-79
Chimpanzee fecal sample	This Study	NNP3-14
Chimpanzee fecal sample	This Study	Onp1-11
Chimpanzee fecal sample	This Study	Onp1-12
Chimpanzee fecal sample	This Study	Onp1-2
Chimpanzee fecal sample	This Study	Onp1-20
Chimpanzee fecal sample	This Study	Onp1-21
Chimpanzee fecal sample	This Study	Onp1-24
Chimpanzee fecal sample	This Study	Onp1-25
Chimpanzee fecal sample	This Study	Onp1-26
Chimpanzee fecal sample	This Study	Onp1-27
Chimpanzee fecal sample	This Study	Onp1-28
Chimpanzee fecal sample	This Study	Onp1-29
Chimpanzee fecal sample	This Study	Onp1-31
Chimpanzee fecal sample	This Study	Onp1-32
Chimpanzee fecal sample	This Study	Onp1-34
Chimpanzee fecal sample	This Study	Onp1-35
Chimpanzee fecal sample	This Study	Onp1-39
Chimpanzee fecal sample	This Study	Onp1-6
Chimpanzee fecal sample	This Study	Onp1-7
Chimpanzee fecal sample	This Study	Onp1-8
Chimpanzee fecal sample	This Study	Onp1-9
Chimpanzee fecal sample	This Study	Rt1-1
Chimpanzee fecal sample	This Study	Rt1-2
Chimpanzee fecal sample	This Study	Rt1-5
Chimpanzee fecal sample	This Study	Rt1-7
Chimpanzee fecal sample	This Study	Rt1-8
Chimpanzee fecal sample	This Study	Rt2-1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Rt2-14
Chimpanzee fecal sample	This Study	Rt2-21
Chimpanzee fecal sample	This Study	Rt2-22
Chimpanzee fecal sample	This Study	Rt2-24
Chimpanzee fecal sample	This Study	Rt2-25
Chimpanzee fecal sample	This Study	Rt2-26_2
Chimpanzee fecal sample	This Study	Rt2-31
Chimpanzee fecal sample	This Study	Rt2-35
Chimpanzee fecal sample	This Study	Rt2-37
Chimpanzee fecal sample	This Study	Rt2-38
Chimpanzee fecal sample	This Study	Rt2-41
Chimpanzee fecal sample	This Study	Rt2-6
Chimpanzee fecal sample	This Study	Rt2-7
Chimpanzee fecal sample	This Study	Rt2-8
Chimpanzee fecal sample	This Study	San1-13
Chimpanzee fecal sample	This Study	San1-17
Chimpanzee fecal sample	This Study	San1-19
Chimpanzee fecal sample	This Study	San1-2
Chimpanzee fecal sample	This Study	San1-20
Chimpanzee fecal sample	This Study	San1-22
Chimpanzee fecal sample	This Study	San1-3
Chimpanzee fecal sample	This Study	San1-32
Chimpanzee fecal sample	This Study	San1-39
Chimpanzee fecal sample	This Study	San1-4
Chimpanzee fecal sample	This Study	San2-1
Chimpanzee fecal sample	This Study	San2-10
Chimpanzee fecal sample	This Study	San2-13
Chimpanzee fecal sample	This Study	San2-16
Chimpanzee fecal sample	This Study	San2-20
Chimpanzee fecal sample	This Study	San2-26
Chimpanzee fecal sample	This Study	San2-48
Chimpanzee fecal sample	This Study	San2-49
Chimpanzee fecal sample	This Study	San2-53
Chimpanzee fecal sample	This Study	San2-59
Chimpanzee fecal sample	This Study	Sob1-24
Chimpanzee fecal sample	This Study	Sob1-27
Chimpanzee fecal sample	This Study	Sob1-31
Chimpanzee fecal sample	This Study	Sob1-32
Chimpanzee fecal sample	This Study	Sob1-33
Chimpanzee fecal sample	This Study	Sob1-4
Chimpanzee fecal sample	This Study	Sob1-47
Chimpanzee fecal sample	This Study	Sob1-5
Chimpanzee fecal sample	This Study	Sob1-56
Chimpanzee fecal sample	This Study	Sob1-57
Chimpanzee fecal sample	This Study	Sob1-6
Chimpanzee fecal sample	This Study	Sob1-7
Chimpanzee fecal sample	This Study	Sob1-77

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Sob1-83_2
Chimpanzee fecal sample	This Study	Sob1-84
Chimpanzee fecal sample	This Study	Sob2-12
Chimpanzee fecal sample	This Study	Sob2-3
Chimpanzee fecal sample	This Study	Sob2-37
Chimpanzee fecal sample	This Study	Sob2-43
Chimpanzee fecal sample	This Study	Sob2-5
Chimpanzee fecal sample	This Study	Tai_R1-23
Chimpanzee fecal sample	This Study	Tai_R1-26
Chimpanzee fecal sample	This Study	Tai_R1-28
Chimpanzee fecal sample	This Study	Tai_R1-4
Chimpanzee fecal sample	This Study	Tai_R2-13
Chimpanzee fecal sample	This Study	Tai_R2-15
Chimpanzee fecal sample	This Study	Tai_R2-16
Chimpanzee fecal sample	This Study	Tai_R2-18
Chimpanzee fecal sample	This Study	Tai_R2-22
Chimpanzee fecal sample	This Study	Tai_R2-30
Chimpanzee fecal sample	This Study	Tai_R2-4
Chimpanzee fecal sample	This Study	Tai_R2-43
Chimpanzee fecal sample	This Study	Tai_R2-5
Chimpanzee fecal sample	This Study	Tai_R2-52
Chimpanzee fecal sample	This Study	Tai_R2-57
Chimpanzee fecal sample	This Study	Tai_R2-6
Chimpanzee fecal sample	This Study	Tai_R2-8
Chimpanzee fecal sample	This Study	Tai_R2-80
Chimpanzee fecal sample	This Study	Tai_R2-88
Chimpanzee fecal sample	This Study	Tai_R2-9
Chimpanzee fecal sample	This Study	Tai-E1-13
Chimpanzee fecal sample	This Study	Tai-E1-42
Chimpanzee fecal sample	This Study	Tai-E1-50
Chimpanzee fecal sample	This Study	Tai-E1-52
Chimpanzee fecal sample	This Study	Tai-E1-54
Chimpanzee fecal sample	This Study	Tai-E1-55
Chimpanzee fecal sample	This Study	Tai-E1-56
Chimpanzee fecal sample	This Study	Tai-E1-58
Chimpanzee fecal sample	This Study	Tai-E1-60
Chimpanzee fecal sample	This Study	Tai-E1-7
Chimpanzee fecal sample	This Study	Tai-E1-8
Chimpanzee fecal sample	This Study	Tai-E2-11
Chimpanzee fecal sample	This Study	Tai-E2-18
Chimpanzee fecal sample	This Study	Tai-E2-29
Chimpanzee fecal sample	This Study	Tai-E2-31
Chimpanzee fecal sample	This Study	Tai-E2-35
Chimpanzee fecal sample	This Study	Tai-E2-48
Chimpanzee fecal sample	This Study	Tai-E2-51
Chimpanzee fecal sample	This Study	Tai-E2-8
Chimpanzee fecal sample	This Study	Uga1-1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee fecal sample	This Study	Uga1-11
Chimpanzee fecal sample	This Study	Uga1-12
Chimpanzee fecal sample	This Study	Uga1-17
Chimpanzee fecal sample	This Study	Uga1-22
Chimpanzee fecal sample	This Study	Uga1-34
Chimpanzee fecal sample	This Study	Uga1-9
Chimpanzee fecal sample	This Study	Uga2-1
Chimpanzee fecal sample	This Study	Uga2-22
Chimpanzee fecal sample	This Study	Uga2-28
Chimpanzee fecal sample	This Study	Uga2-31
Chimpanzee fecal sample	This Study	Uga2-41
Chimpanzee fecal sample	This Study	Uga2-44
Chimpanzee fecal sample	This Study	Uga2-46
Chimpanzee fecal sample	This Study	Uga2-49
Chimpanzee fecal sample	This Study	Uga2-53
Chimpanzee fecal sample	This Study	Uga2-73
Chimpanzee fecal sample	This Study	Uga2-74
Chimpanzee fecal sample	This Study	Uga2-81
Chimpanzee fecal sample	This Study	Uga3-29
Chimpanzee hair sample	This Study	Africa_Mona
Chimpanzee hair sample	This Study	Bea_Mona
Chimpanzee hair sample	This Study	Charly_Mona
Chimpanzee blood sample	This Study	Cheeta_Mona
Chimpanzee hair sample	This Study	Cheeta_Rainfer
Chimpanzee hair sample	This Study	Coco_Mona
Chimpanzee hair sample	This Study	Gombe_Rainfer
Chimpanzee blood sample	This Study	Guille_Rainfer
Chimpanzee blood sample	This Study	Iván_Rainfer
Chimpanzee blood sample	This Study	Jackie_Rainfer
Chimpanzee blood sample	This Study	Judi_Rainfer
Chimpanzee hair sample	This Study	Lulú_Rainfer
Chimpanzee hair sample	This Study	Marco_Mona
Chimpanzee blood sample	This Study	Maxi_Rainfer
Chimpanzee hair sample	This Study	Nico_Mona
Chimpanzee hair sample	This Study	Sammy_Rainfer
Chimpanzee hair sample	This Study	Sandy_Rainfer
Chimpanzee hair sample	This Study	Tico_Mona
Chimpanzee hair sample	This Study	Toni_Mona
Chimpanzee blood sample	This Study	Toti_Rainfer
Critical commercial assays		
QIAamp Fast DNA Stool Mini Kit	Qiagen	cat#51604
High Sensitivity Genomic DNA 50Kb Analysis kit	Advanced Analytical	cat#DNF-488
Bioanalyzer Agilent DNA 7500 kit	Agilent	cat#5067-1506
Bioanalyzer High Sensitivity DNA Analysis	Agilent	cat#5067-4626
Deposited data		
Raw sequencing reads	This study	ENA: PRJEB46115
Chimpanzee genomes	de Manuel et al., 2016[6]	ENA: PRJEB15086

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chimpanzee genomes	Prado-Martinez et al., 2013[3]	SRA: PRJNA18943 and SRP018689
Homo sapiens reference genome (Hg19 or GRCh37)	Church et al., 2011[68]	https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/
Pan troglodytes reference genome (panTro6)	Kronenberg et al., 2018[69]	https://www.ncbi.nlm.nih.gov/assembly/GCF_002880755.1/
Papio anubis reference genome (Panu_3.0)	Roger et al., 2019[70]	https://www.ncbi.nlm.nih.gov/assembly/GCF_000264685.3/
green monkey (Chlorocebus_sabeus_1.1)	Warren et al., 2015[71]	https://www.ncbi.nlm.nih.gov/assembly/GCF_000409795.2/
Colobus angolensis palliatus reference genome (Cang.pa_1.0)	Genereux et al., 2020[72]	https://www.ncbi.nlm.nih.gov/assembly/GCF_000951035.1/
Cercocebus atys reference genome (Caty_1.0)	Genereux et al., 2020[72]	https://www.ncbi.nlm.nih.gov/assembly/GCF_000955945.1/
Gorilla gorilla gorilla reference genome (gorGor4)	Scally et al., 2012[73]	https://www.ncbi.nlm.nih.gov/assembly/GCF_000151905.2/
Mandrillus leucophaeus reference genome (Mleu.le_1.0)	Genereux et al., 2020[72]	https://www.ncbi.nlm.nih.gov/assembly/GCF_000951045.1
Erythrocebus patas reference genome (EryPat_v1_BIUU)	Genereux et al., 2020[72]	https://www.ncbi.nlm.nih.gov/assembly/GCA_004027335.1/
Cercopithecus neglectus reference genome (CertNeg_v1_BIUU)	Genereux et al., 2020[72]	https://www.ncbi.nlm.nih.gov/assembly/GCA_004027615.1
Mandrillus sphinx (BGI_mandrill_1.0)	Yin et al., 2020[74]	https://www.ncbi.nlm.nih.gov/assembly/GCA_004802615.1/
Oligonucleotides		
Univ_Block_P7: 5′-AGATCGGAAGAGCACACG TCTGAACTCCAGTCAC-Pho-3′	Rohland and Reich, 2012;[75] Sigma-Aldrich	N/A
Univ_Block_P5: 5′-AGATCGGAAGAGCGTCGT GTAGGGAAAG-Pho-3′	Rohland and Reich, 2012;[75] Sigma-Aldrich	N/A
P5_Indexing_Primer: 5′-AATGATACGGCGACCA CCGAGATCTACACNNNNNNNACACTCTTTCCC TACACGACGCTCTT-3′	Meyer and Kircher, 2010;[76] Sigma-Aldrich	N/A
P7_Indexing_Primer: 3′-TGTGCAGACTTGAGGT CAGTGNNNNNNNTAGAGCATACGGCAGAAGA CGAAC-5′	Meyer and Kircher, 2010;[76] Sigma-Aldrich	N/A
PreHyb_P5_F 5′-CTTTCCCTACACGACGCTCTTC-3′	Meyer and Kircher, 2010;[76] Sigma-Aldrich	N/A
PreHyb_P7_R 3′-GTGTGCAGACTTGAGGTCAGTG-5’	Meyer and Kircher, 2010;[76] Sigma-Aldrich	N/A
F_P5_7nt_XX Indexed Adapter: 5′-CTTTCCCTACAC GACGCTCTTCCGATCTNNNNNNN-3′	Rohland and Reich, 2012;[75] Teknokroma	N/A
F_P7_7nt_XX Indexed Adapter: 5′-GTGACTGGAGT TCAGACGTGTGCTCTTCCGATCTNNNNNNN-3′	Rohland and Reich, 2012;[75] Teknokroma	N/A
R_P5/P7_7nt_XX Indexed Common Adapter: 5′-NN NNNNNAGATCGGAA-3′	Rohland and Reich, 2012;[75] Teknokroma	N/A
Ns represent indexes	N/A	N/A
Software and algorithms		
Admixfrog	Peter, 2021[40]	https://github.com/BenjaminPeter/admixfrog
AdmixTools	Patterson et al., 2012[77]	https://github.com/DReichLab/AdmixTools
ANGSD v0.916	Meisner and Albrechtsen, 2018[36]	http://www.popgen.dk/angsd/index.php/ANGSD
BBsplit	N/A	https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BEDtools v2.22.1	Quinlan and Hall, 2010[78]	https://bedtools.readthedocs.io/en/latest/
BWA-mem v0.7.12	Li and Durbin, 2009[79]	https://bio-bwa.sourceforge.net/
EEMS	Petkova et al., 2016[42]	https://github.com/dipetkov/eems
FastMe v2.1.5	Lefort et al., 2015[80]	https://www.atgc-montpellier.fr/fastme/
GATK v3.7	McKenna et al., 2010[81]	https://gatk.broadinstitute.org/hc/en-us
HuConTest	Kuhlwilm et al., 2021[33]	https://github.com/kuhlwilm/HuConTest
IBDseq	Browning and Browning, 2013[45]	https://faculty.washington.edu/browning/ibdseq.html
Mapdamage v2.0	Jónsson et al., 2013[82]	https://ginolhac.github.io/mapDamage/
NGSAdmix	Skotte et al., 2013[83]	https://www.popgen.dk/software/index.php/NgsAdmix
ngsDist v1.0.2	Vieira et al.,2016[84]	https://github.com/fgvieira/ngsDist
NgsRelate	Korneliussen and Moltke, 2013[85]	https://github.com/ANGSD/NgsRelate
PCAngsd V0.8	Meisner and Albrechtsen, 2018[36]	https://www.popgen.dk/software/index.php/PCAngsd
Picard v1.95	N/A	https://broadinstitute.github.io/picard
PLINK v1.9	Purcell et al., 2007[86]	https://www.cog-genomics.org/plink/
R pacakge maptools	Bivand and Lewin-Koh, 2013[87]	https://cran.r-project.org/web/packages/maptools/index.html
R package admixr	Petr et al., 2019[88]	https://cran.r-project.org/web/packages/admixr/index.html
R package ape v5.4-1	Paradis and Schliep, 2019[89]	https://cran.r-project.org/package=ape
R package Phangorn	Schliep, 2011[90]	https://cran.r-project.org/web/packages/phangorn/index.html
R package phytools	Revell, 2012[91]	https://cran.r-project.org/web/packages/phytools/index.html
R package rEEMSplots	Petkova et al., 2016[42]	https://github.com/dipetkov/eems
R package sf	Pebesma, 2018[92]	https://cran.r-project.org/web/packages/sf/index.html
R package sp	Pebesma and Bivand, 2005[93]	https://cran.r-project.org/web/packages/sp/index.html
R package Vegan	Oksanen et al., 2020[94]	https://CRAN.R-project.org/package=vegan
R v3.6.3	R Core Team[95]	https://www.R-project.org/
rareCAGA	This study	https://github.com/kuhlwilm/rareCAGA https://doi.org/10.5281/zenodo.6199201
RAxML-NG v0.9.0	Kozlov et al., 2019[96]	https://github.com/amkozlov/raxml-ng
realSFS v0.916	Nielsen et al., 2012[97]	http://www.popgen.dk/angsd/index.php/RealSFS
Sabre	N/A	https://github.com/najoshi/sabre
Samtools v1.5	Li et al., 2009[98]	https://www.htslib.org/
snpAD v0.3.2	Prüfer, 2018[99]	https://bioinf.eva.mpg.de/snpAD/
TreeMix v1.12	Pickrell and Pritchard, 2012[100]	https://bitbucket.org/nygcresearch/treemix/wiki/Home
Trimmomatic v0.36	Bolger et al., 2014[101]	http://www.usadellab.org/cms/?page=trimmomatic
vcftools v0.1.12b	Danecek, 2011[102]	http://vcftools.sourceforge.net/
Other		
SureSelect Custom Array (chr21)	Agilent	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Tomas Marques-Bonet (tomas.marques@upf.edu) or Mimi Arandjelovic (arandjel@eva.mpg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All genomic data generated is available at a public repository (ENA) under the accession code ENA: PRJEB46115. Code for geolocalization is available on a public repository (https://github.com/kuhlwilm/rareCAGA). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All fecal and shed hair samples from wild chimpanzees included in this study were collected in a non-invasive manner, following standard practices and with no animal contact and no direct observation of the animals under study. Fecal samples from a zoo chimpanzee in Ghana were also obtained non-invasively. Full research approval, sample collection approval and research and sample permits of national ministries and protected area authorities were obtained in all countries of study. Sample export was also done with all necessary certificates, export and import permits. Fecal samples are exempt from the Convention on the Trade in Endangered Species of Wild Fauna and Flora (CITES), CITES permits were obtained for all hair samples. For all PanAf samples and research sites, research permits, veterinary certificates, certificates of origin, national export and import (German) permits and CITES import and export permits (when needed by export countries) were obtained by the PanAf through the Max Planck Institute for Evolutionary Anthropology, Department of Primatology, reviewed by the designated department officer and approved by national export and import officials. All documents are permanently stored with the PanAf and copies are electronically archived with the Max Planck Society. Hair and blood samples from chimpanzees in the Spanish rescue centers (Fundació Mona and Rainfer) were collected during a routine veterinary check of the animals. All laboratory work conforms to the relevant regulatory standards of the Max Planck Society, Germany and University Pompeu Fabra, Spain. No experiments were undertaken with live animals.

METHOD DETAILS

Sample selection and sequencing

Fecal DNA was extracted from a total of 5,397 PanAf samples and screened with a microsatellite genotyping assay,[11] leaving only non-related samples, from which a minimum of 20 samples per location were selected for further sequencing whenever possible (Figure S1). Samples were randomized in batches of 24–48 samples and processed on different days for library preparation. A unique double-inline barcoded library was prepared for each sample following the BEST protocol with minor modifications.[25,24] Pooling for capture was devised based on the host DNA content (fraction of chimpanzee DNA, relative to gut microbial and exogenous DNA) (Supplemental Note 1, Figures S3, S4, Tables S1–S3).[28,25] Each pool was divided into several aliquots to perform multiple hybridizations (Figure S5). Afterwards, with predesigned RNA baits (SureSelect Agilent), we captured the non-repetitive regions of chromosome 21 following the protocol provided by the Agilent Sureselect Custom Array, adding two consecutive hybridization rounds for pools containing samples with <5% host DNA. Captured libraries were sequenced on the HiSeq 4000 Illumina platform with 2 × 100 paired-end reads.

Data processing and filtering

We processed the data to demultiplex libraries belonging to the same hybridization pool using Sabre (https://github.com/najoshi/sabre) and reads were trimmed with Trimmomatic (version 0.36).[101] Paired-end reads were then aligned to the human genome Hg19 (GRCh37, Feb.2009 (GCA_000001405.1))[68] using BWA (version 0.7.12).[79] Duplicates were removed using PicardTools (version 1.95) (http://broadinstitute.github.io/picard/) and further filtering of the reads was done using samtools (version 1.5).[98] To retrieve the on-target reads we used intersectBed from the BEDTOOLS package (version 2.22.1).[78] Average coverage of the target space was calculated as the number of bases in the target region divided by the size of the target space (Figures S6 and S7). We obtained genotype likelihoods using ANGSD[35] version 0.916 and genotype calls using snpAD[99] v0.3.2, a software that takes DNA damage and biases into account for genotype calling (Figure S2). Analyses of error damage patterns and genotype discovery can be found in Supplemental Note 3 (Figures S20–S28). Principal Component Analysis (PCA) was performed using PCAngsd[36] (Supplemental Note 3, Figures S9–S13 and S30–S34). Sources of primate contamination in fecal samples were determined using BBsplit (https://sourceforge.net/projects/bbmap/), mapping to 11 different primate genomes (Supplemental Note 3, Figures S14 and S15). Human contamination was estimated as the fraction of the number of observations of human-like alleles across all positions where chimpanzees and humans consistently differ, using the available script HuConTest[33] which has been designed and tested for this purpose, and shown to work on fecal samples at very low coverage (Supplemental Note 3, Figures S16 and S17). Although samples had been

screened prior to library preparation with a microsatellite assay,[103,104] we used NgsRelate[105] (Supplemental Note 3) to identify and remove identical or putative first order relative individuals (Figures S18 and S19). Due to the high variation of sample qualities and specific requirements for the application of different methods, a variety of filtering procedures was applied. In most analyses, samples with evidence of contamination from either human (>1% or >0.5%) or other primate species were removed, as well as first degree relatives and identical samples (n = 89), as well as samples that were found to be most likely mislabeled (n = 2) (Supplemental Note 5, Table S1, Figure S30). Finally, we used samples with different coverage cutoffs for different analyses (0.5-fold, 1-fold or 5-fold), as indicated for each method. The minimum coverage cutoff for the initial PCA was decided at 0.5-fold, so at least half of the chr21 would be covered on average by at least 1 read, which would provide sufficient expected overlap of variants between individuals.

QUANTIFICATION AND STATISTICAL ANALYSIS

Population genetics

To obtain pairwise F_{ST} estimates between sampling sites, we computed the 2-dimensional SFS (2d SFS) between each pair of geographical sites with ANGSD -doSaf 1 and realSFS.[35] The genetic relationships between populations were used to build a matrix, from which we constructed a neighbor-joining tree using the ape package[89] in R (version 3.5.2) (Figures S45–S48). F3 outgroup statistics were calculated between sampling sites using qp3Pop[77] and taking an orangutan (*Pongo pygmaeus*) as the outgroup (pygmaeus_ERS1986511);[106] this also ensures that low remaining amounts of human contamination would not influence the analysis on genotypes called on the human genome (Figures S54–S56). Regions of Homozygosity (RoH) were defined as heterozygous positions with a distance larger than 100 kbp, irrespective of missing information in between, for individuals with more than 5-fold average coverage in the target space (Figures S40–S42). We defined short RoHs as those between 10 and 100 kbp, and long RoHs as those longer than 100 kbp, following a previous approach.[51] Long-term effective migration rates were calculated using EEMS[42] with samples of more than 5-fold coverage (Supplemental Note 10, Figures S82–S85). The same dataset was used to obtain IBD-like tracts using the IBDseq software, which does not require phasing of the data.[45] To increase the power to detect IBD-like fragments in such sparse dataset, we restricted the analysis on samples with >5-fold coverage, we included the genotype data on the chromosome 21 of 59 previously published whole-genome chimpanzees,[6] and we kept only genotypes with a depth of at least eight reads (Supplemental Note 10, Figures S86–S88). We observed an exponential decay of IBD-like tract lengths with geographical distance within eastern and western chimpanzees (Figures S89 and S90), as expected for isolation-by-distance. The number of shared IBD-like tracts is likely the consequence of recent migration events or the shared population history between geographic sites or areas (Supplemental Note 10, Figures S91 and S93). The length of the shared segments is correlated with the time of such genetic exchange, with more recent migration resulting in longer IBD-like tracts. Therefore, a way to estimate the age of an IBD-like segment is by using its length. When the time (in generations g) to the most recent common ancestor (MRCA) is known, the total length (in cM) of a shared IBD segment follows an exponential distribution with rate 100/2g. Therefore, to time the events, we followed this rate of g = 100/(2*cM),[49,107] with cM being the length of the fragments, and g the number of generations. The length in cM was estimated from the length in Mbps by applying the western chimpanzee recombination map[108] to the same subspecies and assuming an effective population size of Ne = 17,378.[6] For the rest of the subspecies we used the Nigeria-Cameroon recombination map,[109] with the following effective population sizes for each subspecies: central Ne = 47,314, eastern Ne = 32,492 and Nigeria-Cameroon Ne = 27,795.[6] For timing the events between subspecies, we assumed a constant recombination map of 1cM/1Mbp since the recombination maps differ substantially between subspecies. We assumed a generation time of 25 years to calculate the time.[110] We took the maximum IBD-like tract length per pair of individuals between sites to estimate the time frame of connectivity per site, and calculated average, maximum and minimum for each subspecies. Our reported expected time to the MRCA, derived from the length of the IBD-like fragments, can encompass large confidence intervals as other factors (technical and biological) could modify it. The log connectivity ratio of each sampling site was calculated as the sum of IBD-like tract counts (normalized by the number of pairwise sample comparisons between sites) that each site shares with the other sites, over the median global average of normalized IBD-like tract counts between all sampling sites (Figures S92, S94, Table S10). Subspecies ancestry introgressed fragments and bonobo introgression were determined with admixfrog.[40] Admixfrog is a newly developed method to reliably infer ancestry fragments even from low-coverage and contaminated data. It uses a Hidden Markov Model (HMM) to infer local ancestry in a target individual from different sources which represent the admixing populations. Here, we use as potential sources of admixture 10 bonobo genomes and 16 chimpanzees (4 of each subspecies) from previous publications. The reference panel on chromosome 21 (*source*) was built using an equal number of individual genomes of each chimpanzee subspecies (16 genomes), 10 bonobo genomes,[6] two human genomes[111] (to serve as potential source of contamination and remove its effect) and 1 orangutan[106] (as ancestral state). We recovered the global simulated runs of ancestry (from .res2 file) (Figures S61 and S63–S65) and the called runs of ancestry (from.rle file) (Figures S62 and S66), following the instructions of the method (link to https://github.com/BenjaminPeter/admixfrog). A Wilcoxon ranked test was performed in R, correcting for multiple testing with p.adjust (method = ”BH”) in R (Supplemental Note 8).

Rare alleles

Rare variation was used to assess connectivity between geographic regions in the recent past (1.5kya-15kya), and to estimate the most probable origin of chimpanzee fecal samples (Supplemental Note 9). For each sampling location (38 locations) with at least one

individual at more than 1-fold coverage and less than 0.5% human contamination, we determined positions that were derived at the location itself and observed with a cumulative frequency of less than 1 across all other sampling locations (434 individuals, on average 11 per location; 963,656 SNPs, on average 26,671 per location) (Figures S67–S72, Table S6). Observation of a quality genotype (Supplemental Note 3) in either allele state (ancestral or derived) was required for at least 2 sampling locations, while missing data was ignored. The proportion of shared near-private sites of all observed near-private sites for each reference population was calculated, with heterozygous and homozygous derived positions equally treated as derived. Spatial modeling and kriging to the chimpanzee range were performed using the R package gstat[112] to create a surface of rare allele sharing (R version 3.5.0). Accuracy was assessed by leaving one location out, calculating rare alleles for the remaining 37 locations, and applying the test to the individuals from the 38th location, analogous to the "leave-location-out" cross validation in Wasser et al., 2015.[21] For comparison with previous work, we calculated the 75th percentiles of distances to the true origin (Figure S73). We applied this method to all remaining samples from this study (low coverage, substantial human contamination, PCA outliers), as well as chromosome 21 from great ape whole genomes[6][3,106,113] and shallow sequencing data (median 0.25-fold coverage, ranging from 0.15-fold to 4.3-fold, Table S8) of blood and hair samples from 20 rescued chimpanzees from two Spanish rescue centers (Supplemental Note 9, Figures S74, S75, S80, S81, Data S1 – Figures S95–S105). Since the rare variants used in our approach are not necessarily fixed at a given location, but can be present at other locations, the pattern of shared rare variants is informative on past connectivity (Figures S76–S79, Data S1 – Figure S106). We calculated the proportion of derived variants in a given population shared with all other populations. We then used these data points to infer a landscape of sharing with other populations, and applied the kriging procedure described above, where we left out the test population from the landscape.

Cell Genomics, Volume *2*

Supplemental information

Population dynamics and genetic connectivity in recent chimpanzee history

Claudia Fontsere, Martin Kuhlwilm, Carlos Morcillo-Suarez, Marina Alvarez-Estape, Jack D. Lester, Paolo Gratton, Joshua M. Schmidt, Paula Dieguez, Thierry Aebischer, Paula Álvarez-Varona, Anthony Agbor, Samuel Angedakin, Alfred K. Assumang, Emmanuel A. Ayimisin, Emma Bailey, Donatienne Barubiyo, Mattia Bessone, Andrea Carretero-Alonso, Rebecca Chancellor, Heather Cohen, Emmanuel Danquah, Tobias Deschner, Andrew Dunn, Jef Dupain, Villard E. Egbe, Olga Feliu, Annemarie Goedmakers, Anne-Céline Granjon, Josephine Head, Daniela Hedwig, Veerle Hermans, R. Adriana Hernandez-Aguilar, Inaoyom Imong, Sorrel Jones, Jessica Junker, Parag Kadam, Mike Kaiser, Mbangi Kambere, Magloire V. Kambale, Ammie K. Kalan, Ivonne Kienast, Deo Kujirakwinja, Kevin Langergraber, Juan Lapuente, Bradley Larson, Anne Laudisoit, Kevin Lee, Manuel Llana, Miquel Llorente, Sergio Marrocoli, David Morgan, Felix Mulindahabi, Mizuki Murai, Emily Neil, Sonia Nicholl, Stuart Nixon, Emma Normand, Chris Orbell, Lucy J. Ormsby, Liliana Pacheco, Alex Piel, Laura Riera, Martha M. Robbins, Aaron Rundus, Crickette Sanz, Lilah Sciaky, Volker Sommer, Fiona A. Stewart, Nikki Tagg, Luc Roscelin Tédonzong, Els Ton, Joost van Schijndel, Virginie Vergnes, Erin G. Wessling, Jacob Willie, Roman M. Wittig, Yisa G. Yuh, Kyle Yurkiw, Klaus Zuberbuehler, Jochen Hecht, Linda Vigilant, Christophe Boesch, Aida M. Andrés, David A. Hughes, Hjalmar S. Kühl, Esther Lizano, Mimi Arandjelovic, and Tomas Marques-Bonet

Document S1

Table of Contents

Note 1. Sample collection and data generation

1.1. Sample collection

A total of 828 non-invasive samples (819 fecal and 9 hair samples) from 53 different sites in Africa (Fig. S1) were collected as part of the Pan African Programme (PanAf): The Cultured Chimpanzee project (http://panafrican.eva.mpg.de) (Table S1). Tai_R and Tai_Eco, all Comoé sites and Bakoun and Sobory are going to be considered the same sampling site since they are separated by less than 15 km. Therefore, a total of 48 sampling sites are considered for the analysis.

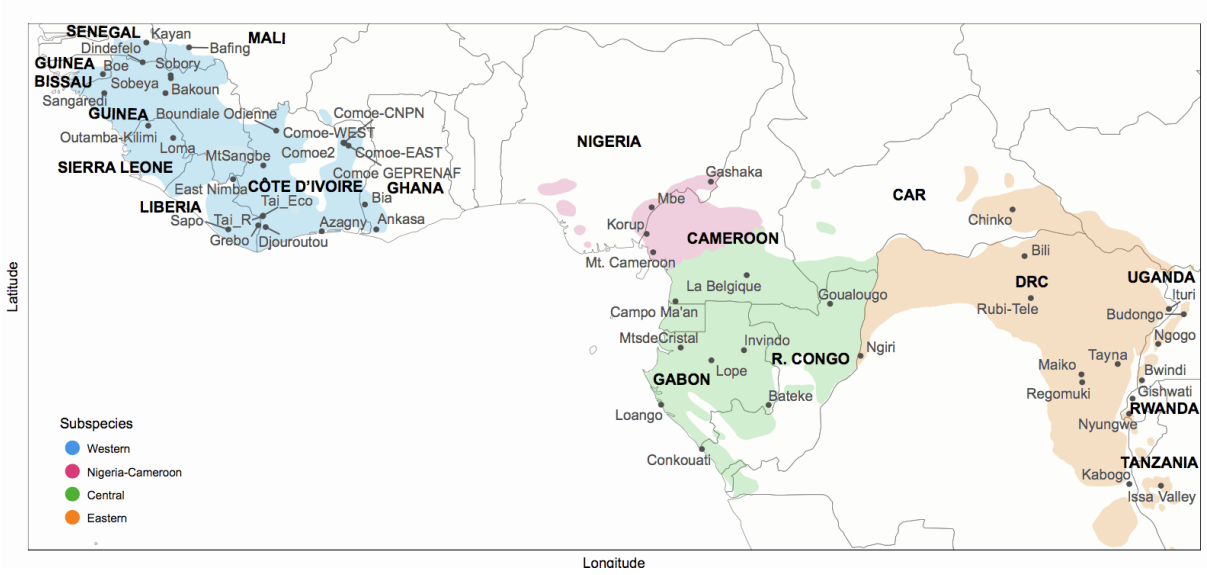


Fig. S1. Map of geographical sites where fecal and hair samples were collected. Western chimpanzee (*Pan troglodytes verus*) distribution is marked in blue, Nigeria-Cameroon chimpanzee (*Pan troglodytes ellioti*) distribution in red, central chimpanzee (*Pan troglodytes troglodytes*) distribution in green and eastern chimpanzee (*Pan troglodytes schweinfurthii*) distribution in orange. Related to Fig. 1.

Approximately 5g ("hazelnut-size") of feces were collected from each chimpanzee fecal sample and stored in the field using the two-step ethanol-silica preservation method[1]. Depending on the density of the sample, between 10 and 80 mg of dry fecal sample were extracted for DNA using a QIAamp Fast DNA Stool Mini Kit (Qiagen). Hair extractions were done using the Buffer Digest Method for hair digestion for amplification without purification. Next, extractions were screened using a microsatellite genotyping assay[2,3] to keep up to 20 samples from PanAf field site. Samples were selected as follows: 1) those that amplified at the most loci of the 15 tested; 2) represented unique individuals and 3) were ascertained to have low probability of being first degree relatives[4]. Total DNA concentration and fragmentation were measured on a Fragment Analyzer using a Genomic DNA 50Kb Analysis kit (Advanced Analytical) and the fragmentation level was calculated with PROSize software. For 269 fecal samples endogenous or host DNA (hDNA) content (fraction of chimpanzee DNA, relative to gut microbial and exogenous DNA) was estimated by qPCR[5] (Table S1, S2). Finally, the percentage of

endogenous content for each sample was calculated by dividing the endogenous chimpanzee DNA concentration by the total DNA concentration. For the rest of samples, we estimated the hDNA after library preparation by shallow shotgun sequencing (Table S1, S3). Previous studies have shown that both estimates, shotgun sequencing and qPCR, positively correlate although shotgun sequencing is considered to be more reliable[6,7].

1.2. Target design

The chromosome 21 RNA baits were designed by SureSelect Agilent using the PanTro4 assembly (Pan_troglodytes-2.1.4), masking repetitive regions (RepeatMasker -s setting) into two Tier 5 custom arrays with 3x tiling density. Total design size is 22.358 Mbp.

1.3. Library preparation and Capture protocol

A unique double-inline barcoded library was prepared for each sample following the BEST protocol[8] with minor modifications presented in Fontsere *et al.* (2021)[4]. Libraries of samples for which hDNA had not been quantified by qPCR (526 out of 828), were pooled in pools of 30 each and adapters were extended to full length with 6 PCR cycles using indexed primers. Pools were sequenced in one HiSeq4000 lane, and we obtained the percentage of hDNA by dividing the number of Reliable reads (mapped reads with mapping quality > 30 and without duplicates and secondary alignment) with insert size > 35 by the total production reads. Next, libraries were pooled approximately equi-endogenously in pools of ~30 samples, with a maximum ratio of 1:2 between the highest % hDNA and the lowest % hDNA among samples within each group[4,6]. If this ratio of 1:2 would have been violated, the pool size was reduced to 20 samples. Also, samples with hDNA values below 2% were captured in pools of 20 samples. Next, each pool was divided into several aliquots to perform additional hybridizations. In case libraries within a pool were used up due to a low DNA concentration, but additional hybridizations were possible for the other samples, we performed a re-pooling of the other samples excluding those. Depending on sufficient amounts of DNA, the starting material for each hybridization was 2 µg, if this was not possible, 1 µg of starting DNA was used. We followed the capture protocol provided by Agilent Sureselect Custom Array, adding two consecutive hybridization rounds with a PCR amplification step in the middle (PCR cycles between 10-13) for the pools containing samples with <5% hDNA[7]. After the second hybridization, a final PCR was performed, using indexed primers[9] to double-index each pool of libraries with a unique pair of indices (PCR cycles between 10-13).

Note 2. Sequencing and data processing.

2.1. Sequencing and mapping

Captured libraries were sequenced on the HiSeq 4000 Illumina platform with 2x100 paired-end reads. Libraries belonging to the same hybridization pool were demultiplexed and the 7bp internal barcodes removed using Sabre (https://github.com/najoshi/sabre). Trimmomatic (version 0.36)[10] was used to trim the Illumina adapters in the FASTQ files (ILLUMINACLIP:2:30:10), and bases with an average quality score of less than 20 (SLIDINGWINDOW:5:20). Paired-end reads were aligned to the human genome Hg19 (GRCh37, Feb.2009 (GCA_000001405.1)) using BWA mem (version 0.7.12)[11]. Duplicates were removed using PicardTools (version 1.95) (http://broadinstitute.github.io/picard/) with MarkDuplicates option. Further filtering of the reads was done to discard secondary alignments and reads with mapping quality lower than 30 using samtools (version 1.5)[12]. Henceforth, we will refer to those reads remaining after filtering as "reliable reads". To retrieve the reliable on-target reads we used intersectBed from the BEDTOOLS package (version 2.22.1)[13] using the designed target regions provided by Agilent. Since the same libraries were captured more than once and sequenced in different lanes, we merged filtered bam files from different hybridizations using the MergeSamFiles option from PicardTools (version 1.95) (http://broadinstitute.github.io/picard/). The merged bam files can still contain duplicates generated during library preparation; thus, we removed duplicates and then retrieved the reliable reads on-target using the same methodology as above. Average coverage of the target space was calculated as the number of bases in the target region divided by the size of the target space (Table S1).

2.2. Genotype calling

We obtained genotype calls using two methodologies: Genotype likelihoods with ANGSD and called fixed genotypes with GATK and snpAD.

2.2.1. ANGSD

We used the program ANGSD version 0.916[14] to obtain genotype likelihoods using GATK model with the following parameters for all the samples: -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minInd 15 -skipTriallelic 1 -GL 2 -minMapQ 30 -doGlf 2 -doMajorMinor 1 -doMaf 2 -minMaf 0.004 -SNP_pval 1e-6 -r chr21: .

2.2.2. GATK

GATK v3.7[15] haplotype caller was used with -ERC GVCF option to call all sites for all samples, a gold-standard method for whole genome data. Next, we used CombineGVCFs to obtain a merged VCF.

2.2.3. snpAD and assessment of DNA damage patterns

We speculated that the DNA obtained from fecal samples might be subject to DNA degradation. DNA quality has been of concern in studying fecal samples genetically for a long time[16], and analyzed regarding the length distribution of DNA fragments, for example, of predator and prey in sea lions[17]. This is particularly relevant for studying STRs and other markers, while the kind of data generated here may be more sensitive to damage patterns in the sequences. However, the actual patterns of DNA damage have, to our knowledge, not been studied as thoroughly as those in ancient DNA[18], for example the deamination of cytosine residues at 5' ends.

We applied the software mapDamage[19], commonly used in ancient DNA studies, to estimate the substitution rates along the sequencing fragments. We find a small increase of T-to-C substitutions at 5' ends, as well as an increase of A-to-G substitutions (Fig. S2). The C-to-T and G-to-A substitutions are generally observed to an elevated degree, with an average increase towards the 5' end, but not the final base. However, there is large variation, with some samples showing an excess of this substitution type towards the 5' end. We note that the error rates observed here are an order of magnitude smaller than those observed in ancient DNA, suggesting an effect of DNA decay to be small (less than 2%). Furthermore, the sequencing libraries have been sheared, causing a great amount of random sequence starting points within the endogenous DNA fragments and reducing the relative abundance of endogenous 5' ends. We also caution that the effect of capture on these patterns is unclear.

We observe that the variability across samples is large, most likely due to a large variation in time until collection, environmental conditions and possibly diet of the sampled individuals. We conclude that it is appropriate to take DNA damage into account for genotype calling, and use the software snpAD v0.3.2[20]. This program has been developed for genotype calling in ancient DNA but is applicable to other types of data that may be subject to systematic biases, specifically since it infers priors for each sample separately. The mapped sequences were transformed from bam-format into snpAD-format files, priors for base composition estimated, and genotypes were called using standard settings.

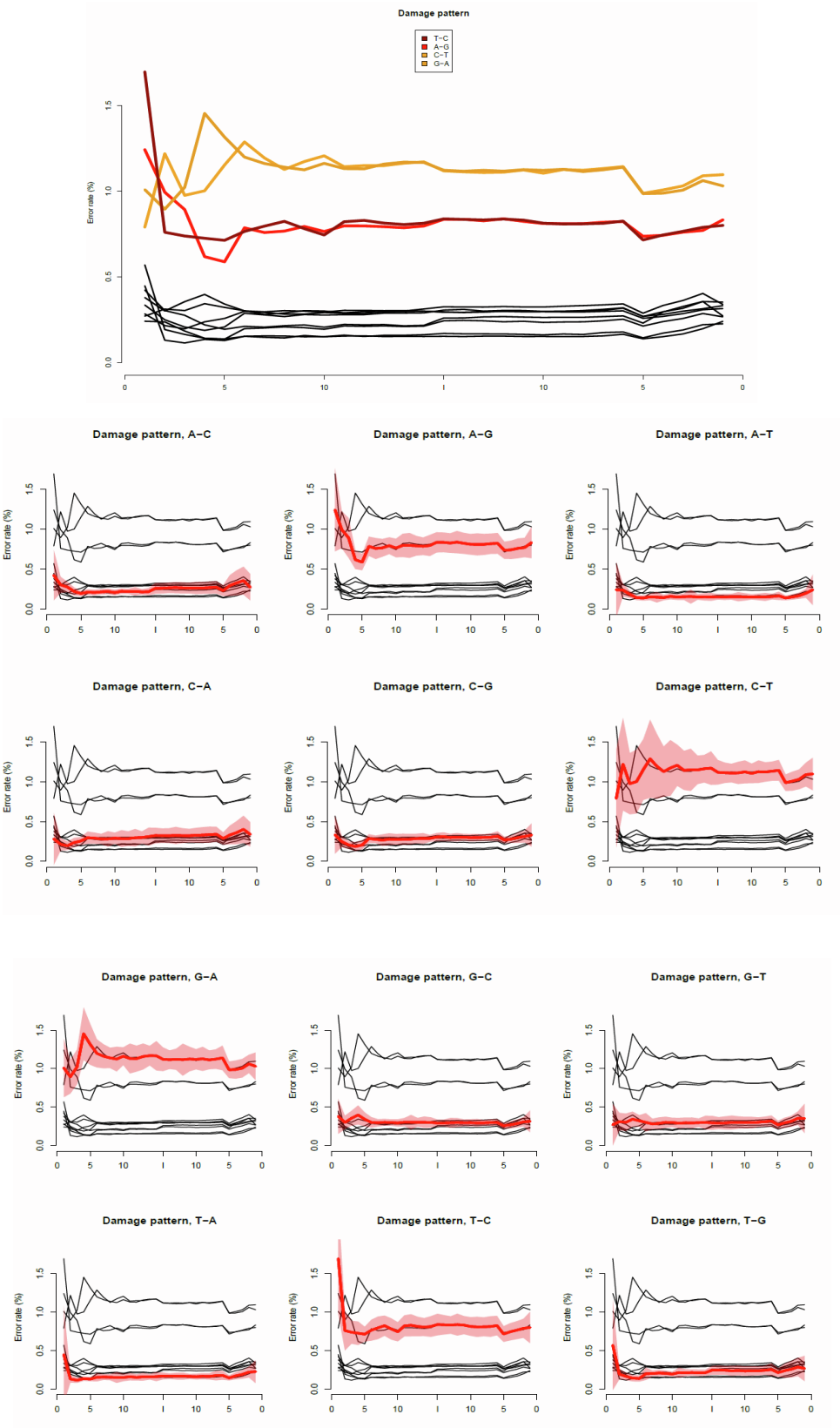


Fig. S2. Aggregated DNA damage patterns across 551 samples with a coverage of more than 0.5-fold (Table S1), and 95% confidence intervals for each substitution type separately. X-axis represents the position in the read. Related to STAR methods.

Note 3. Quality assessment.

3.1 hDNA distribution

To quantify the proportion of endogenous or host DNA (hDNA) in each sample we used two different methods: qPCR and shallow shotgun sequencing, as explained in Note 1.3. We did not pool samples together from which the hDNA estimates were derived from different methods. The hDNA estimates range from 0% to 47.57% with a median of 0.412% (Table S1, Fig. 3).

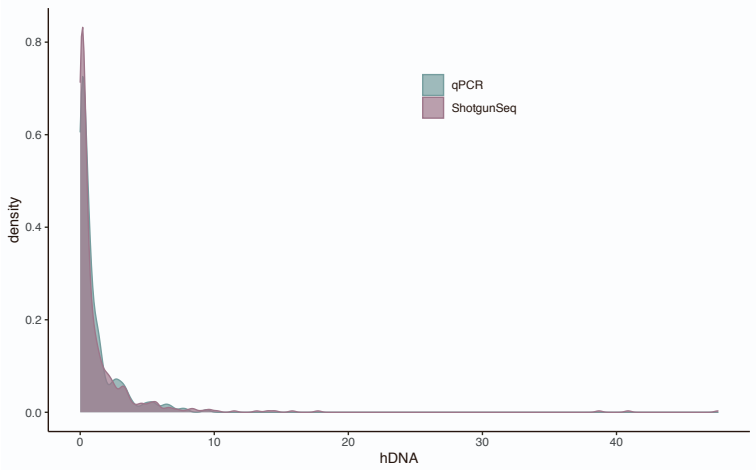


Fig. S3. Host or endogenous DNA (hDNA) estimates distribution obtained with two different methods: qPCR and shotgun sequencing. Related to Fig. 1.

We observed some geographical sites with higher than average and others with lower than average estimates of hDNA (Fig. S4) as previously described[4]. Differences in weather, humidity and temperature conditions on DNA preservation and bacterial growth in the fecal sample before collection as well as a product of sample age and quality of sampling conditions could explain this observation.

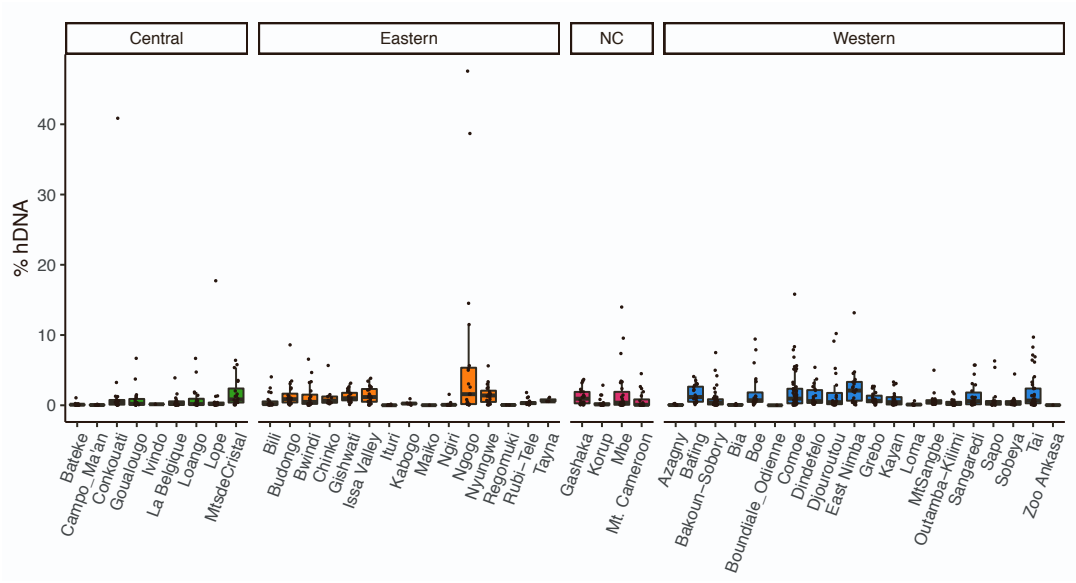
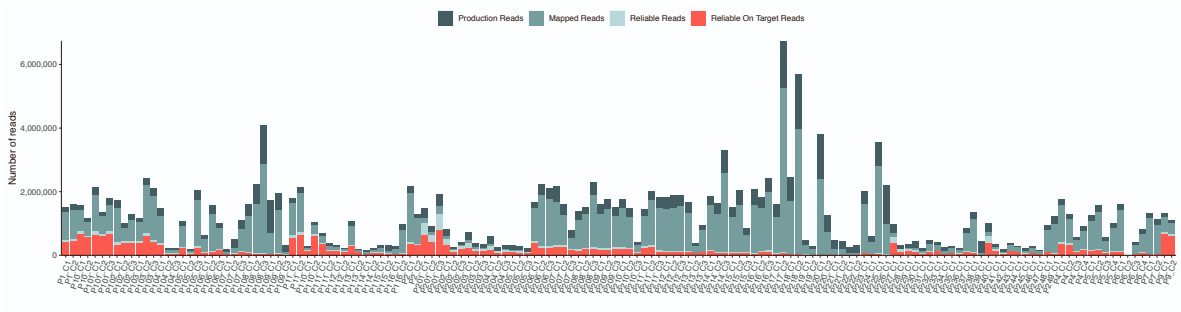


Fig. S4. Percent host or endogenous DNA (hDNA) distribution of all fecal samples from 48 PanAf sampling locations, without filtering contaminated samples. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and the whisker extends from the hinge to the largest and lowest value no further than 1.5 * IQR (inter-quartile range). Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 1.

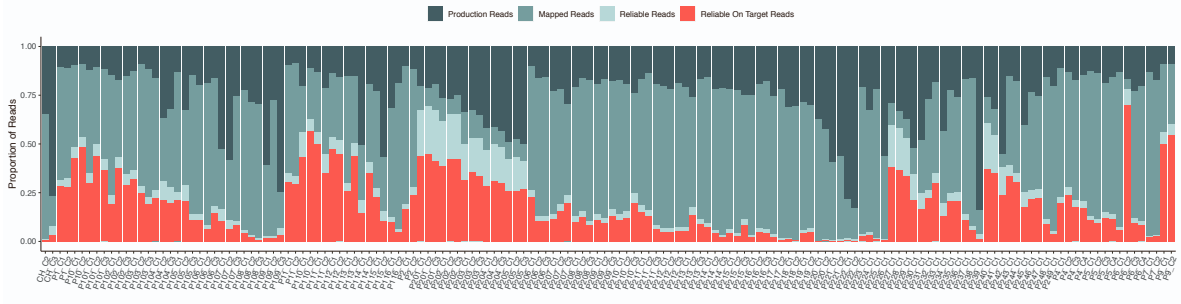
3.2. Capture performance

We quantified the success of capture hybridization by calculating the total number and average percentage of on-target reliable reads, reliable reads and mapped reads (with duplicates), compared to the total production reads at each individual capture hybridization (Fig. S5A, B) and geographical site (Fig. S5C, D).

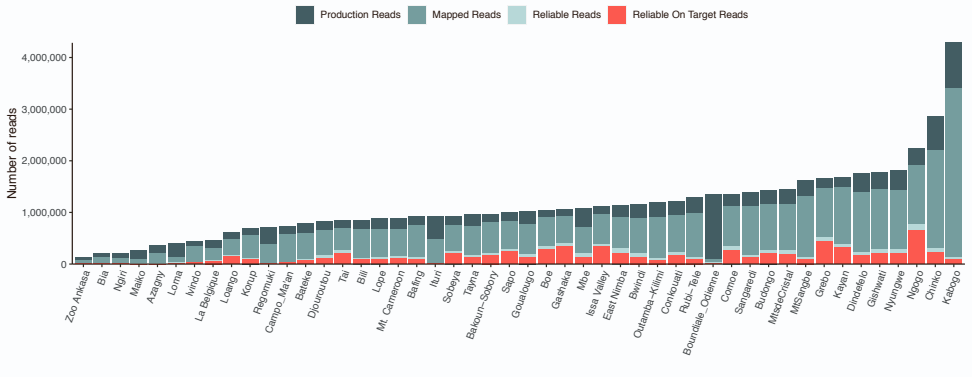
A – Summary Reads per Pool



B – Average Summary Reads



C – Summary Reads per Site



D – Average Summary Reads

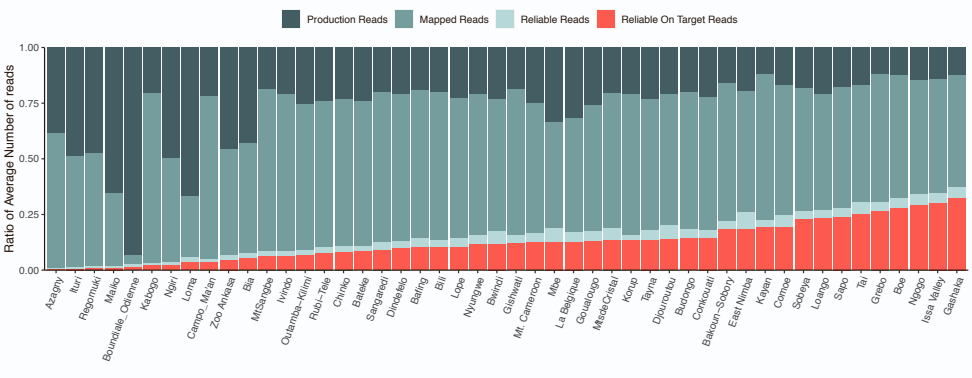


Fig. S5. Capture performance on total reads and on average by hybridization (A and B, respectively) and by geographical site (C and D, respectively). Related to Fig. 1 and STAR methods.

As expected, from pools with samples with higher average hDNA, a higher percentage of on-target reads were obtained. The same can be observed in the final coverage reach for each sample (although coverage is influenced by other factors such as amount of sequencing) (Fig. S6).

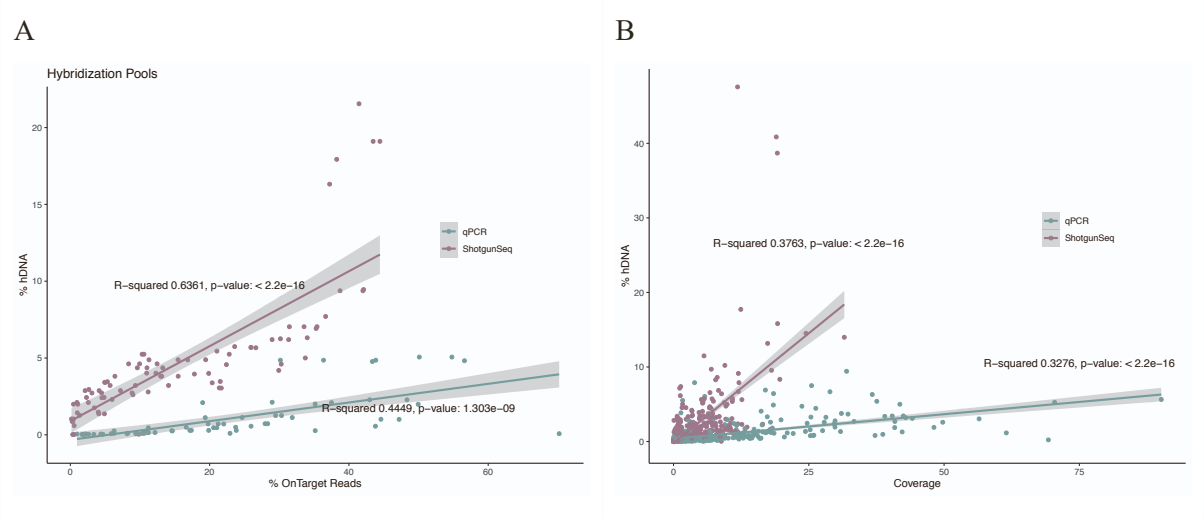


Fig. S6. Correlations between (A) percentage of on-target reads per hybridization pool with average hDNA of each hybridization pool and (B) average coverage per sample with hDNA of each sample. Each color represents a method used to estimate hDNA: qPCR and shotgun sequencing. Related to Fig. 1 and STAR methods.

The median target coverage from the whole dataset is 1.89-fold (0 - 90.14-fold), with variability according to site (Table S1 and S7A), with nearly half of the dataset having a coverage larger than 2-fold (Fig. 7).

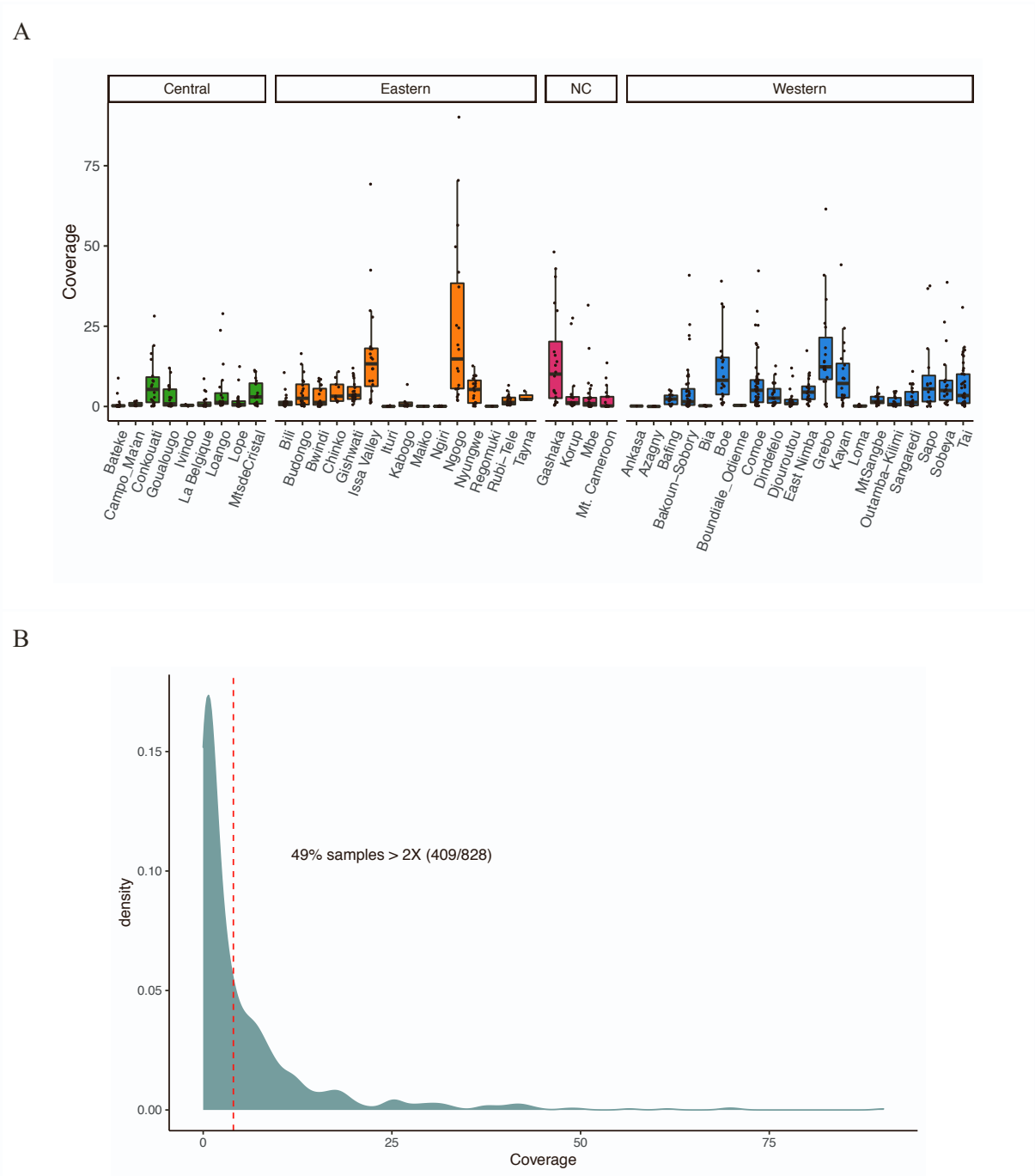


Fig. S7. Coverage. (A) Average coverage on the target region of chromosome 21 from 48 PanAfrican field sites. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and the whisker extends from the hinge to the largest and lowest value no further than 1.5 * IQR (inter-quartile range). Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. (**B**) Overall sample distribution of coverage on chr21, dashed red line marks 2-fold coverage threshold. Related to Fig. 1.

We calculated the amount of data (total sequenced base-pairs) within the target space and on the whole chromosome 21 for each individual (Table S4). We find that the vast majority of the genotype data obtained with snpAD for the captured samples (92-98%) falls within the target space (Fig. S8). In comparison, we find ~34% of the data outside the target space for the 59 individuals sequenced to high coverage from a previous study [21]. We conclude that the capture works efficiently to enrich DNA

fragments in the target space, and that the low-coverage samples in our study do show the same enrichment pattern as the high-coverage capture samples.

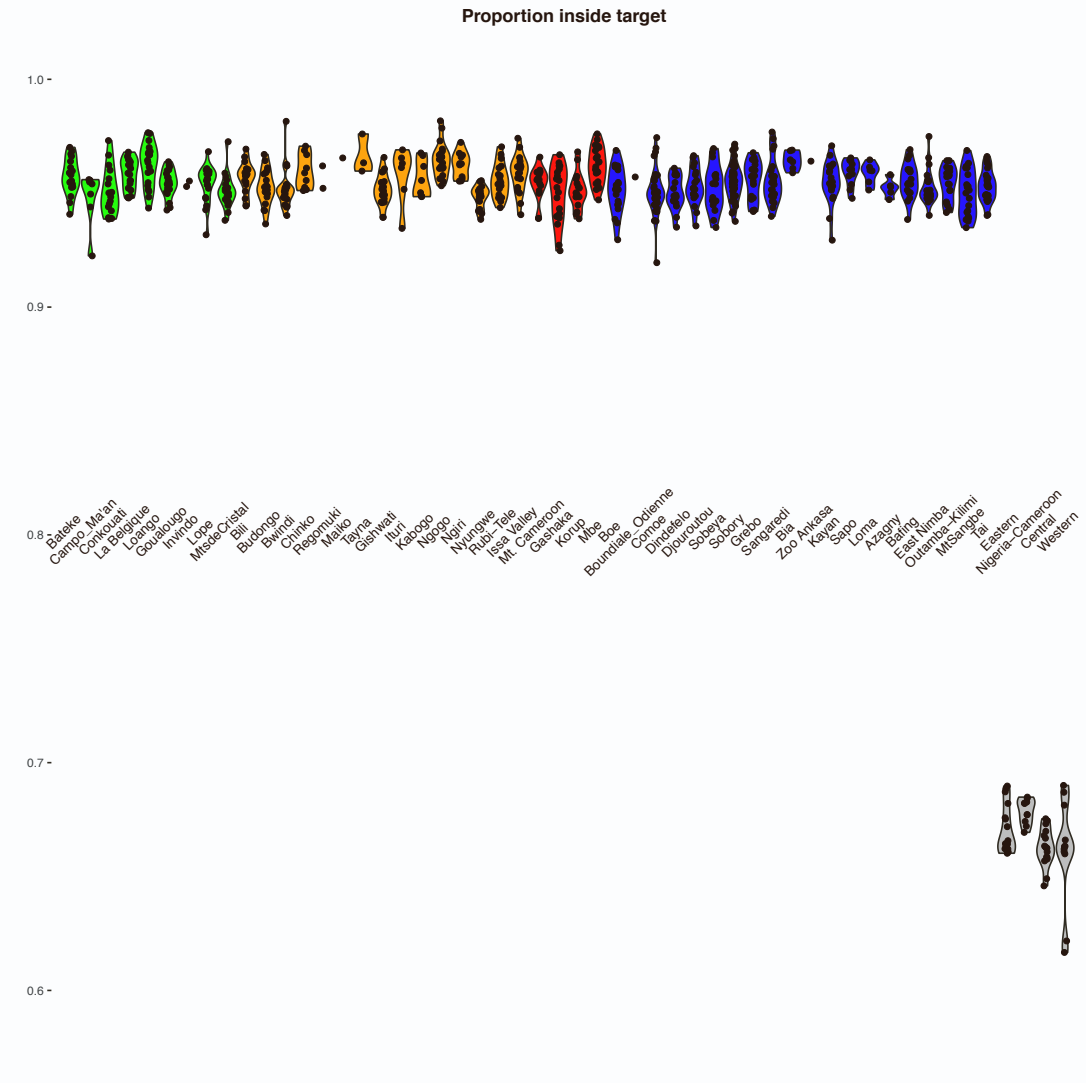


Fig. S8. Proportion of all genotype data on chromosome 21 falling inside the target space for captured samples and the whole genome samples. Western chimpanzees are marked in blue, Nigeria-Cameroon chimpanzee in red, central chimpanzee in green and eastern chimpanzee in orange. Related to Fig. 1.

3.3. Principal component analysis

Principal component analysis (PCA) of all samples (N=828) was done with PCAngsd[22] after obtaining genotype likelihoods with ANGSD, as described previously. We included 4 representatives of each known chimpanzee subspecies from de Manuel *et al.* (2016)[21] of high coverage to ensure the fecal samples analyzed in this study recapitulate the known genomic diversity of chimpanzees. The first component (PC1) is dominated by a small number of samples which are separated from the known variation from whole-genome sequencing [21] (S9A). Most likely, these samples have extremely high

levels of contamination, possibly through diet, or fecal samples from other primate species mistakenly collected for this study, as demonstrated in Note 3.4.

We performed the PCA only with fecal samples and defined a threshold to keep only those samples representative of chimpanzee diversity (Fig. S9B). This threshold was set at -0.01 of the initial PCA, in order to retain the main two clusters of samples, while being permissive at this point (Fig. S9C).

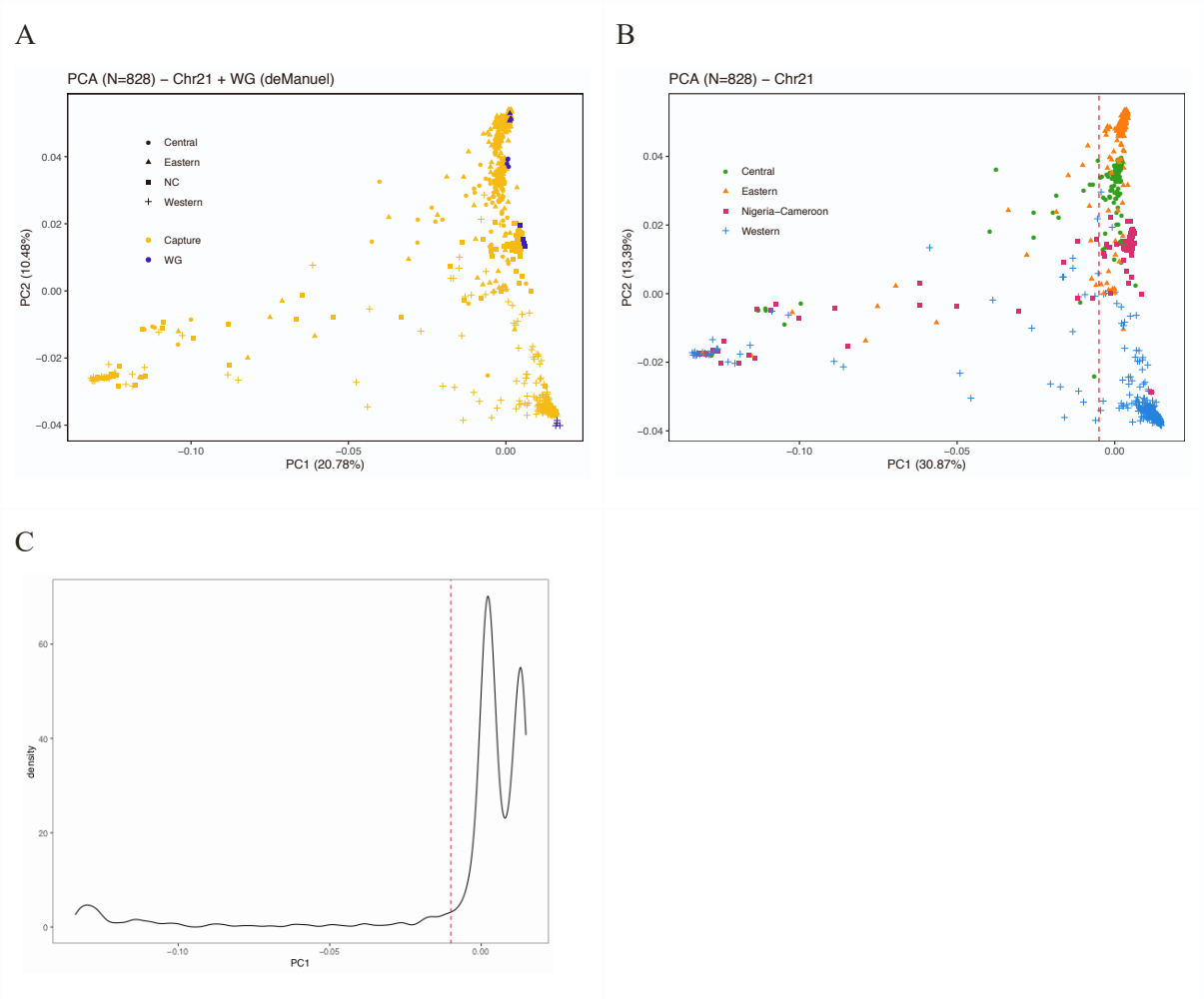


Fig. S9. PCAs of chromosome 21 single-nucleotide polymorphisms on: (**A**), all captured fecal samples (yellow, N=828) and with the data on chromosome 21 for four whole genomes from each subspecies[21] (blue, N=16), using 1,533,092 variable positions in the target space; (**B**) only fecal samples with 1,430,461 markers, threshold is set at -0.01 PC1 to exclude the potentially contaminated samples and (**C**), distribution of samples at PC1. Related to Fig. 1.

The resulting PCA without 100 outlier samples (column *PCA all chimps N=728,* Table S1) results in the clustering of the four chimpanzee subspecies, also confirmed by adding 4 samples of each know subspecies from de Manuel *et al.* (2016)[21] (Fig. S10).

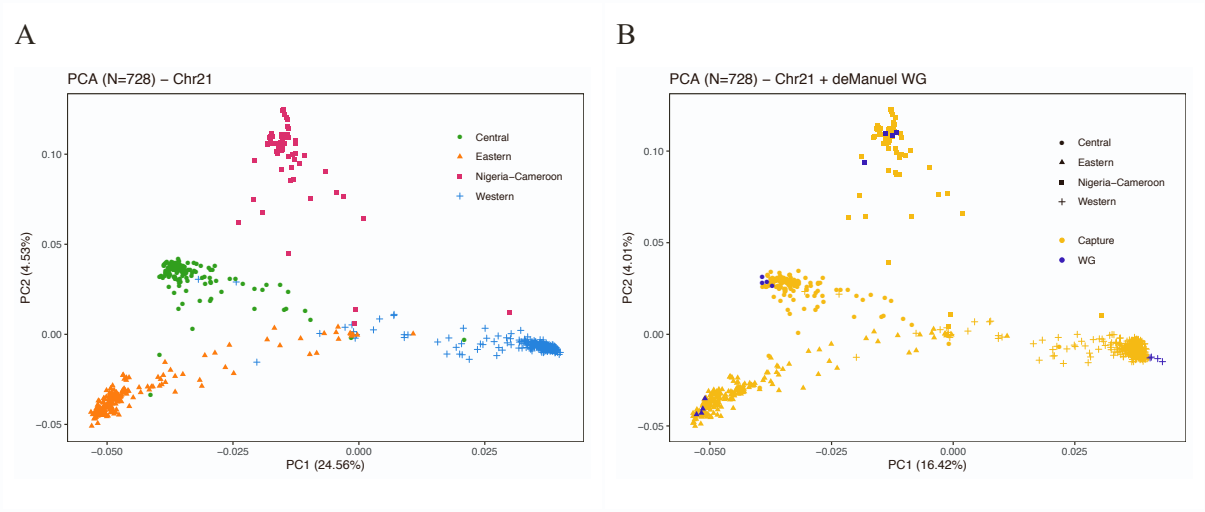


Fig. S10. PCA of chimpanzee fecal samples (without putatively contaminated samples) (A) excluding WG samples from de Manuel *et al.* (403,762 markers) and (B) including WG samples from de Manuel *et al.* (513,406 markers) at chr21. Related to Fig. 1.

We do not observe clusters separating from the four known subspecies, while a number of samples is tending towards the center (0,0). These samples have significantly lower coverage than those falling close to the four major clusters (Fig. S11), but do not belong to specific sites, suggesting that this is due to missing data, rather than a genetic gradient.

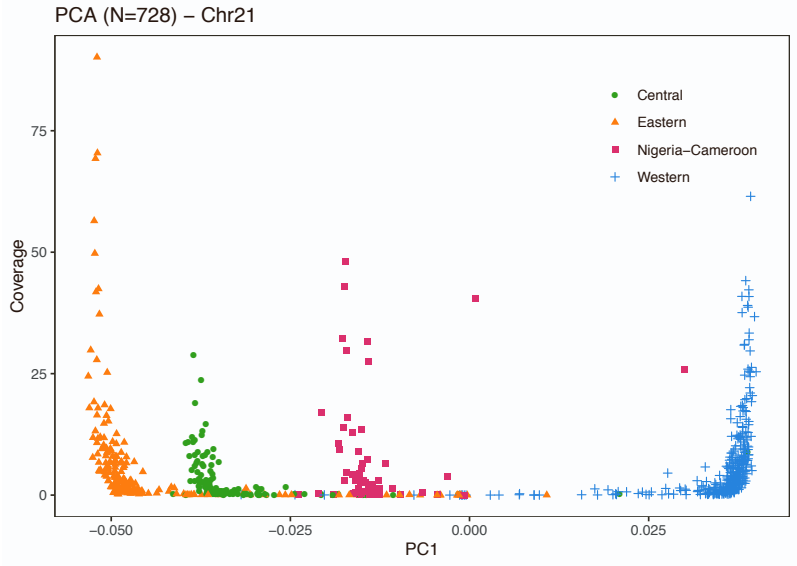


Fig. S11. Coverage at target space with PC1 from Fig.10A. Colors and shapes represent each chimpanzee subspecies. Related to Fig. 1.

When keeping only samples with average coverage > 0.5-fold (N=556) (Table S1), samples belonging to the four known subspecies clearly cluster together, supporting previous evidence for four distinct chimpanzee populations in the wild [21,23]. We find that one Nigeria-Cameroon chimpanzee from Korup (Kor1-35) (Fig. S12) clusters within Western chimpanzee diversity, probably as a result of

mislabeling of samples since those subspecies are separated by several 1000 kilometers without habitat continuity, and a recent migration event between those subspecies would be highly unlikely. We removed this sample from further analyses (Table S1) to obtain the four-known subspecies clustering in a PCA (Fig. S13)

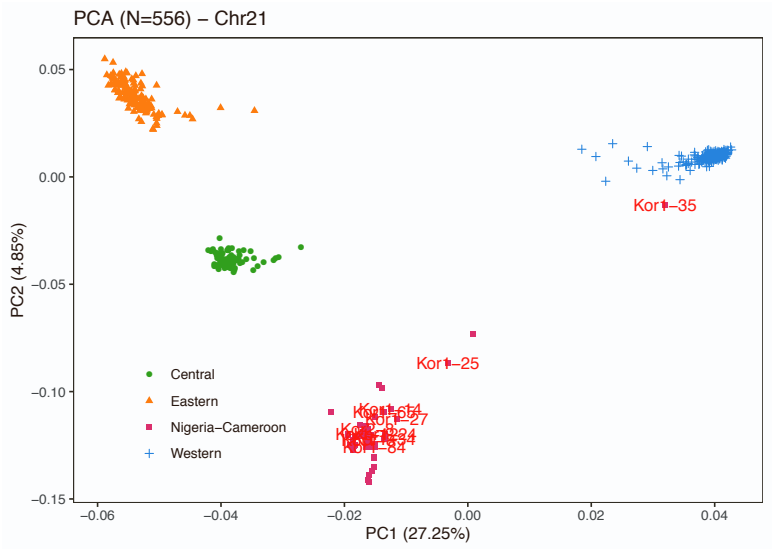


Fig. S12. PCA of samples with average coverage > 0.5X (N=556 samples) using 405,136 markers. An outlier sample detected with PCA: Kor1-35 from Korup, a Nigeria-Cameroon sampling site, falls within western chimpanzee diversity. Related to Fig. 1.

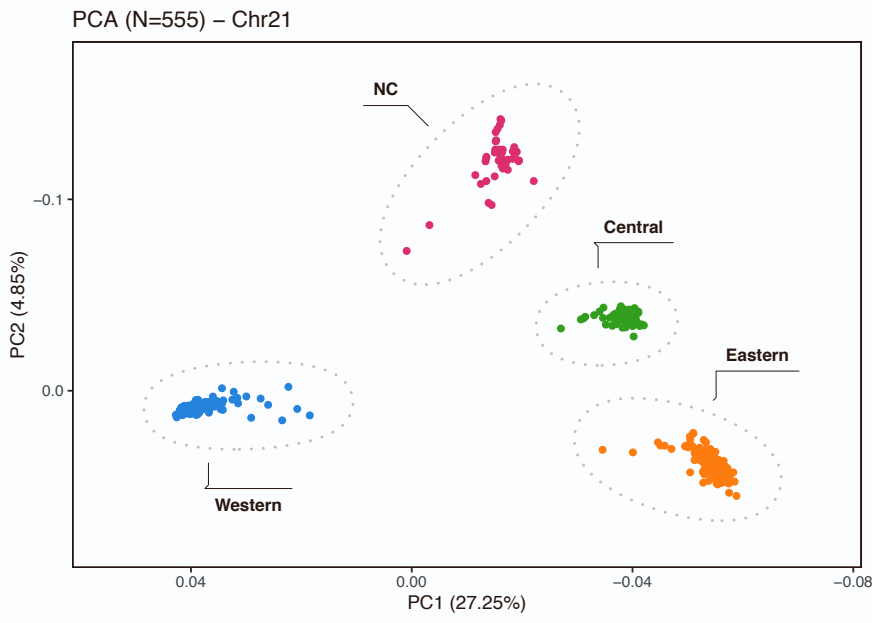


Fig. S13. PCA of chromosome 21 single-nucleotide polymorphisms on chimpanzee fecal samples with more than 0.5-fold coverage (N=555) and using 405,136 markers after removing the outliers. Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 1.

3.4. Sources of contamination

3.4.1. Primate contamination

According to Fig. S9, it is likely that a minority of samples (N=100) had significant amounts of primate contamination, either being samples from a different primate, or chimpanzee samples with some degree of contamination. Since other primates are closely related to chimpanzees and humans, we used the BBsplit (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/) software to competitively map sequencing reads obtained from each the sample to a range of other primate genomes, and thus obtain a summary of unambiguous mappings to each genome. We used the BAM files without duplicates mapped to the human genome (Hg19), converted them into FASTQ and then mapped them to these primate genomes. We restricted the analysis to only primates since we are only looking at the endogenous portion of the DNA that was already mapped to the human genome. The primate assemblies used were: chimpanzee (panTro6), human (hg19), olive baboon (Panu_3.0), green monkey (Chlorocebus_sabeus_1.1), Angolan colobus (Cang.pa_1.0), sooty mangabey (Caty_1.0), gorilla (gorGor4), drill (Mleu.le_1.0), Patas monkey (EryPat_v1_BIUU), Da Brazza monkey (CertNeg_v1_BIUU) and mandril (mandrill_1.0). We used the proportion of unambiguously mapping reads as a proxy to determine whether other primates than chimpanzees are more likely to be the source of the endogenous DNA, not to definitively determine these primate species, since we limited this test to only 11 primate assemblies.

It is worth noting that the following results might be confounded by low coverage data, and since samples with less than 0.5-fold coverage will be excluded from further analysis, we have plotted them separately. We determined that the majority of samples removed following PCA criteria (Table S1, N=100) (Fig. S14) contained a higher proportion of reads mapping to other primate genomes compared to samples retained with that filtering (Fig. S15).

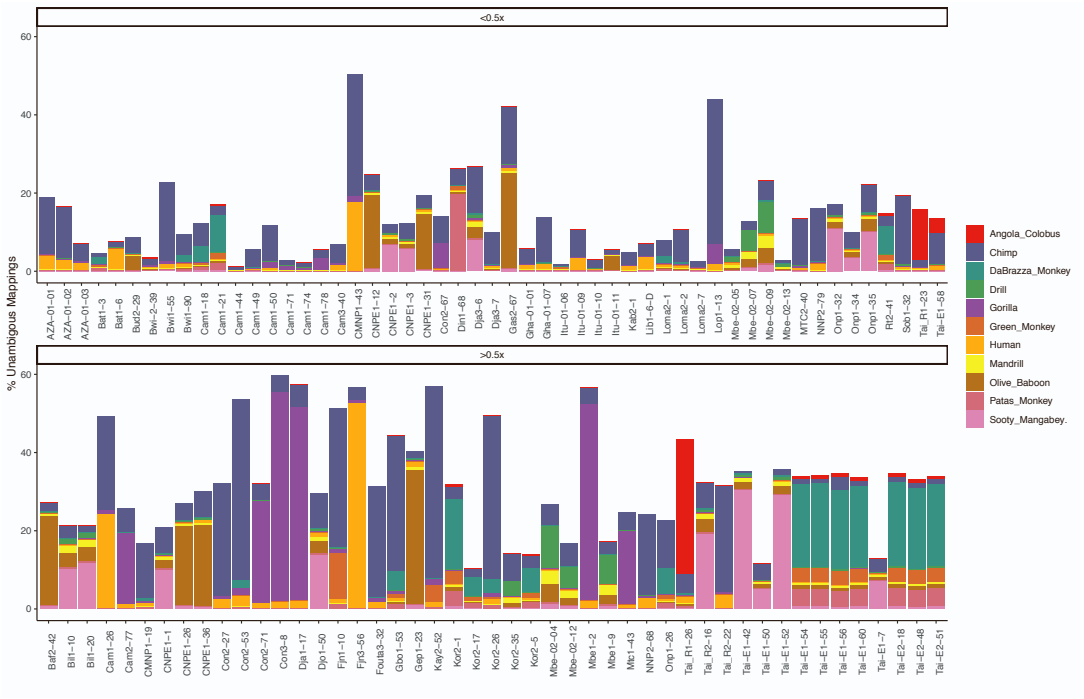
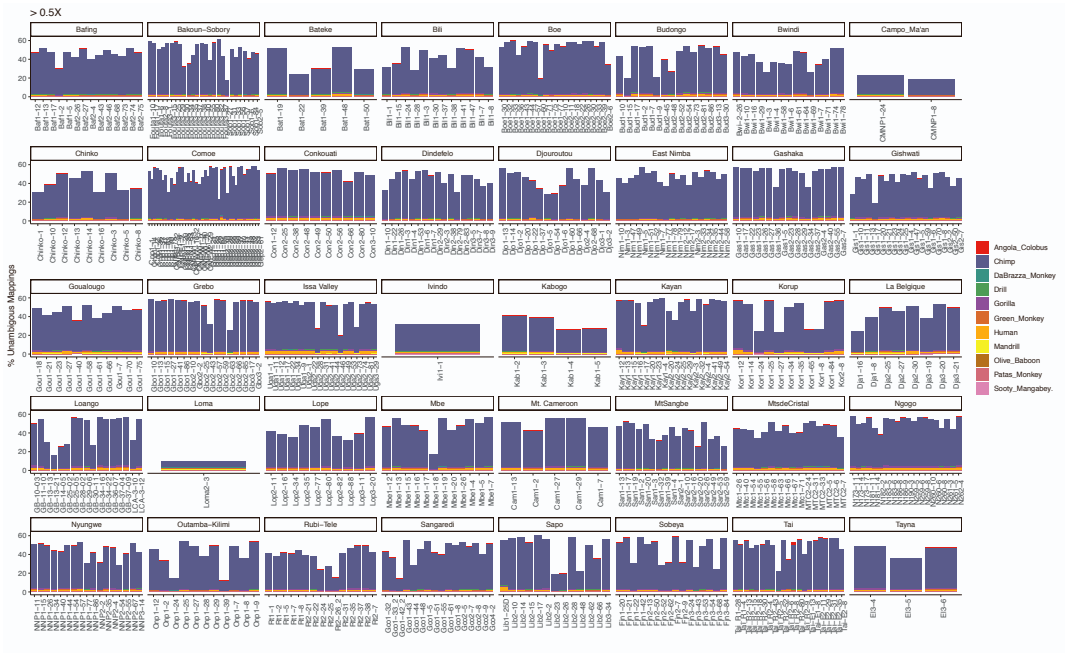


Fig. S14. Proportion of unambiguous mappings to many primate species for samples that have been excluded by PCA filtering (N=100) at different coverage (higher or lower than 0.5-fold coverage). Related to STAR methods.

Some samples have the majority of their reads mapping to the gorilla genome, this is the case of two samples from Conkouati (Con2-71 and Con3-8), one from La Belgique (Dja1-17), one from Mt. Cameroon (Cam2-27), one from Mts de Cristal (Mtc1-43) and one from Mbe (Mbe1-2), all of them from geographical locations where the gorilla species range overlaps with the chimpanzee range. Most probably, these samples are misidentifications at the moment of collection, mistakenly taking a gorilla sample for a chimpanzee sample. We also find samples with a high proportion of olive baboon reads from sites within the geographic range of this species, such as Comoé (CNPE1−26 CNPE1−36, Gep1-23, CNPE1-12, CNPE1-31), Bafing (Baf2-42), Gashaka (Gas2-67) and Ituri (Itu-01-11). Two samples from Tai-Eco (Tai-E1-42 and Tai-E1-52) may belong to sooty mangabey, a primate species that inhabits western Africa and is specifically present at Taï Forest. Many other samples from Tai-Eco seem to be from an unidentified primate species. Surprisingly, one sample, Fjn3-56, is probably a human feces. Other samples had some degree of primate contamination with the majority of reads still mapping to the chimpanzee genome. This might be explained by diet, considering that the chimpanzee diet includes other primate species[24] and thus DNA of the diet can survive the intestinal tract and be found in the feces[6].

On the other hand, for samples kept after initial PCA filtering, the majority of reads were mapping to chimpanzee genome (Fig. S15), although some low levels of reads mapping to the human genome were shared across all samples.

A



B

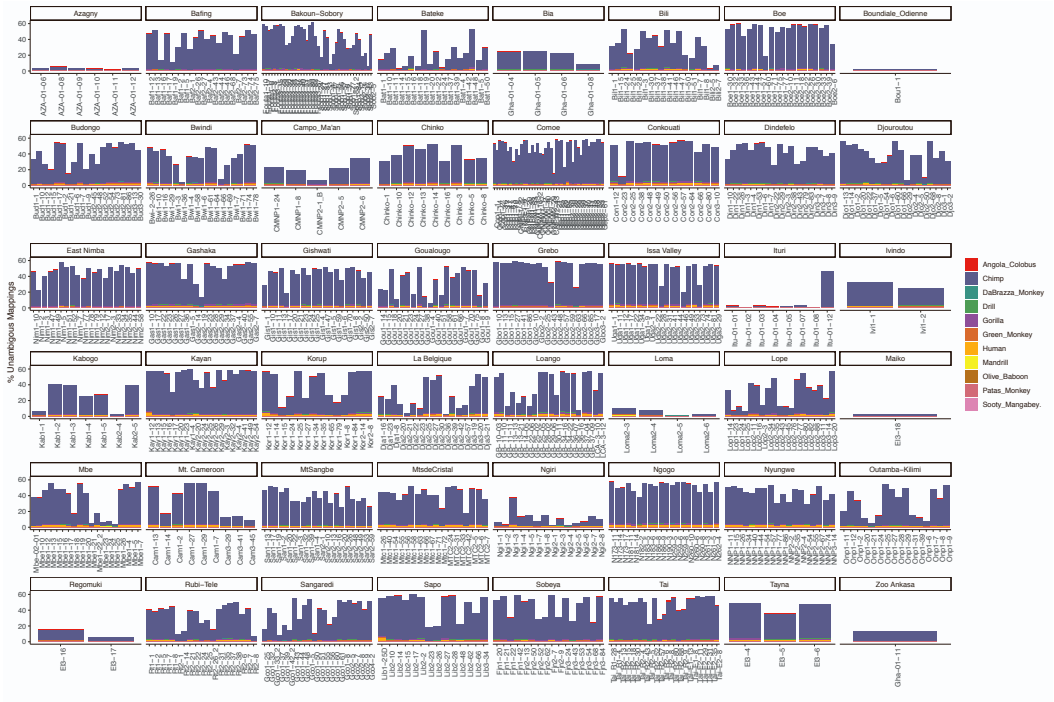


Fig. 15. Proportion of unambiguous mappings to many primate species for samples that have been kept by PCA filtering (N=728) at different coverage cutoff (more (A) or less (B) than 0.5-fold coverage). Related to STAR methods.

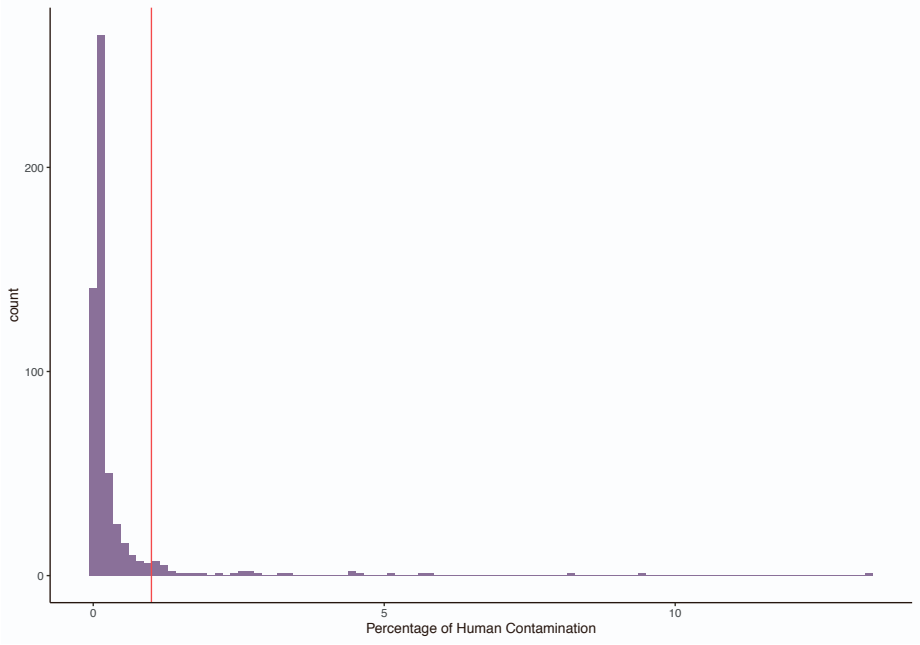
Two samples, Lib1-25D and Gep1-65, show a relatively high amount (5.02% and 3.04%, respectively) of reads mapped to the human genome than the rest of samples, and Tai_R2-8 showed a

larger proportion of Angolan Colobus (3.27%) than the rest. These samples were excluded from further analyses. Since it is difficult to disentangle human contamination from mapping bias, we decided to apply another method to be more rigorous and remove from the dataset those samples with higher levels of human contamination.

3.4.2. Human contamination

Human contamination can occur at different stages, during sample collection, laboratory procedures and sequencing. We devised a human contamination test by using positions where modern humans and chimpanzees consistently differ, as implemented in the HuConTest script [25]. Using samtools mpileup [12], we retrieved the number of observations of human-like and chimpanzee-like alleles at these positions, considering the fraction of observations for the human-like allele across all positions as an estimate for possible human contamination. Using this method, we obtained the estimates for each sample (Table S1). Within the working dataset that passed previous filters of primate contamination and coverage >0.5-fold (N=556) we found that 36 samples had more than 1% human contamination (Table S1, Fig. S16). Samples selection based on human contamination and other parameters can be found in Table S1.

A



B

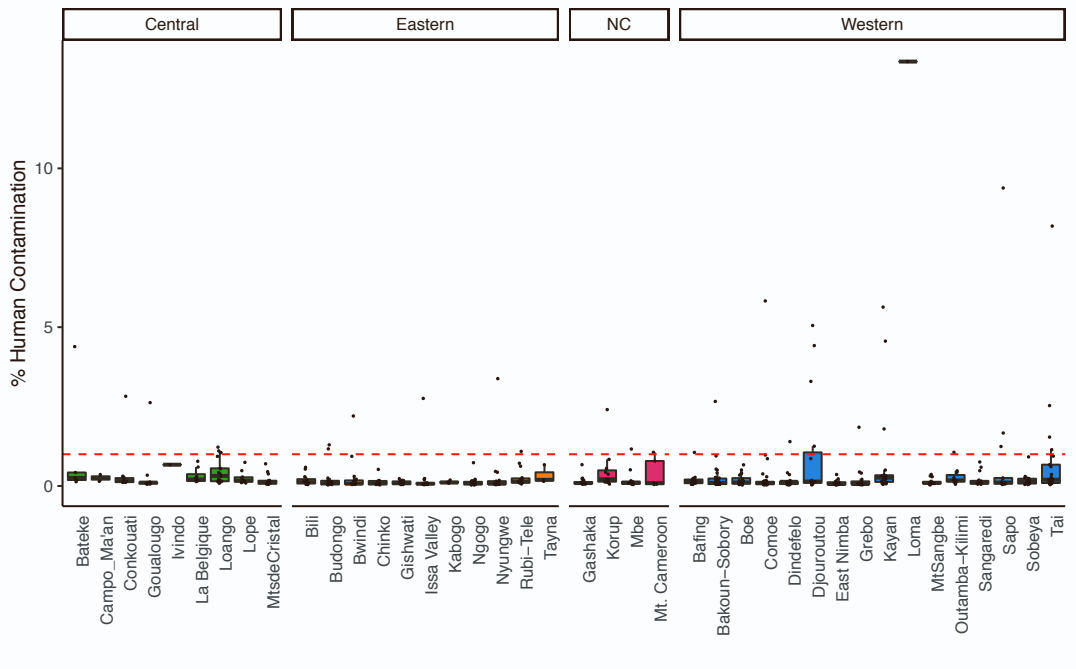


Fig. S16. Global human contamination distribution (A) and by field site (B). Threshold (red line) marks 1% of human contamination. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and the whisker extends from the hinge to the largest and lowest value no further than 1.5 * IQR (inter-quartile range). Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

For more refined analyses, we kept a smaller dataset with more than 5-fold coverage. In this case we also reduced the threshold of allowed human contamination to less than 0.5% (Table S1 and Fig. S17).

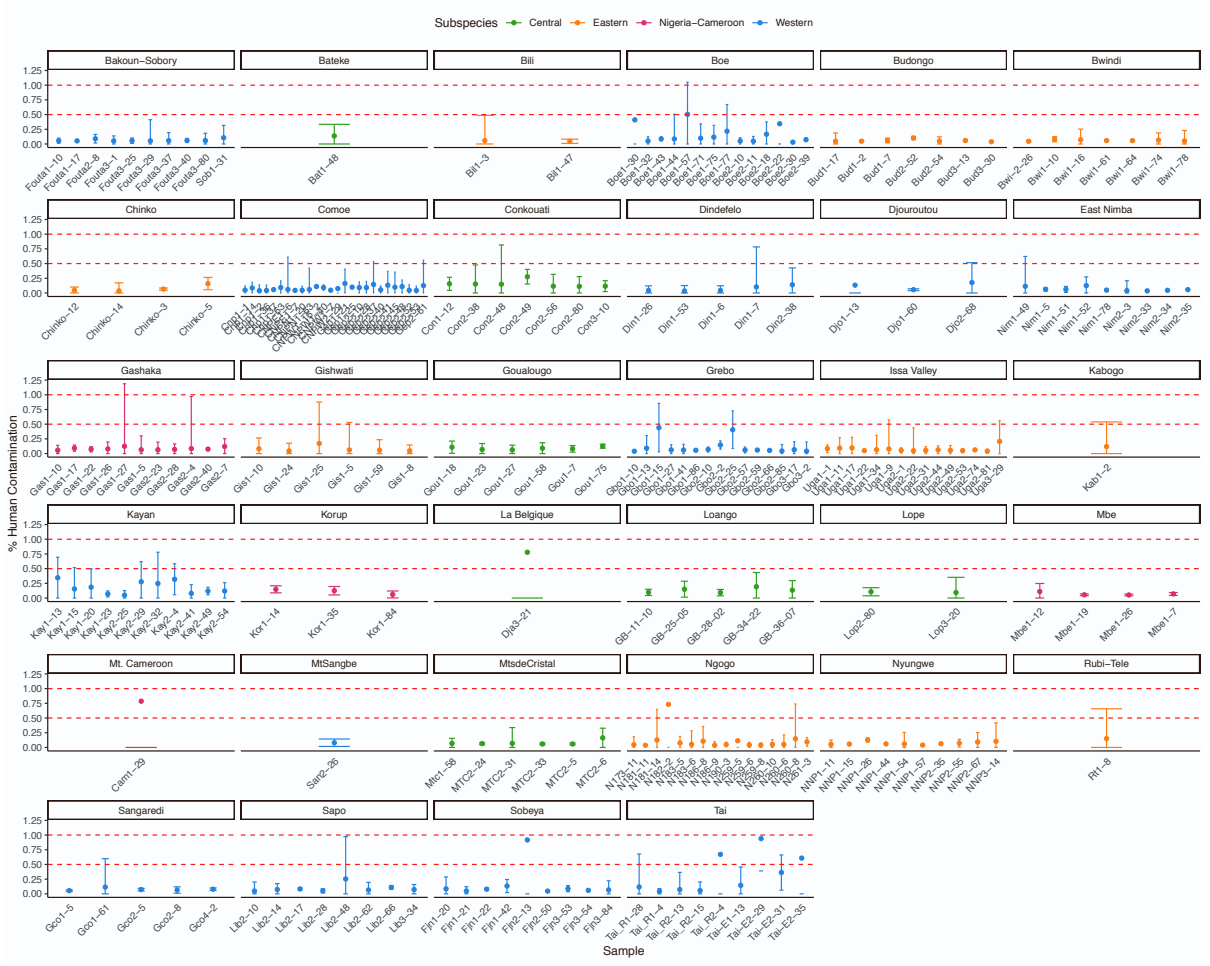


Fig. S17. Human contamination distribution per sample in the dataset with average coverage larger than 5-fold, with error bars representing the standard deviation across chromosomes. Thresholds (red lines) indicate 1% and 0.5% or human contamination. Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

3.5. Relatedness

Fecal samples used in this study had been previously genotyped with microsatellites to discard those samples that belong to the same individual or were 1st order relatives. Still, we tested the dataset to remove any related pair at 1st degree or more that may have remained. We obtained genotype likelihoods for each field site independently using ANGSD [14], extracted allele frequencies at each site and ran NgsRelate [26]. We decided to calculate relatedness at each site separately to avoid population structure bias within each subspecies, that would result in an overestimation of related pairs (Fig. S18).

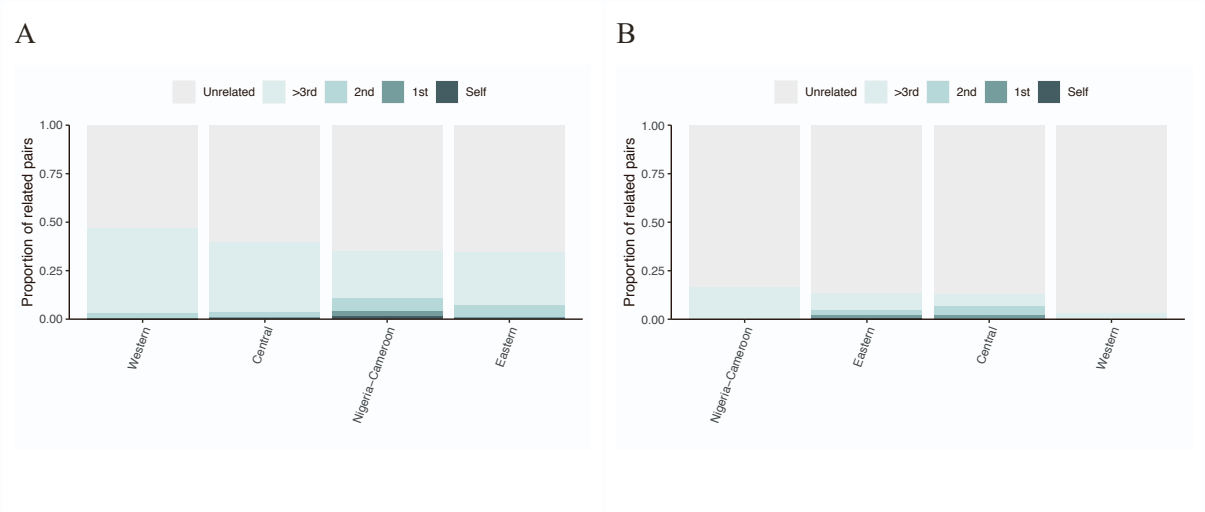


Fig. S18. Proportion of related pairs of each subspecies, by calculating relatedness estimates with (A) all samples of each subspecies together or (B) by restricting the analysis by geographical site independently. Related to STAR methods.

We consider unrelated individuals when their kinship coefficient has a value of 0, third degree or higher when it has a value between 0 and 0.0625, 2nd degree relatives when it fluctuates between 0.0625 and 0.1875 and 1st degree relatives when this coefficient has a value between 0.1875 and 0.375. Samples with a kinship coefficient higher than 0.375 are considered to be identical [27].

Whenever we encountered a pair with kinship coefficient > 0.1875, we kept the sample with the highest coverage. With these criteria, we removed from further analysis a total of 89 samples (Table S1). Out of 3581 total pairs analyzed at each site, only a small fraction (96 pairs) were first order or identical samples (Fig. S19), with the majority being of unrelated pairs.

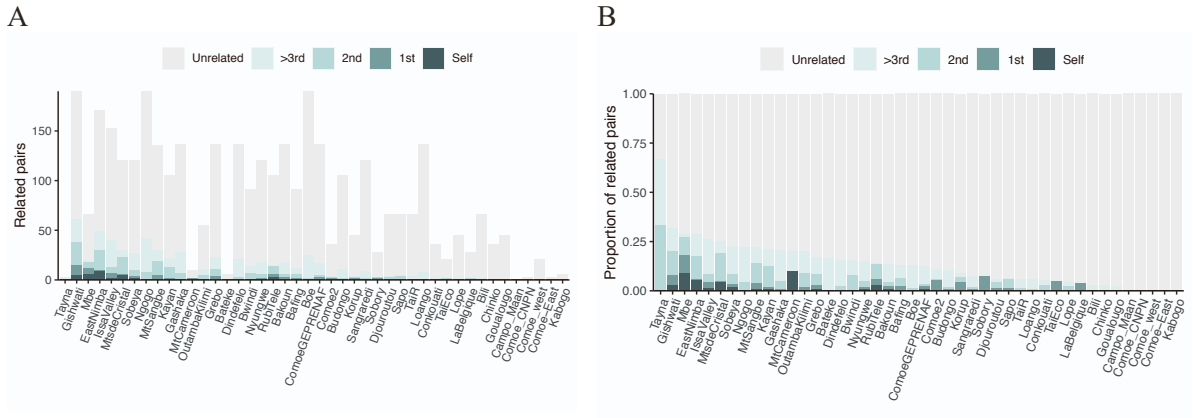


Fig. S19. Relatedness events encountered with the analysis of samples with dataset without primate and human contamination and with > 0.5x coverage. A) Total events B) Proportion of events in each category. Related to STAR methods.

3.6. Comparison of genotype calls between snpAD and GATK

We performed a comparison of genotype calls using GATK and snpAD for one randomly chosen sample (Kor1-65) of medium coverage, restricted to genotypes called by both methods and

within the target space. We observe similar transition to transversion ratios (GATK: 2.12; snpAD: 2.14) and average coverage (snpAD: 2.49; GATK: 2.52), but a much higher average genotype quality for snpAD (35.8) than GATK (8.2) calls. This suggests a much higher confidence in genotype calls using snpAD. We find 0.08% (5,322) of the overlapping genotypes in disagreement between the two methods, most of them (3,979) in heterozygous state when called with GATK, but not with snpAD. These disagreeing heterozygotes have on average lower genotype quality score for snpAD (25.1) than the average score with the same method (35.8), suggesting that a large proportion of these are due to errors or contamination. When estimating heterozygosity from the overlapping genotype calls, we find an estimate of 1.14 heterozygous sites per 1,000 bp (kbp) for GATK, and 0.54 sites per kbp for snpAD genotype calls, suggesting that erroneous heterozygous calls are removed by snpAD.

This individual belongs to the Nigeria-Cameroon subspecies, for which heterozygosity values of ~0.94 per 1,000 bp have been found previously [21]. Considering that the heterozygosity values calculated on likelihood genotypes as well as snpAD genotypes of captured individuals are consistently below the estimated heterozygosity from high-coverage individuals, most likely due to capture bias and missing heterozygotes in regions of low coverage, we conclude that snpAD efficiently removes erroneous heterozygotes, even at low coverage. Most likely, not all errors would be removed from the data, however, the estimate improves using snpAD compared to GATK, and could be further improved by quality filtering. Still, only samples of sufficiently high coverage (more than 3-fold) may be used for reliable estimates [20].

We also performed a PCA analysis, using the best individuals per site and a stringent filtering cutoff of at least 8-fold coverage per genomic position in order to avoid the capture bias. Due to the high missingness when using these filtered genotypes, we only retain 16,838 segregating sites across 27 individuals from different geographic locations. Still, we find a clear separation between the four known subspecies in the first two components (Fig. S20), although central chimpanzees appear to have a larger diversity than the other subspecies, in agreement with previous findings [21] and other aspects of this work (see below).

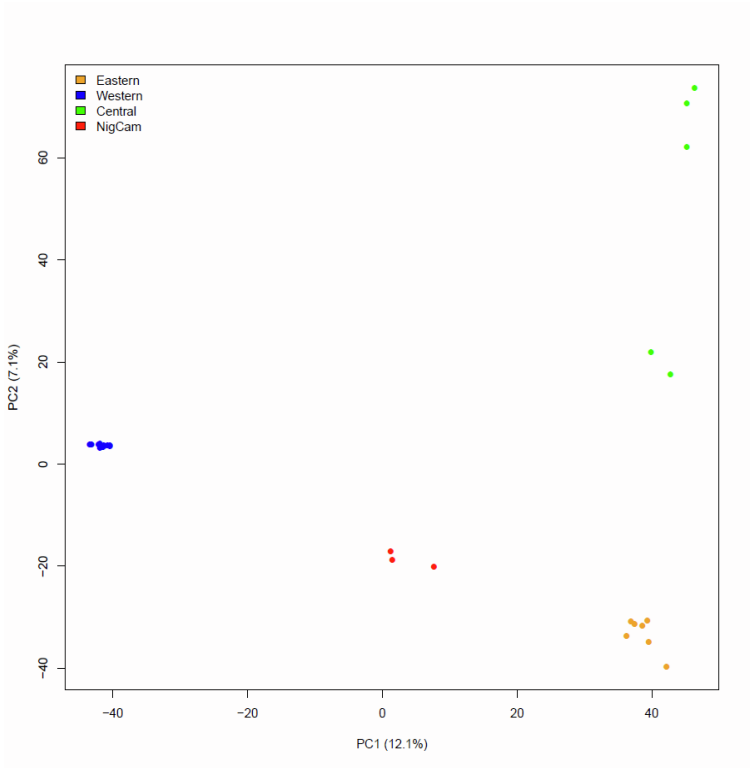


Fig. S20. PCA of 16,838 high-quality genotypes found across 27 individuals. Related to STAR methods.

3.7. Quality assessment of snpAD genotype calls

In order to study biases in the sequencing data, we used the genotype calls obtained with snpAD for samples assigned to chimpanzee by BBsplit. We analyzed the data at three different levels of quality, using bcftools version 1.9 [28] and R (3.2.0) and the R package GenomicRanges [29]. These were:

- "all sites": All sites with a called genotype across the merged VCF file of all individuals, regardless of their quality.

- "quality sites": Subset of these sites with a sequencing depth of at least 3 reads, less than 100 reads, and a genotype quality of more than 20.

- "high quality sites": Subset of these sites with a sequencing depth of at least 8 reads, less than 100 reads, and a genotype quality of more than 40.

We did not apply further filters at this stage given the low coverage and limited amount of data. We performed the same filtering for the genotypes of 59 chimpanzees from whole-genome sequencing [21]. We retrieved the fraction of target space bases covered by genotype calls in each individual for the whole dataset at each of the three quality levels, as well as heterozygosity and the number of homozygous alternative sites.

First, we inspected the fraction of the target space covered at each filtering threshold. We find that samples at high average coverage tend to reach a high degree of covering of the target space, while sites with only low sequencing coverage (like Ngiri) show a small fraction of the target space being covered (Fig. S21). On average, 60% of the target space is covered by a called genotype and supported by at least one read (all sites). When considering samples that are not outliers in the initial PCA (Note 3.3), this number increases to 62.7%. After applying the basic filtering for quality sites, on average 47.6% of the target space is covered across samples (50.2% without PCA outliers). This number drops to an average of 27.3% (29% without PCA outliers) when considering only high quality sites, with 151 samples retaining data for less than 0.1% of the target space (compared to two such samples for quality sites), suggesting that the loss of data is substantial at this stage, and using quality sites might be a good compromise.

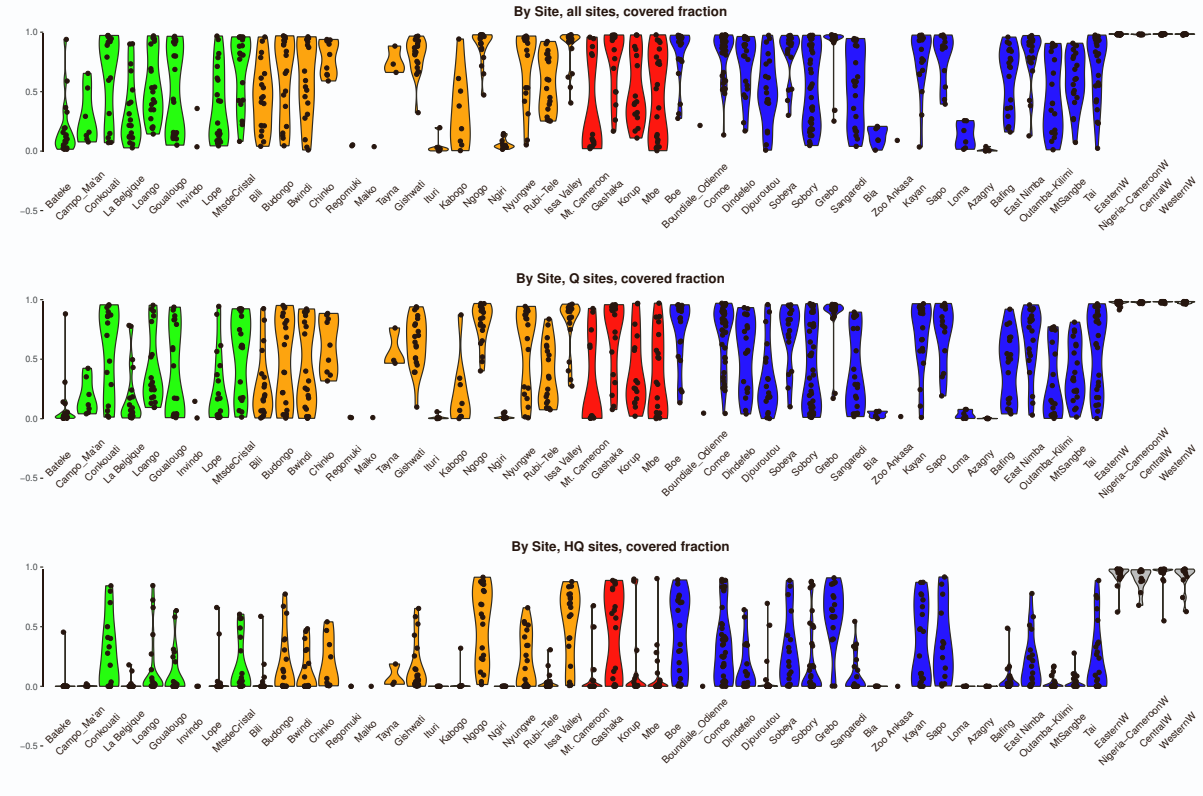


Fig. S21. Fraction of the target space covered before and after filtering of quality and high-quality sites. Color code: green for central chimpanzees, orange for eastern chimpanzees, red for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

As expected, the average sequencing depth of retained sites is mostly affected at the lower bound (Fig. S22). Before filtering, many samples had coverage near or below one-fold, while the remaining sites had coverage mostly near the lower cutoff. A similar pattern can be observed for genotype quality (Fig. S23).

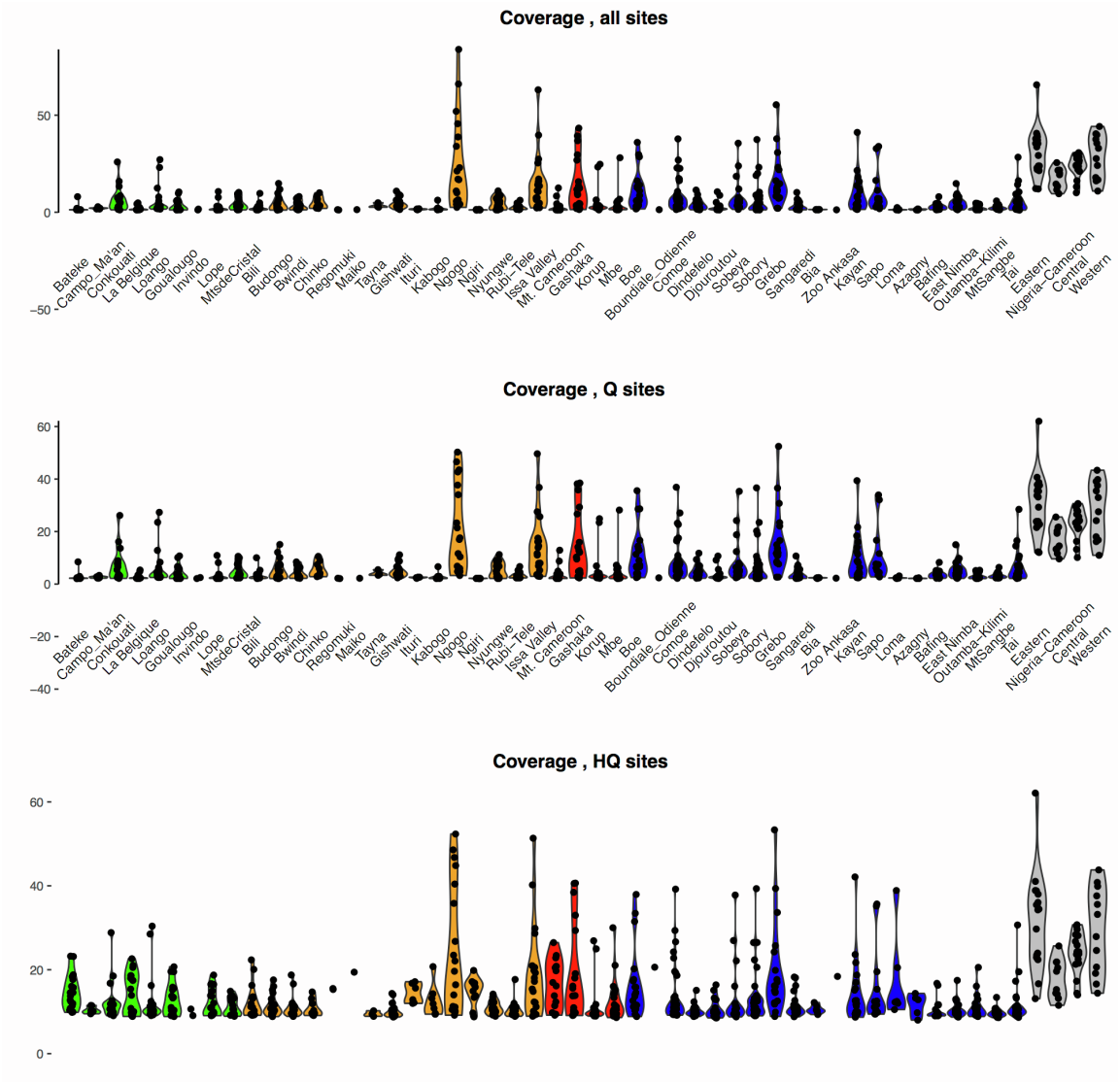
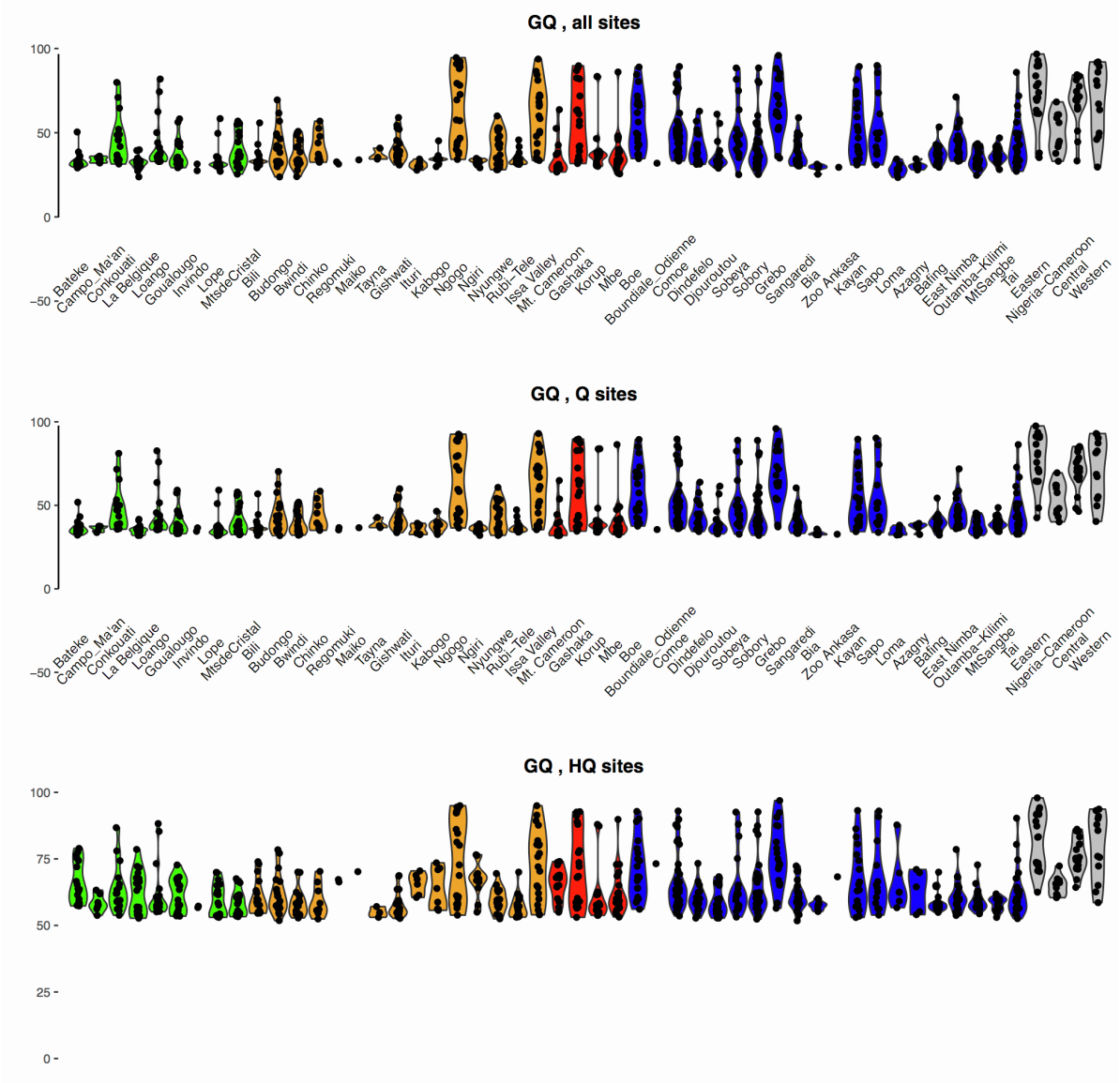


Fig. S22. Average coverage of retained sites in the target space before and after filtering. Color code: green for central chimpanzees, orange for eastern chimpanzees, red for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.



450

Fig. S23. Average genotype quality of samples before and after filtering. Color code: green for central chimpanzees, orange for eastern chimpanzees, red for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

Second, we determined outliers with excessive heterozygosity (Fig. S24). Generally, we find 455 that an increased quality cutoff does not influence the heterozygosity estimates, except for a possible disproportionate removal of false homozygous calls covered by only one sequencing read in samples of very low coverage. Since high quality sites require at least 8 sequencing reads at a given position, these are very rare in such samples, but in contaminated individuals these will more accurately represent the presence of two different classes of reads.

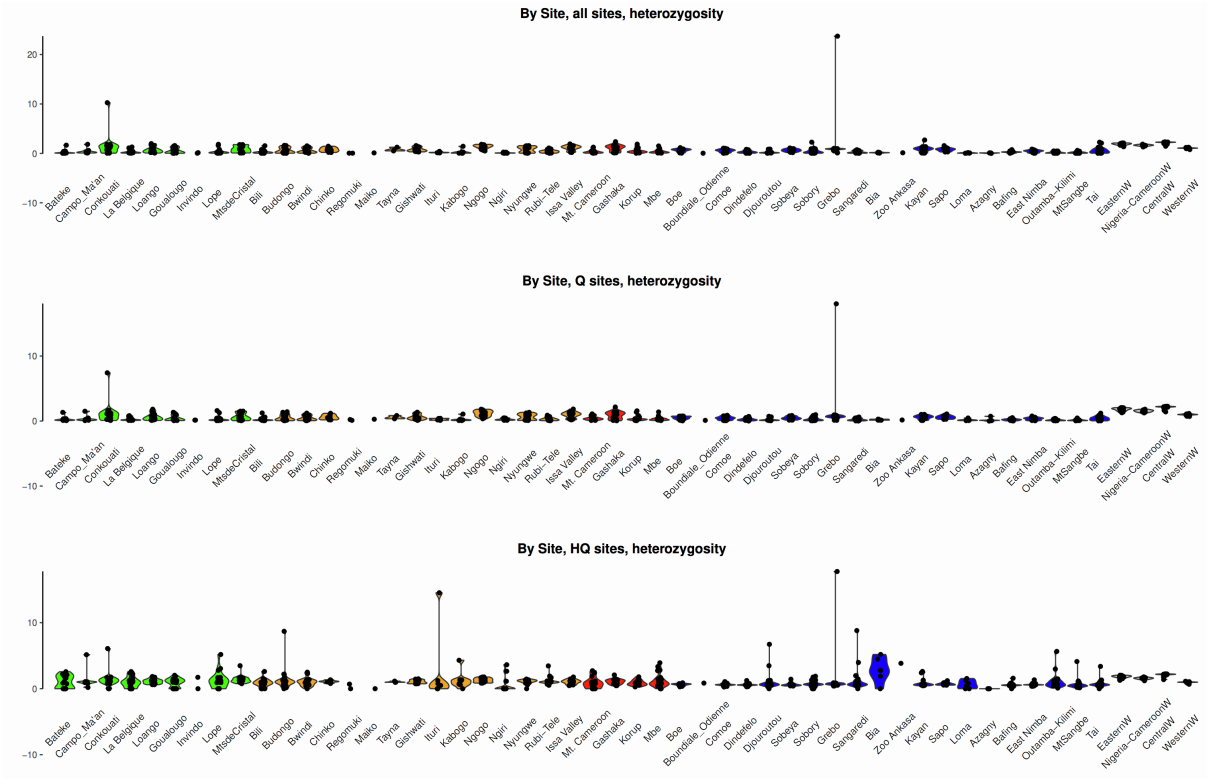


Fig. S24. Estimated heterozygosity per sample and per geographic site. Note that this graph includes all samples, including PCA outliers and contaminated individuals. Color code: green for central chimpanzees, orange for eastern chimpanzees, red for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

Third, when considering homozygous alternative sites, we find samples with an extreme excess of such sites (several-fold compared to other samples), distributed across locations and subspecies (Fig. S25). Again, increased quality filtering does not change the occurrence of this pattern, indicating that it is due to properties of the sample rather than quality of the sequencing data.

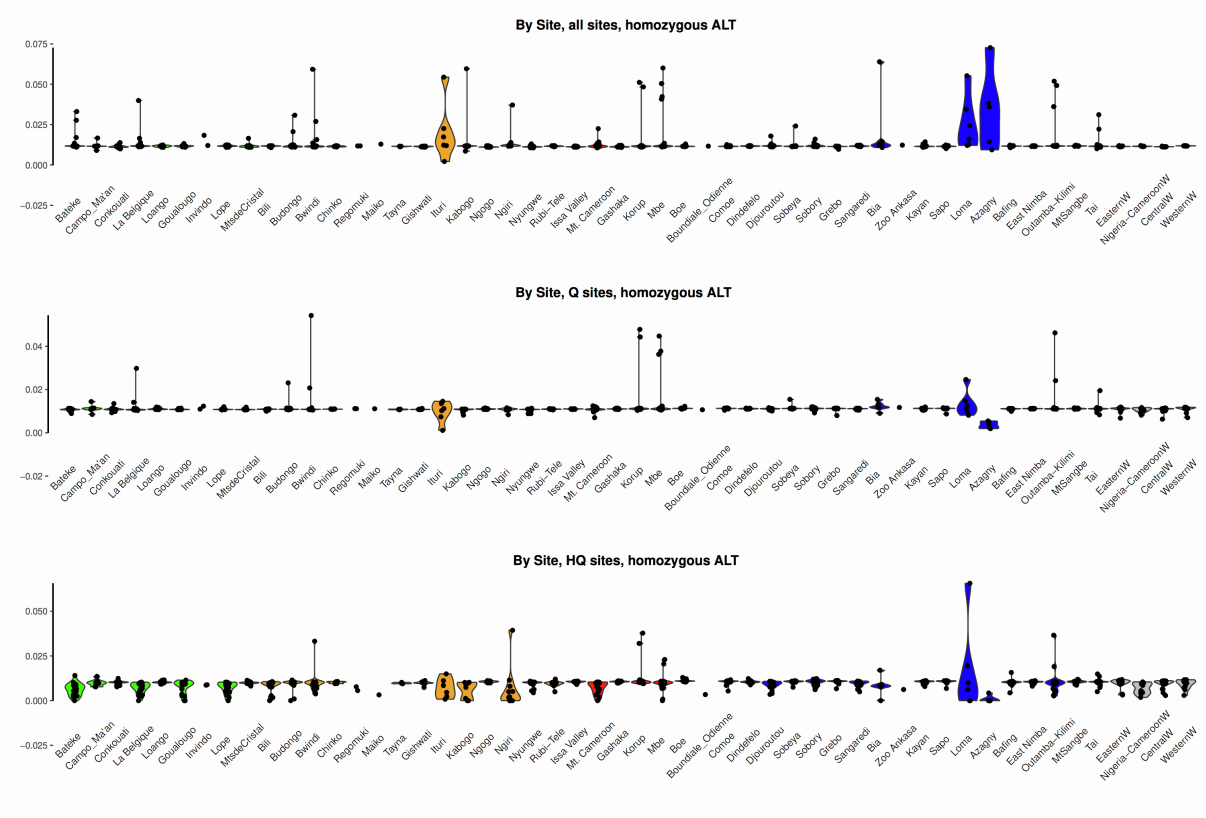
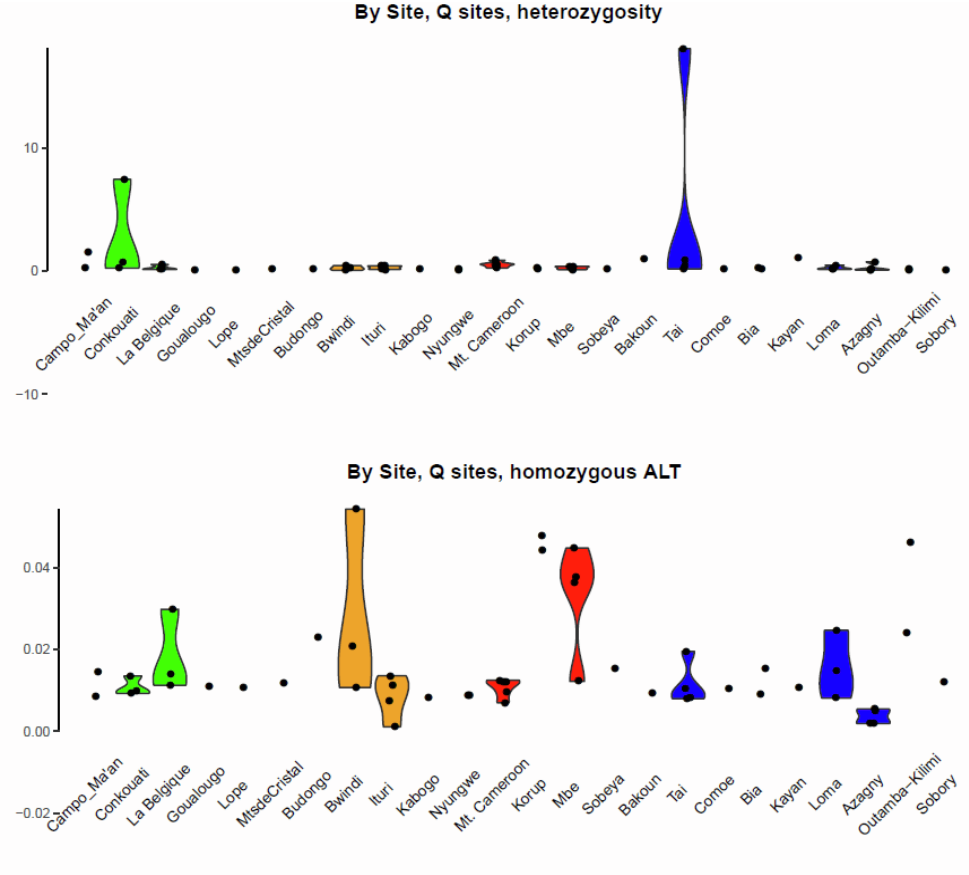
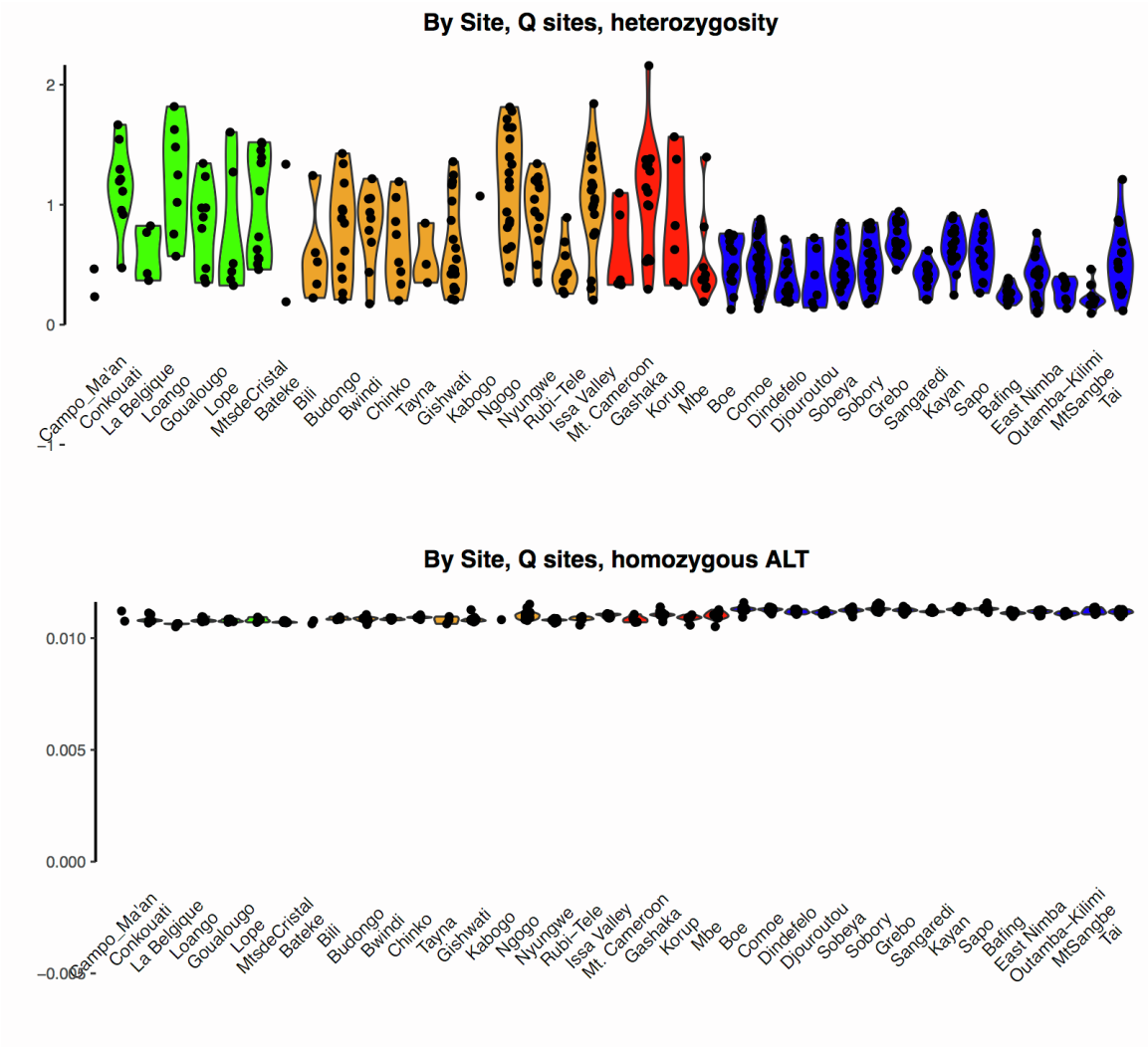


Fig. S25. Fraction of homozygous alternative sites among homozygous sites per sample and per geographic site. Color code: green for central chimpanzees, orange for eastern chimpanzees, red for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

These observations possibly indicate contamination from a non-human source, which, at low coverage, would often be observed in a homozygous alternative state. We find that samples with an excess of either hetero- or homozygous sites were also identified as outliers in the initial PCA (Fig. S26 for quality sites). Using only individuals of high quality (Table S1), we find no individuals with such an excess (Fig. S27 for quality sites), and particularly a flat distribution of alternative homozygosity, clearly showing that these samples are indeed all of chimpanzee origin.



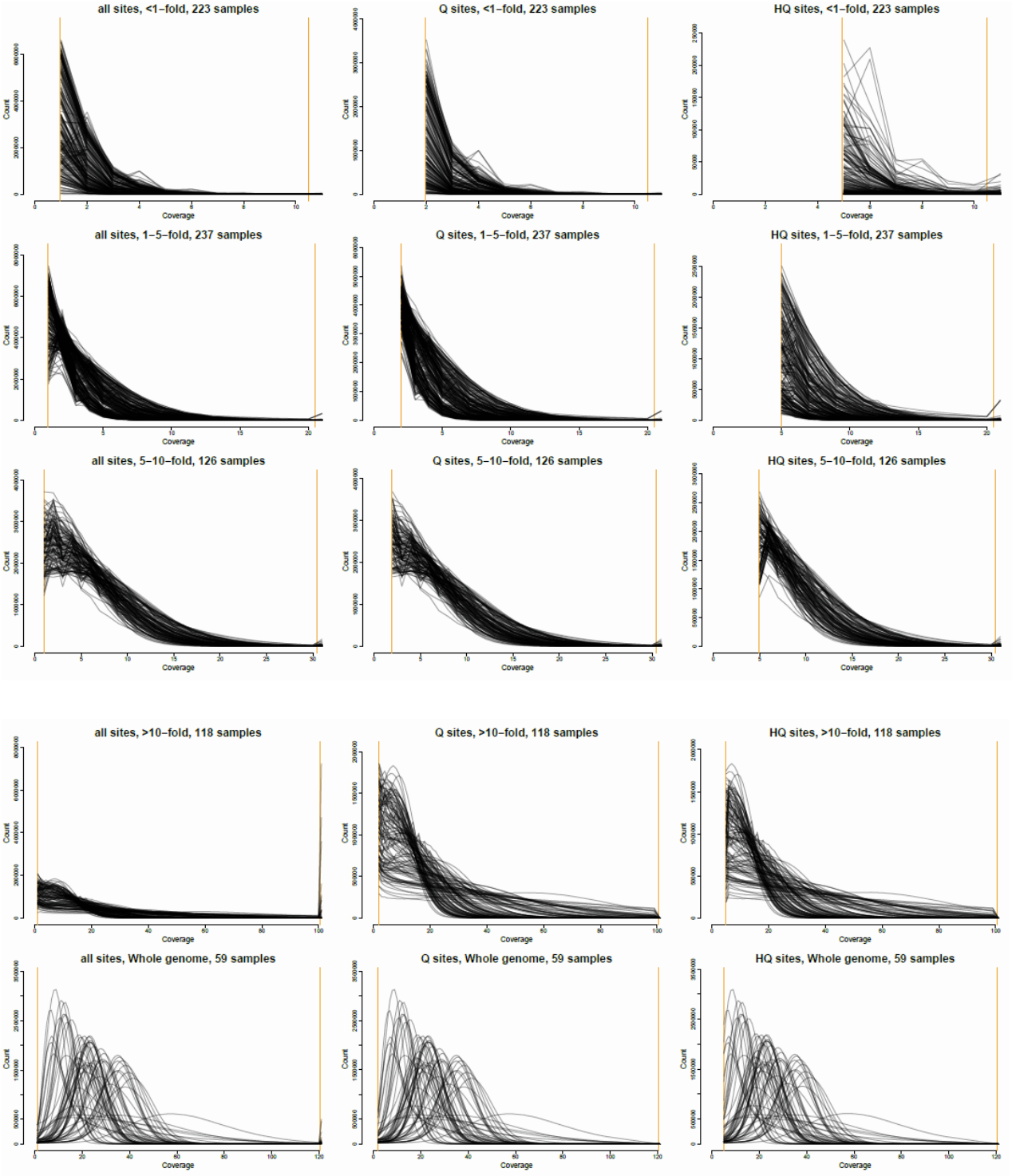
480 **Fig. S26.** Excess of heterozygous and homozygous alternative sites in samples defined as outliers in the initial PCA. Top panel: Heterozygosity per 1,000 bp. Bottom panel: Proportion of homozygous alternative sites among all sites. Color code: green for central chimpanzees, orange for eastern chimpanzees, red for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.



485

Fig. S27. Heterozygous and homozygous alternative sites in kept samples (see text for definition). Top panel: Heterozygosity per 1,000 bp. Bottom panel: Proportion of homozygous alternative sites among all sites. Color code: green for central chimpanzees, orange for eastern chimpanzees, red for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

490 We further examined the coverage distribution in the target space for the samples used in this study. The coverage distribution is skewed towards lower coverages (Fig. S28), as expected for data of average low coverage. Only in samples with an average sequencing coverage of more than 10-fold, we start to see a truncated normal distribution. These patterns are typical for capture data [30,31]. The filtering for quality and high-quality sites largely removes sites at low coverage, while the overall shape of the 495 distribution remains the same. In comparison, the coverage distribution for whole-genome sequencing data does show a normal distribution, which is cut off at low coverage after filtering.



500 **Fig. S28.** Coverage distribution for samples at different coverage levels. Rows show the distribution for samples below 1-fold, between 1-fold and 5-fold, between 5-fold and 10-fold, and above 10-fold average coverage, as well as data from whole genomes. Columns show the distribution before and after quality filtering. The first vertical line represents the lower cutoff applied, the second line represents a cutoff beyond which observations were accumulated into one bin. Related to STAR methods.

505

Note 4. Novel variant discovery.

In order to estimate the amount of variation discovered with new samples, we determined segregating sites (not fixed for the reference or alternative state) within the target space across individuals. Previous studies provided an overall picture of variation across the four chimpanzee subspecies [21,23]. However, the rate at which new variation is discovered does not reach a plateau when randomly sampling individuals (Fig. S29). When considering all sites regardless of quality (see Supplementary Text Note 3.7), we find 1,122,530 variants (including differences to the human reference) across the 59 individuals. When adding all new individuals from this study that were not determined to be another species than chimpanzee and for which at least 10 million positions carry information (723), we find 5,011,732 variants (Fig. S29), largely driven by individuals with extremely high numbers of homozygous alternatives alleles, and observed as outliers in the initial PCA analysis (see Supplementary Text Note 3.3 and Table S1). We then restrict the analysis to the 468 most reliable individuals that do not show an excess of homozygous alternative alleles (ratio of alternative to reference calls < 0.013), that are not outliers in the PCA (see Supplementary Text Note 3.3), have less than 1% human contamination (Table S1), and for which at least 10 million positions carry information. This results in a discovery of a total of 2,328,613 variants (207% more), with an almost linear increase in discovered positions, which does not continue a flattening trend of the high-coverage samples. We conclude that errors and biases contribute largely to this trend (see Supplementary Text Note 3.7), and apply further filtering to the data, as described previously (Supplementary Text Note 3.7).

When using sites that fulfill quality criteria, we find 1,050,120 variants across the previously studied 59 individuals, and 1,585,194 variants when considering the 414 reliable individuals with at least 7.5 million observed sites additional to the conditions stated above, considering that less data is available per individual. This novel discovery rate of 50.9% is likely a reasonable representation of the extent of new variants found by sequencing this number of individuals (Fig. S29). The novel discovery rate per individual is flattening (0-0.16% for the last 20 individuals, Fig. S29). We conclude that when sequencing this number of individuals, a saturation in the discovery of new variation is approached, although not fully reaching a plateau phase. However, only chromosome 21 is considered here, and due to the patchy distribution of the capture data many sites are not covered in all individuals.

Finally, when considering only high quality sites (see Supplementary Text Note 3.7), we find an increase of 10.6% from 1,001,898 to 1,108,367 high quality sites for the 227 reliable individuals with at least 5 million sites (Fig. S29), which is most likely an underestimate due to strict filtering.

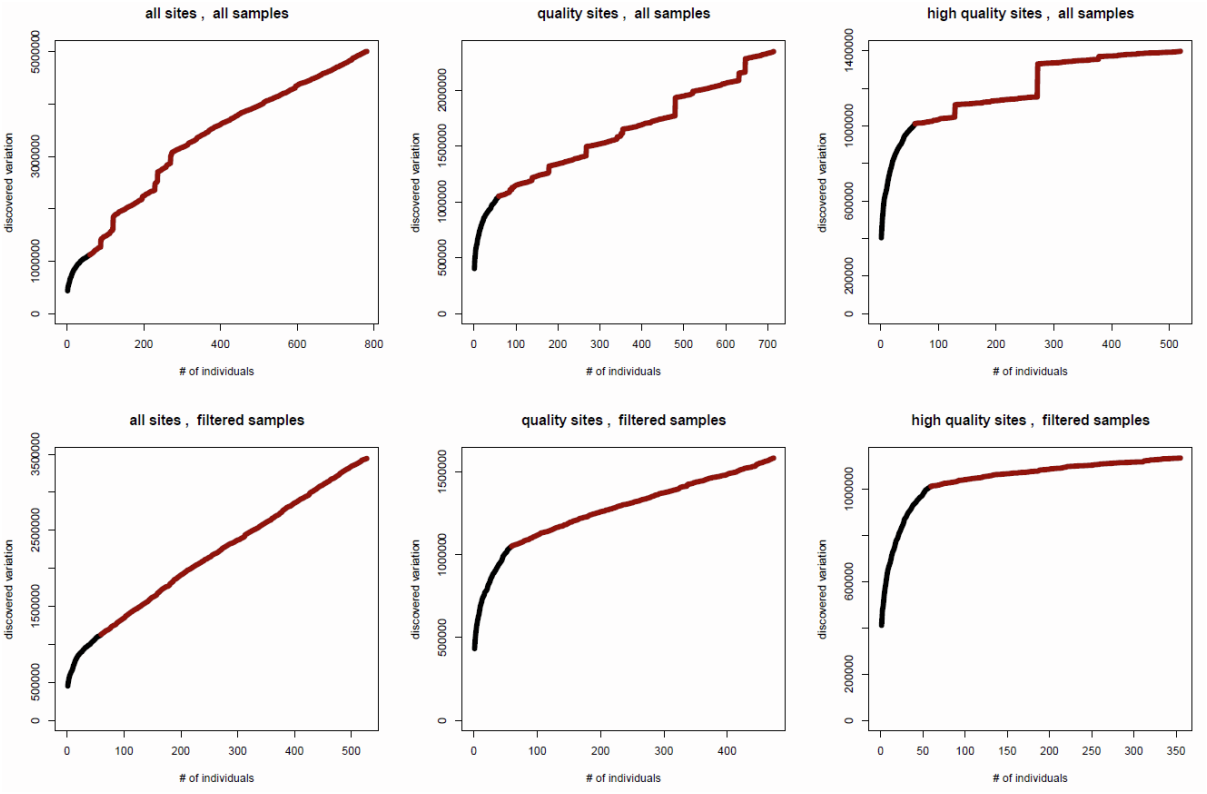


Fig. S29. Novel discovery rate under different conditions. Black line: Cumulative number of variants discovered on chromosome 21 in 59 high coverage genomes. Red line: Cumulative novel variant discovery with the PanAf dataset. Top row: All chimpanzee samples at increasing filtering strength. Bottom row: Filtered samples (non-outliers in PCA, low levels of contamination), at increasing filtering strength. Related to Fig. 1.

Note 5. Population Structure.

5.1. Principal Component Analysis and Procrustes analysis

For the subset of samples that passed PCA, contamination, coverage and relatedness filters (Table S1), we performed a PCA for each subspecies. In the case of eastern (N=120) and central (N=70) chimpanzee subspecies, one sample in each subspecies was not clustering together with the other samples of the same site. For eastern chimpanzees, this was Uga2-81, falling closer to Ngogo instead of the Issa Valley field site, and for central chimpanzees, this was CMNP1-24, falling close to Conkouati instead of other samples from Campo Ma'an. Note that here, we included other samples with lower coverage since there was only one sample with > 0.5-fold coverage, and thus we could not discern which one was the outlier (Fig. S30A, B).

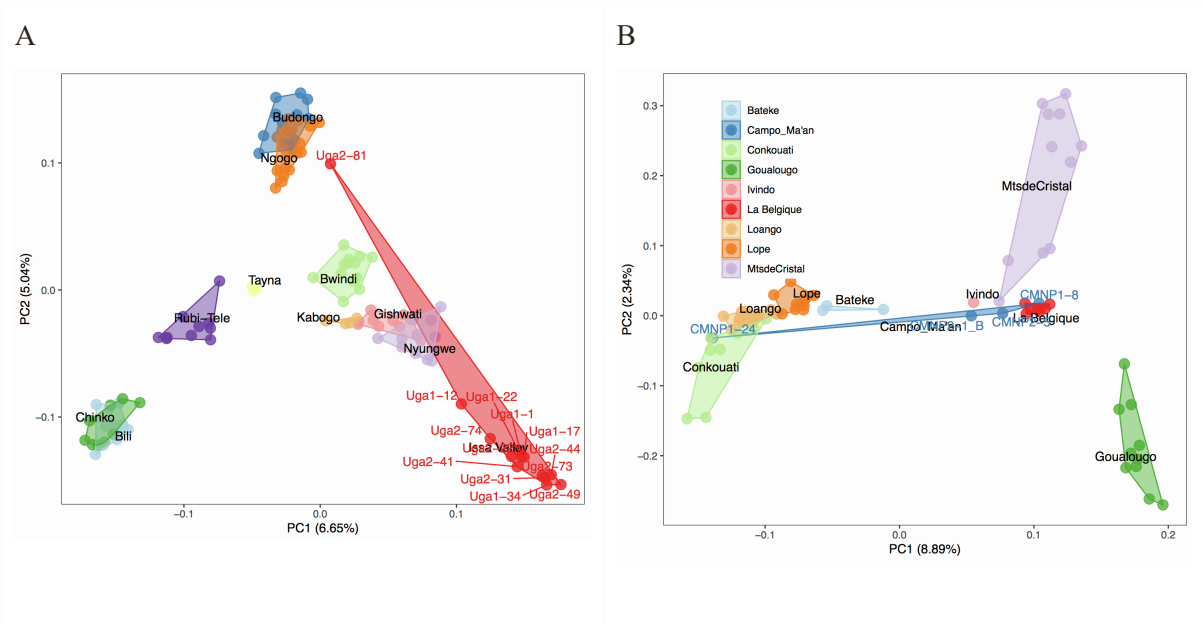


Fig. S30. Discovery of outliers in eastern chimpanzee (**A**) with Uga2-81 not falling within Issa Valley samples; and central chimpanzees (**B**) CMNP1-24 not falling within Campo Ma'an samples. CMNP2-1_B and CMNP2-5 were not previously included in the analysis due to coverage lower than 0.5-fold. However, here they are used as a benchmark to ascertain which of the two Campo Ma'an samples with > 0.5-fold coverage is the outlier. Related to Fig. 1.

We rerun the analysis without those outliers for central (n=69) and eastern (n=119), but also the other two chimpanzee subspecies, westerns (N=227) and Nigeria-Cameroon (n=34) chimpanzees (Fig. S31).

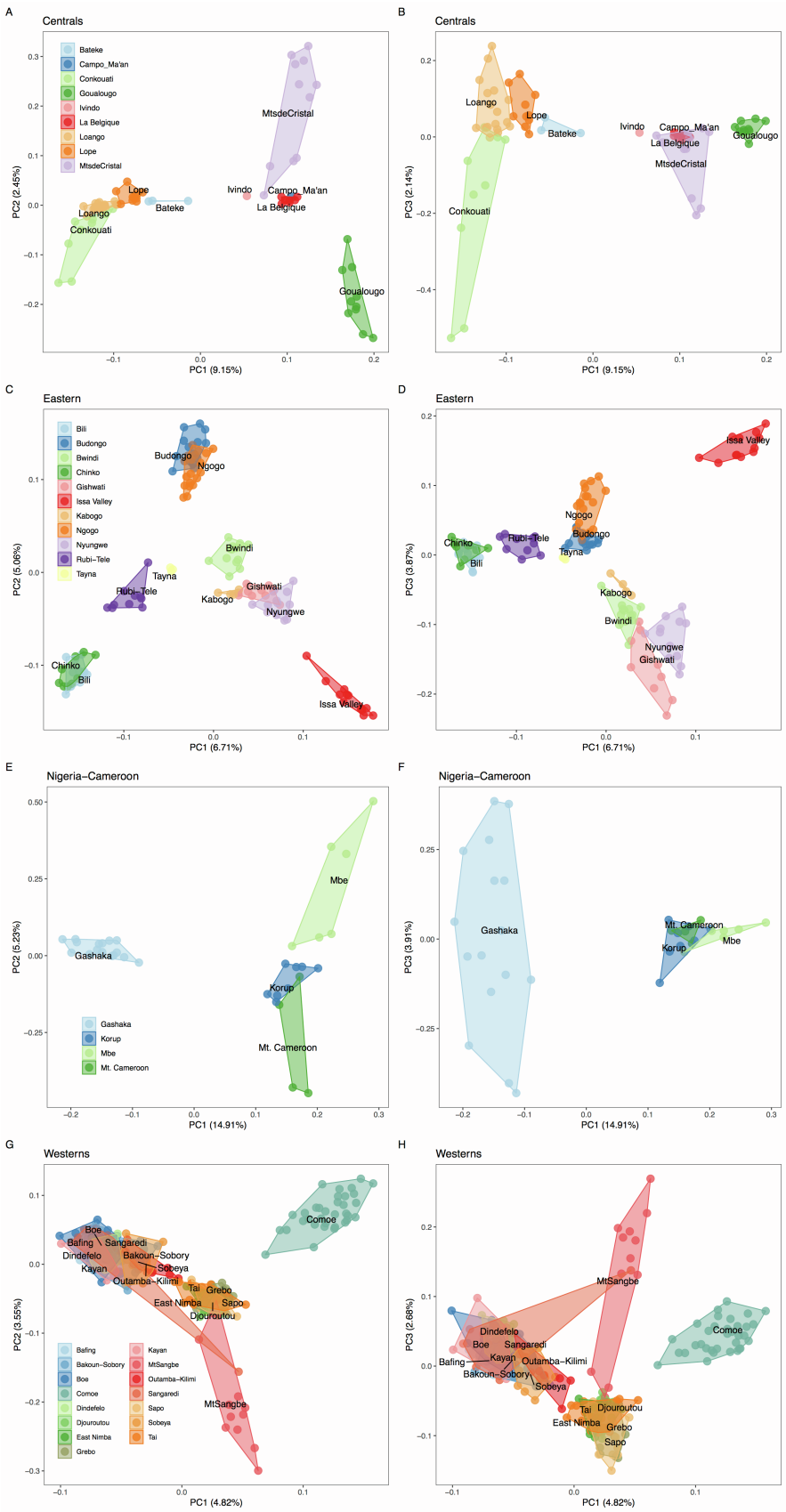


Fig. S31. Principal component analysis (PCA) of chromosome 21 in each chimpanzee subspecies (PC1 vs PC2 and PC1 vs PC3). (A) and (B), Central chimpanzees (N=69); (C) and (D), Eastern chimpanzees (N=119); (E) and (F), Nigeria-Cameroon Chimpanzees (N=34); (G) and (H), Western chimpanzees (N=227). Related to Fig. 1.

For all subspecies, we observe population stratification at geographical sites, more clearly seen in central, eastern and Nigeria-Cameroon chimpanzees. In western chimpanzees, due to our extremely dense sampling we do not see separation by site, however we do see a cline of variation, where PC1 approximately reflects the geographic distribution from east to west, while PC2 approximately reflects a north-to-south distribution. The Comoé sites and MtSangbé are clearly separated from the other sites (Fig. S31G).

Previous studies have shown a correlation between geography and genetic variation [21,32]. In order to test this hypothesis with our data we used Procrustes transformation to allow the comparison of principal components (PC1 and PC2) with the GPS coordinates. To perform the transformation we used MCMCpack [33] in R (version 3.5.2) (Fig. S32).

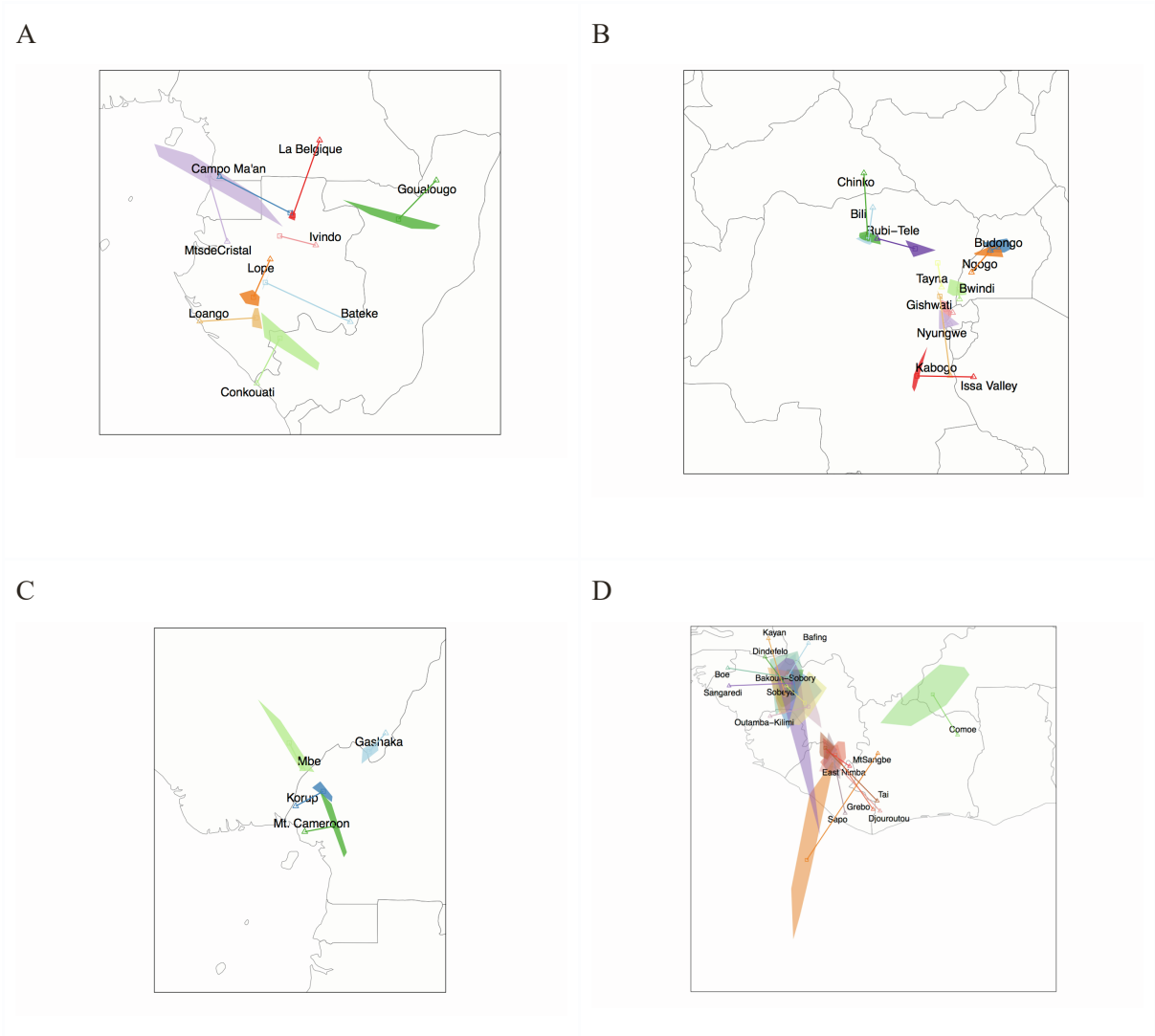


Fig. S32. Procrustes transformation of PC1 and PC2 for (A) central, (B) eastern, (C) Nigeria-Cameroon and (D) western chimpanzees. Triangles represent the geographical coordinates of sites, squares represent the centroid of the PCA coordinates of each site. Polygons enclose all points in the PCA belonging to each site. Related to Fig. 1.

One of the main questions still to be fully uncovered is the relationship between central and eastern chimpanzees. The majority of georeferenced samples of central and eastern chimpanzees from which genomic data has been obtained were not continuously sampled across their extant distribution range [21]. In fact, there was a big sampling gap between central and eastern chimpanzees, with eastern chimpanzees sampled mainly from north, south and east Democratic Republic of Congo (DRC), with no samples originating from the western distribution of eastern chimpanzees. Therefore, the question whether genetic diversity indeed reflects two distinctly separated subspecies, or rather a cline of variation, remained open. In the PanAf dataset we sampled feces from Ngiri, an eastern chimpanzee site from west DRC, at the border between the subspecies separation of central and eastern chimpanzees. Ngiri site is geographically closer to a central chimpanzee site (Goualougo, 281.5 km) than to other eastern chimpanzee sites (Rubi-Télé, 843.6 km and Bili, 898.9 km). These samples were not included in previous analyses due to their low coverage (less than 0.5-fold) and slightly higher human contamination levels. However, three of those samples had a coverage close to 0.5-fold, thus, we used them to test this specific hypothesis (Table S1). We performed PCA and Procrustes analysis on eastern chimpanzees alone, but adding also central chimpanzees to determine if Ngiri genetic diversity would fall between central and eastern chimpanzee sites, or within eastern chimpanzee diversity.

Our analysis clearly indicates that Ngiri chimpanzees fall within eastern genetic diversity (Fig. 2B of the main text and Fig. S33A, B), and not between the central and eastern chimpanzee subspecies clusters, pointing to a clear separation of both subspecies for an extended period of time. From this analysis, we find no particular affinity of this population to central chimpanzees compared to other eastern chimpanzees (Fig. 33C, D).

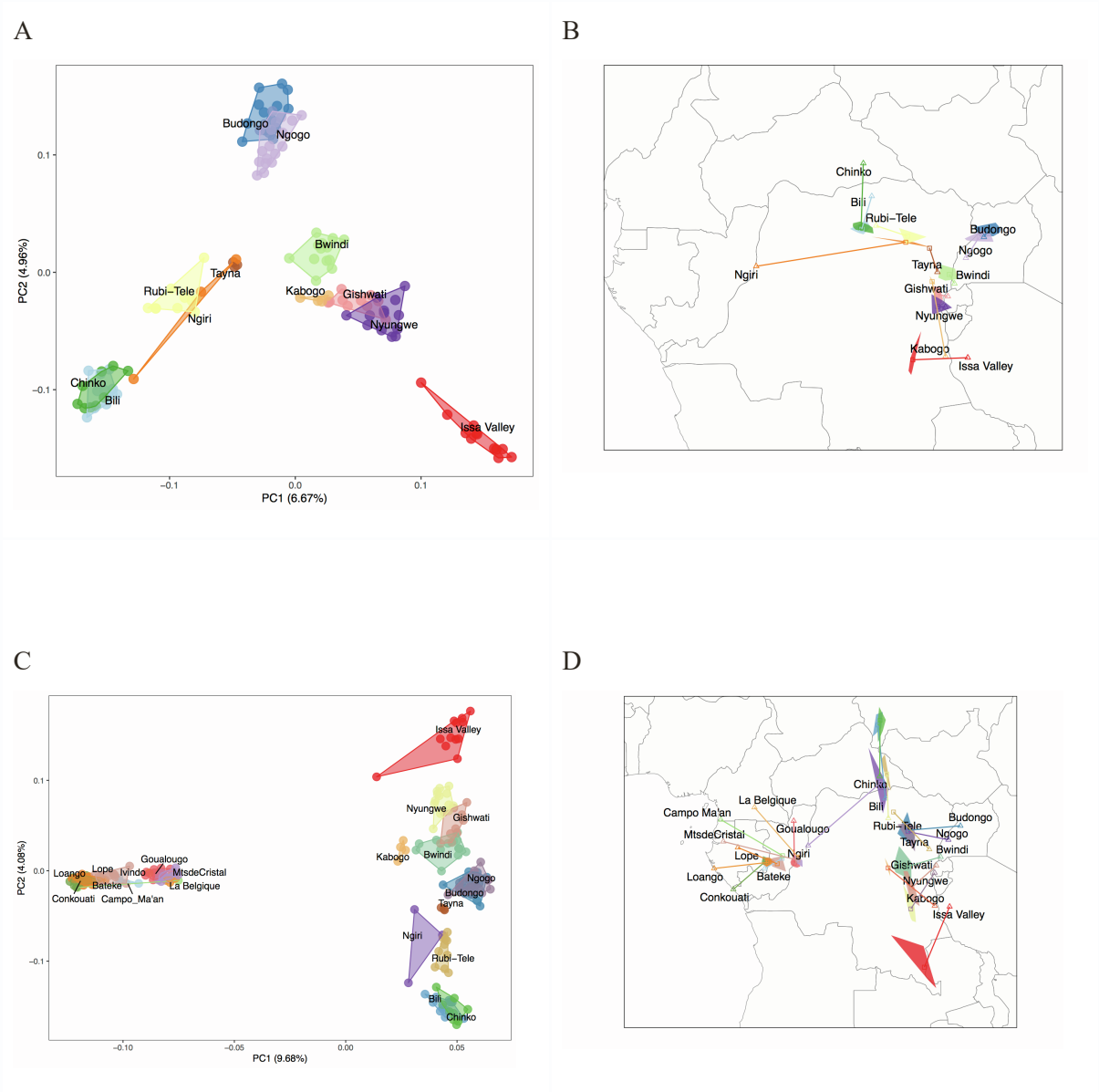
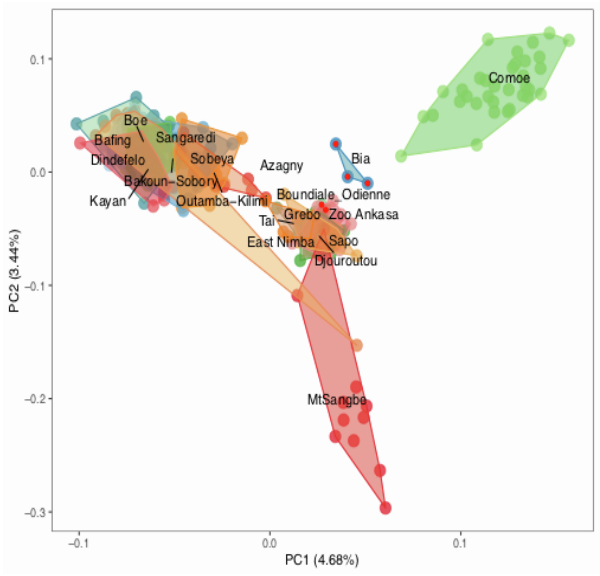


Fig. S33. PCA and Procrustes analysis of eastern (A and B), and eastern and central chimpanzees (C and D). Related to Fig. 1.

For western chimpanzees, we also added other samples to the analysis, since they were collected in regions where the presence of chimpanzee communities is minimal and on the periphery of the distribution (Boundialé-Odienne and Bia). Even though these samples had a coverage lower than 0.5-fold, with slightly higher human contamination, we retained enough data to perform PCA analysis, and determine where would they fall within the diversity of western chimpanzees (Boundialé-Odienne: Bou1-1; Bia: Gha-01-04, Gha-01-05, Gha-01-06 and Gha-01-11, the later coming from Kumasi Zoological Gardens (Ghana) but reported to be the last chimpanzee from the Ankasa Conservation Area, Ghana (Table S1). In a PCA analysis, the Bia samples fall close to the Comoé and southern western sites (Fig. 34), while the Ankasa and Boundiae Odienne samples fall directly within southern western sites.



615 **Fig. S34.** Western chimpanzee PCA, with samples from Bia, Ankasa and Boundialé-Odienne highlighted in red. Related to Fig. 1.

5.2. Ancestry components

From ANGSD genotype likelihoods we computed the admixture model of the STRUCTURE software using NGSadmix [34], with all samples passing filters, setting ancestral populations (K) from 2 620 to 10 with 50 replicates for each K, -minInd 310 and -minMaf 0.05 parameters at a total of 92,731 genotypes after filtering. At each K, we kept the most supported population separation within all runs with the highest likelihood for plotting. These results might be biased towards western chimpanzees, the most abundant population in our dataset, since unequal representations of groups are among the various known factors influencing and biasing this model [35], as well as differences in population history, 625 bottlenecks, admixture from other groups and ancient structure, all of which are known to be part of the complex history of chimpanzees[21]. We also caution that often linked loci in this dataset on only one chromosome may contribute to biases in this analysis as well.

Admixture analysis of 449 samples from 21 western sites, 4 Nigeria-Cameroon sites, 9 central sites and 11 eastern sites, shows that the most likely K is 2 (Fig. S35), which features a clear separation 630 of the western chimpanzee population from central and eastern chimpanzees, with Nigeria-Cameroon being a composite of the two components (Fig. S36). At K=3, Nigeria-Cameroon chimpanzees show their own component, while central chimpanzees appear to be composed of mixed ancestries of Nigeria-Cameroon and eastern chimpanzees.

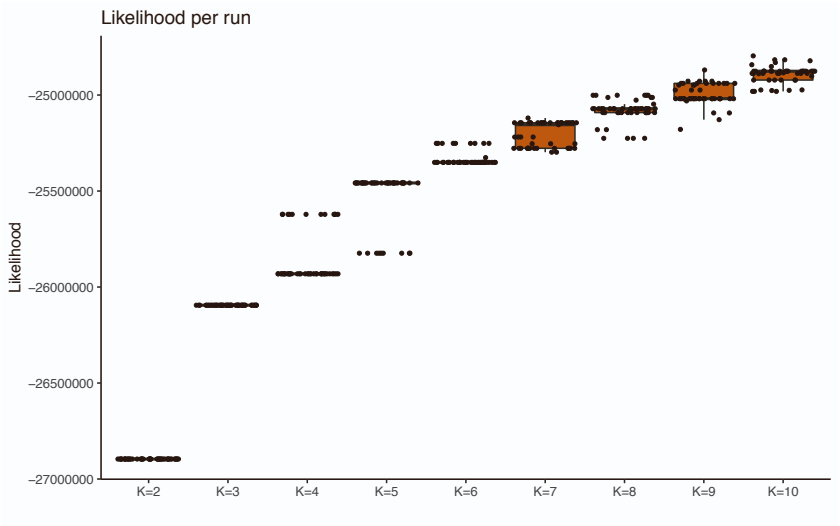


Fig. S35. Likelihood of each run (N=50) at each K. Related to Fig. 1.

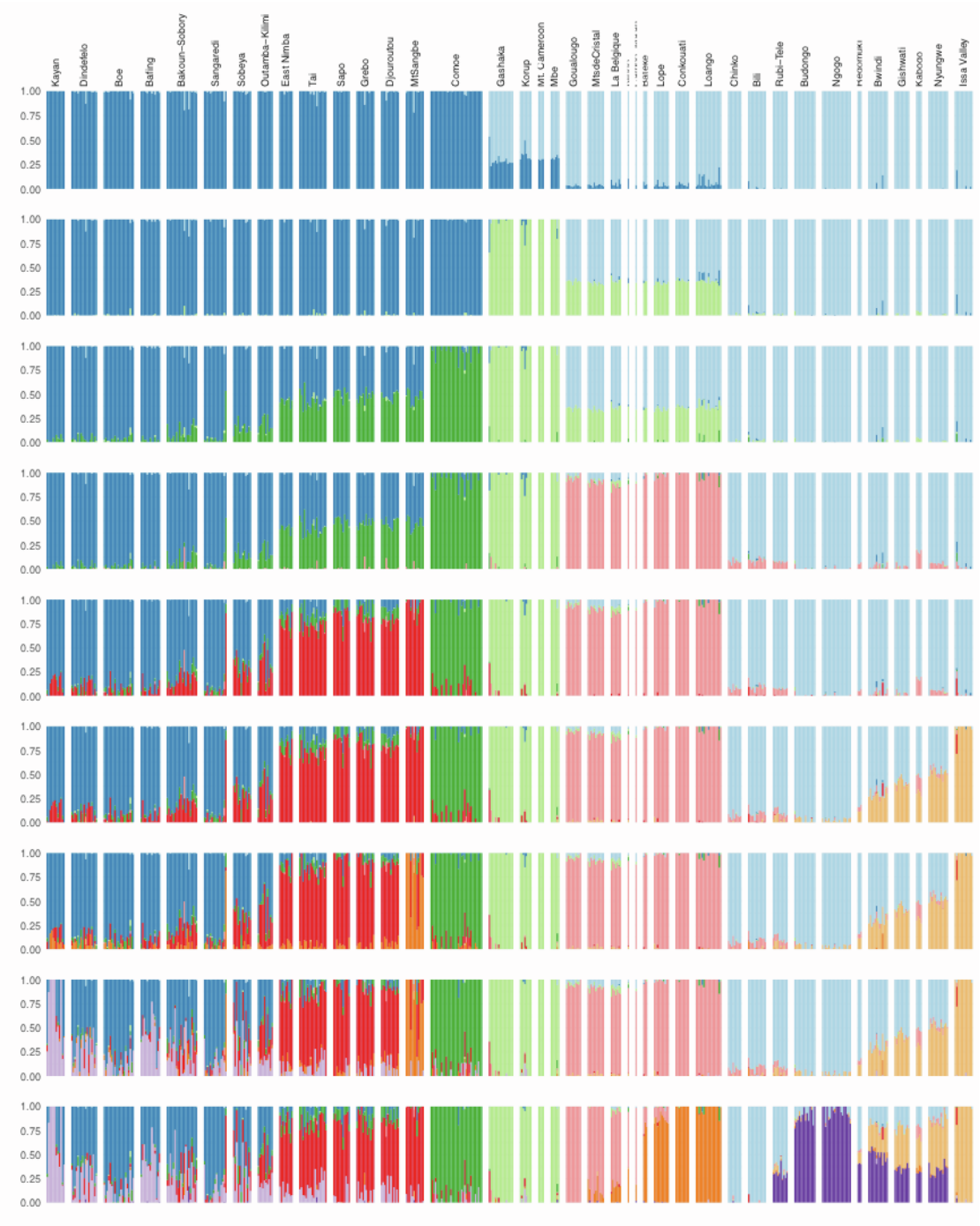
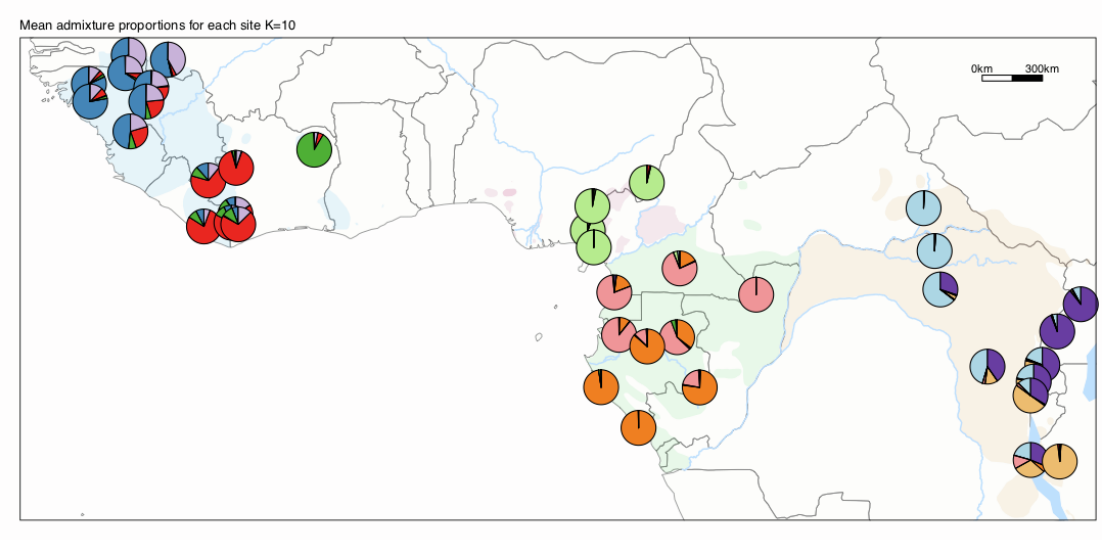


Fig. S36. Ancestry compositions from K=2 to K=10. Related to Fig. 1.

At K=4, due to our larger representation of western chimpanzees, specifically of Comoé, western chimpanzees appear to have two ancestral components, represented by Comoé and the northern clade (Kayan, Dindefelo, Boe, Bafing, Bakoun, Sobory, Sangaredi); Outamba-Kilimi and Sobeya, and at a higher level the southern clade (East Nimba, Mt Sangbé, Tai_Eco, Tai_R, Sapo, Grebo and Djouroutou) appear as a mixture of both ancestral components, while appearing at K=6 as a distinct ancestral component of the southern clade, following a geographical pattern. Central chimpanzees are separated by their own ancestral component at K=5.

At higher Ks, specific sites gain ancestral components, separating these clusters such as Mt Sangbé at K=7, and Issa Valley at K=8. At K=10; central chimpanzees are separated into two components, the northern sites (Goualougo, Mts de Cristal, La Belgique, Ivindo, Campo Ma'an) and the southern sites (Batéké, Lopé, Conkouati and Loango) (Fig. S37).



650

Fig. S37. Mean ancestry proportions for each site at K=10, to see regional clustering. Subspecies distribution color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 1.

Note 6. Population demography.

6.1. Heterozygosity

To assess global levels of heterozygosity, the unfolded SFS was calculated for each sample separately using ANGSD and realSFS with the following quality filter parameters: -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 20 -minQ 20 -setMaxDepth 200 -doCounts 1 -GL 1 -doSaf 1 -r chr21:. We used the human genome (Hg19) to determine the ancestral state. Our results indicate that western chimpanzees have the lowest heterozygosity values, in line with previously published data (Fig. S38) [21,23].

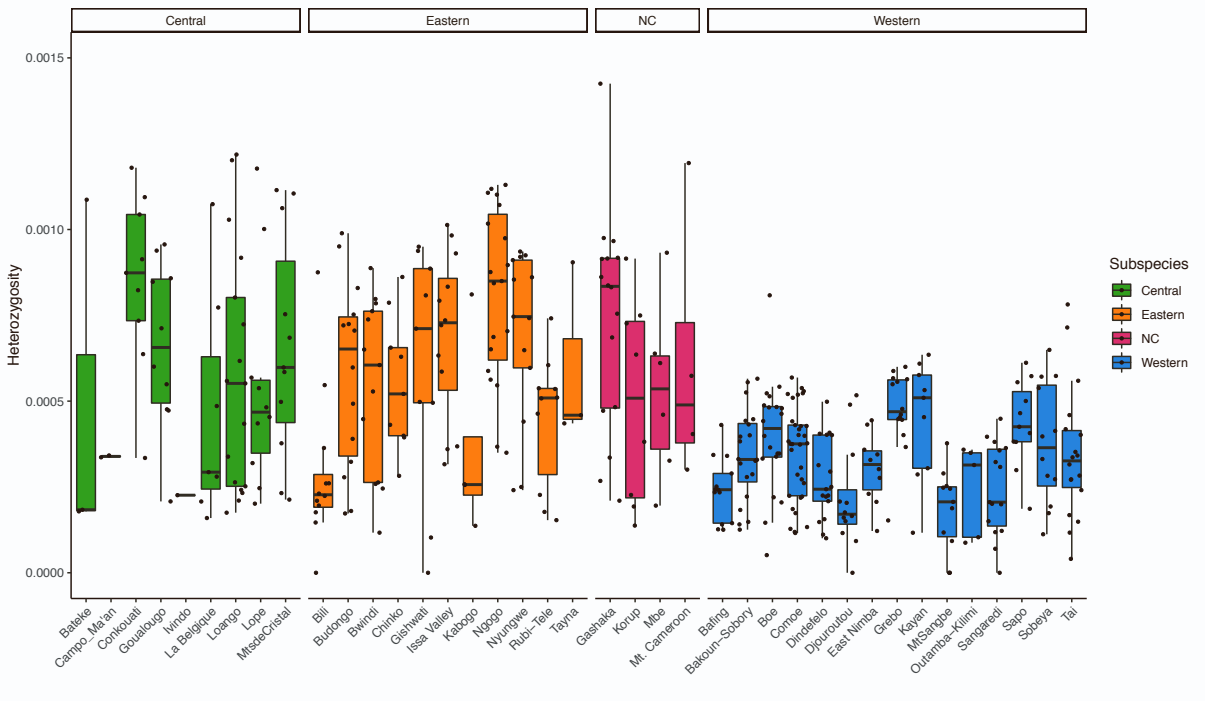


Fig. S38. Heterozygosity estimates per sampling site using genotype likelihoods. The lower and upper hinges of the boxplot correspond to the first and third quartiles (the 25th and 75th percentiles) and the whisker extends from the hinge to the largest and lowest value no further than 1.5 * IQR (inter-quartile range). Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 1.

We also computed heterozygosity estimates from SNP data at different coverage cutoffs, as well as from the chr21 target region from high-coverage whole genomes[21]. As expected, average coverage influences the correctness in the detection of heterozygous sites, and the higher the coverage the higher the global heterozygosity obtained. With captured fecal samples we cannot fully recover the global heterozygosity, but the trend differentiating the four chimpanzee subspecies (central more than eastern/Nigeria-Cameroon more than western) is largely maintained. Therefore, even though the

estimates are not identical, conclusions about relative comparisons are correct (Fig. 1E from the main text and Fig. S39).

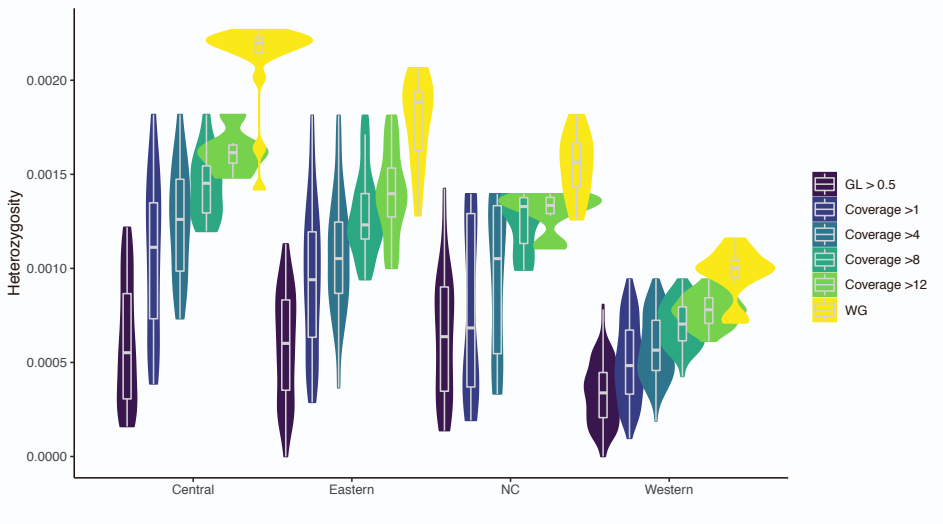


Fig. S39. Heterozygosity estimates per subspecies. Colors represent different datasets, GL (genotype likelihoods) of samples > 0.5-fold coverage, then different coverage thresholds of SNPs from PanAf dataset and WG (Whole Genome) from de Manuel *et al*., on the target region of chr21. The lower and upper hinges of the boxplot correspond to the first and third quartiles (the 25th and 75th percentiles) and the whisker extends from the hinge to the largest and lowest value no further than 1.5 * IQR (inter-quartile range). Related to Fig. 1.

6.2. Regions of Homozygosity

Regions of homozygosity (RoHs) are an informative pattern for population history as well as recent inbreeding in populations [36], and even the limited amount of data on chromosome 21 can be used for such a purpose[31]. We calculated regions of homozygosity, by calculating the distance between consecutive heterozygous sites on chromosome 21, irrespective of the discontinuous nature of the data. We restricted the analysis to the informative subset of samples with more than 5-fold coverage, and where more than 50% of the sites in the target space were observed after quality filtering for snpAD genotype calls (Table S1). We defined short RoHs as those between 10 and 100 kbp, and long RoHs as those longer than 100 kbp.

We find that central chimpanzees carry the smallest cumulative amount of short RoHs (median 5,317 Mbp per individual), western chimpanzees the largest amount (median 13,654 Mbp), and the other subspecies intermediate amounts (median eastern chimpanzees: 9,415 Mbp, median Nigeria-Cameroon chimpanzees: 8,807 Mbp). These distributions are significantly different ($p < 0.001$, two-sided Wilcoxon rank test, Benjamini-Hochberg corrected) between most of the subspecies, except for eastern and Nigeria-Cameroon chimpanzees ($p > 0.05$, two-sided Wilcoxon rank test, Benjamini-Hochberg corrected). These observations (Fig. S40, x-axis) on short RoHs, which are a result of long-term population size correspond very well with the observed differences in overall heterozygosity, and the known long-term differences in population history of these populations[21].

Longer RoHs are informative of more recent inbreeding events. We find no systematic difference between the four subspecies regarding observations of longer RoHs (median central chimpanzee: 254 kbp; median eastern chimpanzee: 397 kbp; median Nigeria-Cameroon chimpanzee: 244 kbp; median western chimpanzees: 379 kbp). None of the subspecies differs significantly from another ($p > 0.05$, two-sided Wilcoxon rank test, Benjamini-Hochberg corrected). However, we find single individuals across the subspecies to carry an excess of long RoHs, suggestive of recent inbreeding within the sites (Fig. S40, y-axis). For example, in the case of the site of Mbe in Nigeria-Cameroon chimpanzees, three out of four tested individuals from this site have the largest amount of long RoHs in this subspecies (cumulative length of more than 900 kbp), increasing the total cumulative homozygous region length of two of these individuals to be among the largest among chimpanzees in this study (Fig. S41). Among central chimpanzees, one out of two individuals from Bili shows a signature of inbreeding, as well as, to a smaller extent, two out of 15 individuals from Issa Valley and three out of four individuals from Chinko. Possibly, some sites become recently more isolated than others, causing an increasing risk of inbreeding. In central chimpanzees, one individual from Goualougo might have experienced a recent history of inbreeding. In western chimpanzees, we find no specific site where long RoHs seem to be enriched, and no individuals with a very strong excess of long RoHs. One individual from Dindefelo and several individuals from Comoé appear to have a slight increase in long RoHs, in addition to a relatively large amount of short RoHs (cumulative length > 15 Mbp), which might be the consequence of elevated relatedness during the past few generations. Contrary to previous studies suggesting inbreeding based on STRs[37], the single sample from the relatively isolated site of Mt Sangbé does not show outstanding amounts of either short or long RoHs, although the total sum of both types ranges among the largest in our dataset.

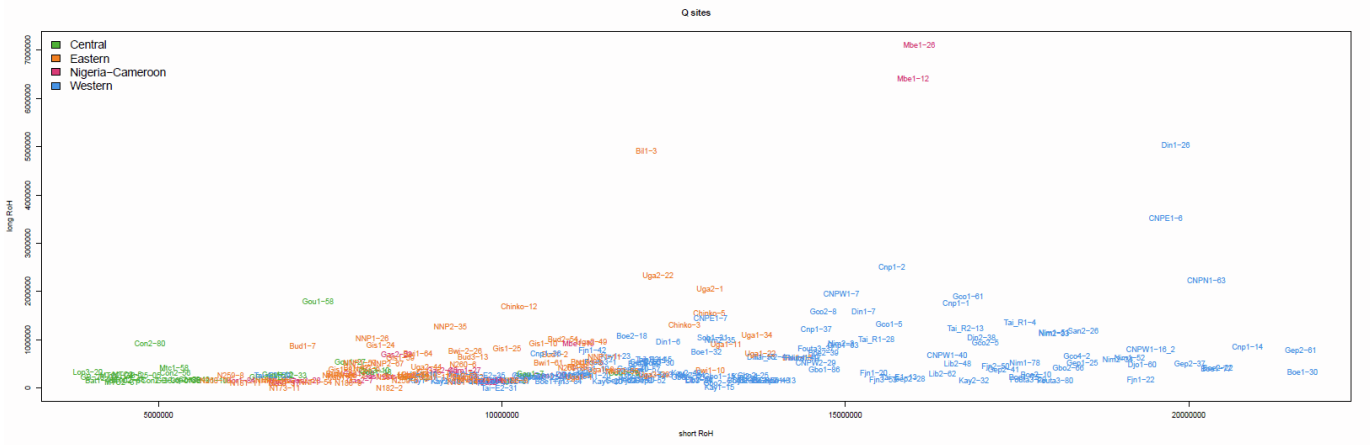


Fig. S40. Short (10-100 kbp, x-axis) and long (>100 kbp, y-axis) RoHs per individual, color-coded by subspecies (green = central, orange = eastern, red = Nigeria-Cameroon, blue = western chimpanzee). Related to Fig. 2.

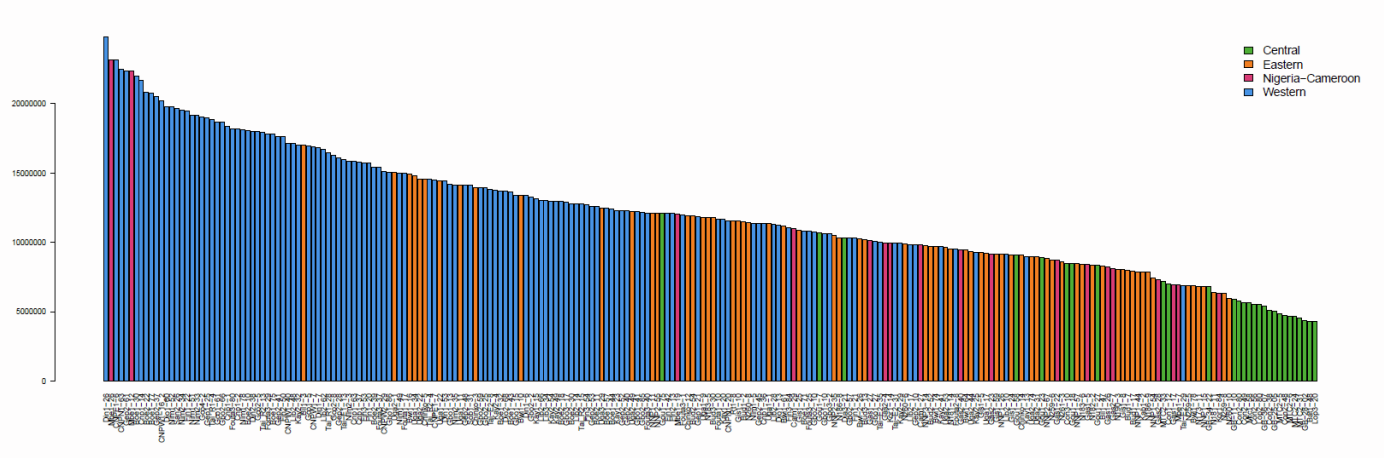


Fig. S41. Sum of cumulative short and long homozygous regions per individual. Related to Fig. 2.

When comparing these individuals to previously published data from whole genomes [21], we find the same pattern of increasing cumulative lengths of short RoHs in lower effective population size in the four populations, and individual outliers in long RoHs across subspecies. Here, the strongest signature of inbreeding is found in the individual central chimpanzee B025_Marlin, for which no data on its origin are available, but which is likely from central Gabon (Note 9) (Fig. S42).

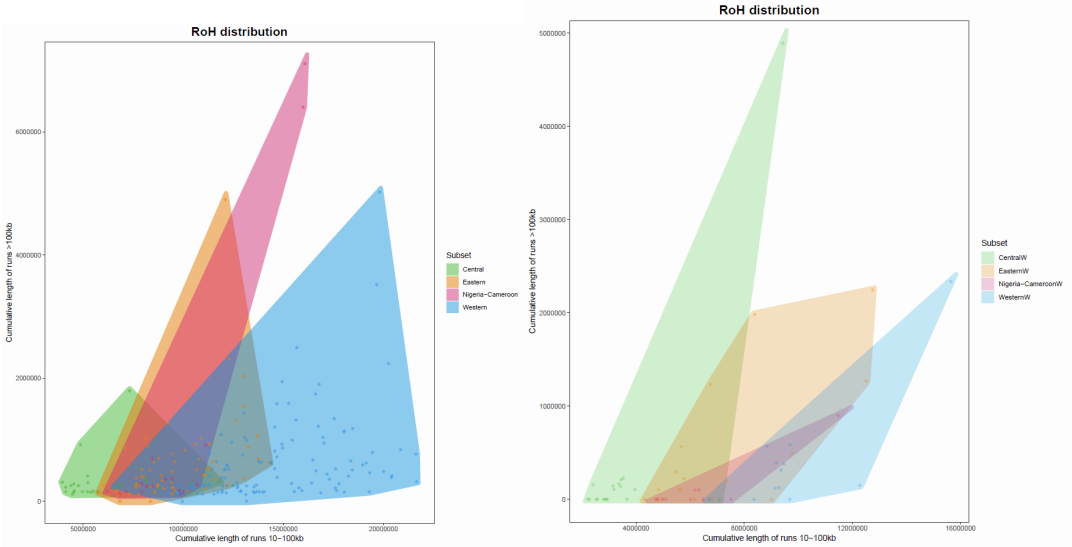


Fig. S42. Summary of RoHs on chromosome 21 in data from whole genomes (right) and captured individuals (left), color-coded by subspecies. Related to Fig. 2.

6.3. Diversity

We calculated the Watterson Estimator of diversity and the Tajima's D neutrality index for each geographical site based on Site Frequency Spectrum (SFS) estimates. SFS was estimated for all geographical sites. We used the ANGSD option -doSaf 1 and then realSFS to obtain the unfolded SFS, also implemented in ANGSD [38]. We used the human genome (Hg19) to polarize the ancestral state, and

we applied the following quality filter parameters: -r chr21: -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 20 -minQ 20 -setMaxDepth 200 -doCounts 1 -GL 1. From the calculated SFS estimates with realSFS at each site, we ran the thetaStat software also implemented in ANGSD [39]. In both cases, we estimated those values in 50kb windows with a sliding window size of 25kb, and for the Watterson Estimator we corrected by sample size and number of basepairs observed.

Central chimpanzee sites are the subspecies with the highest diversity, and western chimpanzees with the lowest diversity (Fig. S43). It is worth noticing that we find Nigeria-Cameroon chimpanzees to have higher diversity than eastern chimpanzees, in disagreement with previous studies. This could be explained because Prado-Martinez *et al.* (2013)[23] and de Manuel *et al.* (2016)[21] had limited information of the geographical origin of Nigeria-Cameroon chimpanzees, and we may have sampled other areas and discovered new variation which was previously unexplored. We found that Ngogo harbors more diversity than any other eastern chimpanzee site. Western chimpanzees have the lowest diversity, but interestingly Dindefelo and Taï Eco have a higher diversity than the rest.

For the neutrality test, we find that Centrals and Nigeria-Cameroons have values near 0, while overall eastern chimpanzees have a Tajima's D larger than 0 (standard deviation from -0.06 to 1.49). We also find more alleles at intermediate frequencies in this population (Fig. S43), altogether possibly explained by a recent bottleneck or population contraction. On the other hand, western chimpanzees have Tajima's D values smaller than 0, which could indicate a recent population expansion after a severe bottleneck, in line with previous demographic models[21].

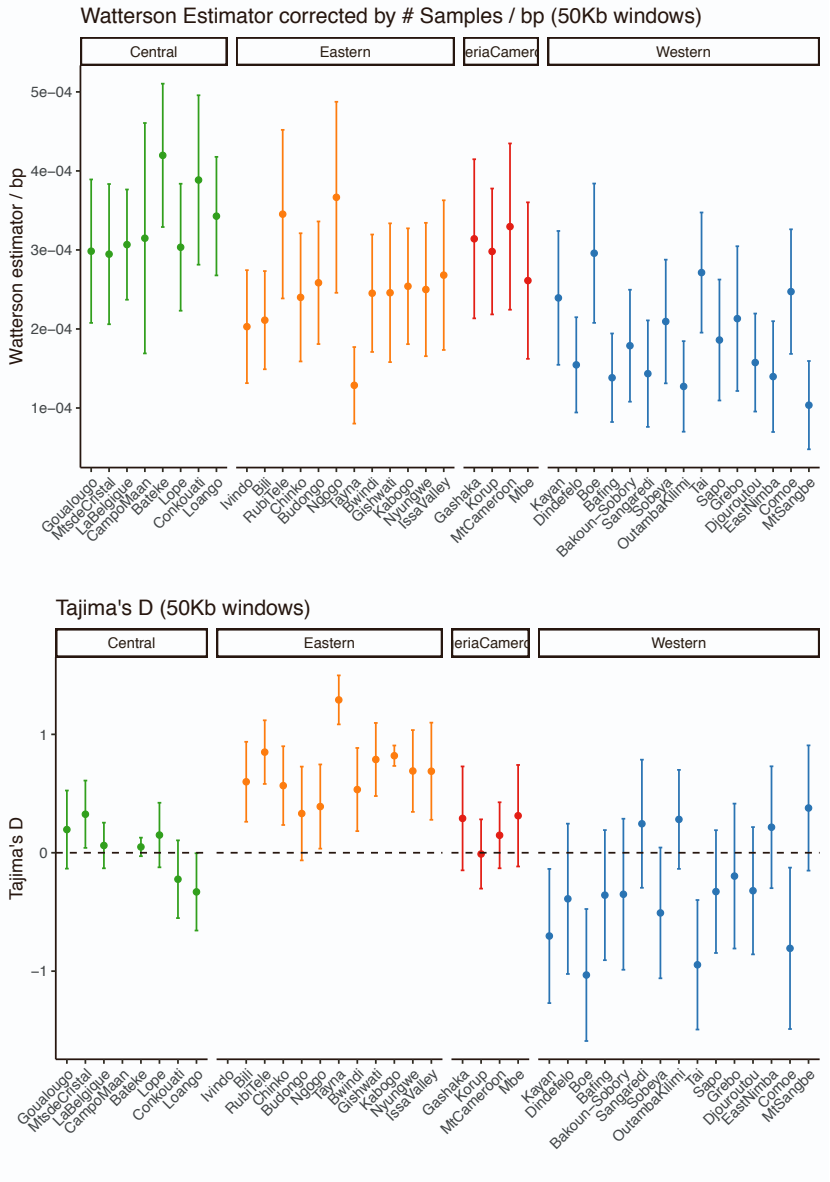


Fig. S43. Diversity and Neutrality at 50kb windows. Watterson estimator corrected by number of samples. Each dot represents the median estimate and the error lines the standard deviation. Since in Ivindo and in Campo Ma'an we had only 1 sample for analysis we could not compute the Tajima's D. Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 2.

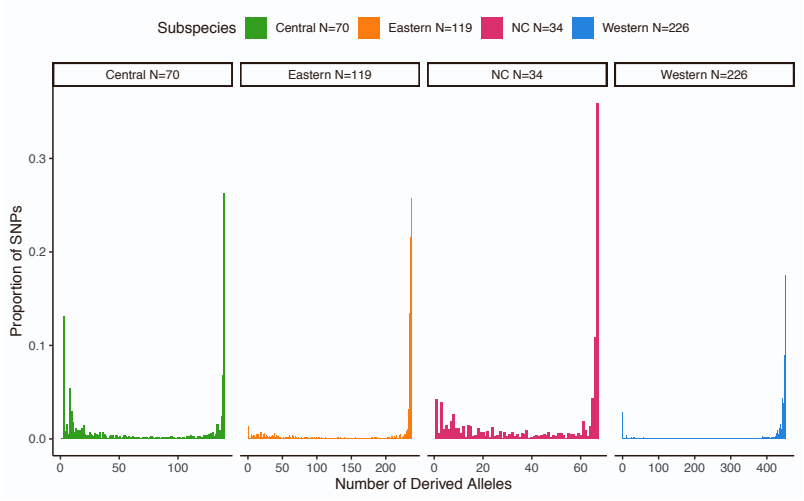
6.4. Site Frequency Spectrum

The Site Frequency Spectrum (SFS) was estimated for all chimpanzee subspecies and geographical sites. We used the ANGSD option -doSaf 1 and then realSFS to obtain the unfolded SFS, also implemented in ANGSD[38]. Since each subspecies had different sample sizes, we estimated first the SFS considering all samples (Fig. S44A), but we also randomly downsampled the same number of individuals per subspecies, dictated by the number of Nigeria-Cameroon samples, the least abundant subspecies (N=34) in our dataset. With this second approach, we did 10 random replicates, sampling different individuals each time, for central, eastern and western chimpanzee subspecies (Fig. S44B).

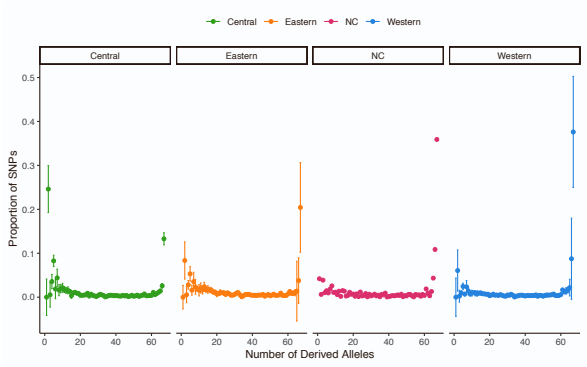
Despite the obvious problems in our data for this approach (sequencing depth, capture bias, high degree of missingness), we were able to make some observations. First, central chimpanzees carry the largest amount of variants at low frequency, as is expected from their demographic history with a large effective population size (*Ne*). Second, western chimpanzee populations carry a larger amount of fixed derived sites, which is the result of increased drift due to repeated bottlenecks. We observe a similar pattern for Nigeria-Cameroon chimpanzees, while eastern chimpanzees show an intermediate pattern. This is clearly in line with previous evidence from high-coverage whole genomes [21,40], where this cline of diversity has been observed from western to central chimpanzees. We find that generally variants at low frequency are highest in central chimpanzees compared to the other subspecies. This is most likely the effect of more neutral variation being observed in populations with large *Ne*, which is subsequently efficiently removed by natural selection, hence not drifting to fixation. This concordance with previous results in chimpanzees [40], demonstrating that the data obtained from chromosome 21 as representative for the population diversity of chimpanzees. We note that there is a drop at singletons across all comparisons, due to low coverage in most of the samples, and consequently low power to detect singletons in a heterozygous state.

Geographical site SFS was estimated with all available samples at each location (Fig. S44D). We do not observe any population with a peak at medium allele frequencies, which would be a pattern of a mix of two subspecies or extreme substructure within a geographic location. We also do not find any population to be deviating in terms of an extreme excess compared to other sites of the same subspecies. However, the variability in sample size per location is large, prohibiting a more fine-grained interpretation.

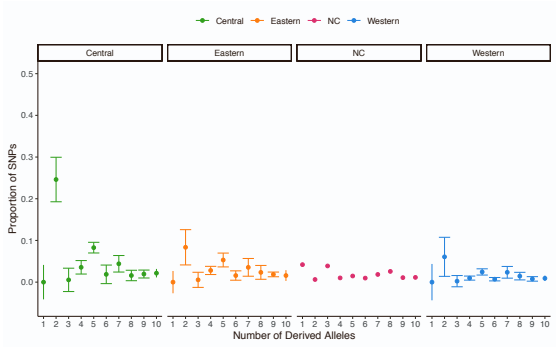
A



B



C



D

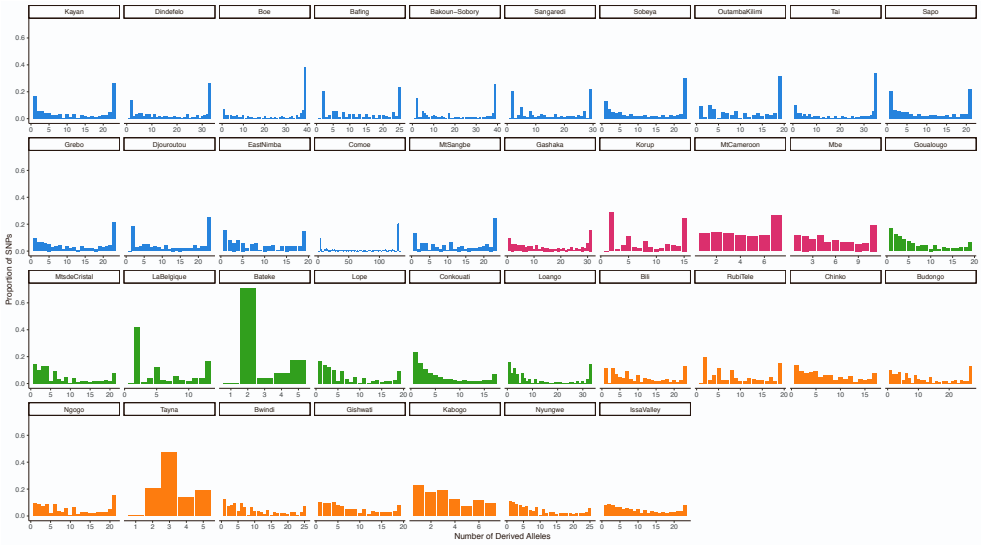


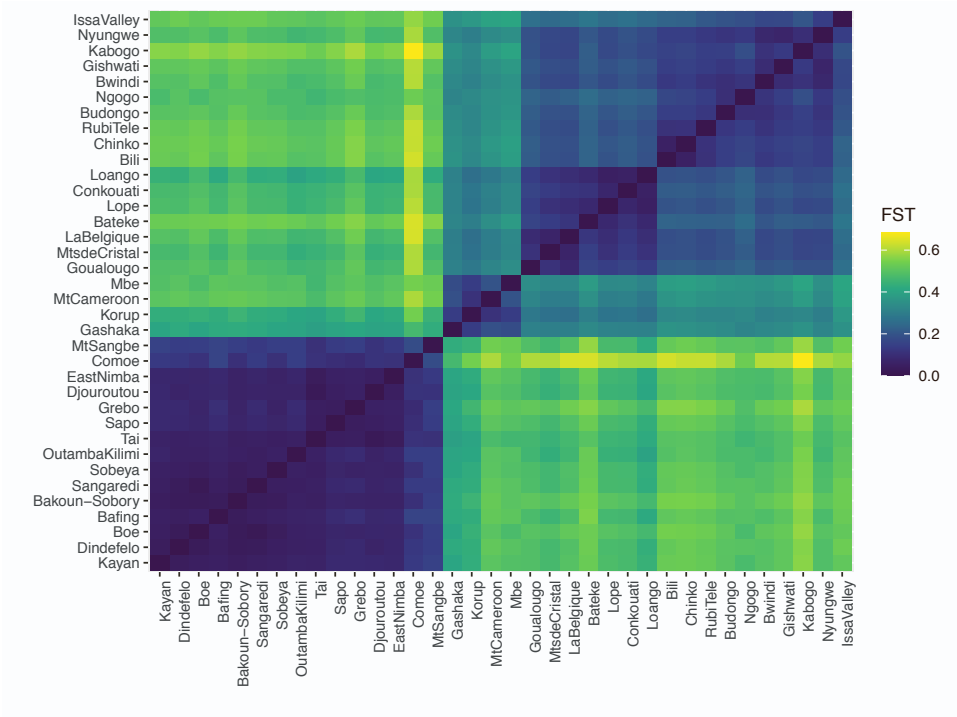
Fig. S44. Site Frequency Spectrum (SFS). (**A**) By subspecies using all samples, (**B**) by subspecies downsampling central, eastern and western chimpanzees to 34 samples and plotting the median and the standard deviation of the 10 replicates, (**C**) the same as B but zoom in at the first 10 alleles, and (**D**) by sampling site using all available samples. Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 2.

6.5. Genomic differentiation

To obtain pairwise F_{ST} estimates between geographical sites, we computed the 2-dimensional SFS (2d SFS) between each pair of geographical sites with ANGSD -doSaf 1 and realSFS (described in Supplementary Text Note 4). The genetic relationships between populations were used to build a matrix, from which we constructed a neighbour-joining tree using the ape package [41] in R (version 3.5.2).

Sites from the same subspecies cluster together (Fig. S45). Western chimpanzees form three clades: Comoé sites, a Southern clade (Taï Eco, Taï R, Sapo, Grebo, East Nimba, Mt Sangbé and Djouroutou) and a Northern clade (Outamba Kilimi, Bafing, Sobeya, Sobory, Bakoun, Kayan, Sangaredi). For Nigeria-Cameroon chimpanzees, Gashaka separates first, while Mbe and Mt Cameroon are very similar. Central chimpanzees form two groups: a northern one (Goulongo, Mts de Cristal and La Belgique) and a southern one (Conkouati, Lopé, Batéké and Loango). Since we only had one sample each from Campo Ma'an and Ivindo, and only three for Tayna (two of them are 2nd degree relatives) they were not included in this analysis. For eastern chimpanzees, we observe Issa Valley as different from the rest, since this site is located at the most extreme southern distribution of eastern chimpanzees, and isolated by the Lake Tanganyika.

A



825

B

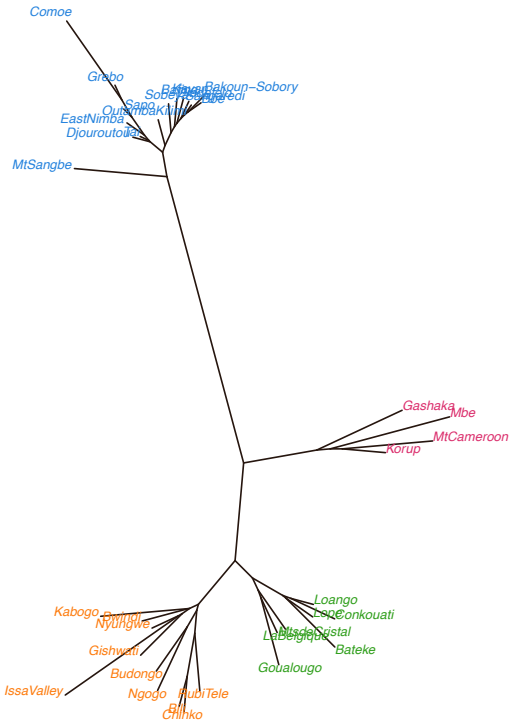


Fig. S45. Pairwise F_{ST} estimates between chimpanzee sampling sites. (**A**) F_{ST} Matrix and (**B**) Neighbor-joining tree constructed from the pairwise F_{ST} estimates. Central chimpanzees in green, eastern in orange, Nigeria-
830 Cameroon in red and western in blue. Related to Fig. 2.

6.6. Pairwise genetic distance, geography and phylogeny

6.6.1. Pairwise genetic dissimilarity

In order to obtain a genetic dissimilarity matrix, we ran ngsDist v1.0.2 [42] on the genotype likelihoods derived from 213 unrelated samples (with >5x coverage and <1% human contamination). We used ANGSD with the following parameters: -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minInd 106 -skipTriallelic 1 -GL 2 -minMapQ 30 -doGlf 2 -doMajorMinor 1 -doMaf 2 -minMaf 0.05 -r chr21: -SNP_pval 1e-6. Next, ngsDist was run with the parameters: --n_sites 127706 --probs TRUE --pairwise_del --n_boot_rep 3 --boot_block_size 1. The obtained distance matrix was plotted with R (version 3.5.2).

The dissimilarity matrix (Fig. S46) shows clustering by subspecies as well as intrasite similarity. We observe that Gas1-10 is a clear outlier, being highly dissimilar from the rest of the Nigeria-Cameroon samples.

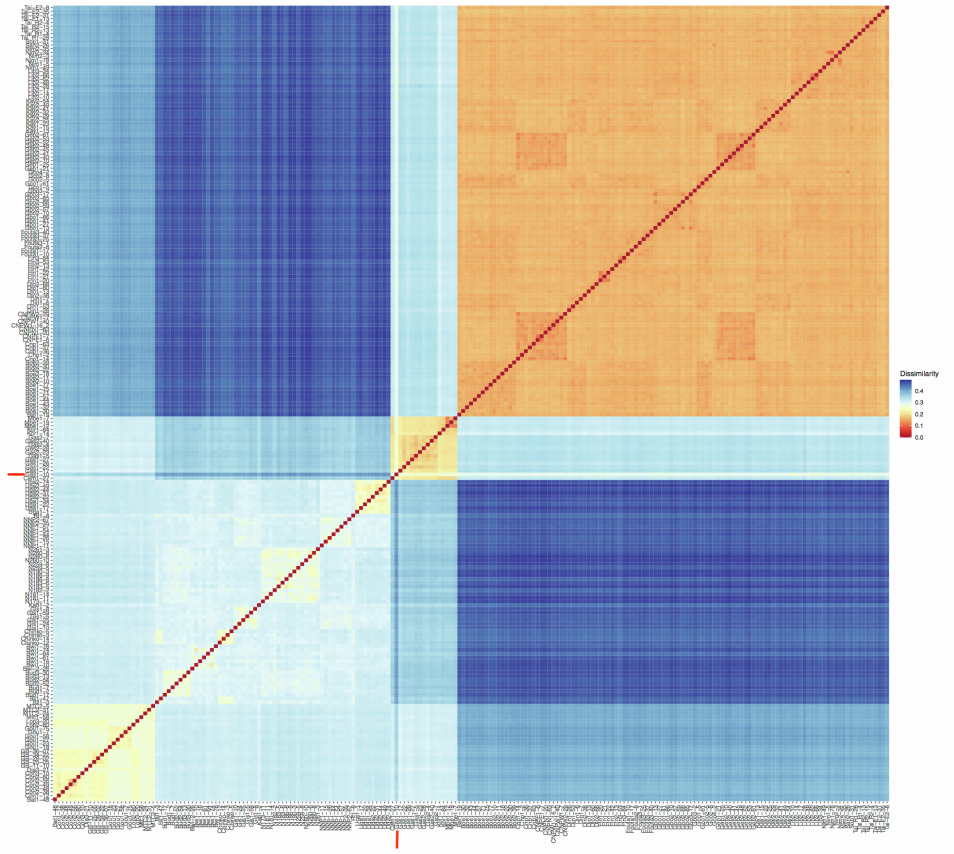


Fig. S46. Dissimilarity matrix derived from genotype likelihoods from ANGSD using ngsDist. We observe one clear outliers Gas1-10 (highlighted in the plot with a red line) and clustering by subspecies as well as between sites. Related to Fig. 2.

6.6.2. Linear regression of genetic distance (F_{ST}) and geographical distance

We then correlated the geographical distance, defined as a straight-line distance in kilometers between the PanAf sites (Fig. S47), with the obtained F_{ST} matrix (Fig. S45).

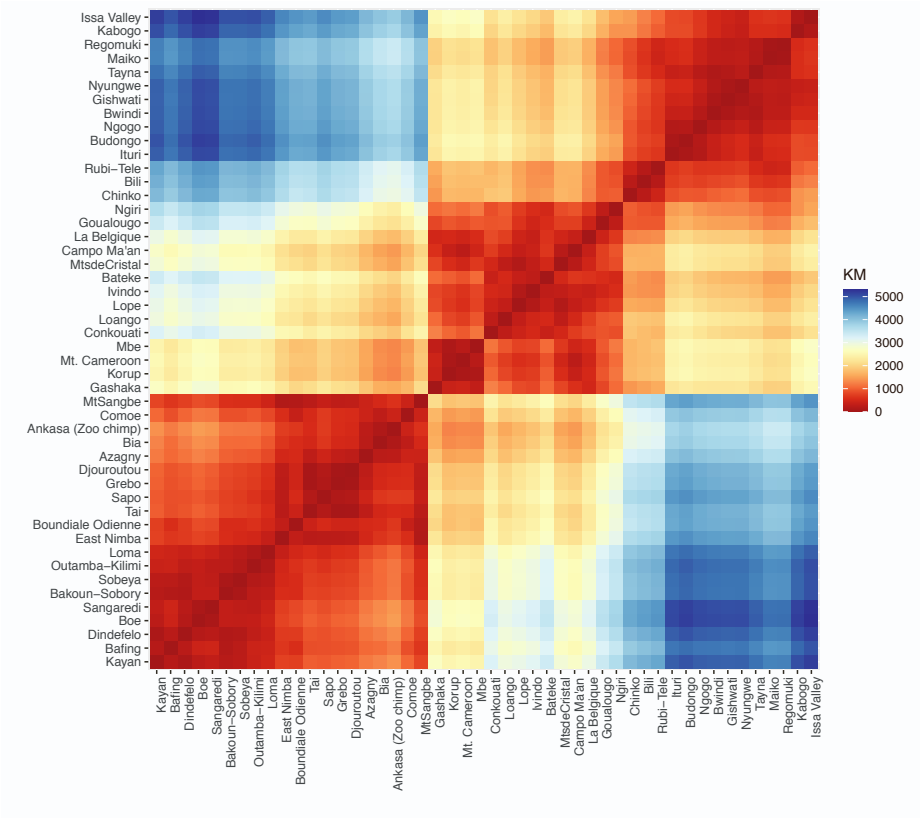


Fig. S47. Pairwise distance in kilometers between each PanAf sampling site calculated in a straight line. Related to Fig. 2.

The geographical distance was plotted against the genetic distance from the F_{ST} values, separately for each subspecies. We tested for correlation between geographic and genetic distance, and testing their significance with a linear model using the lm() function in R. Linear regression within and between subspecies of the genetic distance and geographic distance (Western-Western; Eastern-Eastern; Central-Central; Nigeria-Cameroon - Nigeria-Cameroon) shows different profiles in terms of slope and intercept (Fig. S48A, B). Within-subspecies genetic distance seems to be highly correlated with geographical distance, in accordance with an expectation of isolation-by-distance. On the other hand, when comparing individuals from different subspecies, we do not observe this pattern: inter-subspecies comparisons have lower slope and high intercept (Fig. S48B), pointing to a rather clear boundary between subspecies, instead of a gradient of differentiation, specifically for the case of eastern and central subspecies, with lower intercept but flat slope. There seems to be an interesting case between Nigeria-Cameroon and central chimpanzee subspecies, where distances significantly correlates with genetic distance (p-value = 0.3731), which could be explain with the higher shared ancestry between

central chimpanzee sites located in the northern distribution and Nigeria-Cameroon chimpanzees (see Supplementary Text Note 10).

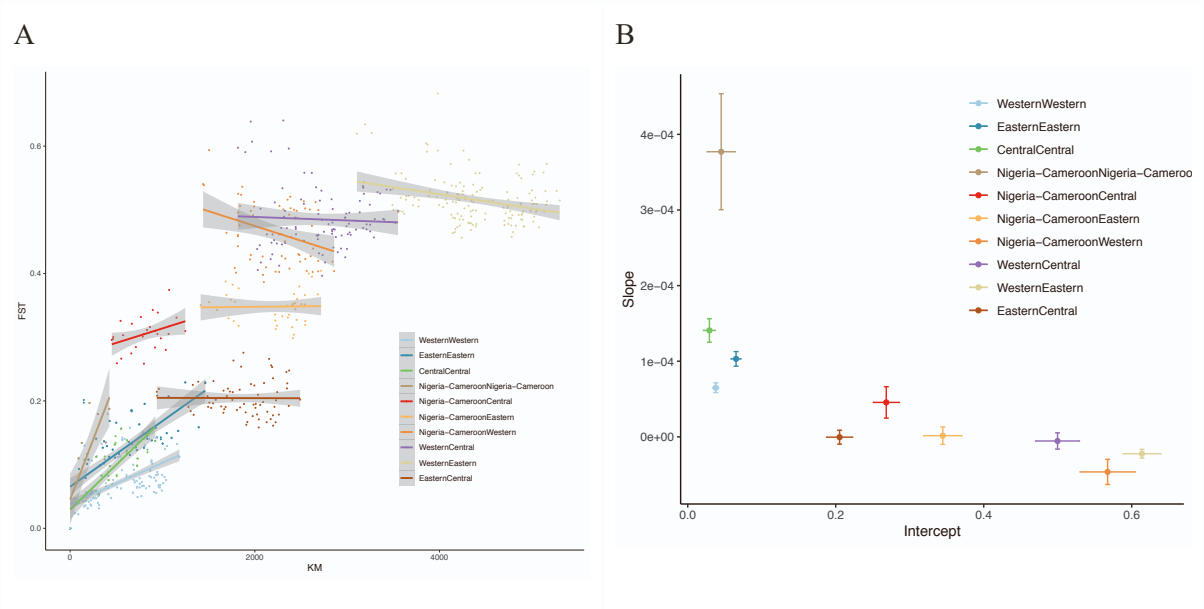


Fig. S48. Linear regression of genetic distance (F_{ST}) as function of geographical distance. **(A)** Linear regressions within subspecies: western-western (adjusted R^2: 0.3129; p-value: < 2.2e-16), eastern-eastern (adjusted R^2: 0.5326; p-value: < 2.2e-16), central-central (adjusted R^2: 0.6289; p-value: 6.667e-12) and Nigeria-Cameroon-Nigeria-Cameroon (adjusted R^2: 0.6072; p-value: 0.0002263); and between subspecies: Nigeria-Cameroon-Central (adjusted R^2: 0.1239; p-value: < 0.03731), Nigeria-Cameroon-Eastern (adjusted R^2: -0.02582 p-value: 0.8934), Nigeria-Cameroon-Western (adjusted R^2: 0.1035; p-value: 0.007019), Western-Central (adjusted R^2: -0.007191; p-value: 0.6129), Western-Eastern (adjusted R^2: 0.08463; p-value: 0.0.0001792) and Eastern-Central (adjusted R^2: -0.01468; p-value: 0.9645). (**B**) Plot of intercept and slope of linear regression between and within subspecies. Within subspecies linear regressions exhibit higher slope and lower intercept, compared to the inter-subspecies comparison, with the exception Nigeria-Cameroon - central subspecies. Horizontal and vertical lines show the standard error of the mean. Related to Fig. 2.

To test the hypothesis of long-term barriers between subspecies, we performed a partial Mantel test using the R package vegan (https://CRAN.R-project.org/package=vegan), between genetic distance and sample subspecies by introducing a binary variable coded as 0 when two samples are from the same subspecies, and as 1 otherwise. We tested for this effect by controlling for geographical distance, finding a significant result (Mantel statistic R: 0.3846; p-value: 0.001). This provides further evidence for the genetic differentiation of the four known chimpanzee subspecies.

6.6.3. Phylogeny

Following the same approach as in Supplementary Text Note 4.7.1, we obtained a genetic dissimilarity matrix with 213 samples including orangutan as outgroup (Pongo_pygmaeus-A939_Nonja) [23]. We run ngsDist version 1.0.2 [42] with 128,251 sites with the same parameters as in Supplementary Text Note 4.7.1. Subsequently, we run FastME version 2.1.5[43] on the dataset with 3 bootstrap replicates to infer the phylogeny and then RAxML-NG version 0.9.0 [44] to place bootstrap supports in the main tree. For the tree reconstruction, Gas1-10 has been excluded due to extreme genetic differentiation observed in the dissimilarity matrix (Fig. S46).

The resulting phylogeny (Fig. S49) shows clustering by subspecies. We observe two clades centrals and easterns (colored in green and orange, respectively) and Nigeria-Cameroon and westerns (colored in pink and blue, respectively. Central chimpanzees are splitted in two groups (Conkouati, Loango and Lopé) as an outgroup of the rest of central chimpanzee samples (Mts de Cristal, La Belgique and Goualougou) and eastern chimpanzee samples. Western chimpanzee internal branches are short due to their shallow population structure.

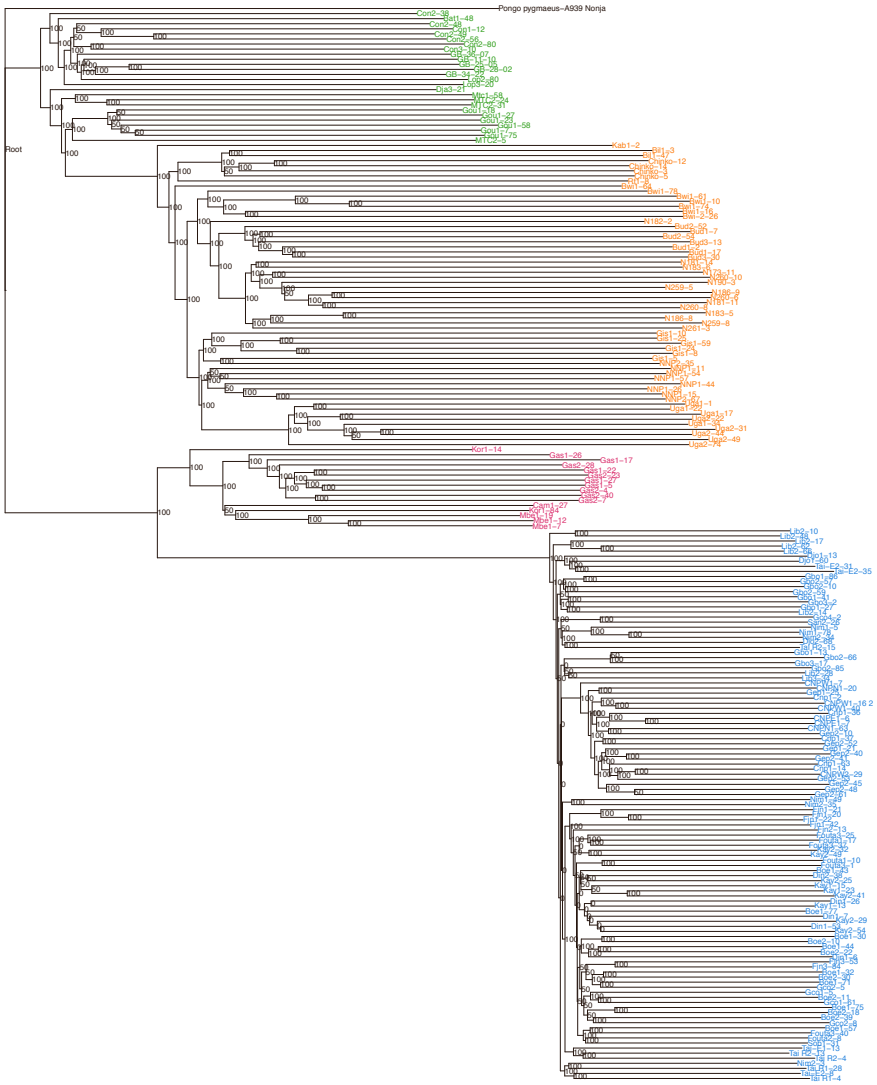


Fig. S49. Phylogeny of 212 PanAf fecal samples and one Orangutan as the outgroup (Pongo_pygmaeus-A939_Nonja). Western chimpanzee samples are in blue, Nigeria-Cameroons are in pink, easterns are in orange and centrals in green. Related to Fig. 2.

To make the phylogeny clearer we decided to include one sample per site and perform as many trees as samples available. With that criteria, Ngogo had 15 unique samples so we repeated the minimum number of samples necessary in each field site to construct 15 trees (Fig. S50). We used the same methodology to construct trees as the one used with all samples.

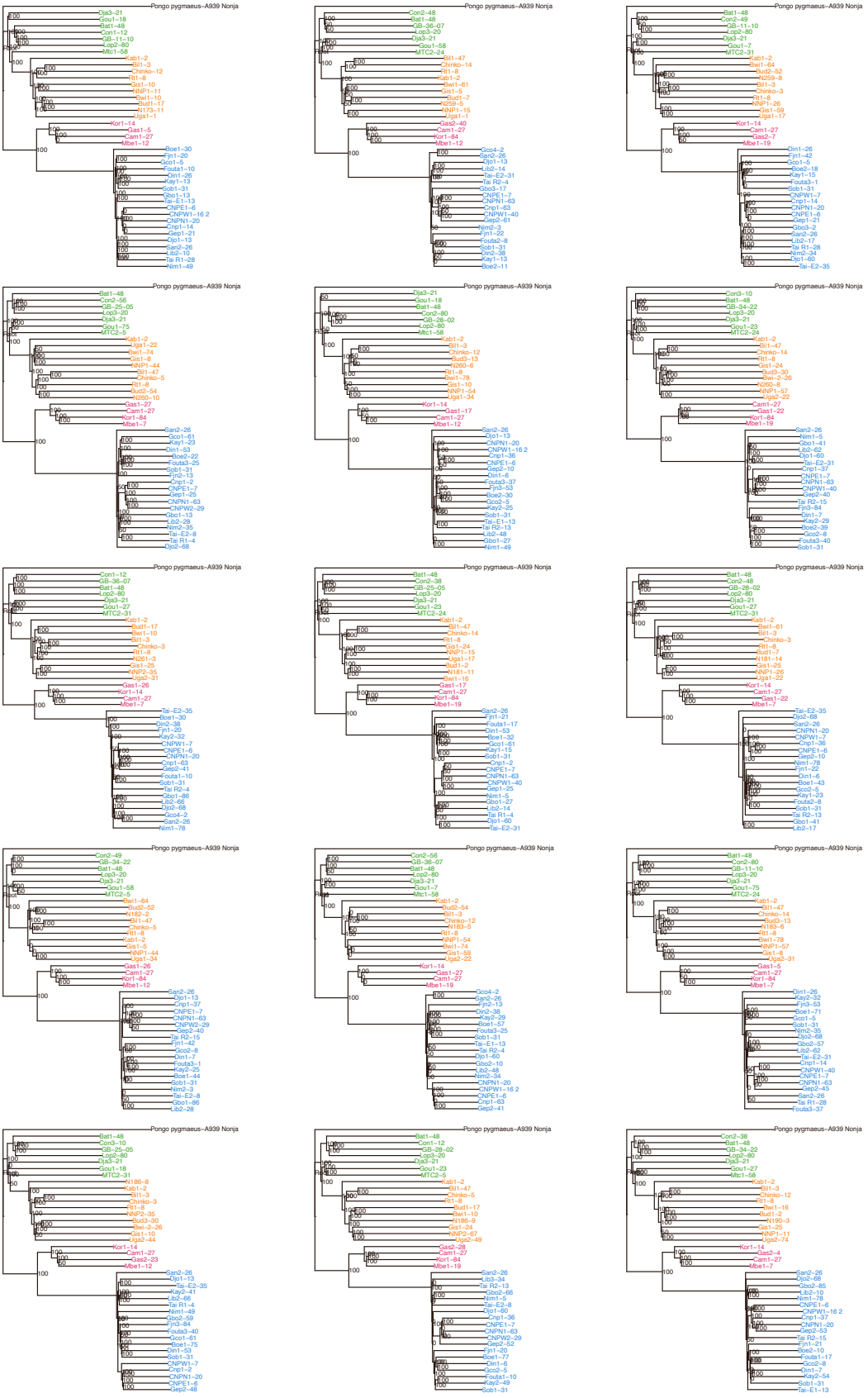


Fig. S50. Tree constructed with one sample per site, up to a total of 15 replicates. Colors define subspecies: central chimpanzees are in green, eastern chimpanzees are in orange, Nigeria-Cameroon chimpanzees are in pink and western chimpanzees are in blue. Related to Fig. 2.

Finally, we generated a single tree from the 15 independent trees, by applying different methods. Using R package ape version 5.3[45] with the consensus function, we constructed a strict-consensus tree with parameter p=1 (Fig. S51A) and a "majority-rule" consensus tree using p=0.5 (50% of trees support the consensus tree) (Fig. S51B). Consensus tree provides an estimate for the level of support for each clade in the final tree. We also obtained the average tree (Fig. S51C) with R package phytools version 0.7-47[46] using averageTree function with method="symmetric.difference", that calculates the tree with minimum sum of squared distances to all other trees. And finally the maximum clade credibility tree using R package phangorn version 2.5.5[46,47] with the maxCladeCred function (Fig. S51D). This last tree is the only that maintains branch lengths since it is the tree with the highest score from the 15 sampled.

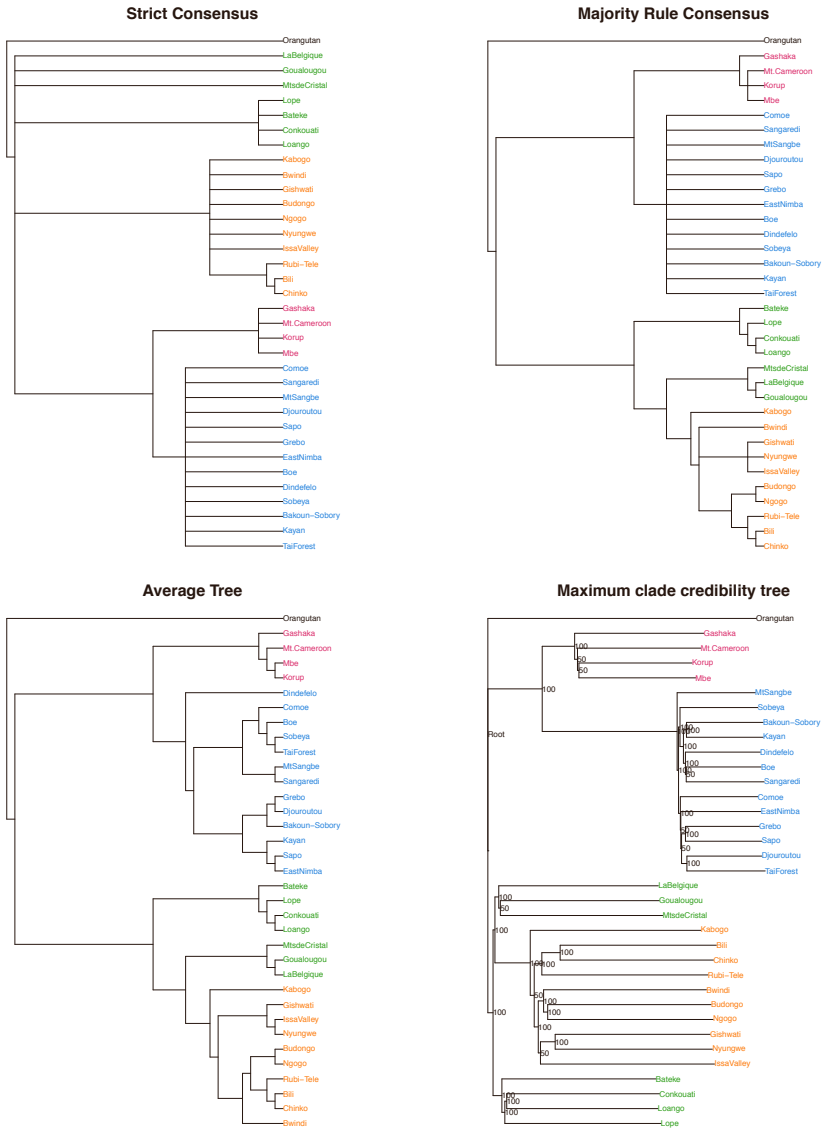


Fig. S51. Strict Consensus tree, majority-rule consensus tree, average tree and maximum clade credibility tree. Colors define subspecies: central chimpanzees are in green, eastern chimpanzees are in orange, Nigeria-Cameroon chimpanzees are in pink and western chimpanzees are in blue. Related to Fig. 2.

Next, we built a single tree using genotype calls derived from snpAD software (Supplementary Text Note 2.2.3) choosing the samples of higher coverage at each site with > 5-fold coverage (Table S1) and genotypes were filtered as follows: indels, missing genotypes, non-biallelic sites, and sites with DP <8, DP >75 and quality <20 were excluded. Next, we built the tree only with the genotypes in the chromosome 21 target space and excluding 50% of the variation private of the orangutan branch. The heterozygous positions were sampled at random to build a phyDat object, to then built a pairwise matrix using dist.dna function from Ape package version 5.4-1[45] in R version 3.6.0 with pairwise.deletion=F and model="K80" parameters. The final neighbour-joining tree (Fig. S52) was built using the NJ function from the Phangorn package version 2.5.5[47] in R version 3.6.0, and the likelihood of the tree was assessed with pml and optim.pml function from the Phangorn package[47] in R version 3.6.0. A hundred tree bootstraps were performed with bootstrap.pml with options optNni=TRUE and control = pml.control(trace = 0) from the Phangorn package version 2.5.5[47] in R version 3.6.0.

Differently to what was observed with the genotype likelihoods (Fig. S50 and Fig. S51), with high-coverage genotype calls (Fig. S52) we reconstructed the known phylogeny of the chimpanzee subspecies.

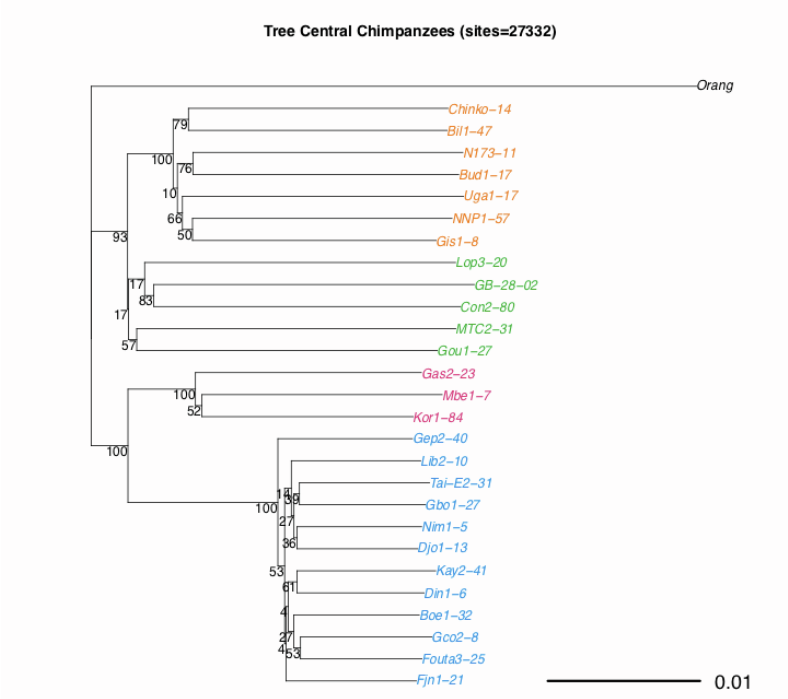


Fig. S52. Neighbour-joining tree based on genotype calls with one sample per site with > 5-fold coverage and < 0.5% human contamination. Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 2.

In the phylogeny obtained from genotype likelihoods, eastern chimpanzees and centrals do not form distinct monophyletic groups but rather eastern chimpanzees are inside the central clade. This result does not support neither previous findings nor the phylogeny of the high-coverage (>5-fold)

genotype calls from snpAD, where we observe that central and eastern chimpanzees are monophyletic. One of the reasons behind this discrepancy could be the uncertainty with genotype likelihoods in low coverage samples when building the genetic distance matrix. Another reason is certainly the deeper population structure within central chimpanzees and differential amounts of bonobo admixture, which may contribute to this uncertainty.

Finally, to increase the number of genotype sites and increase the power to reconstruct the phylogeny in each subspecies independently, we run the same methodology described above to each chimpanzee subspecies separately (Fig. S53). In this case, for the Nigeria-Cameroon subspecies, we included one sample representative of Mt. Cameroon (Cam1-27) although the target coverage was below 5-fold.

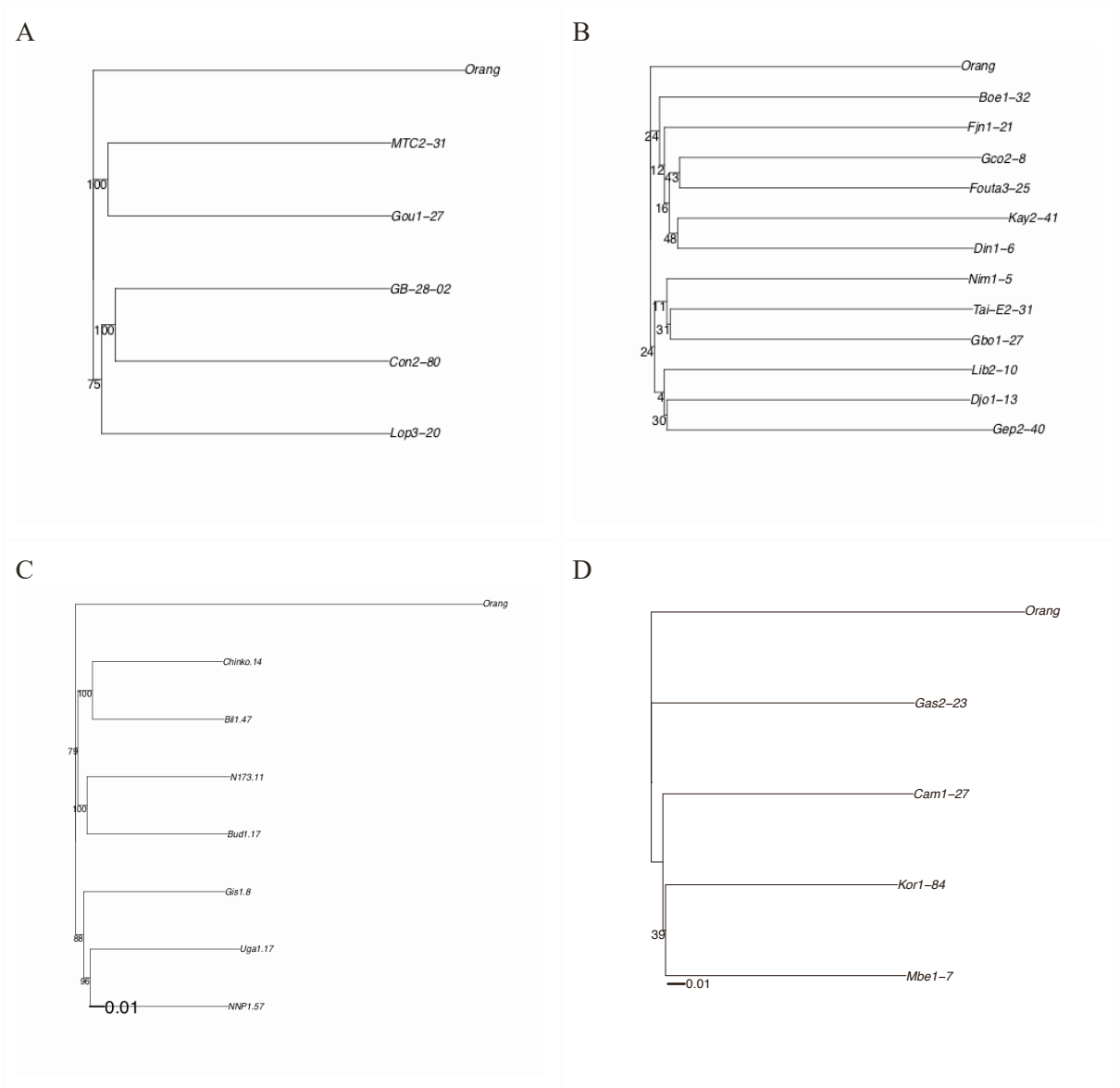


Fig. S53. Neighbour-joining trees based on genotype calls with one sample per site. (A), central chimpanzees (sites=49,644), (B), western chimpanzees (sites=10.050), (C), eastern chimpanzees (sites=22,553) and (D), Nigeria-Cameroon chimpanzees (sites=36,295). Related to Fig. 2.

Note 7. Patterns of genetic drift.

7.1. F3-statistics

We used AdmixTools to compute F3 outgroup statistic and test for shared drift between different chimpanzee populations with orangutan as outgroup (pygmaeus_ERS1986511) [48] using qp3Pop [49]. For this analysis, we used samples with more than 5-fold coverage and less than 0.5% human contamination (Table S1). VCFs (of all populations and by subspecies) were filtered to keep only biallelic sites called with snpAD, remove indels, and keep sites with a minimum depth of 3, minimal genotype quality of 20, MAF 0.01 and allow for 20% missingness.

To visualize shared drift between populations, we constructed pairwise matrices with all subspecies together and separately (Fig. S54). When comparing all communities together, sites from the same subspecies share more drift (Fig. S54A). To analyze the relationships within subspecies in detail, we calculated this statistic at the subspecies level. We observe that central chimpanzees are divided into a northern clade (Goualougo and Mts de Cristal) and a southern clade (Batéké, Lopé, Conkouati and Loango) (Fig. S54B), in agreement with the results in Supplementary Text Note 4.5. Interestingly, the southern clade of central chimpanzees shares less drift with western and Nigeria-Cameroon chimpanzees than the northern clade. Eastern chimpanzee sites form three clades, a northern one (Bili, Rubi Télé and Chinko), an eastern clade (Bwindi, Budongo and Ngogo) and a southern clade (Kabogo, Issa Valley, Gishwati and Nyungwe) (Fig. S54C). Interestingly, Rubi-Télé shares a lot of drift mainly with Ngogo and Budongo but also with sites from the southern one. Since we only have three Nigeria-Cameroon sites, the only conclusion we can retrieve is that Korup and Mbe share more drift with each other than with Gashaka (Fig. S54D). Finally, in western chimpanzee sites, Mt Sangbé stands out as the site that shares the least drift with the other sites, and sites from the northern distribution of western chimpanzees (Kayan, Bakoun, Sobory, Sobeya, Dindefelo, Boe and Sangaredi) share more drift between them compared to the rest. The same scenario is observed with sites from the southern distribution of western chimpanzees (Tai_Eco, Tair_R, Djouroutou, Grebo and Sapo) and Comoé. Comoé sites share a lot of drift among each other, and more with the southern than with the northern sites (Fig. S54E).

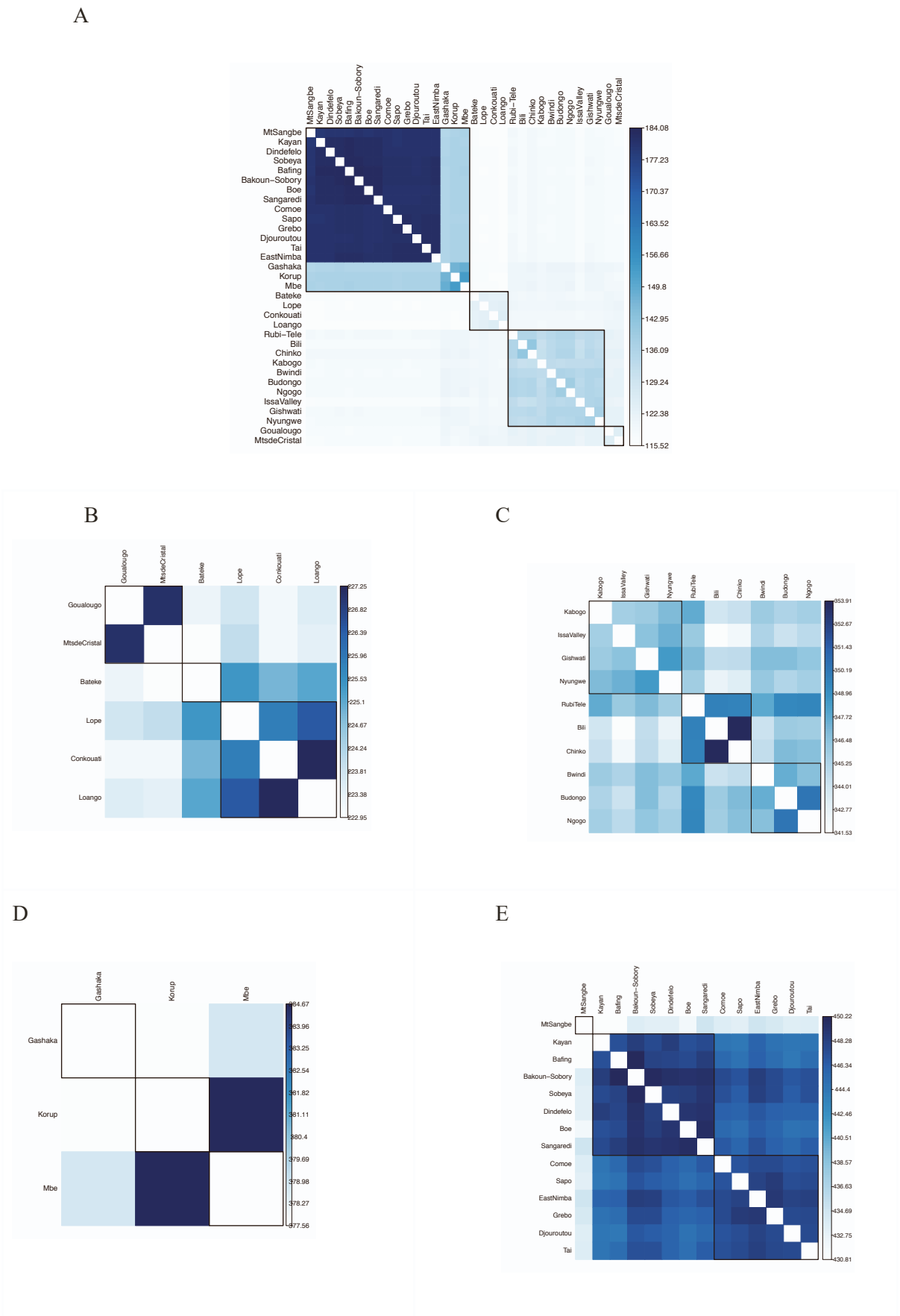


Fig. S54. F3 outgroup statistic (absolute values) between: (**A**) chimpanzee communities at all sites; or between subspecies: (**B**) central chimpanzee; (**C**) eastern chimpanzee; (**D**) Nigeria-Cameroon chimpanzee and **e,** western chimpanzee. Related to STAR methods.

7.2. Population assignment

We also ran qp3Pop to determine sample origin by leaving one sample out at a time, testing with which population a given sample shares more drift. This strategy works well for samples at sufficient coverage (Table S1). As an example, see Fig. S55, where we correctly assign the origin of one sample from Conkouati. Since eastern, central and Nigeria-Cameroon chimpanzees have a higher level of population stratification than western chimpanzees, we correctly assign them to their origin in the majority of cases (Fig. S56A), or otherwise to the most proximal location (Fig. S56A). However, in western chimpanzees this method only has power to assign samples to the northern, southern or Comoé clade (Fig. S56B).

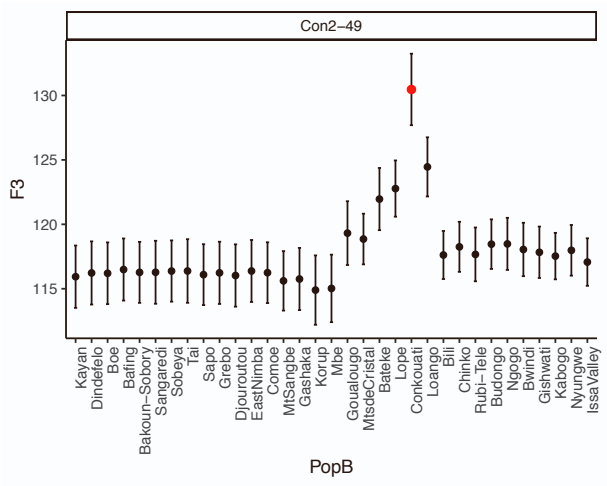
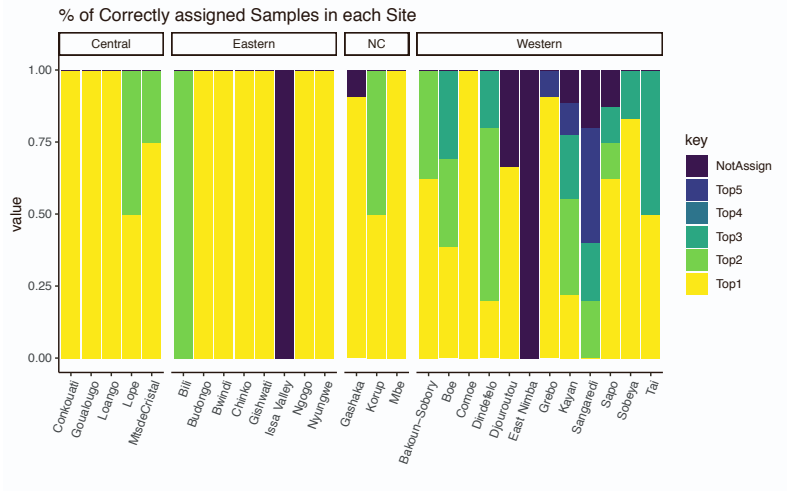


Fig. S55. Population assignment of sample Con2-49. This is a sample from Conkouati, and the F3 statistic is correctly assigning it to its origin. Error bars represent the standard deviation. Related to STAR methods.

A



B

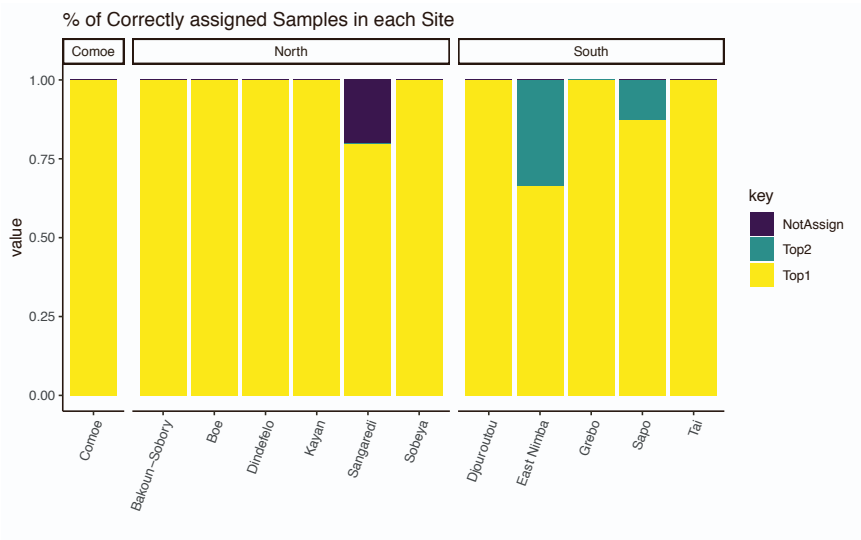


Fig. S56. Accuracy in determining the origin of chimpanzees to their community of origin by leaving one sample out and testing it with the rest of sites. (**A**) We have excluded from the analysis those sites where only one sample was left after filtering (Batéké, Kabogo, Rubi-Télé and Mt Sangbé). For central, eastern and Nigeria-Cameroon chimpanzees, the classification is correct in almost all instances. **(B)** Western chimpanzee classification based on three geographical clusters; here, we classify almost all of them correctly. Related to STAR methods.

7.3. TreeMix

We used TreeMix v1.12[50] to infer patterns of divergence and migration events between the different chimpanzee communities and subspecies. This software requires SNPs, thus we used the genotype calls obtained with snpAD[20] and only those samples with a coverage larger than 5-fold and less than 0.5% human contamination (Table S1). We added the chr21 genotypes obtained with GATK[15] of one orangutan individual, to be used as outgroup (pygmaeus_ERS1986511)[48] instead of the human genome which could introduce a bias due to residual human contamination in the samples. The VCF

file was filtered to keep only biallelic sites, remove indels, and keep sites with a minimum sequencing depth of 3-fold, minimal genotype quality of 20, minor allele frequency of 0.01, and allow for 20% missingness. We run TreeMix v1.12 setting Orangutan as the outgroup and the block size to 500 with subspecies clusters and geographical site clusters. We modelled migration events with 10 iterations with different random seeds. We visualized the resulting most supported tree model using an R script provided in TreeMix, and calculated the variance explained with the get_f() function provided in TreeMix.

First, we assessed the subspecies tree topology with 205 samples at 180,075 SNPs passing the quality filters (Fig. S57). The ML tree model explains 99.94% of the variance in the data.

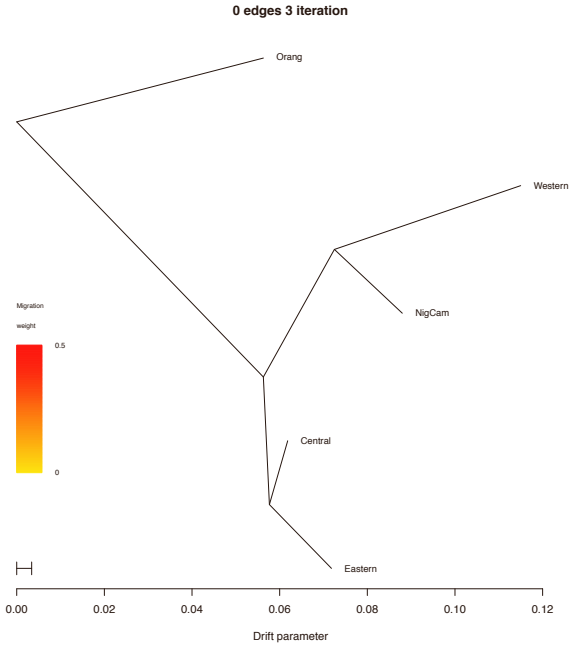


Fig. S57. ML tree by subspecies inferred with TreeMix, explaining 99.94% of the variance. Related to Fig. 2.

Regarding inferred gene-flow, we modelled one migration event between chimpanzee subspecies, using the previously generated ML tree as the user-specified tree. TreeMix results support significant post-divergence gene flow (p-value 1.58462e-11) between Nigeria-Cameroon chimpanzees and eastern chimpanzees, explaining 99.97% of the variance, an event already described in Prado-Martinez *et al*. (2013)[23] (event present in 7 out of the 10 independent runs) (Fig. S58).

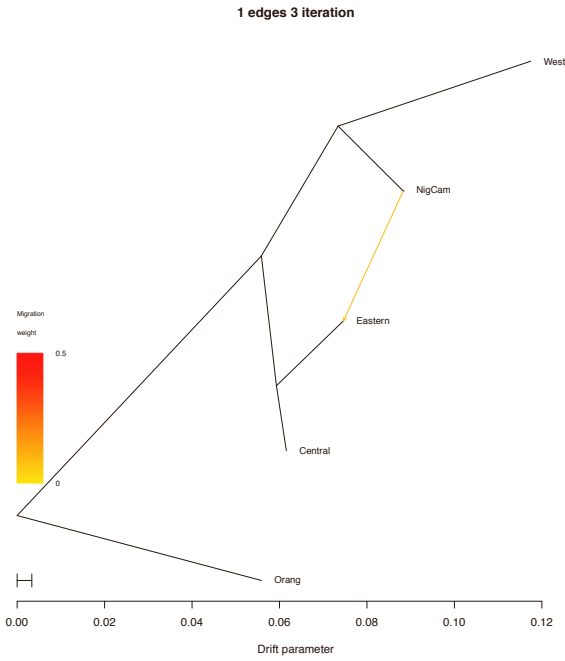


Fig. S58. ML tree with one migration event. It shows a significant migration between Nigeria-Cameroon and eastern chimpanzee populations, explaining 99.97% of the variance. Related to Fig. 2.

We also assess the tree topology of the different chimpanzee communities by running TreeMix with field site clustering (Fig. S59). The ML tree explains 99.85% of the variance. The tree topology supports a separation of sites by subspecies as well as clustering by close geographical location.

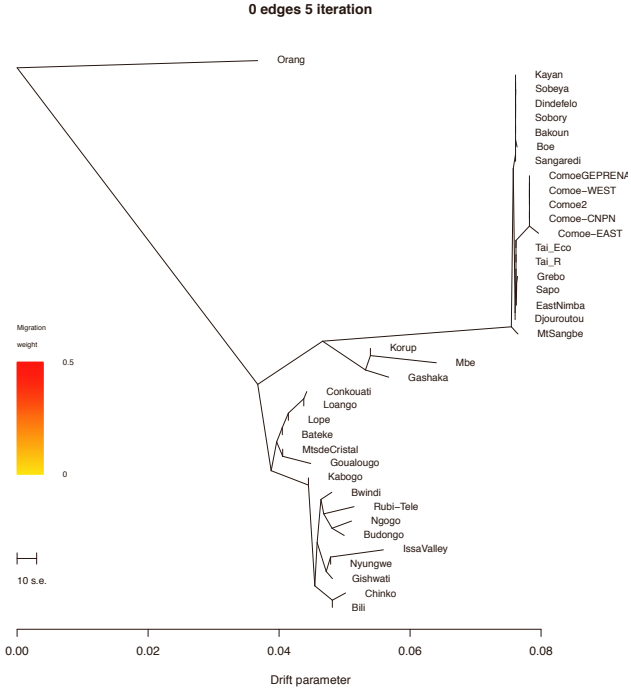


Fig. S59. ML tree of all filed sites inferred with TreeMix, explaining 99.85% of the variance. Related to Fig. 2.

We also modelled one migration event from the ML tree of different field sites starting with the previous ML tree with ten iterations. Five out of ten iterations support one migration event (p-value <2.22507e-308) from the ancestor of eastern and central chimpanzees to Batéké, and the variation explained is 99.88%, slightly higher than without any migration event (Fig. S60). Higher migration events do not result in meaningful conclusions. Thus, we conclude that this method cannot distinguish specific migration events in this dataset.

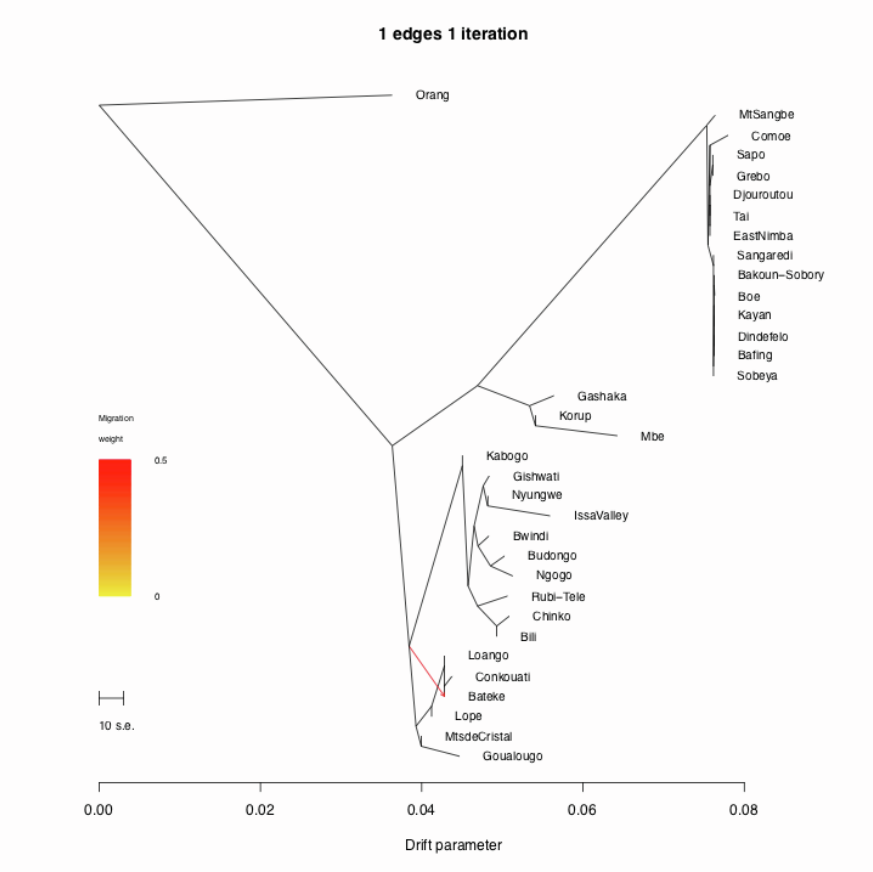


Fig. S60. ML tree with one migration event between chimpanzee field sites. It shows a significant migration event between the ancestor of eastern and central chimpanzees to Batéké, explaining 99.88% of the variance. Related to Fig. 2.

Note 8. Introgression patterns.

8.1. F4-statistics

Admixture from bonobos into the non-western chimpanzee subspecies to a small extent (less than 1%) has been described before[21], most likely as a part of a complex population history of the *Pan* clade[51] and probably with differential consequences in terms of selection in the different subspecies[52]. Unfortunately, tests for a significant enrichment of allele sharing, which are indicative of gene flow between the sampled populations (D-statistic, F-statistics)[53], depend on large numbers of independent (unlinked) loci across the genome.

We calculated F4-statistics of the form f4(non-Western chimpanzee, Western chimpanzee; Bonobo, Orangutan), in order to determine whether the previously described gene flow could be detected here. We use the full data from each population, and the orangutan individual pygmaeus_ERS1986511 as the outgroup, and perform calculations using the admixr package in R version 3.6.0 [54]. We first use data from all 22 autosomes (with strict quality filtering: coverage between 6 and 99, mapping quality more than 20, mappability score of 1, repeat mask) from the previously published high-coverage panel of 59 individuals[21]. We find this statistic to be significant for central (f4=0.000026, Z-score=4.82) and eastern chimpanzees (f4=0.000022, Z=4.14), with a slightly lower amount in the latter, in line with previous results[21]. We do not observe a significant value for Nigeria-Cameroon chimpanzees (f4=0.000001, Z=0.159).

Although with data on only one chromosome, the shortest, the power of such statistics is very limited[31], we find that it is still significant for central chimpanzees (f4=0.000244, Z=3.03), but not for eastern (f4=0.000097, Z=1.761) and Nigeria-Cameroon (f4=-0.000149; Z=-0.655) chimpanzees, when using the data on chromosome 21. However, when analyzing the individuals newly sequenced in this study, a combination of low coverage, smaller target space and increased noise led to non-significant values across all subspecies (central chimpanzee: f4=0.000018, Z=0.441; eastern chimpanzee: f4=0.000022, Z=0.485; Nigeria-Cameroon chimpanzee: f4=0.000047, Z=1.598), despite a larger number of individuals. We conclude that using this approach does not have the power to detect gene flow, particularly at a low extent that is already difficult to determine using high-coverage whole genomes[21,51], and we are not able to meaningfully distinguish differential amounts of gene flow even on the subspecies level.

8.2. Inference of introgressed fragments

Admixfrog[55] is a newly developed method to infer ancestry fragments from low-coverage and contaminated data. It uses a Hidden Markov Model (HMM) to infer local ancestry in a target individual

from different sources which represent the admixing populations. Here we use as sources of admixture bonobo (BON) and all chimpanzee subspecies (EAS, CEN, WES, NIG). First, we tested this method on four chimpanzee whole genomes (WG) of each subspecies[21] as target. The reference panel on the chr21 (source) was built using an equal number of individual genomes of each subspecies (6 genomes) from de Manuel *et al.* (2016)[21] excluding the 16 target individuals, and we included 10 bonobo genomes. We also included two humans (1 African and 1 European)[56] to serve as a source of contamination (HUM) and 1 orangutan (pygmaeus_ERS1986511)[48] as the ancestral state (ORA). A VCF file including all aforementioned samples was filtered to remove indels, sites with genotype quality <20, depth lower than 6 and higher than 100, we applied a mappability filter and kept the heterozygote positions in allele balance. The reference file was built from the filtered VCF using the admixfrog-ref script with the following parameters --states HUM BON EAS NIG CEN WES ORA --state-file data.yaml --chroms chr21 --ancestral ORA. Then, each target file was obtained from bam files using the admixfrog-bam script. Finally, introgressed fragments were obtained by running admixfrog with the following parameters: --infile ${sample}.in.xz --ref ${ref}.xz --states BON NIG CEN WES EAS -P --ancestral ORA --cont-id HUM --ll-tol 0.01 --bin-size 5000 --est-F --est-tau --freq-F 3 --freq-contamination 3 --e0 0.01 --est-error --run-penalty 0.2 --max-iter 250 --n-post-replicates 200 --filter-pos 25 --filter-map 0.000.

We obtained the global simulated runs of ancestry from the *.res2* file and the called runs of ancestry from the *.rle* file. The simulated runs of ancestry recovered each main ancestry component of each target individual (Fig. S61A), with eastern and central chimpanzees exhibiting shared ancestry. We recovered the simulated runs of bonobo ancestry in each subspecies and we obtained values below 1% and with higher values of bonobo introgression in central chimpanzees, compared to the rest of subspecies. Western chimpanzees have the lowest bonobo introgression (Fig. S61B). This pattern has already been proposed by de Manuel *et al.* (2016)[21].

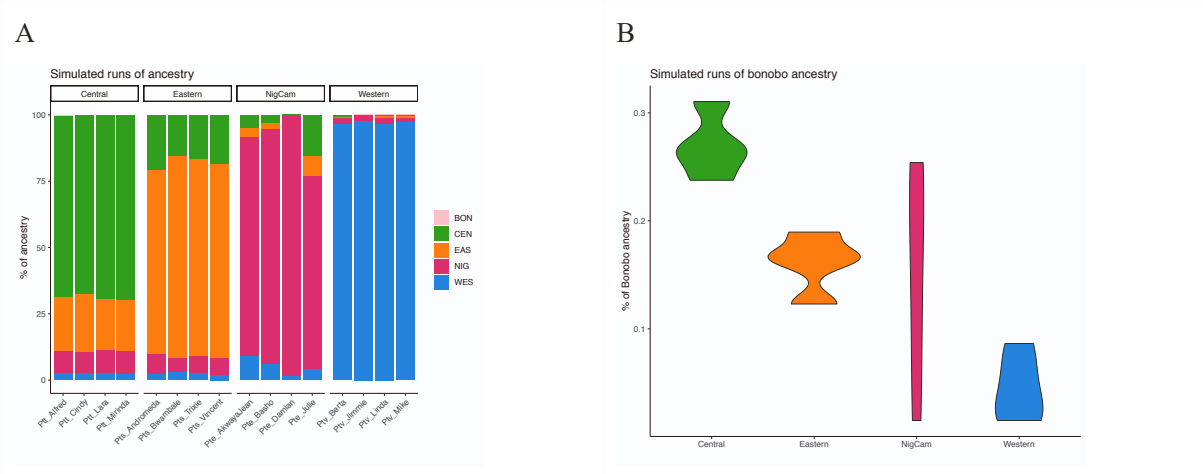


Fig. S61. Simulated runs of ancestry. **(A),** Percentage of ancestry from all chimpanzee subspecies and bonobo. **(B),** Percentage of bonobo ancestry in each chimpanzee subspecies. Related to STAR methods.

When looking at the called fragments of ancestry we detect that in all subspecies chromosome 21 is largely assigned to the corresponding subspecies, with small interspersed fragments putatively informative of gene flow between subspecies (Fig. S62A). We also detect bonobo introgressed fragments (in black) in all subspecies except western chimpanzees (Fig S62A, B).

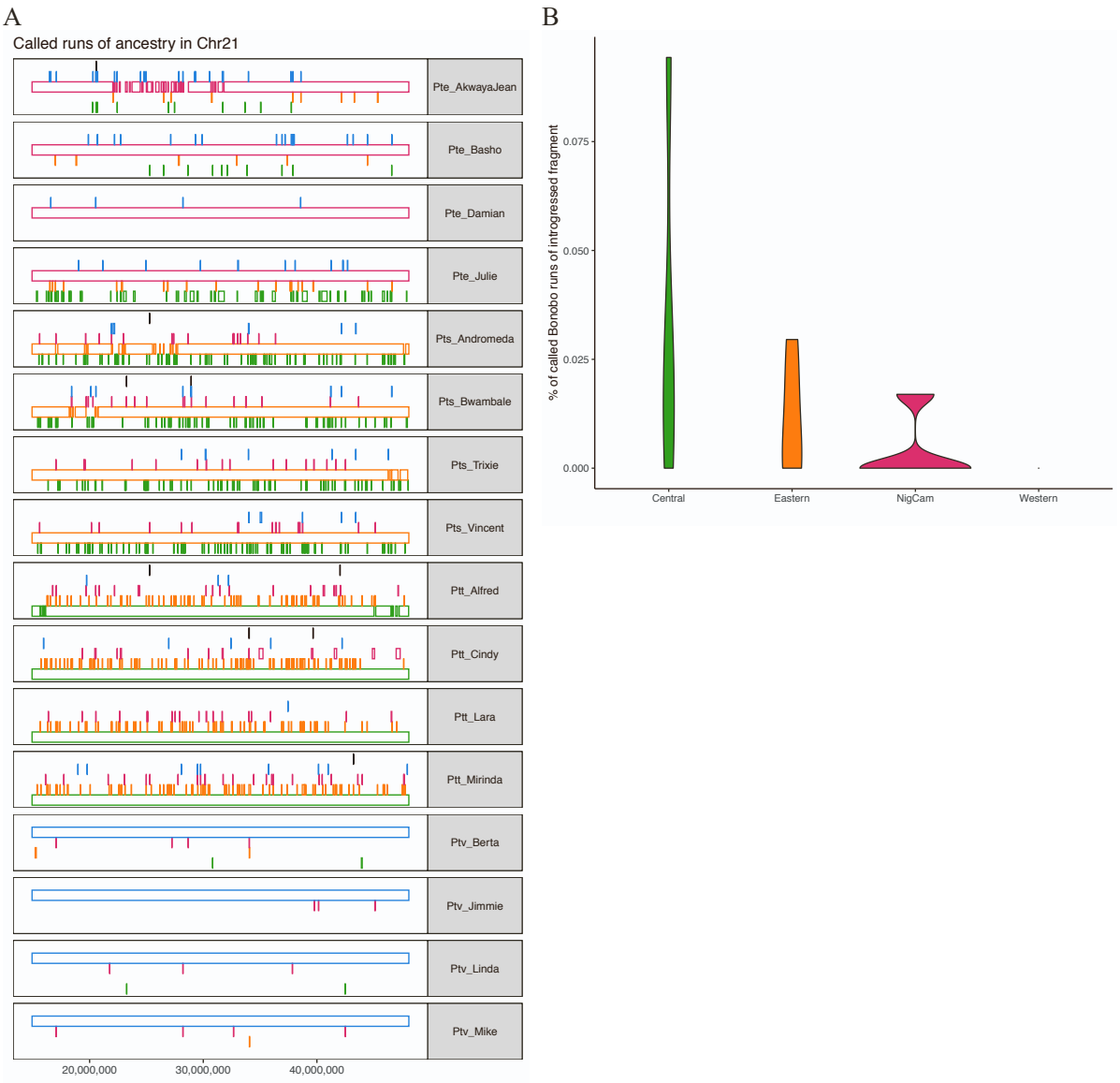


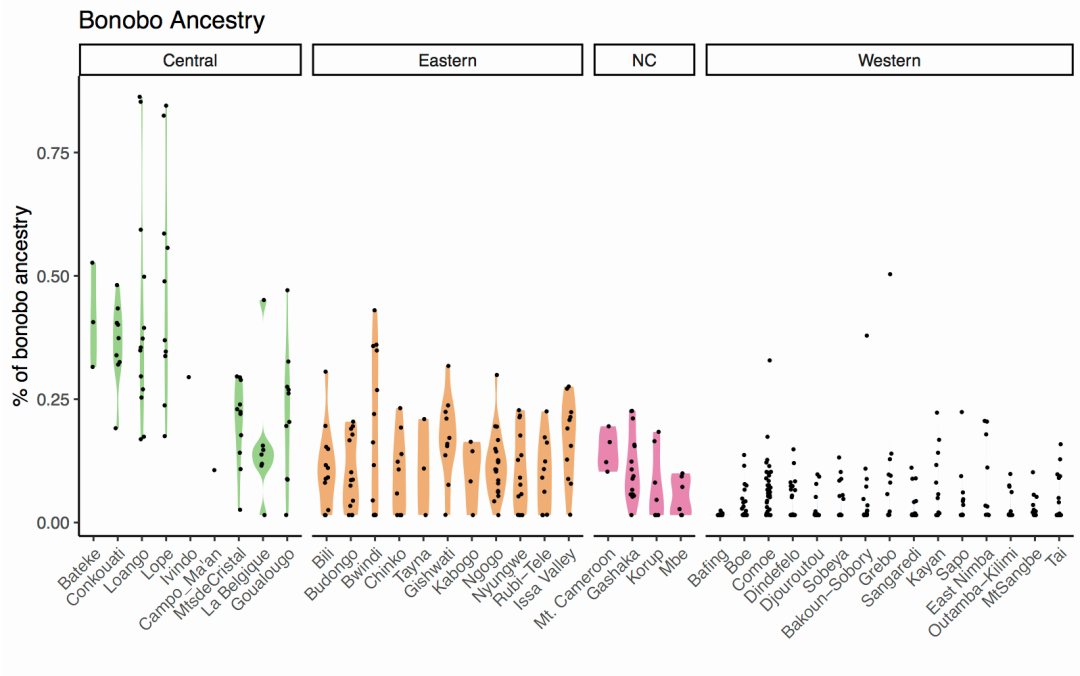
Fig. S62. Called runs of ancestry **(A)** along the chr21 and **(B)** percentage of bonobo introgression. Related to STAR methods.

Next, we applied the same methodology to recover global bonobo simulated ancestry as well as called runs of bonobo introgression in the PanAf dataset, including 449 samples with > 0.5-fold coverage and without related samples (Table S1). Similar to what we obtained with the WG dataset; we recover the ancestral components associated with each population (Fig. S63).

1125 **Fig. S63.** Simulated runs of ancestry. (**A**) Western chimpanzees, (**B**) Nigeria-Cameroon chimpanzees, **(C)** central chimpanzees and **(D)**, eastern chimpanzees. Related to STAR methods.

After removing some samples considered to be outliers in various other analyses (Gbo2-43, Boe1-36, Fouta3-55, Boe2-33, San1-4, Gco2-9, Gas1-10, Sob1-27, Tai_R2-9, Kor1-25, Bwi1-7, Uga1-12, LCA-3-10, GB-13-13, GB-25-02, GB-29-06), we estimated the amount of simulated runs of bonobo ancestry per site and subspecies, considering two groups for central chimpanzee populations: northern to Ogooué river (Campo_Ma'an, Mts de Cristal, Goualougo, Ivindo and La Belgique) and southern to this river (Conkouati, Loango, Batéké and Lopé) (Fig. S64). We find the same overall pattern observed with WG samples, but we find that the two central chimpanzee groups show differences in bonobo introgression, with a larger amount of bonobo ancestry in southern central chimpanzees. We tested for significance using wilcoxon.test() in R and adjusting by multiple testing using p.adjust (method="BH"), finding that the southern group of central chimpanzees have significantly more bonobo introgression than the northern group (two-sided Wilcoxon test p-value= 5.73e-8) (Table S5). We hypothesize that this group has experienced another additional gene flow event from bonobos.

A



B

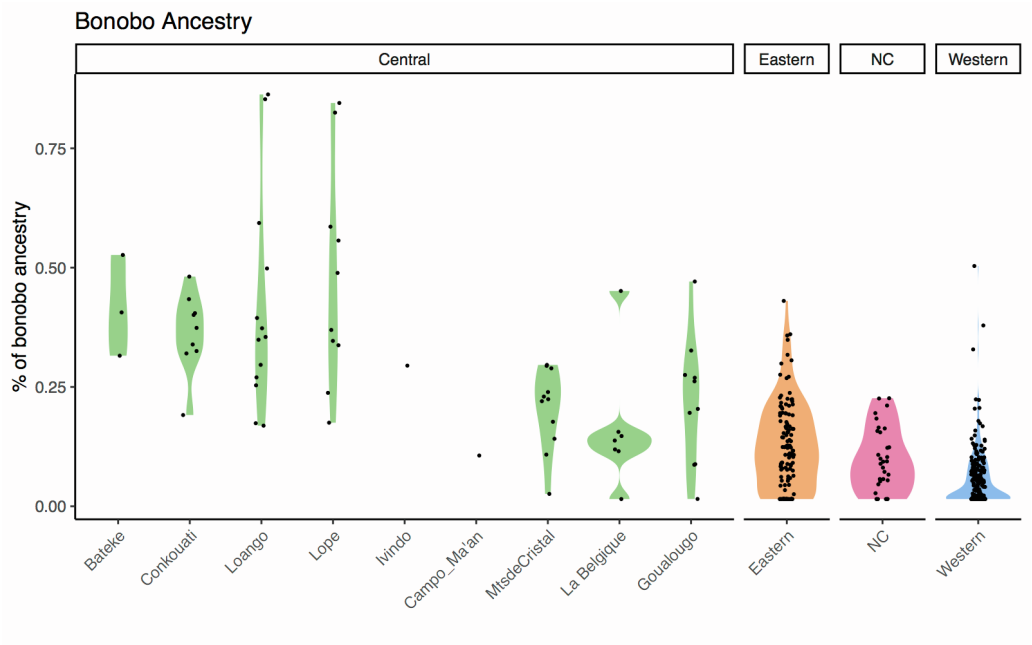


Fig. S64. Percentage of bonobo ancestry present in different chimpanzee sampling sites (A) and by subspecies (B) highlighting the two groups in central chimpanzees: north to the Ogooué river (Campo_Ma'an, Mts de Cristal, Goualougo, La Belgique, Ivindo) and south to the river (Loango, Conkouati, Lopé, Batéké). Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 2.

It is also important to note that since this software is designed to work with low coverage data, the final coverage does not influence the power to detect bonobo introgression (Fig. S65) in our dataset.

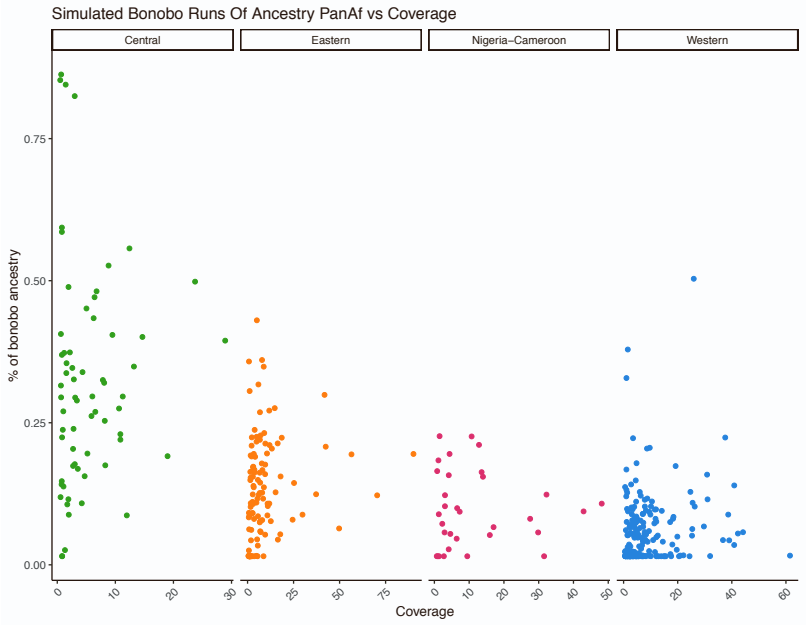


Fig. S65. Coverage does not have a big influence in the detection of bonobo simulated ancestry calls. Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to STAR methods.

Next, we analyzed the called runs of ancestry (percentage of shared ancestry and fragment length), taking into consideration not only the bonobo ancestry but also each chimpanzee subspecies origin (Fig. S66). Eastern chimpanzees have high levels of central ancestry, with Issa Valley being the eastern chimpanzee population with the highest levels (S66A) and also with longer fragments together with Gishwati and Nyungwe (S66B). In Nigeria-Cameroon chimpanzees, Gashaka has more central ancestry with longer introgressed fragments (S66C,D) than any other site; interestingly, this is the same site in which we detected an IBD-like shared fragment with Goualougo (Fig. 3 in the main text and Fig S92). Similarly, in central chimpanzees we detect eastern chimpanzee ancestry in similar amounts in all field sites (S66E), but with Goualougo being the site with longer fragments of central ancestry (S66F). Gashaka from Nigeria-Cameroon chimpanzees also shows increased eastern ancestry with longer fragments, and we also detected IBD-like tracts linking this population to some eastern field sites (Fig. 3 and Fig. S66C,D). None of the non-western chimpanzee sites present an increase in western chimpanzee ancestry (Fig. S66G,H). Finally, regarding Nigeria-Cameroon ancestry, we see that Goualougo and La Belgique have increased Nigeria-Cameroon ancestry, with longer introgressed fragments, the same way that Gashaka had more central ancestry, pointing to recent migration between both subspecies, a pattern also supported by with IBD-like tracts (Fig. 3 and Fig. S66E,F). However, such recent gene flow is only seen in with the northern group of central chimpanzees (Campo Ma'an, Goualougo, La Belgique and Mts de Cristal), while the southern central chimpanzee group (Loango,

Lopé, Ivindo, Conkouati), has much lower Nigeria-Cameroon ancestry. The populations which have more and longer called bonobo introgressed fragments are in central chimpanzees, specially within the southern group (Fig. S66I,J). We estimated the time of coalescence of the fragments shared between subspecies by extracting the maximum shared fragment length between two subspecies, which would represent the most recent coalescent event in time, and then we calculated the time by assuming a generation time of 25 years[57] and using the rule of g=100/(2*cM)[58], with cM being the length of the fragments, and g the number of generations (Table S7). Here we assume a constant recombination rate of 1cM-1Mbp and thus we use the physical distance instead of the genetic distance (see discussion of this in Supplementary Note 10).

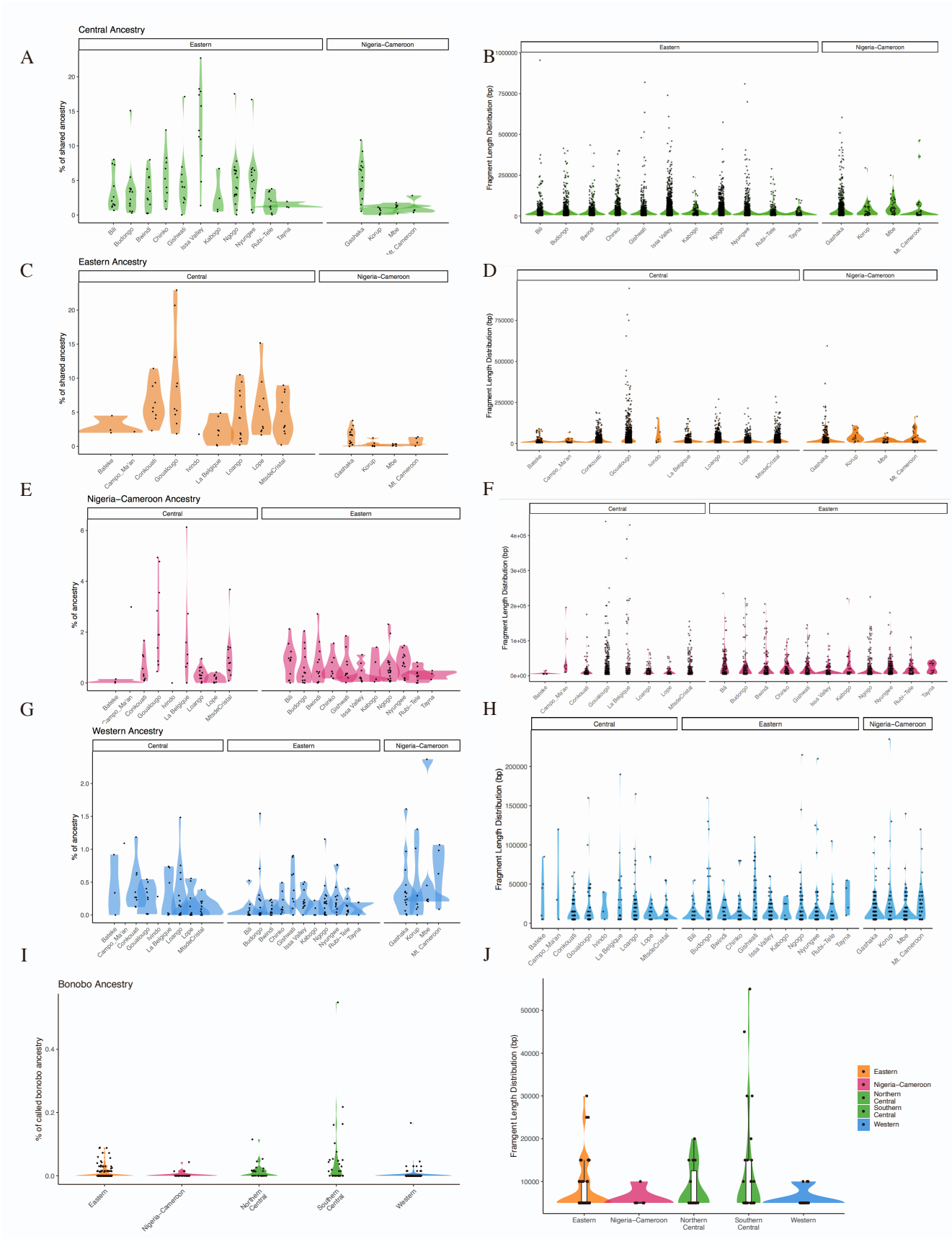


Fig. S66. Called runs of ancestry with the proportion and the distribution of the introgressed fragments per site: (**A**) Cumulative percentage of central called ancestry and (**B**) fragment length distribution (bp) in eastern and

Nigeria-Cameroon chimpanzees. (**C**) Cumulative percentage of eastern called ancestry and (**D**) fragment length distribution (bp) in central and Nigeria-Cameroon chimpanzees. (**E**) Cumulative percentage of Nigeria-Cameroon called ancestry and (**F**) fragment length distribution (bp) in central and eastern chimpanzees. (**G**) Cumulative percentage of western called chimpanzee ancestry and (**H**) fragment length distribution (bp) in non-western chimpanzees. (**I**) Cumulative percentage of bonobo called ancestry and (**J**) fragment length distribution (bp) in chimpanzee subspecies (separating central chimpanzee subspecies in central and southern centrals). The lower and upper hinges of the boxplot correspond to the first and third quartiles (the 25th and 75th percentiles) and the whisker extends from the hinge to the largest and lowest value no further than 1.5 * IQR (inter-quartile range). Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 2.

Note 9. Rare alleles and geolocalization.

9.1. Patterns of rare allele sharing

In previous studies on other species like elephants, microsatellite markers have been proven useful for geolocalization of individuals, a powerful tool for conservation purposes[59]. However, this type of data does not show sufficient clines of variation in chimpanzees, at least at the current state of knowledge[37]. Rare alleles have been a useful resource for studying human populations[60], since these variants were arising during the past few 100 or 1,000 years. Although this study provides the largest dataset of variation in chimpanzees so far, there are limitations to the sample size compared to human studies for defining rareness of an allele, and to the data quality (low coverage, DNA degradation, use of a single chromosome) that prevent the use of existing methods exploiting rare variants for ancestry estimation, like rarecoal[61]. However, here we approached a noise-tolerant strategy for detecting variants which are almost location-specific. These should represent the recent variation emerging locally in chimpanzee groups which were only loosely connected after their split, as suggested by the patterns of differentiation in the PCA analysis above and other measures of connectivity (Supplementary Text Notes 5 and 10).

We defined a reference panel of sufficient quality, using only individuals with more than 1-fold coverage in the target space, less than 0.5% human contamination, which were not determined as non-chimpanzee samples, not outliers in the PCA (see above, Note 3), not determined as belonging to another population in the PCA (Kor1-35, CMNP1-24, Uga2-81), and not carrying an excess of heterozygous alleles (Gas1-10). This yields a geographically distributed reference set of 434 reliable individuals across 38 sampling locations (Table S6, on average 11 individuals per location), with the Comoé, Taï and Bakoun-Sobory sites each being merged. We used only bi-allelic on-target quality-filtered sites for these individuals, and calculated the site-specific alternative allele frequency for each variant. We then defined near-private variants per site (instead of per individual), where a given variant had to be observed a) at a frequency larger than zero at one site, b) at a cumulative frequency of less than 1 when calculating the sum of frequencies across all other sites (for example, the variant may be observed at a frequency of 0.4 at two different sites, or 0.8 at only one site), c) with data for at least two sites (not NA values). We find a total of 963,656 of such variants on chromosome 21, with on average 26,671 variants per site. As expected, the number of variants correlates with the number of individuals per site (Pearson's correlation 0.53, p= 0.0006).

This panel of near-private variants can be used to estimate the matching of additional individuals, even when sequenced at shallow depth. In order to do this, we use the quality genotype calls for each individual (Supplementary Text Note 3.7), and overlap these with the known near-private variants. Alternative alleles in heterozygous and homozygous states were counted as derived alleles.

For each comparison, we counted the number of positions that were near-private in a given reference population and carried information in the test sample, and the subset of these positions that carried the alternative allele in the test sample. We then calculated the proportion of shared near-private sites of all observed near-private sites for each reference population as the summary statistic of shared rare alleles. This strategy has the advantage that even a large fraction of false alternative alleles in the test sample (sequencing error, contamination from other sources) would not necessarily cause an enrichment of shared alternative alleles with any reference population.

We first tested this strategy in a panel of 99 samples sequenced to less than 1-fold coverage, with less than 0.5% contamination, and with known locations within the reference panel. Out of these, 67 (68%) were correctly assigned to their reference population, and for 91%, the correct population was among the top three ranked populations (Fig. S67). We conclude that this method can be used to determine the approximate origin of samples even at a coverage of substantially less than 1-fold on only a single chromosome. Another panel of 139 samples with more than 0.5% human contamination, with known locations within the reference panel gave similar results (86% within the top three locations). Among these, only two samples (Bat1-14, 0.04-fold coverage, 0.9% contamination; Lop1-25, 0.22-fold coverage, 3.4% contamination) were assigned to the wrong subspecies (eastern instead of central chimpanzee). Hence, even samples with substantial human contamination (up to 21.2% for Kab2-4) can be used for this approach, even though spurious allele sharing with other sampling locations increases (see below). We find that neither endogenous DNA content of the sample, nor final sequencing coverage, nor contamination was a main factor for the correct assignment of the best matching site ($P > 0.05$, two-sided Wilcoxon rank test). The overall matching to sites from the respective subspecies is very high across samples (Fig. S68-S69), suggesting that the four known subspecies were sufficiently separated during the recent past, allowing for variation to arise within restricted geographic areas. Due to the size of these heatmap plots, we provide these in original resolution in an additional file (Data S1 – Fig. S95).

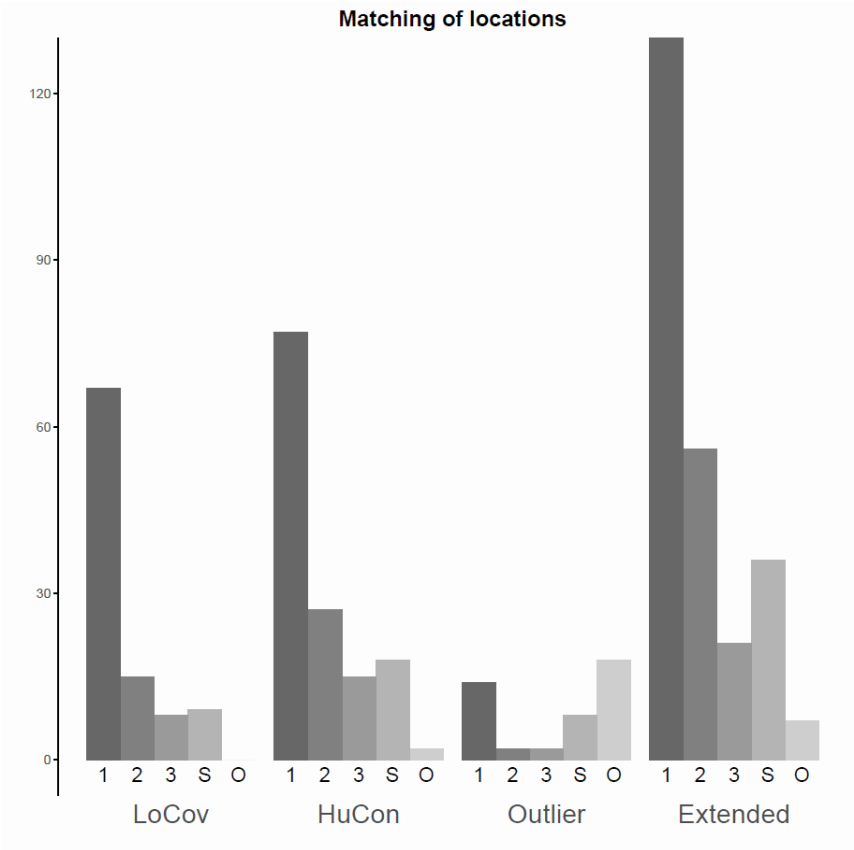


Fig. S67. Number of samples for which the first-, second- or third-best match was correctly assigned, and for the remainder whether or not the subspecies was correctly assigned (S) or not (O), stratified by low coverage samples (LoCov, n=99), human contaminated samples (HuCon, n=139), outlier samples (Outlier, n=44) and all samples in an extended reference set ("Extended", see next Supplementary Text Note, n=250). Related to Fig. 4.

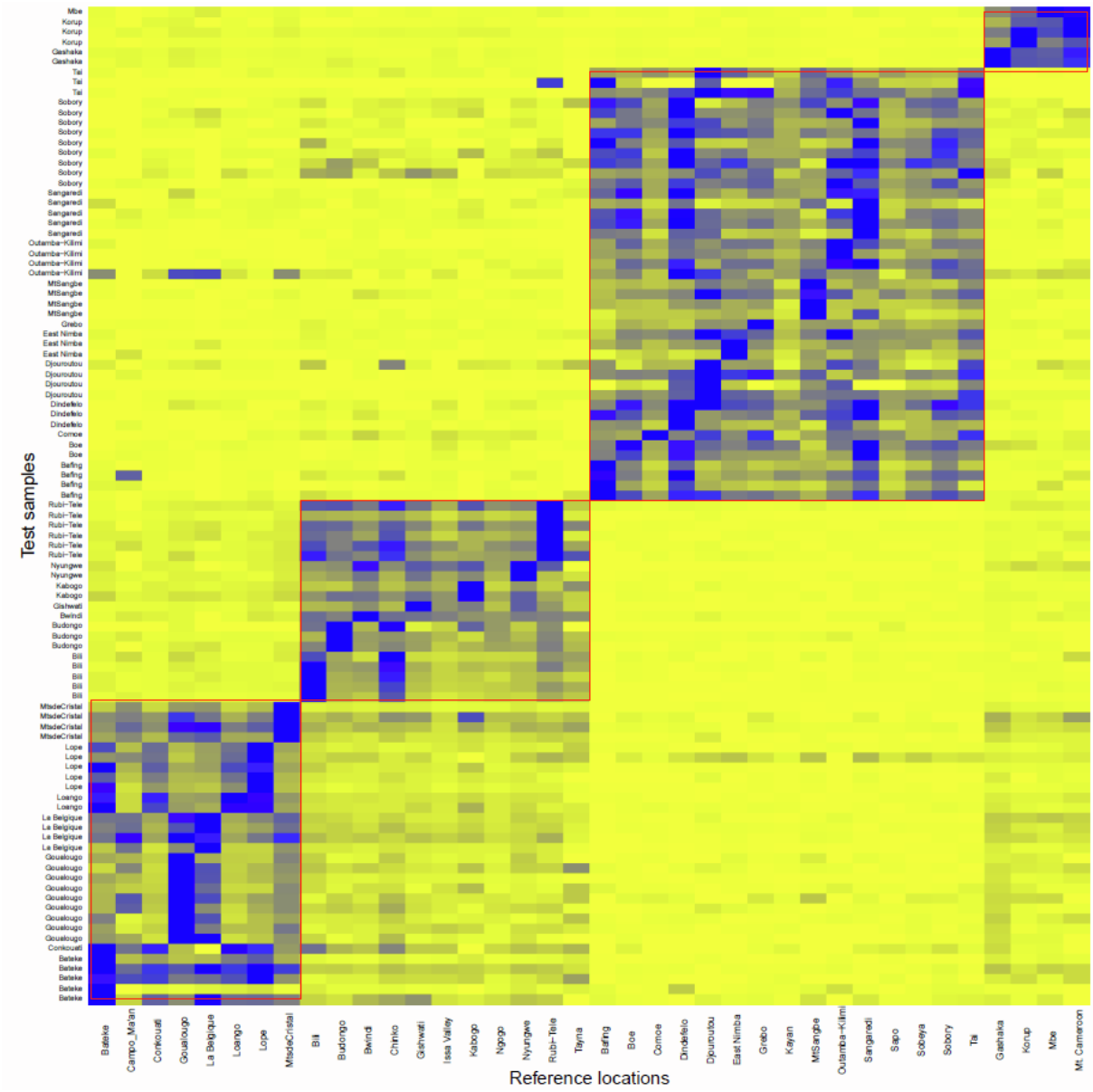


Fig. S68. Matching scores for each low-coverage test sample (y-axis) to all reference populations (x-axis). The four subspecies form clear clusters. Related to Fig. 4.

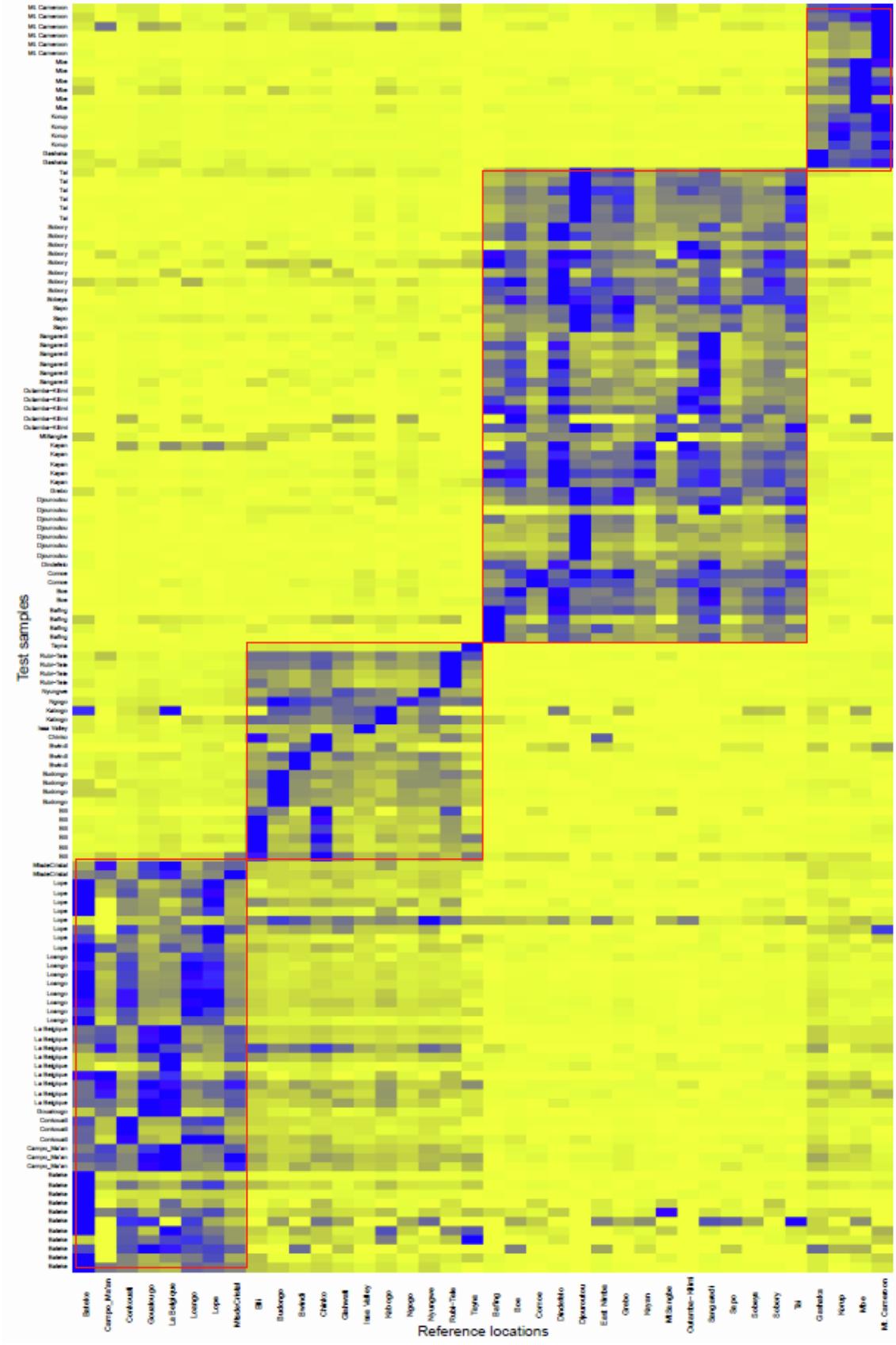


Fig. S69. Matching scores for each contaminated test sample to all reference populations. Again, the four subspecies form clear clusters. Related to Fig. 4.

This is also supported by observations from 26 low-coverage samples (lower than 1-fold) from sites which were not included in the reference panel since no individual was of sufficient quality (Ankasa, Azagny, Bia, Boundialé-Odienne, Ivindo, Ituri, Loma, Maiko, Ngiri, Regomuki). These samples follow the general pattern of rare allele sharing with populations of their subspecies (Fig. S70), suggesting that this method is, in principle, able to identify a proximate origin of samples whose populations were not included in the reference panel. Particularly samples of the eastern chimpanzee population of Ngiri show a clear match to other eastern chimpanzee populations, but not central chimpanzee populations, despite being geographically much closer to those. This suggests that recent genetic exchange between central and nearby eastern chimpanzee populations was at most very limited, as well as in the more distant past, reflected by the clustering in the PCA (Fig. S33), which relies rather on common variation. Samples from Bia in Ghana show the best match to populations in the neighbouring country Cote d'Ivoire.

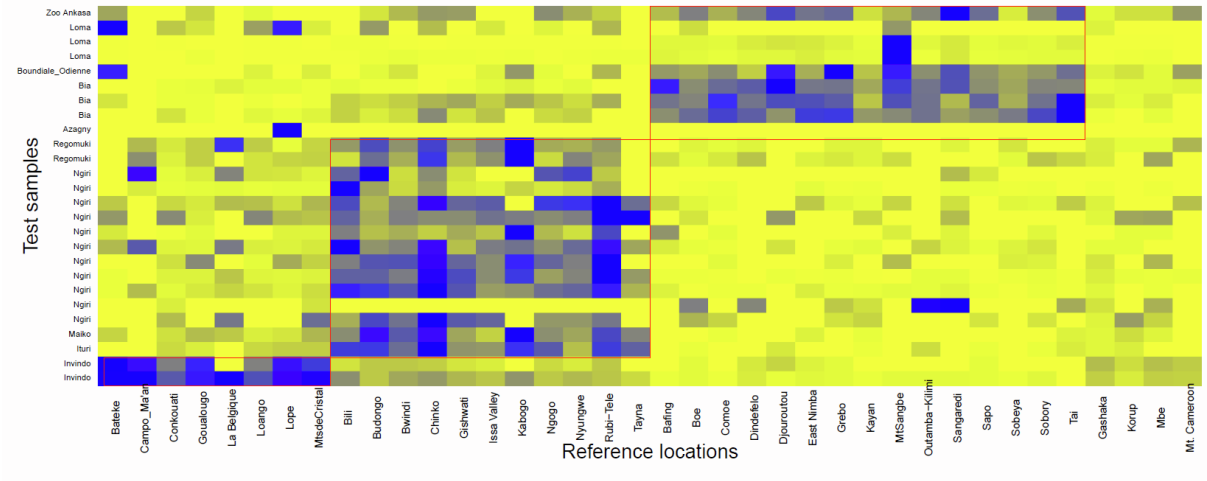


Fig. S70. Matrix of matching scores for samples not included in the reference panel (low coverage). Samples usually share more with their respective subspecies than with other subspecies. Related to Fig. 4.

We also computed the matching score for chromosome 21 of the 59 high-coverage chimpanzee genomes from previous publications[21]. These samples appear to represent the variation of extant chimpanzees, with samples from each subspecies matching strongly to sites belonging to their subspecies, and regional differences (Fig. S71). We find several individuals for which the country of confiscation or proximate origin is not represented by the best match, but rather in neighboring countries, likely because their population of origin was near the borders but not represented here or they were confiscated in a country different from their true origin: A912_Nakuu, B002_Padda, N013_Tongo (DRC) as well as B011_Frederike (Rwanda) are best matched by sites in Uganda; A960_Clara and A957_Vaillant are likely from Gabon, but best match sites in Congo; N016_Alice from Cote d'Ivoire matches best to Grebo in Liberia at the border to Cote d'Ivoire; A992_Annie from Guinea-Bissau is of doubtful origin and might be from a site in Liberia or Cote d'Ivoire according to our results; and

100040_Andromeda from Gombe in Tanzania matches Nyungwe in Rwanda slightly better than Issa Valley in Tanzania, suggesting that the genetic composition differs within Tanzania.

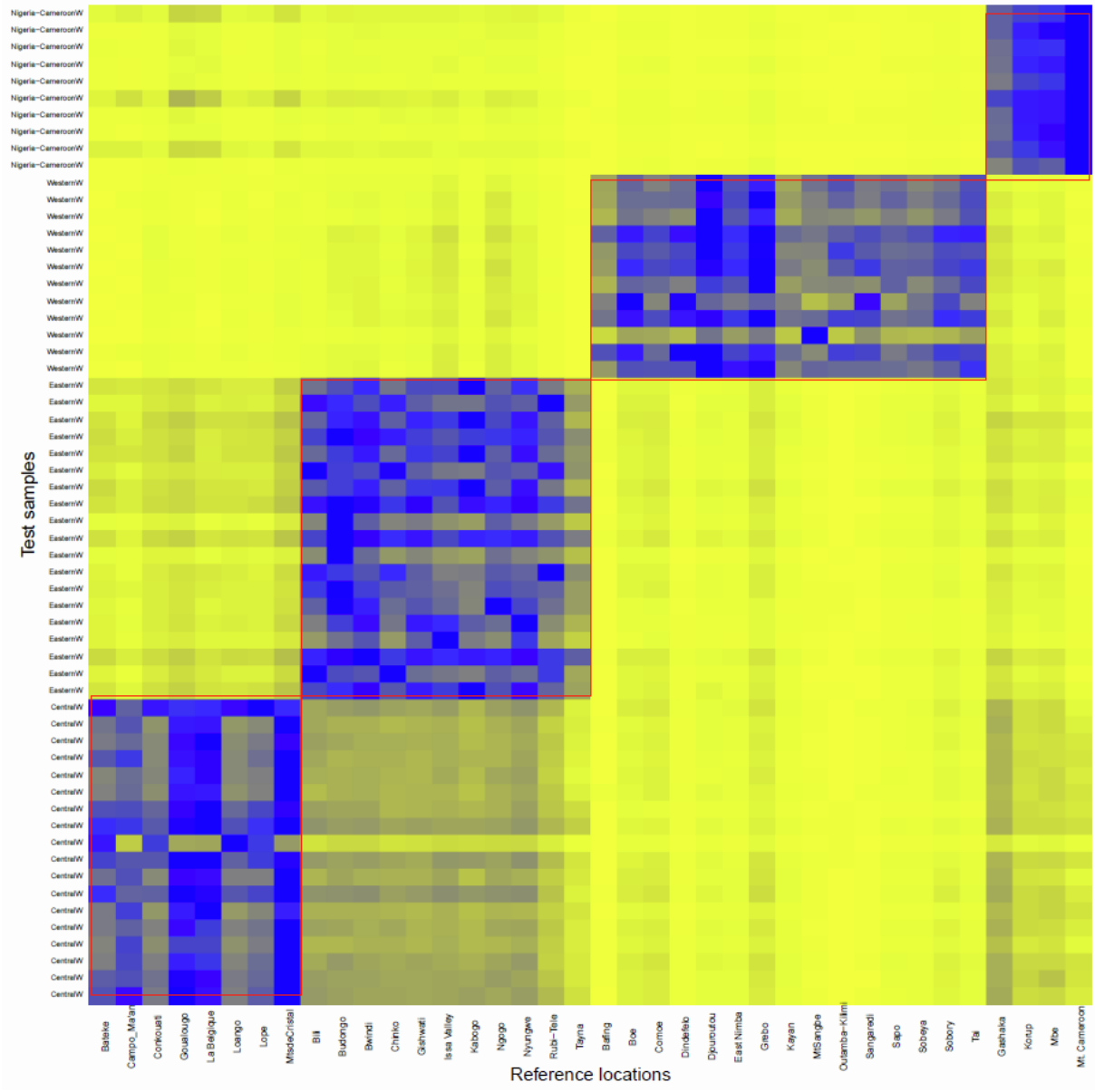


Fig. S71. Matrix of matching scores for high coverage genome samples from previous studies. Samples generally share more with their respective subspecies than with other subspecies. Related to Fig. 4.

9.2. Spatial representation

Based on the observation that near-private allele sharing is high between the different western chimpanzee populations, it seems possible that regional substructure within the range of this subspecies could be used to determine a region of likely origin, rather than relying on the best match only. The dense sampling scheme in this project allows the use of the known geographic coordinates of the sites for an explicit spatial model of these matching scores using the R packages sf[62], sp[63,64], and maptools[65]. We used the known current distribution of chimpanzees according to the IUCN[66], and expanded this

range by 1.5° in all directions to create a spatial grid in which to estimate the geographic areas. We then used the geographic coordinates of the known locations to create a simple variogram of the formula (match score) ~ X + Y, using the R package gstat[67], and fit the variogram using a spherical model and standard parameters, but without fitting ranges. Finally, we performed kriging to the expanded geographic range using the krige function in gstat. The focal point was estimated as the highest point within the surface (usually near the best matching location), in order to calculate the distance to the correct origin of samples. All maps included in this supplementary text are available in original resolution in an additional file (Data S1 – Fig. S95).

We provide maps of this match score surface for the 99 low-coverage samples (Data S1 – Fig. S96), as well as 139 samples with human contamination of more than 0.5% (mean 2%, 0.5-21%) (Data S1 – Fig. S97). The visual inspection of these maps clearly shows an alignment with the effective migration matrices (Supplementary Text Note 10.1), shared drift statistics (Supplementary Text Note 7.1) and general patterns of allele sharing. Samples from the northern part of the central chimpanzee range (Goualougo, La Belgique, Mts de Cristal and Campo Ma'an) tend to share rare alleles, where the power is high for samples from Goualougo, but less so for La Belgique and Camp Ma'an, which are rather assigned to this strongly connected region. The same is true for the southern part of their distribution (Batéké, Lopé, Loango, Conkouati). In eastern chimpanzees, we can determine the origin of samples from the Bili/Chinko area in the north, distinguished from Rubi-Télé/Budongo, distinguished from Bwindi in Uganda; Nyungwe and Gishwati in Rwanda are fairly well distinguishable, was well as Kabogo in the southern DRC. In Nigeria-Cameroon chimpanzees, samples are usually assigned well, despite their relative geographic proximity. Finally, in western chimpanzees, we are able to identify regions of origin and sites for which samples match strongly. Samples from Bafing in Mali are identified with high confidence, while in a north-western cluster samples from the sites of Boé, Dindefelo, Kayan and Sangaredi are often matching similarly well to all locations in this region, and particularly Sobory-Bakoun stands out as apparently highly connected to these sites. Another coastal southern cluster of Taï, Djouroutou, Grebo and Sapo consists of a highly connected area. Interestingly, samples from Sierra Leone and East Nimba in Liberia appear to share rare variants with both clusters, possibly as a result of past connectivity with both. Samples from the sites of Comoé in north-eastern Cote d'Ivoire and Mt Sangbé in north-western Cote d'Ivoire seem to share less rare variation with others, likely due to past isolation. These patterns agree well with observations from the PCA (Supplementary Text Note 5.1).

Generally, samples at very low coverage (less than 0.1-fold) tend to be poorly assigned. This is confirmed by analyzing the average distance to the true location in bins of coverage (Fig. S72). For samples below 0.1-fold coverage, the average distance is 241km, while samples at a coverage between 0.1-fold and 1-fold are on average determined within 81 km of the true origin. Contaminated samples

at any coverage are located, on average, within 157 km, even though higher sequencing coverage seems to increase accuracy as well (Fig. S72). We also find that sampling locations with few individuals of relatively low coverage used are performing poorly in the assignment (Campo Ma'an, Tayna), suggesting that increasing the sample sizes to ~10 individuals per site with a coverage of substantially more than 1-fold may greatly improve the accuracy of this geolocalization approach.

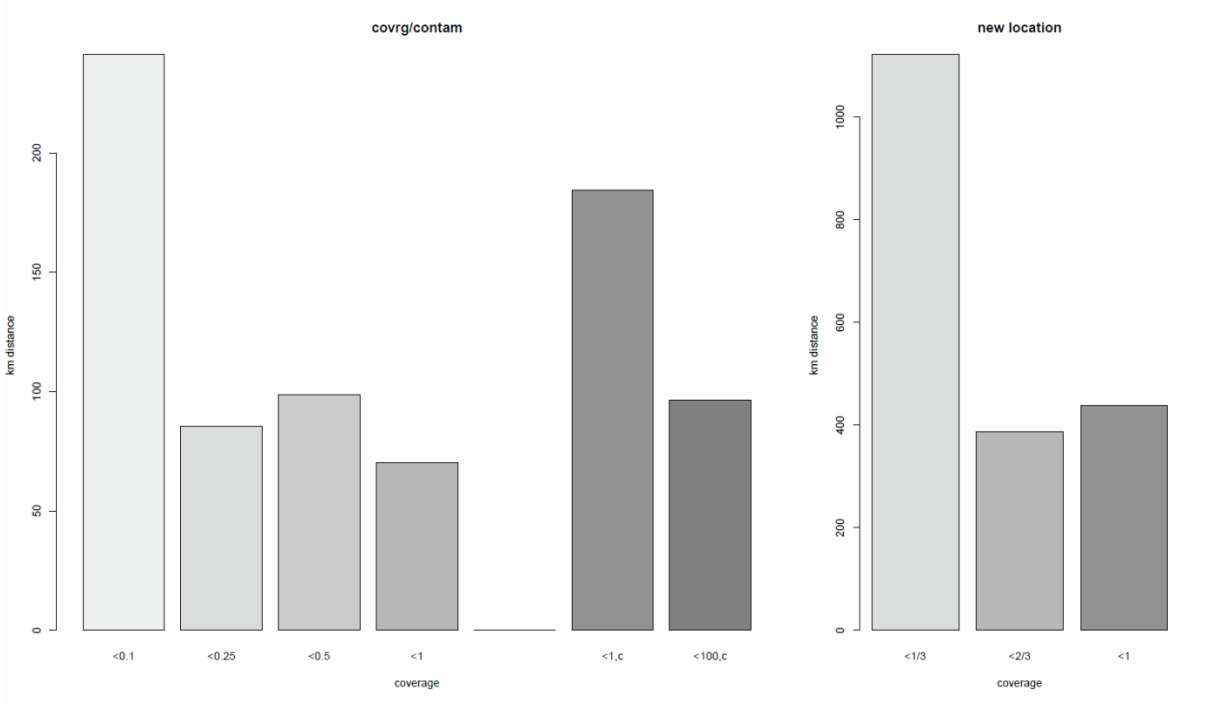


Fig. S72. Average distance of best matching to true location in bins of coverage, for samples with human contamination (">c"), and for locations not included in the reference panel. Related to Fig. 4.

We compared our findings to results from a previous study in elephants to geolocalize confiscated samples using allele frequency data of microsatellites[59]. The closest comparison here would be to the "Single Sample" test (Table S3 in Wasser *et al*.[59]), in which a single sample was left out of the calculation, and subsequently assigned to a geographic origin. Above, we add new samples from the same location, notably those that did not meet the quality criteria for inclusion the reference set. In elephants, 75% of samples were assigned within 552km (forest elephant, F) and 840km (savannah elephant, S). Here, across all subspecies, we assign 75% of low coverage samples within 144km, and 75% of contaminated samples within 217km, considerably closer than in elephants. We also note that the median (distance for 50% of tested samples) is ~10km in our case (exactly matching the sampling location), while it is 349km (F) and 491km (S) for elephants. To further compare our approach to the one taken in elephants, we re-calculated private alleles with leaving whole sampling locations out, i.e. determining those alleles that are rare in 37 locations, and testing the individuals from the 38th location,

which is analogous to the "leave-full-location-out" cross validation in Wasser *et al*. (2015)[59]. This analysis provides estimates for the expectation of new sampling locations (within the range of the sampling scheme). Again, we use the 75th percentile of distances to the real known origin and compare this to the observations in elephants (Table S3 in Wasser *et al.*[59]). We find that this metric is 378km for all chimpanzee samples, while it is 822km (F) and 1026km (S) for single elephant samples, and still 648km (F) and 557km (S) when using a method that exploits much larger groups of elephant samples simultaneously. We note that even the 90th percentile of distances from the true location is smaller for our approach in chimpanzees (537km) than the lowest 75th percentile in elephants (557km). The samples are usually assigned to the closest sampling location in the remaining dataset, and generally to the broader geographic regions described above (Data S1 – Fig. S98).

The 26 samples belonging to sites not included in the reference panel show similar patterns (Data S1 – Fig. S99). However, these samples are peculiar in their low coverage and unusual distance to the closest sampling location. Western chimpanzee samples from Loma in Sierra Leone have the highest affinity, perhaps surprisingly, to Mt Sangbé. Samples from the Bia site in Ghana show varying affinities to other western chimpanzees, with the strongest matching to those from the southern coastal cluster. Hence, the landscape of matching is somewhat informative whether a sample clearly originates from a specific geographic area, or shares more generally genetic material with a broader area. Other samples from Ghana (Ankasa - Zoo chimpanzee, Azagny) show less clear patterns, likely due to high contamination and extremely low coverage. The sample from Boundialé-Odienne appears to share alleles across subspecies, suggesting a substantial fraction of non-chimpanzee DNA, but within western chimpanzees matches best to the southern coastal cluster. Samples from the central chimpanzee population of Ivindo show affinities to other nearby central chimpanzee sites, Ituri to eastern chimpanzee sites from northern DRC, Maiko and Regomuki from central DRC to the southern DRC. Finally, samples from the site of Ngiri at the western fringe of the eastern chimpanzee distribution show a strong rare allele sharing with other eastern chimpanzee populations, particularly from the northern DRC, but with an increased sharing also with sites in the southern DRC. This suggests that in the recent past Ngiri was not strongly connected to geographically close central chimpanzee populations, but remained genetically equidistant to other eastern chimpanzee sites. Small amounts of allele sharing with central chimpanzees in some of the samples from Ngiri may point to a small degree of connectedness in the past, although we caution that this is difficult to interpret at the very low coverage of these samples (lower than 0.1-fold), with patterns similar to those seen in other samples. The large distance between Ngiri and other eastern chimpanzee sites causes a large average of 1011 km between best matching and true location for all non-reference panel samples. Filling in gaps in the sampling will be important to improve the precision of this approach at the fringes of the subspecies, considering that local genetic variation outside the range of the covered distribution within each subspecies cannot be assigned.

An additional batch of 59 samples was removed from previous analyses due to the filtering, mostly being identified as outliers in the initial PCA (Supplementary Text Note 3.3), with usually large amounts of human contamination (0.4-91%, on average 24.6%). Many of these samples may also carry contamination from a non-human source (through diet or environment), or other factors like DNA degradation, but it seems possible that they still contain sufficient amounts of endogenous DNA to be exploited. We find that 18 out of 44 samples (41%) within the reference panel could be correctly assigned to the top three best matching sites (Data S1 – Fig. S100). We identify samples that show allele sharing across the different subspecies, most likely due to a non-chimpanzee nature of these samples. This interpretation is supported by calculating the rare variant sharing with a panel of all other great apes (bonobo, gorilla, orangutan, human; Data S1 – Fig. S103), where we find that gorilla genomes match well to different sites across subspecies, particularly in central chimpanzees, and human genomes specifically well to some western and some eastern chimpanzee sites. This lends additional support to the notion that the initial PCA filters out non-chimpanzee samples (or contaminated samples), while this method using rare alleles can be used to identify mixed samples. Another 15 samples outside the reference panel show similar patterns (Data S1 – Fig. S101) as the 26 samples which were assigned with higher confidence (see above). Other samples from Bia in Ghana possibly share rare variants with sites in western chimpanzees, while all samples from Azagny seem to be largely of non-chimpanzee origin (mostly human, with contamination estimates >10%).

We also inferred the approximate origin of previously published chimpanzee whole-genomes, for some of which information was available[21], and which we complement using our analysis (Data S1 – Fig. S102). We find a good agreement with the known place of origin or confiscation. For 42 individuals, geographic coordinates with a certainty of less than 5° were provided[21]. We calculated the distance of the inferred origin to these coordinates, finding that 75% of samples lie within 452km, which is considerably closer than for elephants, where 75% of tested samples from known locations were inferred within 557km (S) and 648km (F) when large groups of samples are available, which is not expected for confiscated chimpanzees. Note that these coordinates sometimes lie outside the chimpanzee range, for example, the eastern chimpanzee Diana was confiscated in Zambia, but most likely originates from the southern range in the DRC, and we find it most closely related to the sampling location Kabogo.

Finally, we attempted to incorporate locations with at least three individuals with at least 0.1-fold coverage (less than 1.5% contamination, not PCA outliers) into the model (one sample from Ivindo, Boundialé-Odienne and Loma, two from Ngiri and Bia), since some of these are at the fringes of the distribution of their respective subspecies. We calculated the near-private variants and inferred the origin of the remainder of 250 samples of low coverage or with high contamination and 7 samples from other locations (Data S1 – Fig. S104) as well as the data from whole genomes (Data S1 – Fig. S105).

We find a correct assignment for 128 (53%) of these samples, and 211 (84%) with a correct assignment among the top 3 locations (Fig. S67), suggesting that the inclusion of these samples may introduce a higher level of noise to the model and, on average, larger distances to the correct origin (Fig. S73). Seven out of nine remaining samples from Ngiri were correctly assigned (Fig. S74), despite being at less than 0.1-fold coverage. Interestingly, samples from Maiko and Regomuki in central DRC might have an unexpectedly higher affinity to Ngiri than samples from other parts of the DRC. Only two samples from Bia (Gha-01-08, Gha-01-04) were left for testing, one of which was assigned to other western chimpanzee populations. The sample from Ivindo was located to surrounding locations. One sample from Loma was assigned correctly, while another was assigned to Lopé in Gabon, likely a result of mislabeling. Several PCA outlier samples for Loma and Bia were correctly identified (Fig. S75), despite high levels of human contamination (>10%). We conclude that it would be desirable to obtain samples of sufficient quality for populations at the fringes of the chimpanzee distribution, and that our model will most likely have a similar power to correctly assign individuals from these populations. However, the overall performance of the model is better when considering only individuals of a coverage of more than 1-fold, as expressed by numbers of best matching samples and distance to the true known origin (mean 372km, median 19km, 75th percentile 398km, 90th percentile 1083km for 250 samples in the extended set vs. mean 130km, median 9km, 75th percentile 208km, 90th percentile 362km for 238 samples in the high-quality set above).

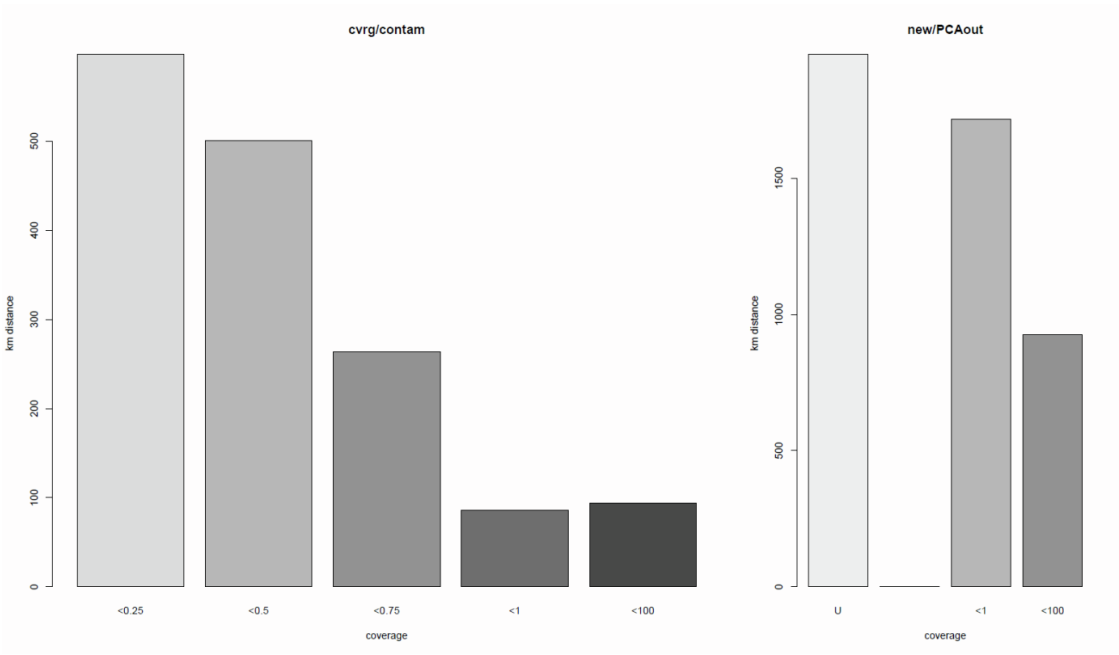


Fig. S73. Average distance to known origin for samples at low coverage or high human contamination when using low-coverage samples from locations without high-coverage samples (Ngiri, Bia, Ivindo, Loma). Related to Fig. 4.

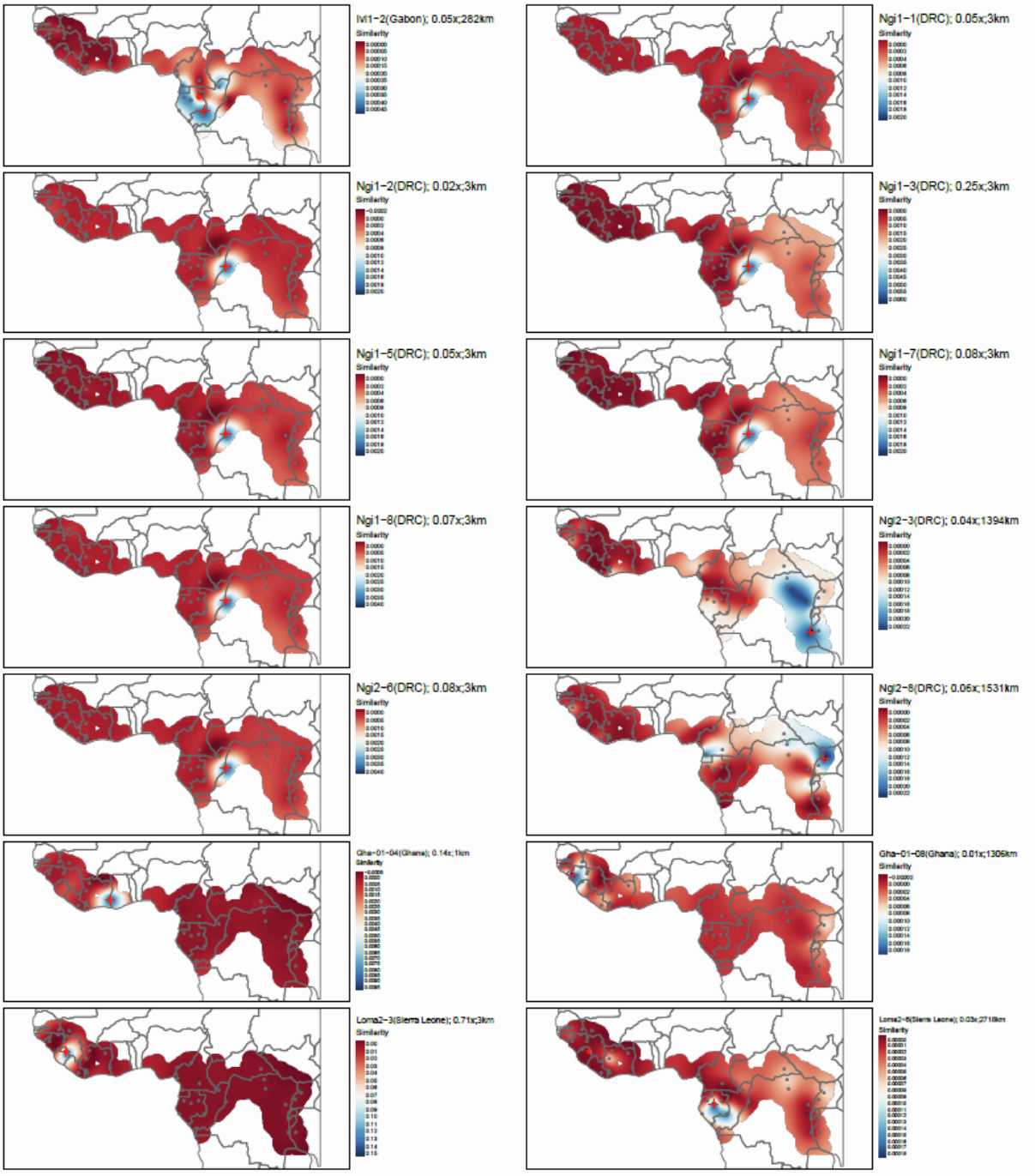


Fig. S74. Spatial representation of near-private allele sharing for samples from Ngiri, Ivindo, Bia and Loma in a model that includes the best samples from each of these locations, additionally to the samples from the previous model (compare Data S1 – Figs. S98 and S104). Related to Fig. 4.

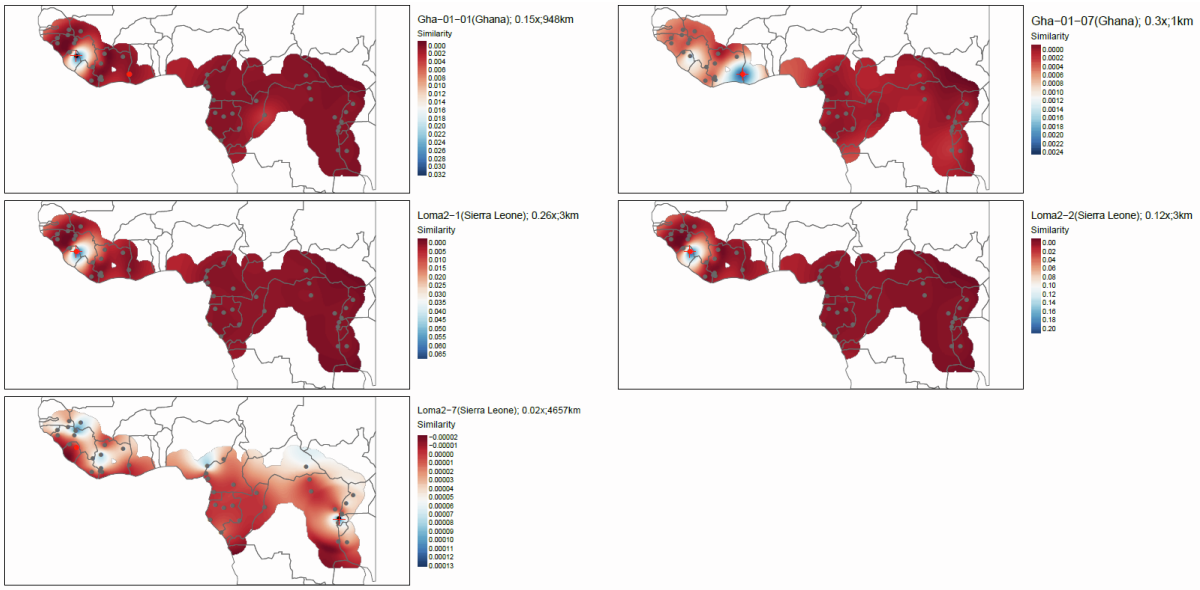


Fig. S75. Spatial representation of near-private allele sharing for PCA outlier samples from Bia and Loma in the extended model. Related to Fig. 4.

9.3. Connectivity based on rare variants

Rare alleles can be used to assess the connectivity between geographic regions in the past. While most likely connectivity patterns obtained with EEMS represent migration landscapes before the last 6,000 years[68], and IBD segments more recent genetic exchange represented by shared haplotypes (for example, ~50 generations/~1250 years for segments of 1 cM/Mbp[58]), rare variants likely allow an intermediate perspective, overlapping with both timeframes, between 1.5kya and 15kya[61].

Since the rare variants used in our approach are not necessarily fixed at a given location, but can be present at other locations (see above), the pattern of shared rare variants is informative for past connectivity. We calculated the proportion of derived variants in a given population shared with all other populations. We then used these data points to infer a landscape of sharing with other populations, and applied the kriging procedure described above, where we left out the test population from the landscape. We excluded Campo Ma'an, for which only one individual was observed, with no rare allele sharing with any other site. We also performed the analysis with the extended reference dataset, which allows including Ngiri and Bia at the fringes of the respective distributions of their subspecies. The locations of Boundialé-Odienne, Loma and Ivindo with only one sample were excluded as well due to absence of sharing.

When applying this approach to central chimpanzees (Fig. S76), it appears that the southern region (Batéké, Conkouati, Loango, Lopé) forms an area of extensive allele sharing, while the northern group (Goualougo, La Belgique, Mts de Cristal) shares more rare variants among each other, but possibly to differential degrees according to their distribution. This agrees well with the results from

other methods in this manuscript, notable IBD-like tracts (Supplementary Text Note 10). Furthermore, the northern group appears to have a relatively larger number of shared alleles with particularly the Gashaka site from Nigeria-Cameroon chimpanzees. This might be the result of a recent pulse of migration, even though we cannot determine the direction of such an event. We also see the complementary pattern when analyzing the shared rare variants in Nigeria-Cameroon chimpanzees (Fig. S77), where only Gashaka shows an increased pattern of sharing with central chimpanzees, but not the other Nigeria-Cameroon sites, suggesting that this event might have been limited to this particular area.

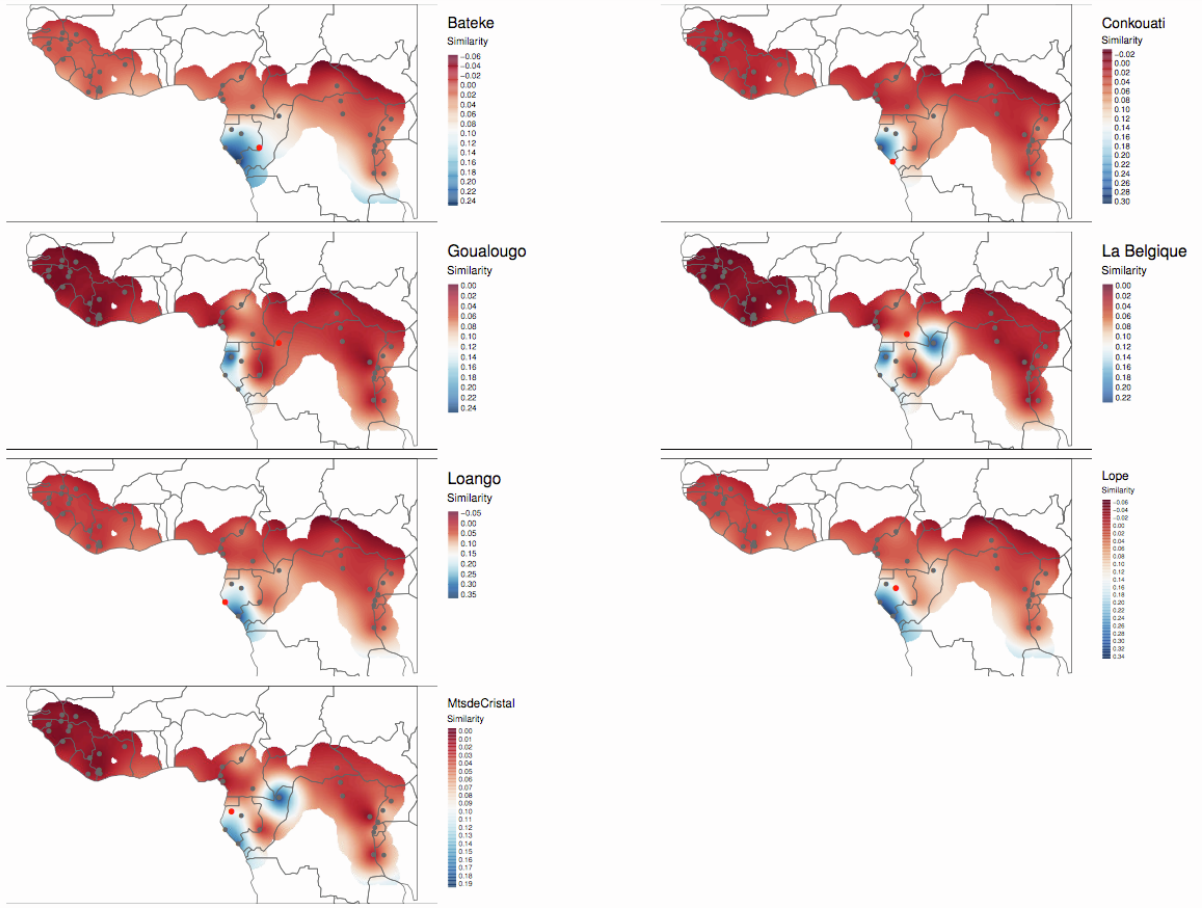


Fig. S76. Rare allele connectivity in central chimpanzees. Blue color represents higher affinity and red color represents lower affinity. Black dots: Sampling sites used as reference. Red dots: Tested population. Related to Fig. 2.

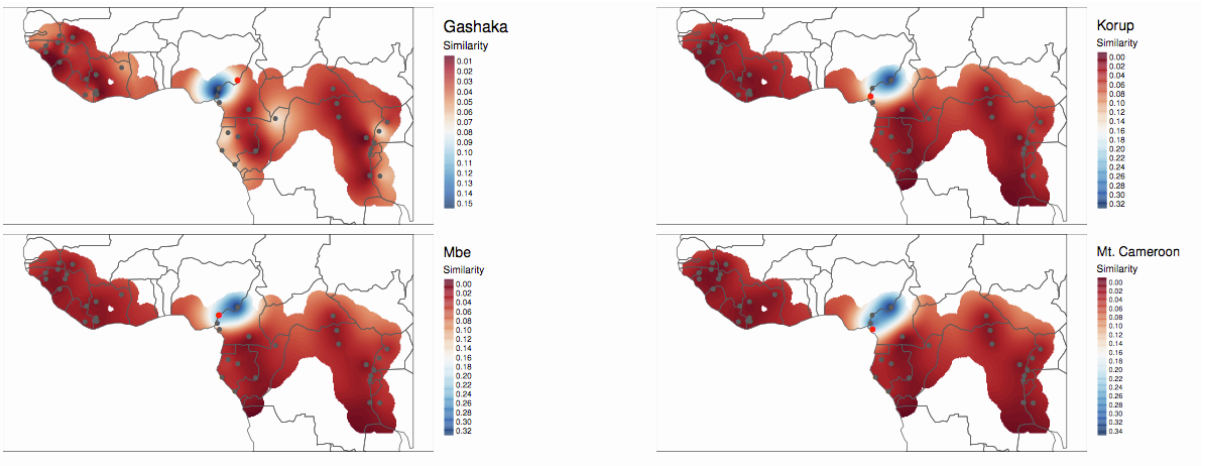


Fig. S77. Rare allele connectivity in Nigeria-Cameroon chimpanzees. Blue color represents higher affinity and red color represents lower affinity. Black dots: Sampling sites used as reference. Red dots: Tested population. Related to Fig. 2.

In eastern chimpanzees, we see rare alleles common across populations in the central-eastern part (Budongo, Bwindi, Gishwati, Ngogo, Nyungwe), while other populations to the south (Issa Valley, Kabogo), to the west (Tayna) and to the northwest (Rubi-Télé, Bili, Chinko) share alleles with this "core" of variation (Fig. S78). Likely, these populations received alleles from these eastern rift valley populations, but contributed less recent variation back to them, which may suggest a rather recent population expansion from the central areas of the eastern chimpanzee population. These observations provide a complementary perspective to the more recent fine-scaled landscape of connectivity from IBD segments (Supplementary Text Note 10). In the extended dataset including the sampling site of Ngiri, these general patterns do not change (Data S1 – Fig. S106). The population from Ngiri does share rare variants with northwestern (Chinko, Bili, Rubi-Télé) as well as central areas of the eastern chimpanzees, but not with nearby central chimpanzee populations to a greater extent than other eastern chimpanzees. This lends further evidence to a scenario in which the western fringe of eastern chimpanzees did not constitute a gradient nor a hybridization zone, in line with results from other analyses.

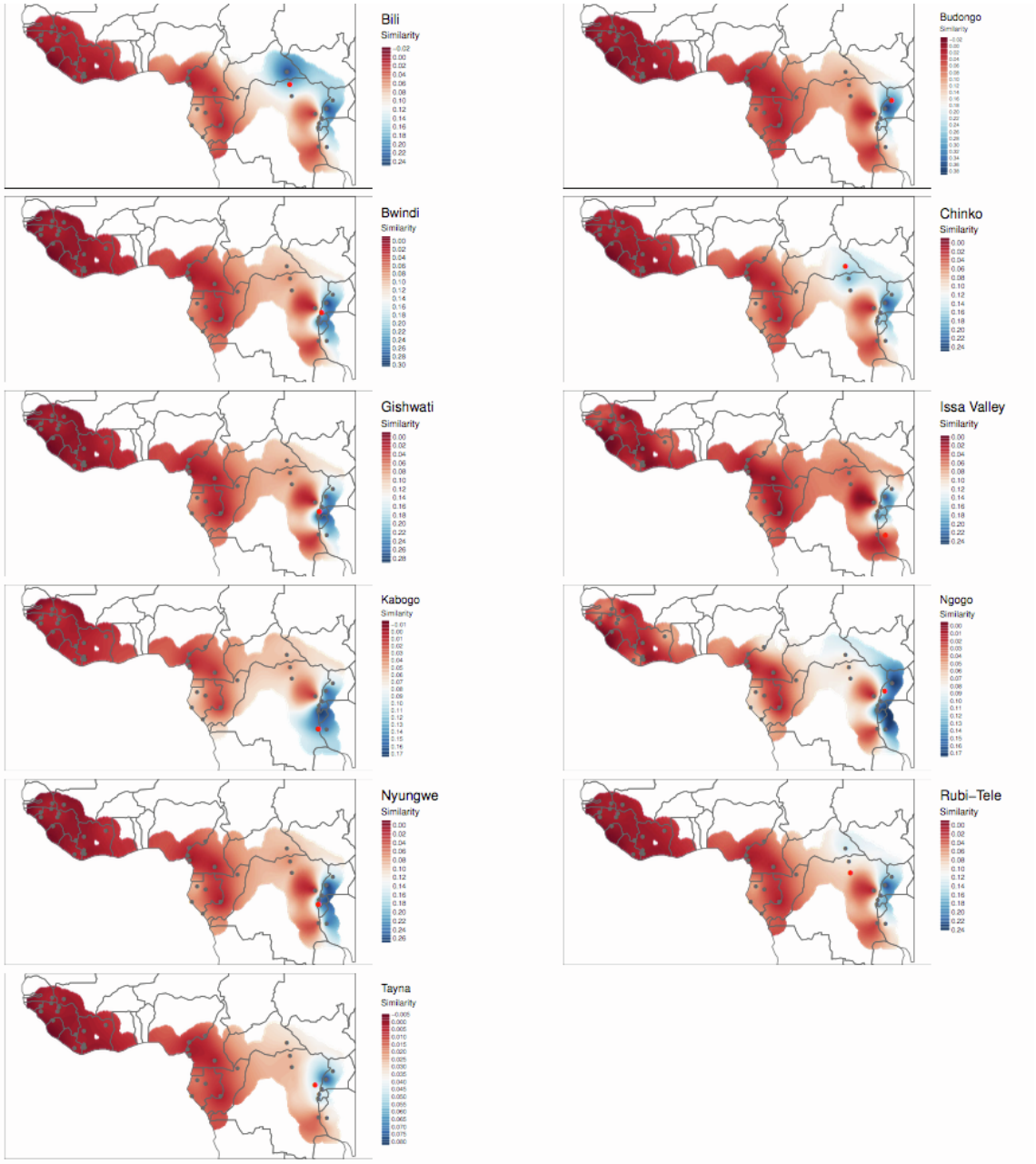


Fig. S78. Rare allele sharing in eastern chimpanzees. Blue color represents higher affinity and red color represents lower affinity. Black dots: Sampling sites used as reference. Red dots: Tested population. Related to Fig. 2.

Finally, western chimpanzees show areas of connectivity, resembling results from other methods (Supplementary Text Note 10). Specifically, almost all areas seem to have been connected with Comoé, while the nearby site of Bia particularly shares alleles mostly with Comoé. The southern area (Taï, Grebo, Sapo, Djouroutou) seems to show a corridor of connectivity to parts of the northwestern area (Boe, Dindefelo, Kayan, Sobory-Bakoun, Sobeya), but less so to Bafing and Sangaredi in that area. Possibly, parts of the northwestern range have been subject to recent expansions into the fringes, where rare variants are shared more with nearby locations, which is the case for Bafing and Sangaredi (Fig. S79). The intermediate area shows differences between the locations: Outamba-

Kilimi in Sierra Leone appears to share rare variants with other locations, while these locations share less with that site than with others. In contrast, northwestern and southern sites both share more rare variants with East Nimba than Outamba-Kilimi, suggesting that the central area might have been a corridor of genetic exchange between the two areas. This is supported by shared IBD segments, which point to two strongly connected regions, both of which are represented in East Nimba (Supplementary Text Note 10). The nearby location of Mt Sangbé seems to have been isolated recently, as suggested by a depletion in IBD-like segments, as well as poor allele sharing of other sites when compared to East Nimba, while the EEMS algorithm does not show a signature of long-term isolation.

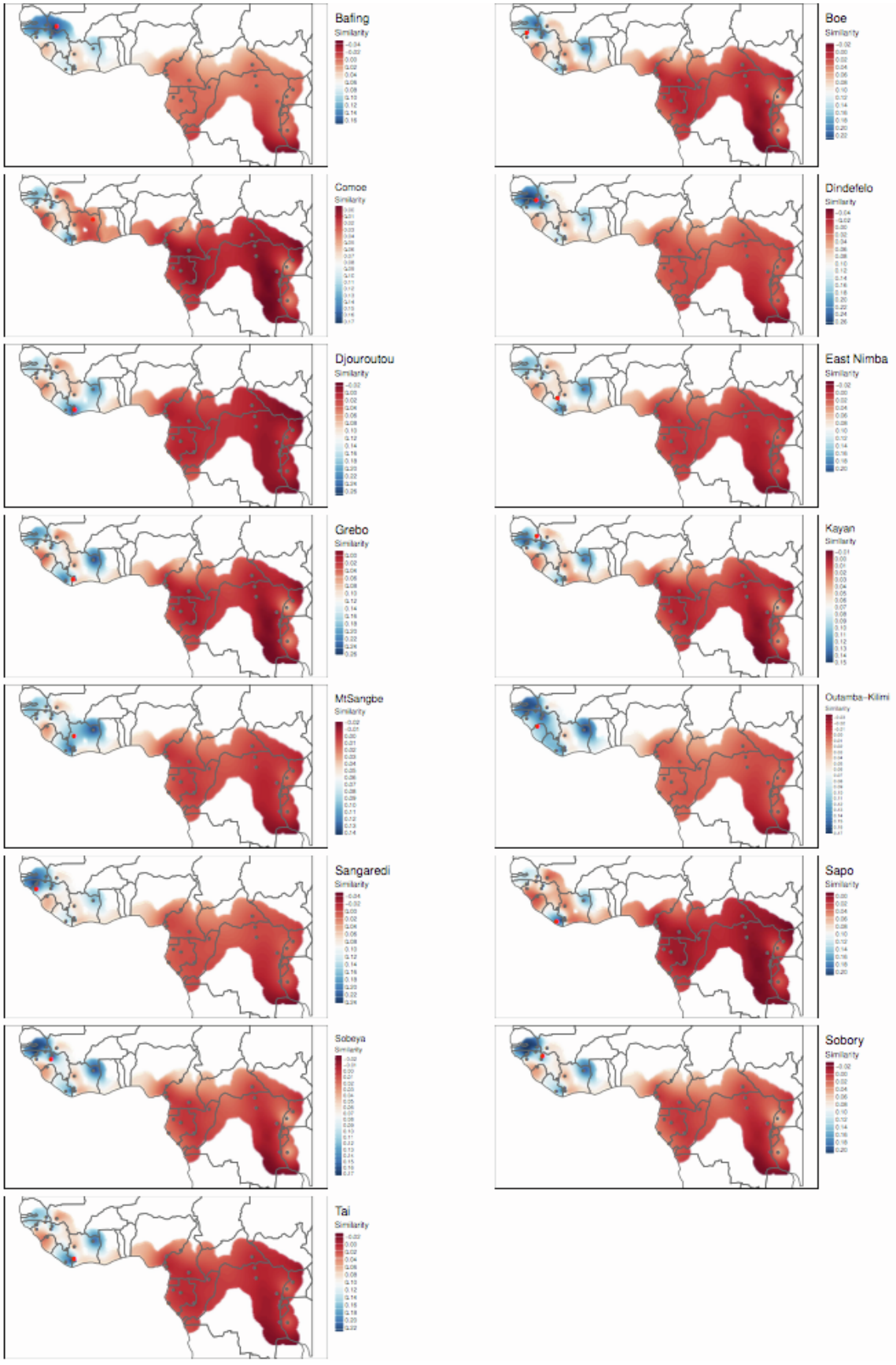


Fig. S79. Rare allele sharing in western chimpanzees. Blue color represents higher affinity and red color represents lower affinity. Black dots: Sampling sites used as reference. Red dots: Tested population. Related to Fig. 2.

9.4. Geolocalization of samples from sanctuaries

We tested this approach for geolocalization based on rare alleles on other chimpanzee samples. We sequenced at low coverage 20 chimpanzees from two rescue centers in Spain: Centro de Rescate de Primates Rainfer (http://rainfer.org) and Fundació Mona (https://fundacionmona.org/).

We extracted DNA from 20 samples, using the standard phenol-chloroform protocol for 7 blood samples and 13 hair samples (Table S8). Samples were quantified using BioAnalyzer and DNA was sheared using a Covaris S2 ultrasonicator. Next, we prepared a single library for 15 samples following the same protocol as explained in Note 1.3. For the remaining five samples, we followed the custom dual-indexed protocol[9] (Table S8). Libraries were amplified with PCR and pooled equi-molarly to be sequenced into a 1 NovaSeq lane 2x100bp. The methodology described in detail in Supplementary Text Note 2.1 for data processing was also applied here: FastQ files were trimmed to remove adapters, and sequences mapped to the human genome (Hg19); duplicated reads, secondary alignments and reads with mapping quality lower than 30 were removed. The median mapping rate of the reliable reads (after filtering) was 75.54% (range 33.90%-77.62%), blood samples had a high median mapping rate (76.92%) compared to hair samples (47.92). We reached a median coverage of 0.35-fold, ranging from 0.15-fold to 4.3-fold (Table S8). We assessed human contamination, finding an average of 0.254% (0.055-0.972%) in blood samples, and slightly higher estimates of 1.57% on average (0.078-4.297%) in hair samples (Table S8).

Genotypes were called on chr21 using the software snpAD, as described in Supplementary Text Note 2.2, due to the low coverage and the possibility of biases due to DNA degradation particularly in the hair samples. Genotypes from each sample were merged using GATK 'CombineVariants' with option -genotypeMergeOptions UNIQUIFY. We excluded genotypes with less than 2-fold or more than 99-fold coverage, genotype quality less than 20 and non-biallelic SNPs from the analysis. We applied the procedure described above to data from these samples: We obtained the overlap of each individual with the near-private alleles from each location, calculated the proportion of matching variants, and applied the spatial model fitting. We find good matches for each sample, clearly detecting a best subspecies and a region within: Cheeta_Mona and Ivan_Rainfer appear to originate from the Southern Cote d'Ivoire/Liberia region, Guille_Rainfer and Maxi_Rainfer from the northwestern range of western chimpanzees, Judi_Rainfer from the northern group and Toti-Rainfer from the southern group of central chimpanzees, and Jackie-Rainfer from the northern DRC (Fig. S80).

Most likely due to the lower coverage (<0.5-fold), lower complexity and possibly biases from DNA degradation, we find more noise in modeling for hair samples. However, we are still able to infer a proximate region of origin for each of these samples (Fig. S81): One sample (Lulu_Mona) likely originates from the northern range of central chimpanzees, while eight samples likely originate from

their southern range (possibly near Equatorial Guinea, the only sub-saharan African state with recent colonial ties to Spain). Two individuals likely belong to Nigeria-Cameroon chimpanzees (Toni_Mona and Bea_Mona), and two most likely to the northwestern range of western chimpanzees (Africa_Mona and Sammy_Rainfer).

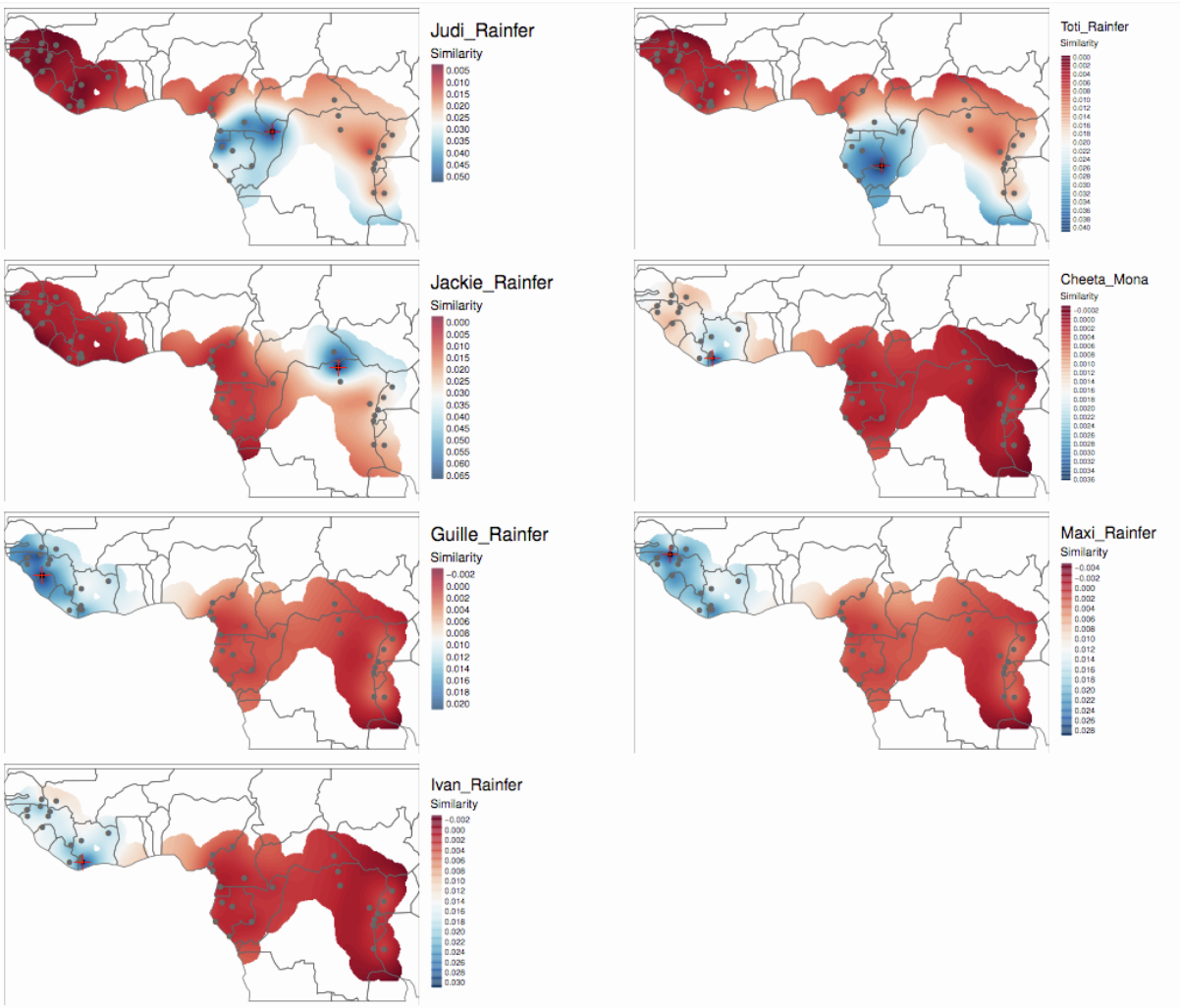


Fig. S80. Spatial matching of 7 chimpanzee blood samples from sanctuaries. Related to Fig. 4.

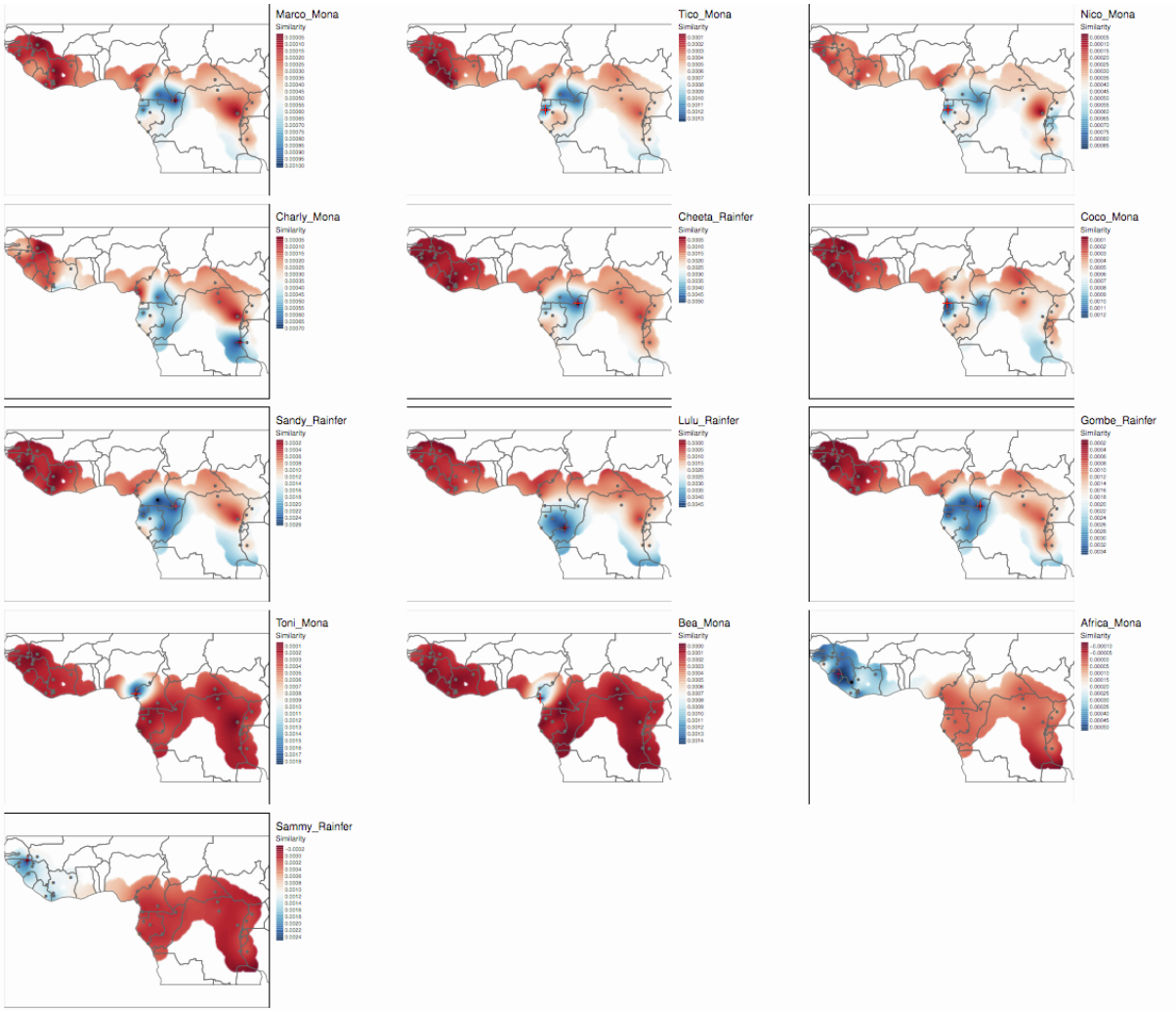


Fig. S81. Spatial matching of 13 chimpanzee hair samples from sanctuaries. Related to Fig. 4.

Note 10. Connectivity, isolation and migration.

10.1. Estimating Effective Migration Surfaces (EEMS)

We applied EEMS[69] to infer past patterns of migration between chimpanzee populations. This program calculates effective migration surfaces, and provides a visualization of potential regions of higher-than-average and lower-than-average historical migration between those sites. Here, we used only samples with more than 5-fold coverage (Table S1). The VCF file was filtered to keep only biallelic sites, at a minimum depth of 3, minimum genotype quality of 20, and allowing for 20% missing sites. We ran the program with all samples to obtain the overall measurement taking into consideration all subspecies. We obtained ten replicate runs of EEMS with the following parameters: nIndiv = 213, nSites = 1112443, nDemes = 2000, diploid = TRUE, numMCMCIter = 2000000, numBurnIter = 1000000, numThinIter = 9999. Since EEMS outputs a relative measure of migration and diversity compared to the rest of samples, we also used this program to study each subspecies independently, as well as all non-western chimpanzee sites together, given that there has likely been connectivity in the past[21,37]. Next, we plotted the results in R with the package rEEMSplots[69]. We also replicated this 10 times, using the same parameters as before except for nIndiv = 102 and nSites = 1098640. When plotting the whole range of the chimpanzee distribution, we observed that there is more effective migration than average between western chimpanzee sites when compared to the other subspecies (Fig. S82A). We also observe significant barriers of gene flow between subspecies, as well as within non-western chimpanzee subspecies (Fig. S82B).

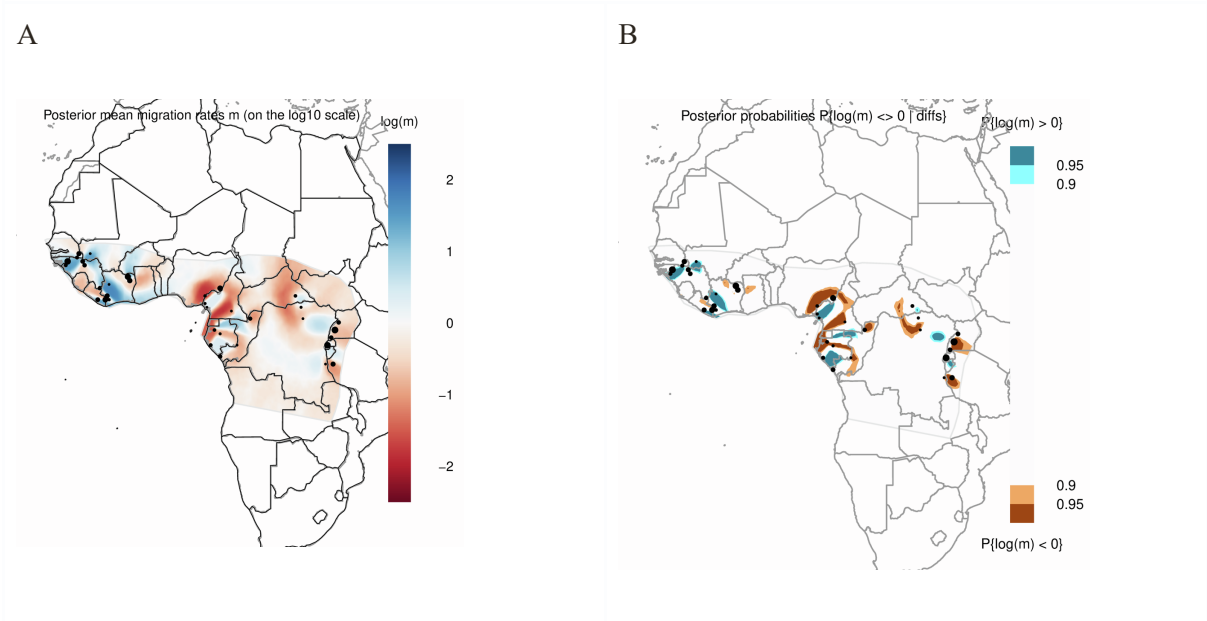


Fig. S82. EEMS between all samples and subspecies in the PanAf dataset. (A) Posterior mean migration rate (B) Posterior probabilities. Related to Fig. 2.

We ran EEMS on non-western chimpanzee sites, since their geographical distribution is connected, and to determine if there were any significant barriers separating them, or on the contrary if there was gene flow between nearby communities. There appears to be a clear barrier separating Nigeria-Cameroon and central chimpanzees, which is overlapping with the Sanaga River (Fig. S83). Between central and eastern chimpanzees, we do not observe such a clear barrier, although the Goualougo site, close to Ubangi river, shows less migration towards eastern chimpanzees. We also observe another barrier at the west of Chinko, Bili and Rubi-Télé eastern chimpanzee sites that may represent low levels of migration of the Nigeria-Cameroon and central chimpanzees with eastern chimpanzee subspecies.

Some areas appear more isolated, while others seem more connected within each subspecies: In central chimpanzees (nIndiv = 25, nSites = 1067355, nDemes = 500), there is a significant barrier which overlaps with the Ogooué River crossing Gabon, separating Mts de Cristal and La Belgique from Lopé, Loango, Conkouati and Batéké, with both groups having more connectivity within their respective communities (Fig. S82 and Fig. S84A, B). Mbe in Nigeria-Cameroon chimpanzees (nIndiv = 17, nSites = 1258968, nDemes = 500) appears to be relatively isolated from the rest (Fig. S84C, D). For eastern chimpanzee communities (nIndiv = 60, nSites = 1098162, nDemes = 600), Bili and Chinko are more connected between them, while Issa Valley seems to have been more strongly isolated from the rest within the timeframe considered by EEMS (Fig. S82 and Fig. S84E, F)

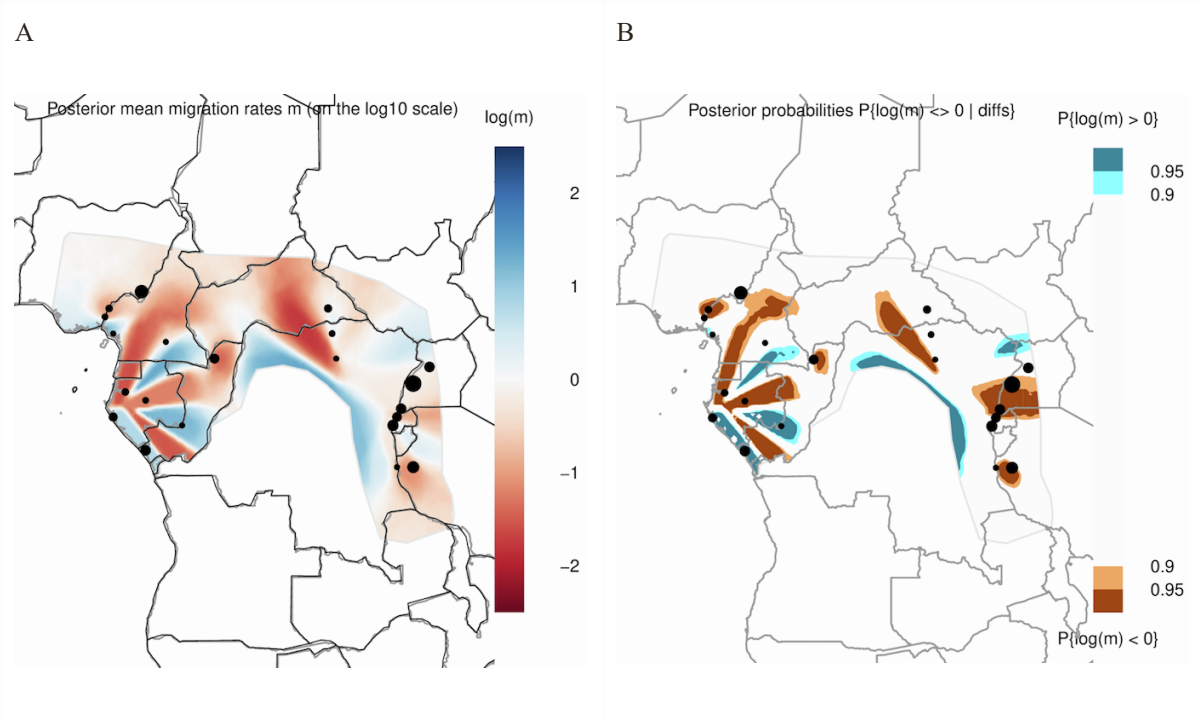


Fig. S83. EEMS for only eastern, central and Nigeria-Cameroon chimpanzees. (A) Posterior mean migration rate (B) Posterior probabilities. Related to Fig. 2.

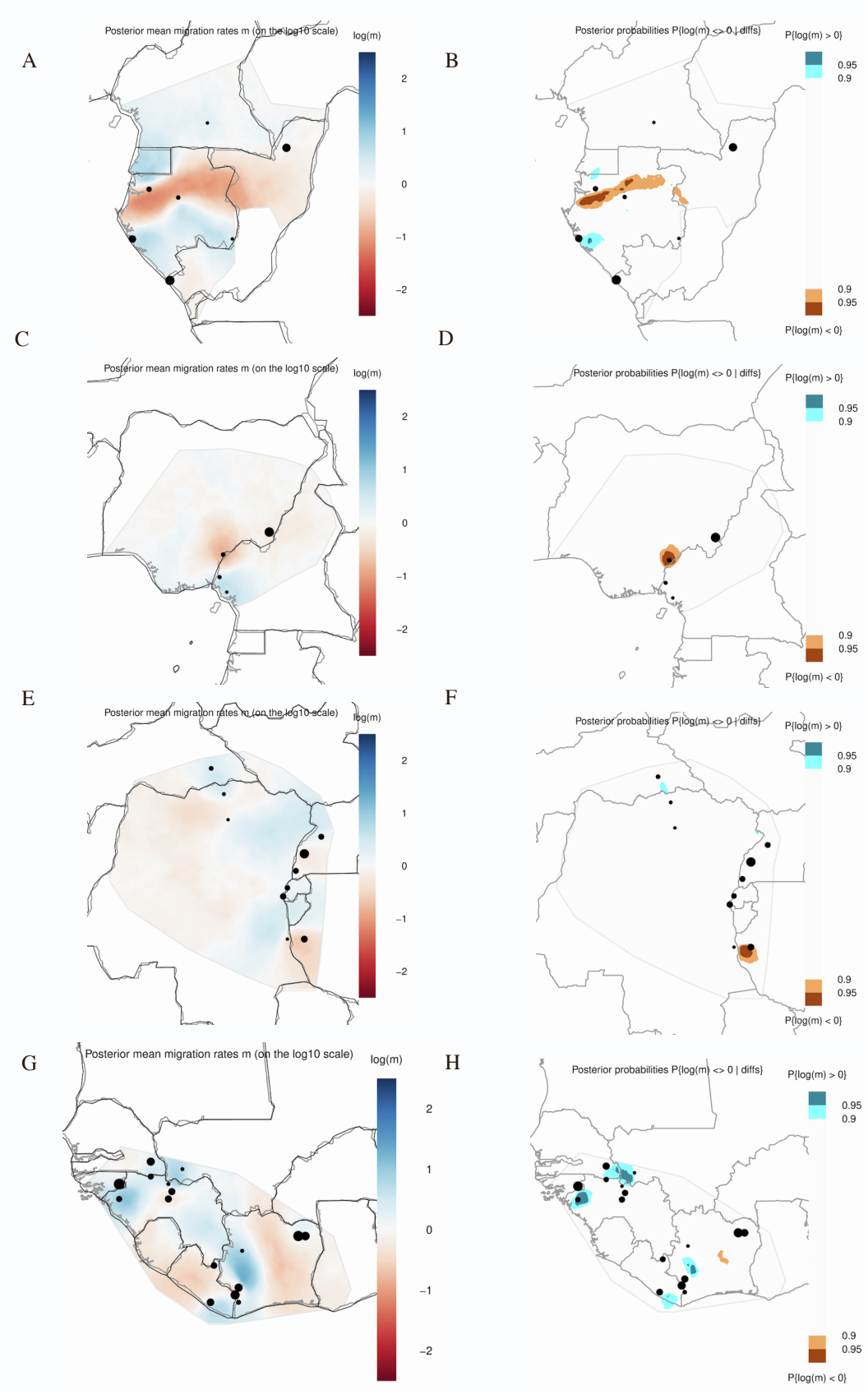


Fig. S84. EEMS running by subspecies individually: (A) and (B) central chimpanzee; (C) and (D) Nigeria-Cameroon chimpanzee; (E) and (F) eastern chimpanzee; and (G) and (H) western chimpanzee. (A), (C), (E) and (G) are posterior mean migration rate; and (B), (D), (F) and (H) are posterior probabilities. Related to Fig. 2.

In western chimpanzee sites (nIndiv = 110, nSites = 1136865, nDemes = 600), we observe a rather high connectivity across their range in comparison to non-western chimpanzees, with no site being significantly more isolated than others. Our data suggests that corridors of genetic connectivity existed between Boe and Sangaredi, but also between Kayan, Bafing, Dindefelo, Bafing and Sobory, and finally between Mt Sangbé and the southern clade of Taï and Grebo sites (Fig. S84G, H).

Finally, we also run EEMS on 59 chimpanzees from de Manuel *et al.* (2016)[21], using all autosomes. For some of these genomes, the exact GPS coordinates were unknown, or only approximate origins were known, we used the best matching geographic location based on rare alleles (Supplementary Text Note 9) as a proxy of the origin of those genomes (Data S1 – Fig. S100). The VCF file was filtered to keep only biallelic sites, at a minimum depth of 3, minimum genotype quality of 20, and allowing for 20% missing sites. We fed the inferred coordinates to EEMS and ran the program with the following parameters: nSites = 36,594,575, nDemes = 1000, diploid = TRUE, numMCMCIter = 2000000, numBurnIter = 1000000, numThinIter = 9999.

The effective migration surfaces of this dataset discover the same barriers we found above using the PanAf dataset, separating central from Nigeria-Cameroon and eastern chimpanzees (Fig. S85), however with lower resolution due to a reduced number of sampling locations. We conclude that chromosome 21 capture of a large number of samples provides power to detect past migration barriers between chimpanzee populations on a fine-grained scale.

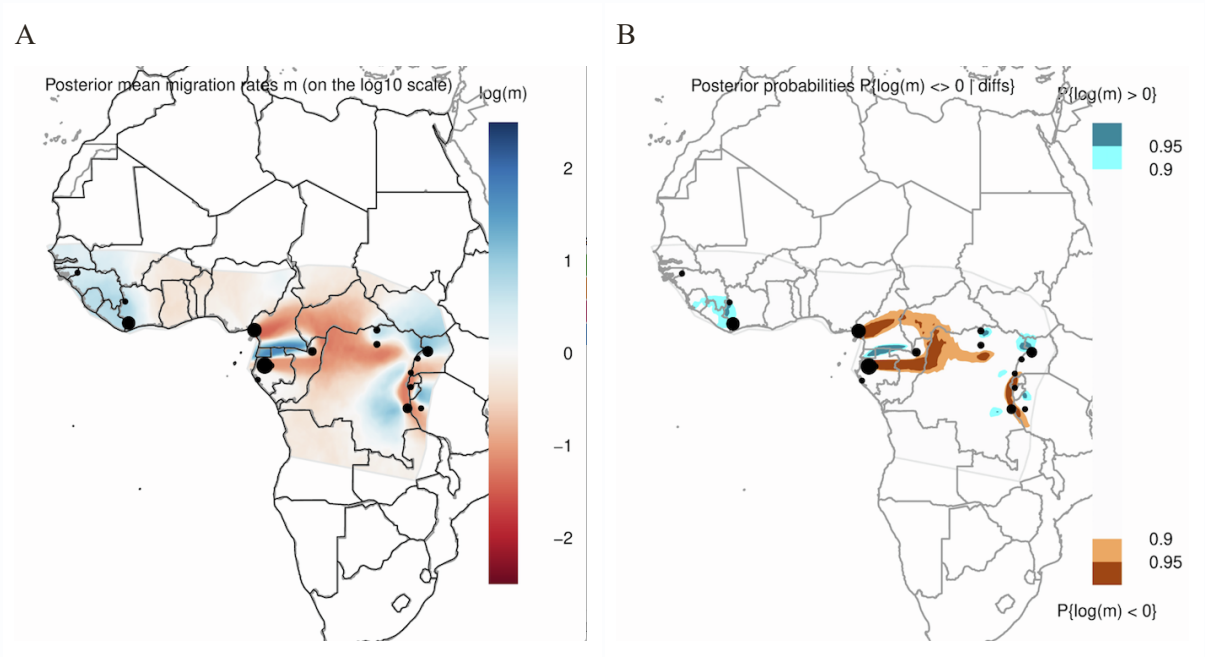


Fig. S85. EEMS between all samples and subspecies in WG dataset[21]. **(A)** Posterior mean migration rate and **(B)** posterior probabilities. Related to Fig. 2.

10.2. Fragments of shared ancestry

10.2.1. Detection of IBD-like segments

In order to identify Identical-By-Descent-like (IBD-like) segments between individuals within and between sites, we used IBDseq[70], since this program does not require phasing of the data. We applied this method to the dataset of samples with more than 5-fold coverage but including pairs of related samples (Table S1). To increase the sample size and thus the power for IBD-like detection, we merged this present dataset with data on chromosome 21 from 59 whole genomes[21]. We removed indels and kept only biallelic positions with minimum genotype quality of 20, minor allele frequency of 0.01 and excluded variants with a missingness of more than 0.6.

We explored the effect of using genotypes at two different thresholds of minimum depth: 3 or 8 reads, since the accuracy in detecting true heterozygous positions likely influences the power to detect IBD-like tracts. We assessed the impact of the depth of coverage on mean heterozygosity, considering global missingness and coverage. Mean heterozygosity and missingness were computed using plink [71] parameters -missing and -het, respectively. At a minimum depth of 3 reads at each genotype, the mean heterozygosity expected by subspecies (taking whole genomes as benchmark) is hardly maintained in samples when the mean coverage is low and the genotype missingness high. This implies that at lower coverage, we lose power to detect heterozygous calls, which at the same time would decrease our sensitivity to detect IBD-like tracts. On the other hand, to increase the power to detect heterozygous calls, we increased the threshold of minimum depth of 8 reads at each genotype. As a consequence, while the mean heterozygosity resembles the values of the whole-genome dataset, the missingness is increased (Fig. S86).

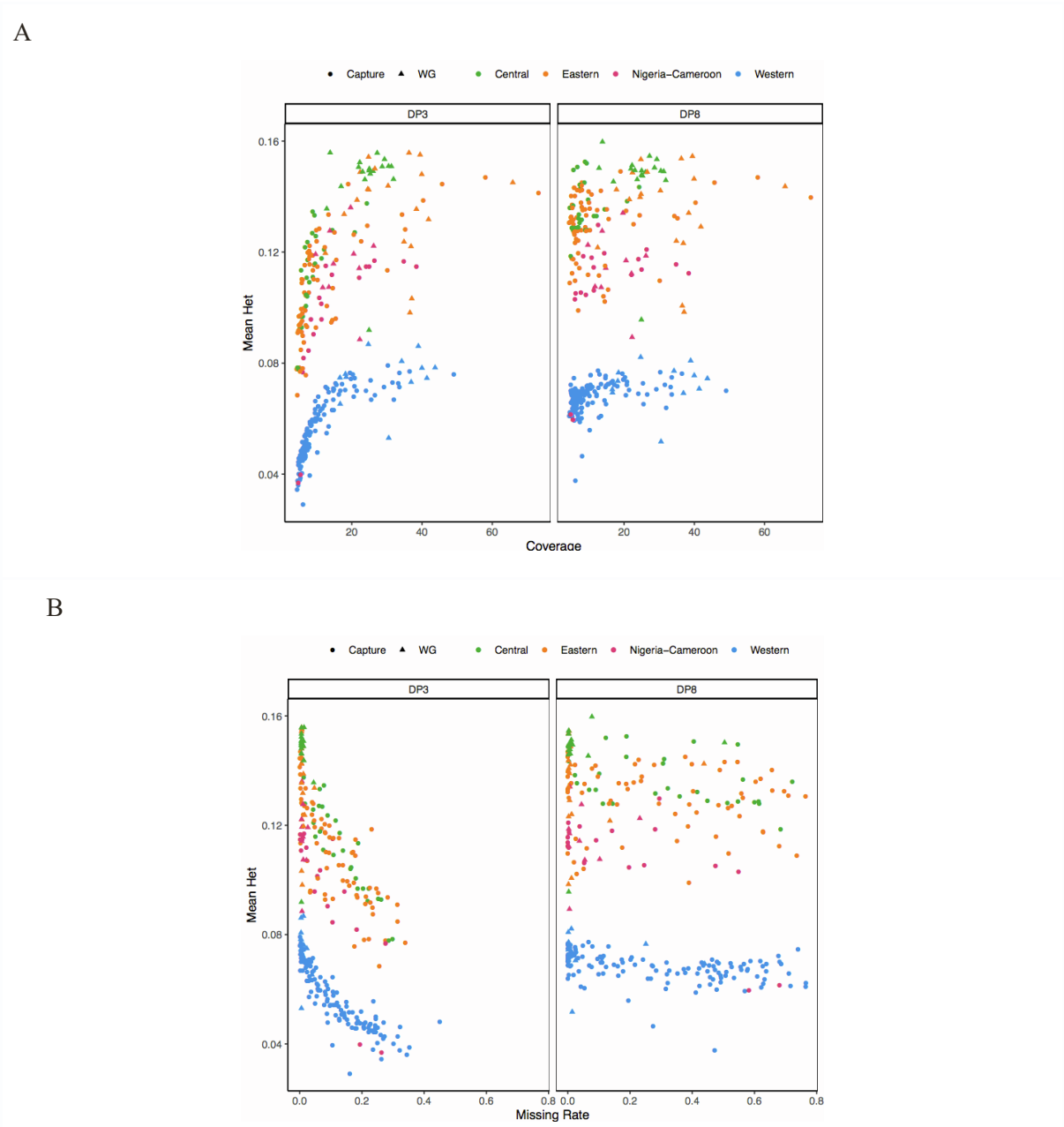


Fig. S86. Impact of depth of coverage at each genotype (3 or 8 reads) on mean heterozygosity in the PanAf dataset (240 samples) and whole genome dataset from de Manuel 2016 (59 samples). **(A)** Mean heterozygosity vs. mean coverage and **(B)** Mean heterozygosity vs. missing rate. Related to Fig. 3.

The IBDseq software was applied to all samples (PanAf and Whole-genome samples) with standard parameters (ibdlod=3.0, ibdtrim=0.3, r2window=500, r2max=0.15), but exploring a range of errorProp and errorMax values. Since the PanAf dataset is of rather low coverage, and although we limited the discovery of SNPs with a minimum of 8 reads, any misassigned genotype may cause a break of any hypothetical IBD-like between two samples. Instead of using the standard parameters (errorProp=0.25 and errorMax=0.001), we tested 5 different errorProp values (0.1, 0.25, 0.5, 0.6 and 0.8) and 12 errorMax values (0.00025, 0.001, 0.004, 0.008, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 and 0.8) on samples previously determined to be from 14 identical individuals (Note 3.5) among the PanAf samples,

processed and sequenced independently, and with of sufficient coverage (more than 5-fold). Theoretically, the whole chromosome 21 should be detected as one perfect IBD-like segment. However, with some combinations of errorProp and errorMax we did not observe this expected result for some samples (Fig. S87). After inspection across the 14 comparisons, we find that the empirical thresholds yielding a maximized length for all identical pairs were 0.1 for errorMax and 0.5 for errorProp.

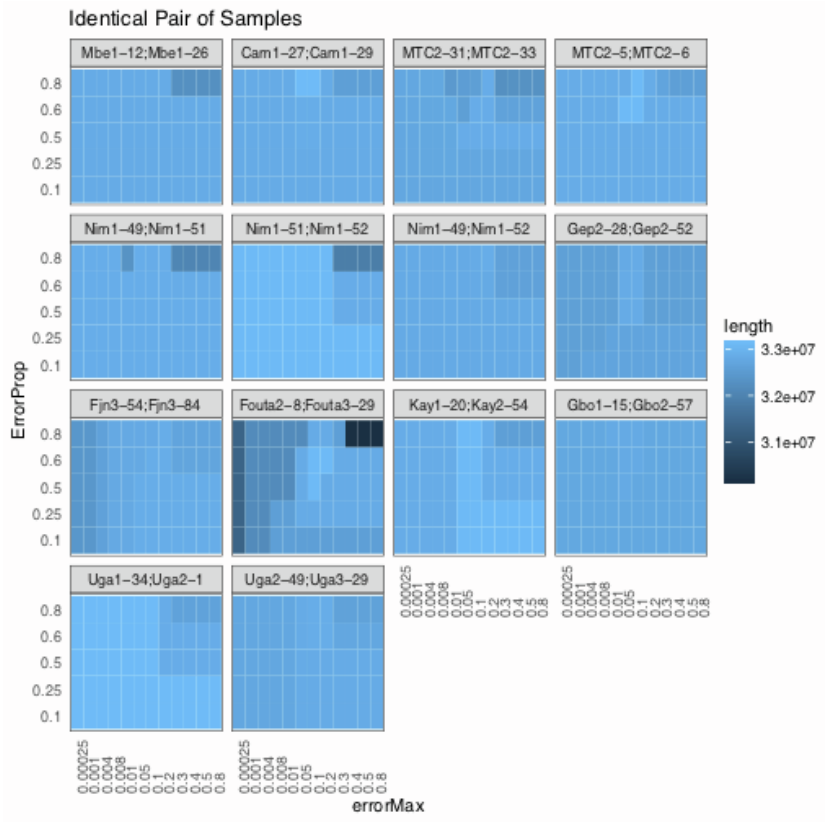


Fig. S87. Cumulative length in IBD-like detected with different errorMax and errorProp values between independently sequenced pairs of samples from the same individuals. Related to Fig. 3.

We applied IBDseq to the PanAf dataset using these parameters. As a quality control of the detected IBD-like segments between PanAf sites (segments involving data from whole genomes were excluded), we correlated their length with the LOD score provided by the software. Those IBD-like segments with large length but small LOD scores are more likely false positives, thus they were excluded from the next analysis (Fig. S88).

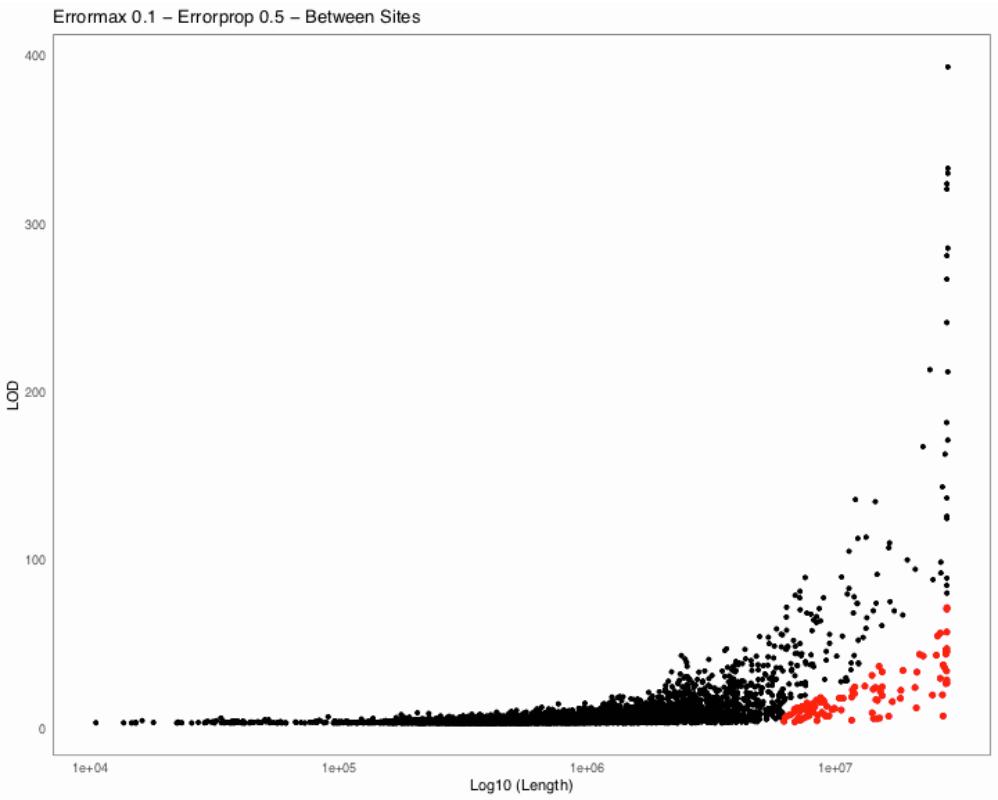


Fig. S88. Correlation of log10 IBD-like length and IBD-like LOD score of pairs of samples within the PanAf dataset. Red dots show the IBD-like tracts between pairs of samples that are excluded from further analysis due to their length not correlating with the LOD score. Related to Fig. 3.

We then correlated the IBD-like segment lengths and LOD scores with their geographical distance (Fig. S47). At this step, we removed related pairs of samples to avoid any bias for detecting an excess of IBD-like tracts between two sites (Table S1).

As probably expected, the larger the distance between geographical sites, the shorter the IBD-like tracts detected between individuals, and the lower the LOD scores. We observe this pattern of a decay of IBD-like tract length with geographical distance in all subspecies. We also detect a small number of IBD-like tracts between subspecies (Fig. S89).

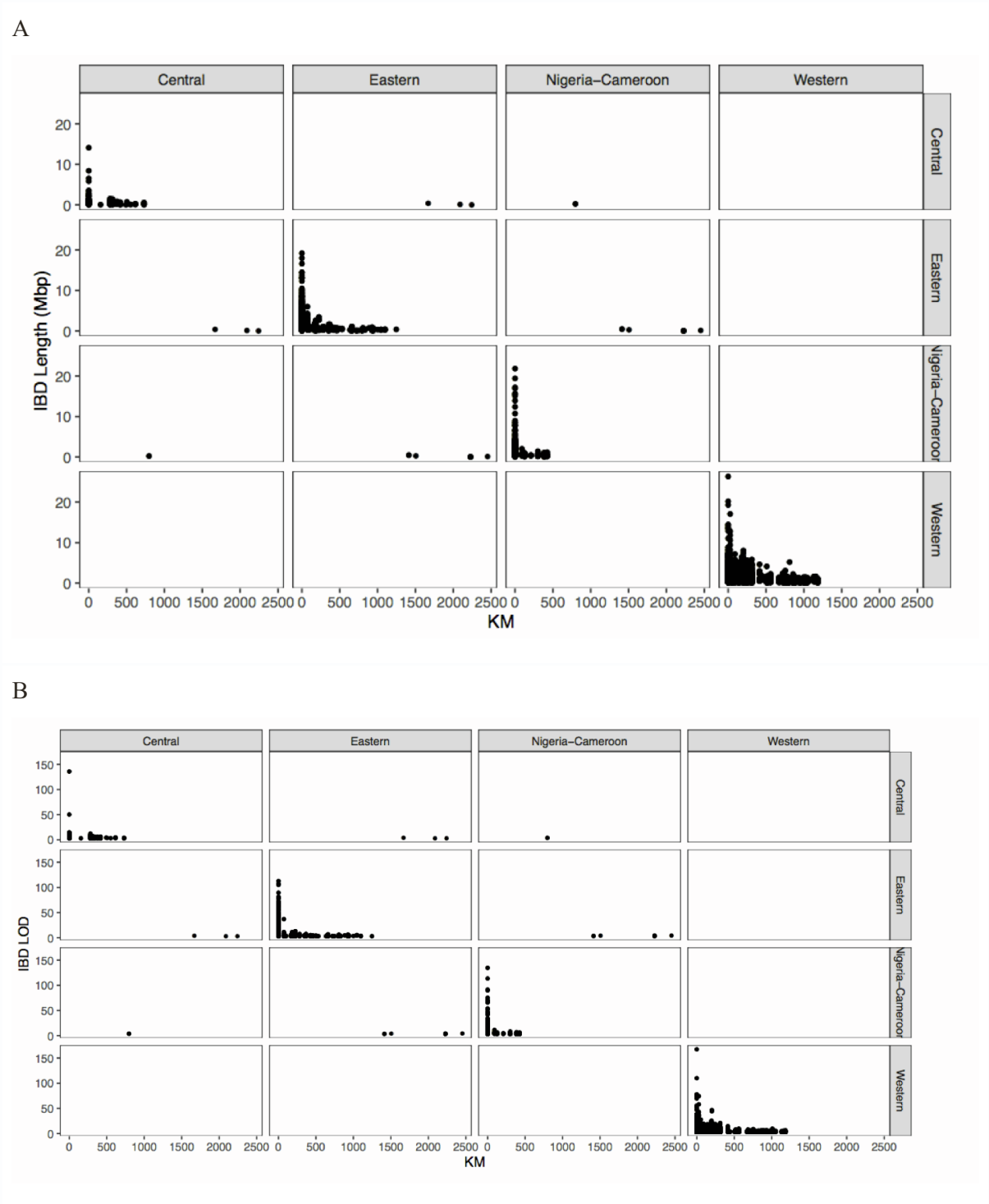


Fig. S89. Correlation of IBD-like tracts detected between geographical sites with their geographical distance in kilometers. (**A)** IBD-like length (in Mbp) and **(B)** LOD scores correlated with geographical distance (km). Related to Fig. 3.

We log-transformed the x axis (km) and obtained the mean IBD-like length at each geographical distance. We observe a linear significant decay for all subspecies but central and Nigeria-Cameroon chimpanzees, after log transformation of the X axis, pointing to an exponential decay (Fig. S90).

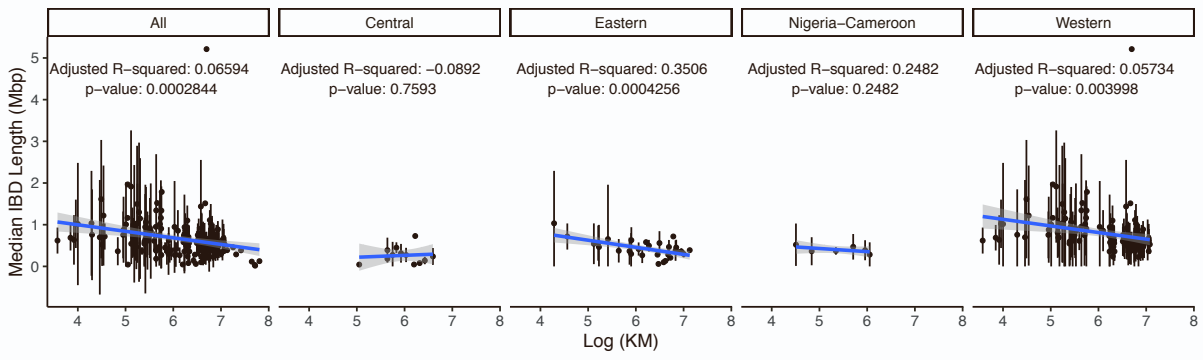


Fig. S90. Median length (Mbps) and standard deviation of IBD-like shared tracts between chimpanzee sampling sites. Western and eastern chimpanzee subspecies show an exponential decay of length with geographical distance. Error bars represent the standard deviation. Related to Fig. 3.

Since we have different numbers of samples per site, we have summarized the IBD-like detected by computing the number of IBD-like segments per pair of samples and also the average length of pairs of samples between sites. Also, sites that are in close proximity (within 15 km) were merged into a single site, as described above (Tai_Eco and Tai_R; Sobory and Bakoun; Comoé Geprenaf, Comoé-WEST, Comoé-East, Comoé2 and Comoé-CNPN) (Fig. S91).

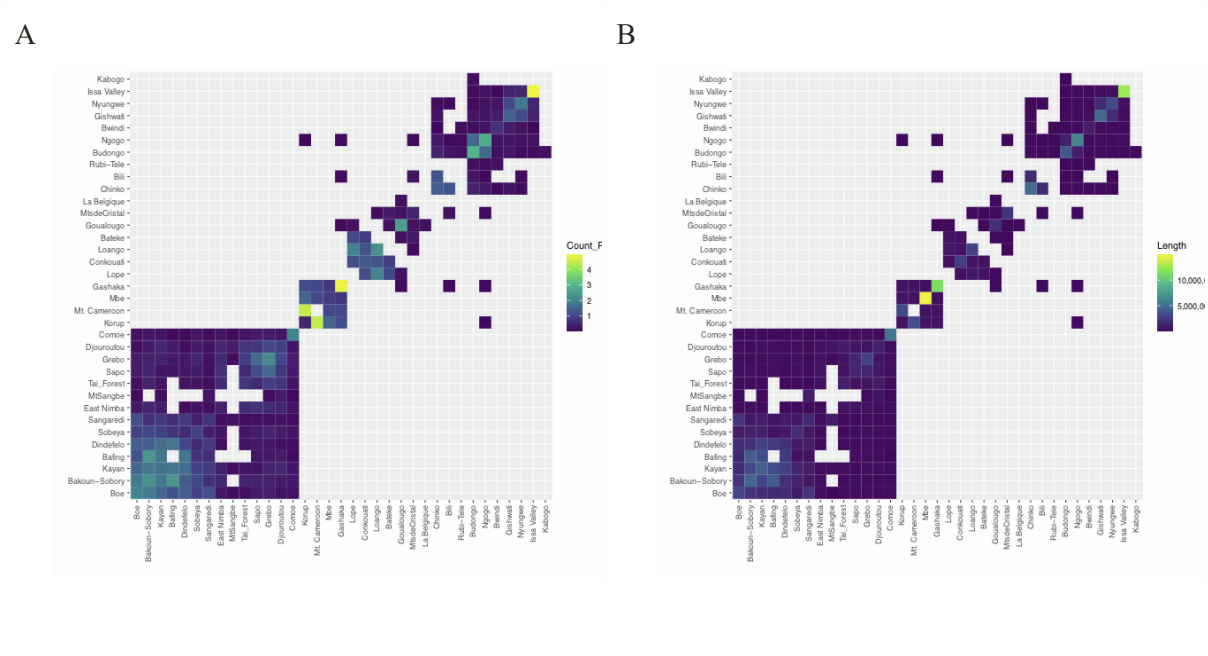


Fig. S91. IBD-like tracts between sampling sites. **(A),** Count of IBD-like tracts normalized by the number of pairs in each comparison. **(B),** Average length (bp) of IBD-like tracts between sites. Related to Fig. 3.

The amount of shared IBD-like segments is likely the consequence of recent migration events between geographic sites or areas, probably happening in the Holocene or Late Pleistocene. The length of the shared segments would be correlated to the time of such genetic exchange, with more recent migration causing longer IBD-like tracts. The intra-subspecies IBD-like tracts follow a pattern of exponential decay with distance, as expected for a scenario where isolation by distance took place. The IBD-like segments shared between individuals from different subspecies are very short (less than 0.5 Mbp), suggesting that this observation may be the consequence of connectivity further back in time, rather than recent events. Genetic exchange between chimpanzee subspecies and possible corridors of migration in the past have been suggested before[21,37], supporting such a scenario. However, we also note that the confidence in specific connections between the geographic locations included in this study is lower than that for the more recent connectivity within subspecies, given the short lengths and low LODs of the inter-subspecies IBD-like tracts. A way to estimate the age of an IBD-like segment is by using the length of each segment[72]. When the time (in generations g) to the most recent common ancestor is known, the total length (in cM) of a shared IBD segment follows an exponential distribution with rate 100/2g. Therefore, to time the events, we followed this rate of $g=100/(2*cM)$[58,72], with cM being the length of the fragments, and g the number of generations. The length in cM was estimated from the length in Mbps by applying the western chimpanzee recombination map[73] to the same subspecies and assuming an effective population size of Ne=17,378[21]. For the rest of the subspecies we used the Nigeria-Cameroon recombination map[74] with the following effective population sizes for each subspecies: central Ne=47,314, eastern Ne=32,492 and Nigeria-Cameroon Ne=27,795[21]. For timing the events between subspecies, we assumed a constant recombination map of 1cM/1Mbp since the recombination maps differ between subspecies. We assumed a generation time of 25 years to calculate the time[57]. We took the maximum IBD-like length per pair of individuals between sites to estimate the timeframe of connectivity per site and perform the average and maximum and minimum for each subspecies (Table S9). The natural log connectivity ratio of each sampling site (Table S10 and Fig. S92) was calculated as the sum of IBD-like segment counts (normalized by the number of pairwise sample comparisons between sites) that each site shares with the other sites, over the median global average of normalized IBD-like segment counts between all sampling sites.

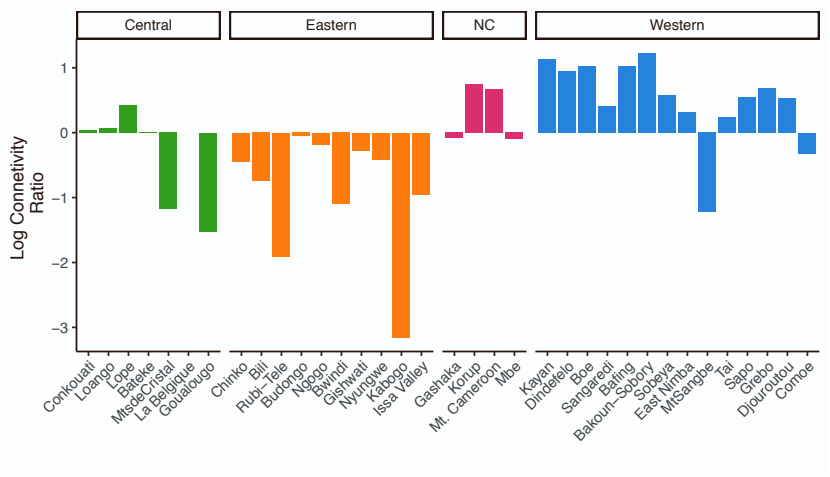


Fig. S92. Natural log connectivity score of each sampling site as a measure of connectivity/isolation. Related to Fig. 3.

Western chimpanzee communities appear to have high levels of connectivity between them, represented by more and longer IBD-like segments than any other subspecies, especially within their northern range (Kayan, Dindefelo, Boe, Sangaredi, Bakoun-Sobory and Sobeya). The EEMS method (Supplementary Text Note 10.2) also detects high effective migration (Fig. S84G, H), which suggests a more long-term connectivity between western chimpanzee communities. Interestingly, we find Mt Sangbé to be isolated recently, with few shared IBD-like segments with nearby sites, while the longer-term historical connectivity appears to have been higher. Comoé sites share few and short IBD-like segments with all other western chimpanzee sites, pointing to a history of high connectivity in the past, but not so more recently (Fig. 3C, in the main text).

Central chimpanzee sites show recent connectivity within their southern range, south of the Ogooué river (Loango, Lopé, Batéké and Conkouati), with more and slightly longer IBD-like segments compared to the connectivity within their northern range (Mts de Cristal, La Belgique and Goualougo), and between both groups (Table S9, Fig 3A, in the main text). Effective migration surfaces showed a strong barrier between the northern and southern clade, which seems to have been largely maintained until recently, although some recent migration appears to have happened across the barrier.

In Nigeria-Cameroon chimpanzees, all sites are quite connected to each other in recent times, sharing many IBD-like tracts. Although historically Mbe was rather isolated, as seen with the high amounts of short and long ROHs (Fig. S40), this seems to have been less the case in recent times. The connectivity between Mt. Cameroon and Korup apparently has been very high, as they share a large number of long IBD-like segments (Table S9 and Fig 3D, in the main text).

Within eastern chimpanzee sites, we observed three clusters of connectivity: Bili-Chinko, Budongo-Ngogo, and, to a lesser extent, Gishwati-Nyungwe. With past effective migration surfaces (Table S9

and Fig. 3B, in the main text) we also detected increased migration between Bili and Chinko, which seems to have continued until more recent periods. Interestingly, although Bili and Rubi-Télé are located in close proximity, Rubi-Télé shares no IBD-like segments with Bili, but is more recently connected with Ngogo, Bwindi and Budongo to the east. Kabogo is the most isolated site of all eastern chimpanzees, while the geographically close Issa Valley is highly connected with Gishwati, Nyungwe and other sites to the north. Here, Lake Tanganyika seems to have been a barrier to gene flow.

To account for the difference in sampling density in each subspecies, we repeated the analysis above by only incorporating samples from six western chimpanzee sites that cover the extension of this subspecies range (Kayan, Sangaredi, Comoé, East Nimba, Mt Sangbé and Grebo) and the rest of sites and WG samples. The results (Fig. S93, Fig. S94) are comparable with the previous findings (Fig. 3 and S92) and we do not observe a reduction or increase in the detection of connectivity in westerns or any other sampling site. Mt Sangbé, Comoé, Kabogo, Rubi-Télé and central sites above the Ogooué River (Mts de Cristal, La Belgique and Goualougo) continue to be detected as isolated.

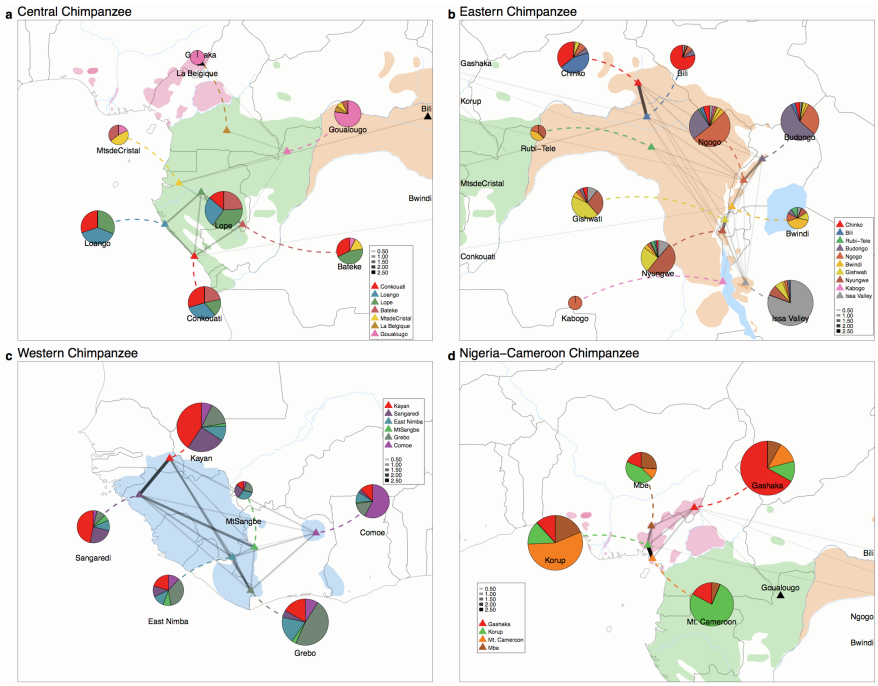


Fig. S93. Recent connectivity between chimpanzee populations. The size of the pie charts represents the pairwise number of shared fragments, normalized by the number of pairs. Thickness of lines indicates the average length of IBD-like tracts (in Mbps). Triangles show the location of sites. Colors in pies indicate the origin of IBD-like tracts. **(A)**, Central chimpanzees **(B)**, eastern chimpanzees **(C)**, western chimpanzees and **(D)**, Nigeria-Cameroon chimpanzees. Subspecies distribution color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 3.

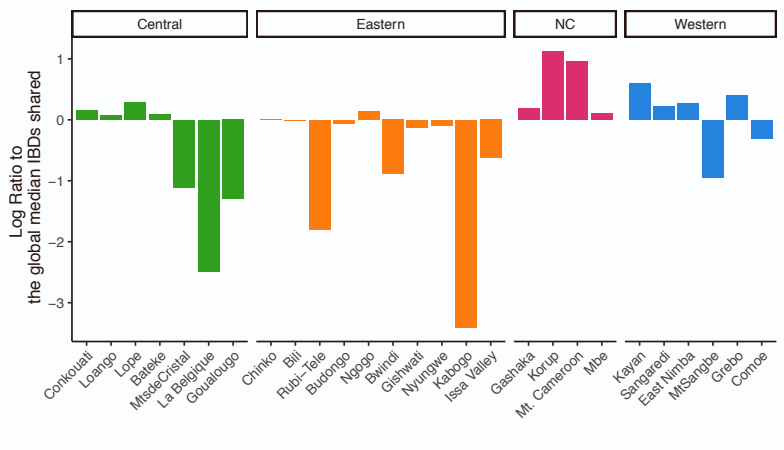


Fig. S94. Natural Log connectivity score of each sampling site as a measure of connectivity/isolation, only including six western chimpanzee sites. Color code: green for central chimpanzees, orange for eastern chimpanzees, pink for Nigeria-Cameroon and blue for western chimpanzees. Related to Fig. 3.

10.3. Comprehensive analysis of the chimpanzee population demography after the LGM and during the Holocene

In western chimpanzees, we found a variation cline from east to west (Fig. 1F, main text), from the north-western fringe of the distribution of western chimpanzees in Senegal and Mali to Côte d'Ivoire, and generally higher levels of connectivity across their range and across timescales than in the other subspecies (Fig. 2B,C and 3C in the main text, S82 and Supplementary Text Note 10). Remarkably, western chimpanzee communities have more and longer IBD-like segments than the other subspecies (Fig. 3C, main text), especially within their northern range in Senegal, Guinea and Guinea-Bissau, which may indicate recent connectivity within the past ~780 years (according to the IBD-like fragment length; range 117-2,200 years) (Table S9), or a range expansion colonizing into the fringe areas of the chimpanzee habitat[75] (Fig. 2C, main text). Interestingly, we found the geographically intermediate sampling site of Mt. Sangbé (log connectivity ratio = -1.22) and, to a lesser extent, Comoé (log connectivity ratio = -0.32), to have been isolated recently, with fewer shared IBD-like segments with nearby sites, while no ancient barrier to connectivity was found (S82 and 3B, main text). Mt. Sangbé had previously been described as an outlier population of western chimpanzees[37]. However, this regional isolation did not lead to excessive inbreeding in these populations, which would have been indicated by an excess of long runs of homozygosity (RoHs) (Supplementary Text Note 6). In addition, we found allele sharing consistent with recent migrations into the fringes in Senegal and Mali, while intermediate sampling locations (East Nimba and Outamba-Kilimi) carry the traces of a corridor of genetic exchange between the northwestern and southern areas (Fig. S79).

In central chimpanzees, we detected two population clusters, clearly separated by the Ogooué river in Gabon (Fig. 2B, main text and S82, Supplementary Information Note 6, 7 and 10), which appears to

have acted as a strong barrier reducing migration between these communities since the LGM. Southern central chimpanzees show recent connectivity until ~2,000 years ago (range 1,333 – 2,777) (Loango, Lopé, Batéké and Conkouati), with more and longer IBD-like segments (normalized count of 7.24 IBD-like tracts of 0.5 Mbp median length, log connectivity ratio >0, S92, Table S10) compared to the connectivity within the northern group (Mts de Cristal, La Belgique and Goualougo; normalized count of 0.22 IBD-like tracts of 0.08 Mbp median length (0.04cM), log connectivity ratio <0, S92, Table S10), and between these two groups (Fig. 2C, Fig. 3A, main text; normalized count of 0.98 IBD-like tracts of 0.01 Mbp median length (0.01cM). Locations in these two regions also show differences in rare allele sharing, with more rare alleles shared within each region (Supplementary Text Note 9 Fig. S76), and a strong genetic differentiation measured by F_{ST} (S45).

In eastern chimpanzees, a north-south cline can be observed from the northern Democratic Republic of Congo (DRC) and Central African Republic (CAR) to samples from Tanzania (Fig. 1G, main text), with most sampling sites being in isolation-by-distance to each other (S90). We observed three clusters of recent connectivity (Fig. 3B, main text): Bili-Chinko, Budongo-Ngogo, and Gishwati-Nyungwe (Table S9, S92). Dispersal corridors have been suggested for populations in western Uganda[76,77] and between western Uganda and eastern DRC[77] (Fig. 2B, main text). Past connectivity (S90) between Bili and Chinko has to have persisted until at least as recently as 450 years ago (mean of 2700 years up to 8000 years ago), while Rubi-Télé, located in close proximity (~198 km), shares no IBD-like segments with Bili, but with other locations towards the east. Concordant with observed behavioral differences to the North and South of the Uélé River[78,79], this river may have been a barrier to gene flow between the northern DRC and all other eastern chimpanzee populations (Fig. 2C and Fig. 3B, main text). Kabogo, while geographically close to Issa Valley, is genetically connected with other sites to the north, but separated from Issa Valley across Lake Tanganyika, which seems to have been a barrier to gene flow in the south (Fig. 2C, main text). Finally, all eastern chimpanzee populations share rare variation with those communities in the area of Pleistocene refugia (Budongo, Bwindi, Gishwati, Ngogo, Nyungwe), suggesting an expansion into the south (Issa Valley, Kabogo), west (Regomuki) and northwest (Rubi-Télé, Bili, Chinko, Ngiri) after the LGM (Fig. 2B, main text and Supplementary Text Note 9, Fig. S78).

All four sampling sites of Nigeria-Cameroon chimpanzees seem to have been connected within the past 2,500 years (mean 1,600 until 1000 years ago) (Fig. 3D, main text, Supplementary Text Note 9 Fig. S77), suggesting that Mbe was isolated only very recently, as evidenced by long RoHs resulting from recent inbreeding (Supplementary Text Note 6 and Fig. S40[31]). The connectivity, represented by a large number of long IBD-like tracts between Mt. Cameroon and Korup, has been very high until at least ~1000 years ago (mean 2,300 years, up to 4,300 years ago), in line with previous work suggesting a close relationship in this area (Fig. 2C, main text)[80].

References

1. Nsubuga, A.M., Robbins, M.M., Roeder, A.D., Morin, P.A., Boesch, C., and Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. Mol. Ecol. *13*, 2089–2094.

2. Arandjelovic, M., Guschanski, K., Schubert, G., Harris, T.R., Thalmann, O., Siedel, H., and Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. Mol. Ecol. Resour. *9*, 28–36.

3. Arandjelovic, M., Head, J., Rabanal, L.I., Schubert, G., Mettke, E., Boesch, C., Robbins, M.M., and Vigilant, L. (2011). Non-invasive genetic monitoring of wild central chimpanzees. PLoS One *6*, e14761.

4. Fontsere, C., Alvarez-Estape, M., Lester, J., Arandjelovic, M., Kuhlwilm, M., Dieguez, P., Agbor, A., Angedakin, S., Ayuk Ayimisin, E., Bessone, M., et al. (2021). Maximizing the acquisition of unique reads in noninvasive capture sequencing experiments. Mol. Ecol. Resour. *21*, 745–761.

5. Morin, P.A., Chambers, K.E., Boesch, C., and Vigilant, L. (2001). Quantitative PCR analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees. Mol. Ecol., 1835–1844.

6. Hernandez-Rodriguez, J., Arandjelovic, M., Lester, J., de Filippo, C., Weihmann, A., Meyer, M., Angedakin, S., Casals, F., Navarro, A., Vigilant, L., et al. (2018). The impact of endogenous content, replicates and pooling on genome capture from faecal samples. Mol. Ecol. Resour. *18*, 319–333.

7. White, L.C., Fontsere, C., Lizano, E., Hughes, D.A., Angedakin, S., Arandjelovic, M., Granjon, A., Hans, J.B., Lester, J.D., Rabanus-Wallace, M.T., et al. (2019). A roadmap for high-throughput sequencing studies of wild animal populations using noninvasive samples and hybridization capture. Mol. Ecol. Resour. *19*, 609–622.

8. Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. Methods Ecol. Evol. *9*, 410–419.

9. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, 1–8.

10. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

11. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

12. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

13. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

14. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*.

15. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 254–260.

16. Buchan, J.C., Archie, E.A., Van Horn, R.C., Moss, C.J., and Alberts, S.C. (2005). Locus effects and sources of error in noninvasive genotyping. Mol. Ecol. Notes *5*, 680–683.

17. Deagle, B.E., Eveson, J.P., and Jarman, S.N. (2006). Quantification of damage in DNA recovered from highly degraded samples - A case study on DNA in faeces. Front. Zool. *3*.

18. Dabney, J., Meyer, M., and Pääbo, S. (2013). Ancient DNA damage. Cold Spring Harb. Perspect. Biol. *5*.

19. Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E., and Orlando, L. (2011). mapDamage: Testing for damage patterns in ancient DNA sequences. Bioinformatics *27*, 2153–2155.

20. Prüfer, K. (2018). SNPAD: An ancient DNA genotype caller. Bioinformatics *34*, 4165–4171.

21. De Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V.C., Desai, T., Prado-Martinez, J., Hernandez-Rodriguez, J., Dupanloup, I., Lao, O., Hallast, P., et al. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science (80-.). *354*, 477–481.

22. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. Genetics *210*, 719–731.

23. Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature *499*, 471–475.

24. Stanford, C.B., Wallis, J., Matama, H., and Goodall, J. (1994). Patterns of predation by chimpanzees on red colobus monkeys in gombe national park, 1982–1991. Am. J. Phys. Anthropol. *94*, 213–228.

25. Kuhlwilm, M., Fontsere, C., Han, S., Alvarez-Estape, M., and Marques-Bonet, T. (2021). HuConTest: Testing Human Contamination in Great Ape Samples. Genome Biol. Evol. *13*.

26. Korneliussen, T.S., and Moltke, I. (2015). NgsRelate: A software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics *31*, 4009–4011.

27. Hanghøj, K., Moltke, I., Andersen, P.A., Manica, A., and Korneliussen, T.S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. Gigascience *8*.

28. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

29. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. *9*.

30. Castellano, S., Parra, G., Sánchez-Quinto, F.A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., et al. (2014). Patterns of coding variation in the complete exomes of three Neandertals. Proc. Natl. Acad. Sci. U. S. A. *111*, 6666–6671.

31. Kuhlwilm, M., Gronau, I., Hubisz, M.J., De Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H.A., Lalueza-Fox, C., De La Rasilla, M., et al. (2016). Ancient gene flow from early modern humans into Eastern Neanderthals. Nature *530*, 429–433.

32. Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., et al. (2008). Genes mirror geography within Europe. Nature *456*, 98–101.

33. Martin, A.D., Quinn, K.M., and Park, J.H. (2011). MCMCpack: Markov chain Monte Carlo in R. J. Stat. Softw. *42*, 1–21.

34. Skotte, L., Korneliussen, T.S., and Albrechtsen, A. (2013). Estimating individual admixture

proportions from next generation sequencing data. Genetics *195*, 693–702.

35. Lawson, D.J., van Dorp, L., and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat. Commun. *9*.

36. Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., Mckeigue, P.M., and Wilson, J.F. (2010). Genomic runs of homozygosity record population history and consanguinity. PLoS One *5*, e13996.

37. Lester, J.D., Vigilant, L., Gratton, P., McCarthy, M.S., Barratt, C.D., Dieguez, P., Agbor, A., Álvarez-Varona, P., Angedakin, S., Ayimisin, E.A., et al. (2021). Recent genetic connectivity and clinal variation in chimpanzees. Commun. Biol. *4*, 283.

38. Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., and Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS One *7*.

39. Korneliussen, T.S., Moltke, I., Albrechtsen, A., and Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics *14*.

40. Han, S., Andre, A.M., Marques-Bonet, T., and Kuhlwilm, M. (2019). Genetic variation in pan species is shaped by demographic history and harbors lineage-specific functions. Genome Biol. Evol. *11*, 1178–1191.

41. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. Bioinformatics *20*, 289–290.

42. Vieira, F.G., Lassalle, F., Korneliussen, T.S., and Fumagalli, M. (2016). Improving the estimation of genetic distances from Next-Generation Sequencing data. Biol. J. Linn. Soc. *117*, 139–149.

43. Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol. Biol. Evol. *32*, 2798–2800.

44. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455.

45. Paradis, E., and Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.

46. Revell, L.J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.

47. Schliep, K.P. (2011). phangorn: Phylogenetic analysis in R. Bioinformatics *27*, 592–593.

48. Nater, A., Mattle-Greminger, M.P., Nurcahyo, A., Nowak, M.G., de Manuel, M., Desai, T., Groves, C., Pybus, M., Sonay, T.B., Roos, C., et al. (2017). Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. Curr. Biol. *27*, 3487-3498.e10.

49. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

50. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. *8*.

51. Kuhlwilm, M., Han, S., Sousa, V.C., Excoffier, L., and Marques-Bonet, T. (2019). Ancient admixture from an extinct ape lineage into bonobos. Nat. Ecol. Evol. *3*, 957–965.

52. Nye, J., Laayouni, H., Kuhlwilm, M., Mondal, M., Marques-Bonet, T., and Bertranpetit, J. (2018). Selection in the introgressed regions of the chimpanzee genome. Genome Biol. Evol. *10*, 1132–1138.

53. Peter, B.M. (2016). Admixture, population structure, and f-statistics. Genetics *202*, 1485–1501.

54. Petr, M., Vernot, B., and Kelso, J. (2019). Admixr-R package for reproducible analyses using ADMIXTOOLS. Bioinformatics *35*, 3194–3195.

55. Peter, B. (2020). 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. bioRxiv, 2020.03.13.990523.

56. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.

57. Langergraber, K.E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J.C., Muller, M.N., et al. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc. Natl. Acad. Sci. U. S. A. *109*, 15716–15721.

58. Thompson, E.A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. Genetics *194*, 301–326.

59. Wasser, S.K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C., and Weir, B.S. (2015). Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. Science (80-.). *349*, 84–87.

60. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., and Bustamante, C.D. (2011). Demographic history and rare allele sharing among human populations. Proc. Natl. Acad. Sci. U. S. A. *108*, 11983–11988.

61. Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., et al. (2016). Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nat. Commun. *7*.

62. Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. R J. *10*, 439–446.

63. Pebesma, E.J., and Bivand, R.S. (2005). Classes and methods for spatial data in {R}. R News *5*, 9–13.

64. Bivand, R.S., Pebesma, E., and Gómez-Rubio, V. (2013). Applied Spatial Data Analysis with R: Second Edition (Springer, NY).

65. Bivand, R., and Lewin-Koh, N. (2013). maptools: Tools for Handling Spatial Objects. R Packag. version 0.8.

66. Humle, T., Maisels, F., Oates, J.F., Plumptre, A., and Williamson, E.A. (2016). Pan troglodytes (errata version published in 2018). IUCN Red List Threat. Species, e.T15933A129038584. https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T15933A17964454.en.

67. Gräler, B., Pebesma, E., and Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. R J. *8*, 204–218.

68. Al-Asadi, H., Petkova, D., Stephens, M., and Novembre, J. (2019). Estimating recent migration and population-size surfaces. PLoS Genet. *15*.

69. Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. *48*, 94–100.

70. Browning, B.L., and Browning, S.R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. Am. J. Hum. Genet. *93*, 840–851.

71. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

72. Nait Saada, J., Kalantzis, G., Shyr, D., Cooper, F., Robinson, M., Gusev, A., and Palamara, P.F.

(2020). Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. Nat. Commun. *11*, 6130.

73. Auton, A., Fledel-Alon, A., Pfeifer, S., Venn, O., Ségurel, L., Street, T., Leffler, E.M., Bowden, R., Aneas, I., Broxholme, J., et al. (2012). A fine-scale chimpanzee genetic map from population sequencing. Science (80-.). *336*, 193–198.

74. Stevison, L.S., Woerner, A.E., Kidd, J.M., Kelley, J.L., Veeramah, K.R., McManus, K.F., Prado-Martinez, J., Sudmant, P.H., Li, H., Lorente-Galdos, B., et al. (2016). The time scale of recombination rate evolution in great apes. Mol. Biol. Evol. *33*, 928–945.

75. Wessling, E.G., Kühl, H.S., Mundry, R., Deschner, T., and Pruetz, J.D. (2018). The costs of living at the edge: Seasonal stress in wild savanna-dwelling chimpanzees. J. Hum. Evol. *121*, 1–11.

76. McCarthy, M.S., Lester, J.D., Howe, E.J., Arandjelovic, M., Stanford, C.B., and Vigilant, L. (2015). Genetic censusing identifies an unexpectedly sizeable population of an endangered large mammal in a fragmented forest landscape. BMC Ecol. *15*.

77. Rich, A.M., Wasserman, M.D., Hunt, K.D., and Kaestle, F.A. (2020). Chimpanzee (Pan troglodytes schweinfurthii) population spans multiple protected areas in the albertine rift. Folia Primatol. *91*, 595–609.

78. Hicks, T.C., Tranquilli, S., Kuehl, H., Campbell, G., Swinkels, J., Darby, L., Boesch, C., Hart, J., and Menken, S.B.J. (2014). Absence of evidence is not evidence of absence: Discovery of a large, continuous population of Pan troglodytes schweinfurthii in the Central Uele region of northern DRC. Biol. Conserv. *171*, 107–113.

79. Hicks, T.C., Kühl, H.S., Boesch, C., Dieguez, P., Ayimisin, A.E., Fernandez, R.M., Zungawa, D.B., Kambere, M., Swinkels, J., Menken, S.B.J., et al. (2019). Bili-Uéré: A Chimpanzee Behavioural Realm in Northern Democratic Republic of Congo. Folia Primatol. *90*, 3–64.

80. Mitchell, M.W., Locatelli, S., Ghobrial, L., Pokempner, A.A., Clee, P.R.S., Abwe, E.E., Nicholas, A., Nkembi, L., Anthony, N.M., Morgan, B.J., et al. (2015). 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. *15*, 3.

81. Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., et al. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science (80-.). *348*, 242–245.