Current Biology

Report

Deep genetic substructure within bonobos

Highlights

- We discover the presence of three genetically distinct bonobo populations
- The deepest estimated split time is ~145,000 years ago
- Genetic divergence among bonobo populations is thus deeper than previously thought
- The smallest effective population size is only ~3,000, suggesting strong isolation

Authors

Sojung Han, Cesare de Filippo,
Genís Parra, ...,
Tomas Marques-Bonet,
Martin Kuhlwilm, Aida M. Andrés

Correspondence

so.jung.han@univie.ac.at (S.H.),
a.andres@ucl.ac.uk (A.M.A.)

In brief

Bonobos, an endangered primate, live exclusively in the Congo basin. Their population structure is poorly understood, despite its importance for conservation. Han et al. find evidence of substructure within the species, with substantial genetic differentiation and estimated divergence times as deep as those of the closest chimpanzee subspecies.

Han et al., 2024, Current Biology *34*, 1–8
November 18, 2024
https://doi.org/10.1016/j.cub.2024.09.043

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

Deep genetic substructure within bonobos

Sojung Han,[1,2,3,12,*] Cesare de Filippo,[2,4,12] Genís Parra,[4,5] Juan Ramon Meneu,[4] Romain Laurent,[4] Peter Frandsen,[6] Christina Hvilsom,[6] Ilan Gronau,[7] Tomas Marques-Bonet,[1,8,9,10] Martin Kuhlwilm,[1,2,3] and Aida M. Andrés[4,11,13,*]

[1]Institut de Biologia Evolutiva, Consejo Superior de Investigaciones Científicas, Universitat Pompeu Fabra, 08003 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]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany
[5]Centre Nacional d'Anàlisi Genòmica (CNAG), Baldiri Reixac 4, 08028 Barcelona, Spain
[6]Conservation, Copenhagen Zoo, Roskildevej 38, 2000, Frederiksberg, Denmark
[7]The Efi Arazi School of Computer Science, Reichman University, 4610101 Herzliya, Israel
[8]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08003 Barcelona, Spain
[9]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, 08193 Barcelona, Spain
[10]Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010, Barcelona, Spain
[11]UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK
[12]These authors contributed equally
[13]Lead contact
*Correspondence: so.jung.han@univie.ac.at (S.H.), a.andres@ucl.ac.uk (A.M.A.)

https://doi.org/10.1016/j.cub.2024.09.043

SUMMARY

Establishing the genetic and geographic structure of populations is fundamental, both to understand their evolutionary past and preserve their future. Nevertheless, the patterns of genetic population structure are unknown for most endangered species. This is the case for bonobos (*Pan paniscus*), which, together with chimpanzees (*Pan troglodytes*), are humans' closest living relatives. Chimpanzees live across equatorial Africa and are classified into four subspecies,[1] with some genetic population substructure even within subspecies. Conversely, bonobos live exclusively in the Democratic Republic of Congo and are considered a homogeneous group with low genetic diversity,[2] despite some population structure inferred from mtDNA. Nevertheless, mtDNA aside, their genetic structure remains unknown, hampering our understanding of the species and conservation efforts. Mapping bonobo genetic diversity in space is, however, challenging because, being endangered, only non-invasive sampling is possible for wild individuals. Here, we jointly analyze the exomes and mtDNA from 20 wild-born bonobos, the whole genomes of 10 captive bonobos, and the mtDNA of 136 wild individuals. We identify three genetically distinct bonobo groups of inferred Central, Western, and Far-Western geographic origin within the bonobo range. We estimate the split time between the central and western populations to be ~145,000 years ago and genetic differentiation to be in the order of that of the closest chimpanzee subspecies. Furthermore, our estimated long-term N_e for Far-West (~3,000) is among the lowest estimated for any great ape lineage. Our results highlight the need to attend to the bonobo substructure, both in terms of research and conservation.

RESULTS AND DISCUSSION

Three genetically differentiated groups of bonobos

Bonobos are assessed as endangered on the IUCN Red List, with less than 20,000 extant wild individuals. Their recent historical range spans an area[3] of 564,542 km^2, encompassing African forests within the Democratic Republic of Congo (DRC), where social unrest has limited research activities for decades.[4] Wild bonobos are thus less studied than other great apes. However, their biological similarity to humans, multi-male/multi-female societies, use of non-reproductive sexual behaviors, inner-/intergroup tolerance, and high social status of females make them uniquely interesting.[5] Even though some cultural differences have been observed among communities,[6,7] from the point of view of their genomes, bonobos have long been considered a homogeneous group[2,8] with low genetic diversity due to population bottlenecks.[9–12] However, mitochondrial DNA (mtDNA) from seven wild communities revealed six haplogroups associated with specific geographic areas,[9,13] with the time to the most recent common ancestor inferred from the six mtDNA haplogroups being at least 380 thousand years ago (kya) (260–530 kya). mtDNA is highly informative, but it is a single locus and its genealogy reflects only the history of the maternal lineages. Because bonobos are primarily patrilocal, the evolutionary history of females does not represent that of the whole population—especially when it comes to gene flow and population subdivision. It is thus important to also generate demographic inferences based on the nuclear genome. Here, we use genomic datasets to assess the assumption of genetic homogeneity: the exomes[14] and mtDNA[2] from 20 wild-born bonobos residing in an African sanctuary, together with 10 full genomes from wild-born captive bonobos[10] and the mtDNA of 136 wild individuals.[13]

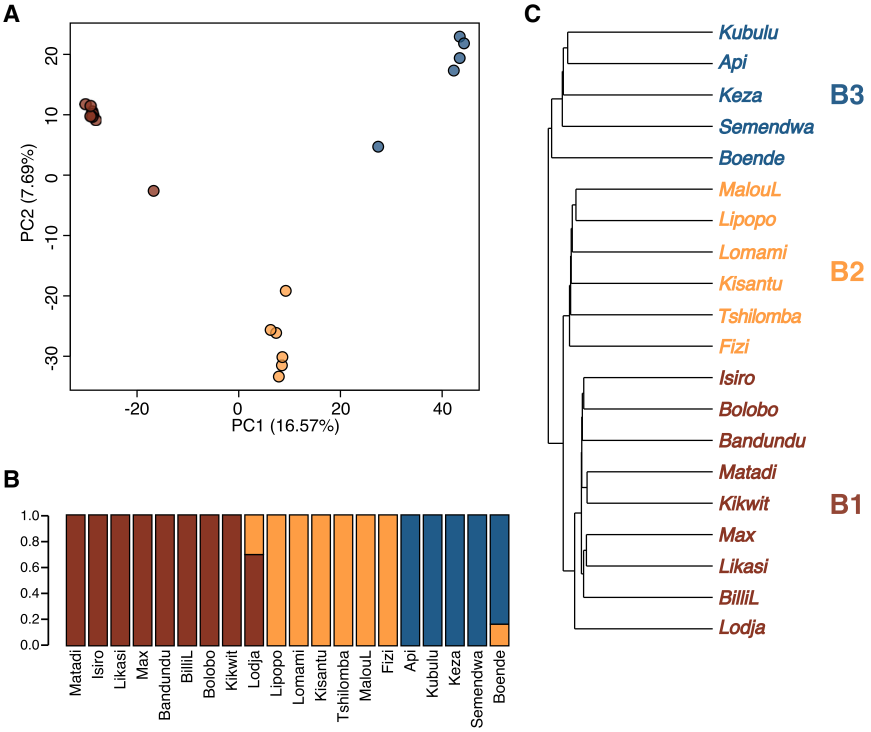

Figure 1. Three distinct groups of bonobos, using whole exomes

(A) PCA, (B) ADMIXTURE clustering of bonobos (with y axis being the proportion of k = 3 ancestry), and (C) NJ tree. Colors for the three groups are B1 maroon, B2 orange, and B3 blue, as labeled in the NJ tree. Comparable plots for central chimpanzees and Yoruba humans can be found in Figure S1. We note that the names of the individuals (as in B and C) do not necessarily reflect the place of origin.

A principal-component analysis (PCA) of the 20 exomes separates three distinct groups (Figure 1A), with two individuals (Lodja and Boende) falling in between groups. We initially label these groups as B1, B2, and B3. The 20 exomes come from individuals of unknown geographic origin, sampled in a sanctuary, so we can ensure that they are unrelated,[2,14] but we are not able to assess sampling representation across the three groups. An ADMIXTURE analysis (Figure 1B) shows $K = 3$ as the best fit, and clustering of individuals is in agreement with the PCA and neighbor-joining (NJ) groups. An NJ phylogenetic tree separates B1 and B2, and both from B3 (Figure 1C). Thus, PCA, NJ, and ADMIXTURE indicate the presence of three distinct groups of bonobos in our 20 exomes. For context, both PCA and ADMIXTURE reveal higher evidence of genetic differentiation among these 20 bonobos than among 20 central chimpanzees (the chimpanzee subspecies with the highest genetic diversity[10,11]) and 20 Yoruba humans (an African population with high genetic diversity[15]) (Figures S1A and S1B).

We quantified the degree of population differentiation among the groups using pairwise F_{ST} (Table S1). The highest differentiation is between B1 and B3 (average $F_{ST} = 0.145$), followed by B1 and B2 ($F_{ST} = 0.093$) and B2 and B3 ($F_{ST} = 0.088$). To interpret these values, we put them in the context of humans and chimpanzees, the other great ape species for which we have fully comparable exome data. The F_{ST} values are in the range of those between chimpanzee subspecies: F_{ST} between B1 and B3 is slightly higher than the F_{ST} between the closest chimpanzee subspecies, Central and Eastern, and lower than between the other chimpanzee subspecies. Compared with human populations, the bonobo F_{ST} values are within the range between African and non-African humans (Table S1). Of note, F_{ST} values do not directly reflect split times because other demographic factors that shape N_e (and thus drift) impact F_{ST}. However, we suggest that the F_{ST} values reveal substantial genetic differentiation among the bonobo groups.

The geographic origin of the three bonobo populations

Unfortunately, we lack precise information on the geographic origin of individuals from sanctuaries, hindering the geographic assignment of the three inferred groups (B1–B3). Nevertheless, we were able to take advantage of 61 mtDNA haplotypes from seven wild bonobo populations of known location from Kawamoto et al.[13] to make inferences about the geographic origin of our samples. In a multidimensional scaling (MDS) analysis, the populations are not completely differentiated, with a few haplotypes from the Central populations and the Eastern population close to the Western group, likely due to female migration and/or higher effective population size (N_e) of Central populations. However, the mtDNA sequences of our 20 sanctuary bonobos fall perfectly within the MDS clusters of the Kawamoto et al.[13] mtDNA sequences (Figure 2A). Because the Kawamoto et al.[13] samples are placed geographically, and there is a strong correlation between autosomal and mitochondrial pairwise nucleotide distances (Mantel test based on Spearman rank correlations = 0.73; $p = 0.001$), this analysis provides the likely geographic origin of the bonobos in the B1, B2, and B3 groups.

In a phylogenetic tree, the Kawamoto et al.[13] mtDNA samples show geographic substructure, with a western, a central, and an eastern clade.[13] Incorporating the B1, B2, and B3 groups in this mtDNA tree does not affect its topology. In agreement with the PCA, NJ, and F_{ST} results, B1 and B2 are closer to each other and fall within the same cluster, the western clade (Figure 2A). B3 falls in a separate cluster, the central clade, and has mtDNA $\Phi_{ST} = 0$ when compared with the central Salonga and Iyondji, suggesting that B3 is closely related to these populations ($\Phi_{ST} = 0$ refers to cases where diversity is equal or larger within than between populations). Interestingly, B1 clusters with the Malebo population ($\Phi_{ST} = 0$), which lives at the western periphery of the bonobo range and is characterized by low levels of mtDNA nucleotide diversity and higher genetic distance to the other western population LacTumba. The good correspondence between genetics and geography in Kawamoto et al.[13] samples, and between the mtDNA of our samples and of the Kawamoto et al.[13] ones, allows us to infer the approximate geographic

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

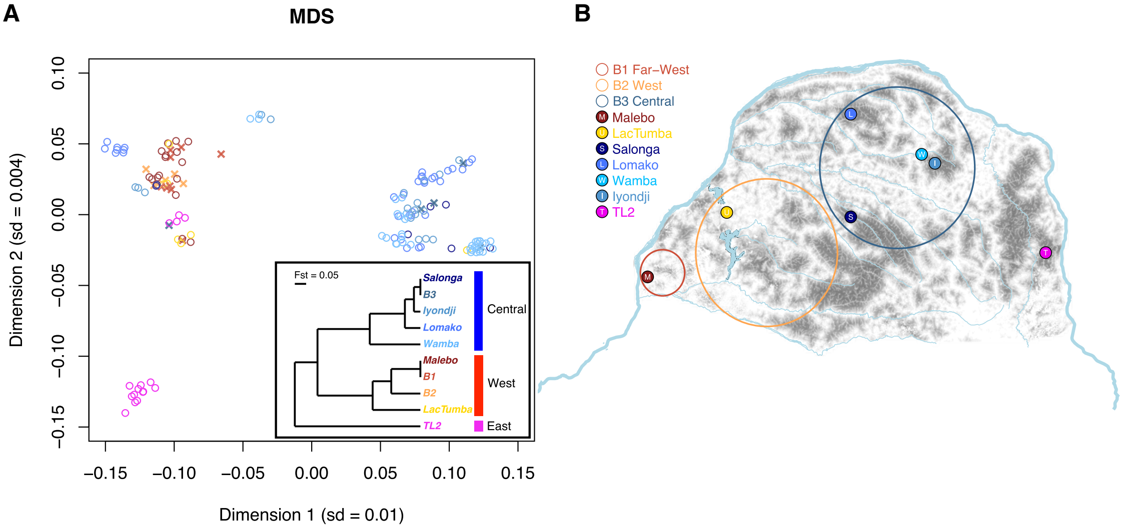

Figure 2. MDS plot and approximate geographical origins on a map

(A) MDS of bonobo individuals, combining the mtDNA sequences from our samples (x symbols), published previously[2,13] (circles). For a better visualization, we added jittering. Inset: unweighted pair-group method of analysis (UPGMA) tree computed for the ten different sample groups using Φ_{ST}.[16]

(B) Sampling location of the bonobo populations from Kawamoto et al.[13] in filled circles and a potential area of origin of our inferred populations in empty circles. The gray areas reflect the probability of suitable habitat for bonobos according to Hickey et al.,[16] where the darker the area the higher the probability. The main rivers and lakes of the Democratic Republic of Congo are shown in light blue.

See also Figure S1.

location of the three bonobo groups. Therefore, we hereafter refer to the B1 group as the "Far-West" population, B2 as the "West" population, and B3 as the "Central" population, based on their likely geographic distribution according to these genetic relationships, and to the Far-West and West populations together as the Western group. These inferences agree with the patterns of genetic differentiation among the three inferred bonobo populations, giving confidence that they reflect the differentiation of geographically separated natural populations. None of our samples fall within the eastern group (namely TL2 in Kawamoto et al.[13]), so we are unable to make inferences about this population. Our analyses and inferences thus concern only the central and western parts of the bonobo geographic range. Based on this analysis, a likely scenario of the geographic range of the three bonobo populations is shown in Figure 2B.

An old split between the western and the central populations

The high level of genetic differentiation among bonobo populations makes estimating split times interesting. Such estimates are most accurate when based on whole genomes, where many neutral loci can be analyzed, so we used ten published high-coverage whole genomes from Prado-Martinez et al.[10] First, to explore the genetic relationships between these and the sanctuary samples, we extracted the exomes of the Prado-Martinez genomes and combined them with the 20 exomes. In a PCA analysis of these 30 exomes, the ten Prado-Martinez individuals cluster well with the three populations we identified (Figure S1), with at least one individual falling in each cluster. Hence, we used one genome, chosen randomly from each population, to make genome-wide inferences about the Far-West, West, and Central populations. Full genomes massively increase the amount of information we have per individual and have the added advantage of containing mostly non-coding, neutral sites. G-PhoCS[17] was then used to infer key demographic parameters, such as population divergence times, effective population sizes, and migration rates. G-PhoCS makes inferences based on unlinked, short, putatively neutral loci from a small number of genomes, making it suitable when sample size per population is small, including when it is restricted to a single individual per lineage, as in this study.[17–20]

For the G-PhoCS analysis, we assumed a phylogenetic tree with the three bonobo populations identified and a western chimpanzee as an outgroup, allowing migration among bonobo populations shortly after their split time (Figure S2B). We chose a western chimpanzee as an outgroup, as previous evidence suggests that this subspecies has not received direct gene flow from bonobos.[11] We confirm that the three groups belong to discrete populations and infer varying levels of gene flow among them (Figures 3, S2B–S2D, and S3; Table 1). The estimated split time between Central and Western bonobos is ~145 kya (95% Bayesian credible interval [CI], 116–302 kya; Table 1), and between West and Far-West ~60 kya (95% CI, 23–116 kya), whereas between chimpanzees and bonobos it is ~1.29 million years ago (mya) (95% CI, 1.21–1.37 mya). When performing the analysis multiple times, CIs overlap between runs, reflecting high concordance among G-PhoCS runs (STAR Methods; Figure S3). As always, such inferences should be considered approximate estimates of population split times because population splits are rarely instantaneous, often being rather complex. For context, the estimated split time between the Central and Western bonobo populations is close to the estimated split time between the Central and Eastern chimpanzee subspecies (~139 kya[11]) and considerably

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

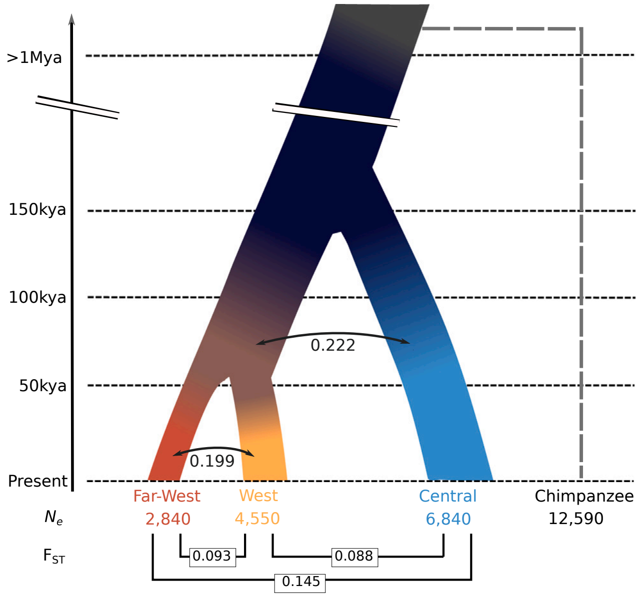

Figure 3. Conceptual demographic model

Split times and effective population sizes (N_e) for the three bonobo populations, as well as migration probabilities (arrows between populations), as estimated by G-PhoCS. One western chimpanzee individual was used as an outgroup. kya, thousand years ago; Mya, million years ago. The values reported here are the estimates from the inference, with full details and Bayesian CIs in Table 1. Distribution of estimates are found in Figures S2 and S3. F_{ST} values are noted at the bottom. Colors for the three groups are maroon for Far-West, orange for West, and blue for Central. See also Table S1.

lower than that between any other chimpanzee subspecies pair (>250 kya[11]). The estimated split time between the two Western populations, on the other hand, is closer to the inferred split time between East African and non-African human populations (65,000 years ago[21–23]), which reflects the genetic differentiation after the out-of-Africa event.

We calculated "mutual migration probability" as a measure of gene flow based on the parameters inferred by G-PhoCS (STAR Methods), which indicates the probability that two populations exchanged genetic material, irrespective of the direction, since their divergence and until the next population split. The calculated probabilities of migration across the 20 replicate runs appear to fluctuate more across runs than other demographic parameters, such as split time and N_e. Migration probabilities estimated between West and Far-West populations in particular show high variation across replicates (Figure S2D). This is likely due to the algorithm being unable to differentiate between deeper divergence with frequent introgression or more recent divergence with sporadic introgression, as shown by the positive correlation between estimated migration probabilities and estimated divergence times (Figure S2D). As expected, a model without migration (Figure S2B) provides similar, albeit slightly younger, estimates: the split time between the Central and Western populations is estimated to be ~127 kya (95% CI, 116–140 kya; Table 1), between the West and Far-West populations to be ~37 kya (95% CI, 23–47 kya), and between chimpanzees and bonobos to be ~1.20 mya (95% CI, 1.14–1.28 mya).

Different genetic diversity, effective N_e, and inbreeding

The levels of genetic diversity differ substantially among populations (Figure S1G), with the mean number of pairwise differences in Central being almost 1.4 times that in Far-West (Far-West has 3.28, West 4.03, and Central 4.53 in windows of 10,000 base pairs [kb]). Heterozygosity (Figure S1F) shows similar patterns (Far-West has, on average, 2.16, West 2.59, and Central 2.56 heterozygous sites in 10 kb), and both measures agree well with mtDNA haplotype diversity,[13] where central populations show higher levels of mtDNA diversity than western populations. Differences in genetic diversity often reflect differences in demographic history. G-PhoCS's estimated N_e is lowest for the Far-West bonobos (2,840, 95% CI, 1,600–4,010) compared with West (4,550, 95% CI, 1,600–7,220) and Central (6,840, 95% CI, 4,810–8,020) (Figures 3, S2C, and S3; Table 1). The estimated Central N_e is more than two times larger than the Far-Western N_e, suggesting substantial differences in the demographic history of the populations. Further, some of these N_e estimates are even lower than previously reported for bonobos as a whole, ranging from 5,000 to 29,129.[10,12] This is not surprising, as these were estimates of a long-term N_e for bonobos as a species, whereas we are estimating specific populations since their divergences. However, it is notable that the N_e estimate of Far-West bonobos is among the lowest N_e estimates across all great ape groups, comparable only with Mountain gorillas and Tapanuli Orangutans, with ~2,200[24] and ~2,500,[25] respectively. Of note, the estimated N_e of the common ancestor of the three bonobo populations is 23,230 (Table 1), which is in line with the estimated N_e of bonobos from Kuhlwilm et al.[12]

A possible effect of small populations is inbreeding, the mating of closely related individuals. Inbreeding increases homozygosity, which can expose deleterious recessive alleles in inbred populations. To investigate putative recent inbreeding, we examined the distribution of runs of homozygosity (ROHs), which are continuous segments depleted of heterozygous positions in an individual due to a shared recent ancestor. Bonobos as a group have on average longer ROHs than our samples of humans and chimpanzees, as expected from their low N_e (Figures S2C and S3). Among bonobos, the Far-West individuals have the longest ROHs, which are, on average, 29% longer than in the West and the Central bonobos. Strikingly, these are 61% longer than those in humans and 139% longer than those in chimpanzees (Figure S1D). Not surprisingly, Far-Western bonobos also have the highest level of identity-by-descent (IBD) segments. IBD segments are inferred from identical chromosome fragments between two individuals sharing the same recent common ancestor, resulting from mating in small and/or inbred populations. All bonobos have IBD segments, shared predominantly with other members of their population, as is the case in our samples of chimpanzees and humans (Yoruba). Still, bonobos in the Far-West show longer IBD segments than bonobos in the other populations, with a total IBD length twice as long as the other two populations (Figure S1D). Far-West bonobos are also related to a much greater extent than Central and even Western populations, which suggests that they are more inbred (Figure S1E). G-PhoCS infers migration between Central bonobos and the common ancestor of the Western bonobos and migration between West and Far-West populations (mutual migration probability of 0.22 and 0.20, respectively; Table 1). Yet our results suggest that

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

Table 1. Estimate summaries

Parameters	With migration			Without migration			Units
	estimate	low	high	estimate	low	high	
T_farWest-West	59.93	23.26	116.28	36.76	23.26	46.51	kya
T_Western-Central	144.66	116.28	302.33	127.46	116.28	139.53	kya
T_Bonobo-Chimp	1,289.20	1,209.30	1,372.09	1,198.44	1,139.53	1,279.07	kya
Ne_farWest	2.84	1.60	4.01	2.64	1.60	4.01	k individuals
Ne_West	4.55	1.60	7.22	3.69	2.41	5.61	k individuals
Ne_Central	6.84	4.81	8.02	7.53	7.22	8.02	k individuals
Ne_Chimp	12.59	12.03	13.63	13.43	12.83	14.43	k individuals
Ne_Western	28.68	8.82	65.76	26.33	17.64	37.69	k individuals
Ne_Bonobo	23.23	21.63	24.86	22.28	20.85	23.26	k individuals
Ne_Ancestral	217.00	211.71	222.13	218.34	213.31	223.74	k individuals
m_farWest-West	0.199	0.007	0.497	N/A	N/A	N/A	probability
m_Western-Central	0.222	0.000	0.979	N/A	N/A	N/A	probability
m_Bonobo-Chimp	0.051	0.027	0.085	N/A	N/A	N/A	probability

Average values for each parameter across 20 G-PhoCs runs using 900 MCMC samples per run. Mean and 95% Bayesian credible intervals (CIs) of split times (T), effective population sizes (Ne), and migration probabilities (m) between populations.

even in the presence of some gene flow, the Far-West is genetically differentiated from the other populations, likely as a result of their fragmented habitat (see the area surrounding the Malebo population in Figure 2B).

Potential signatures of positive selection

Genetic adaptation has contributed to the genetic differentiation among chimpanzee subspecies, as shown by analyses of whole genomes from these four subspecies[26–28]; it has also contributed to differentiation within subspecies, as shown by the analysis of exomes of multiple populations.[29] We are not able to test the evidence of differential local adaptation here, but the three identified bonobo populations are genetically distinct and appear to be geographically differentiated, so they might also have adapted to their fine-scale local environment. We thus used GRoSS[30] to identify the SNPs with the largest allele frequency differentiation among the three populations (STAR Methods). In the presence of local positive selection, these SNPs are best candidates to have mediated such potential adaptations. This empirical approach does not demonstrate that local adaptation among bonobo populations has taken place, but it allows us to identify the genes with the strongest allele frequency change, which are the most likely targets of putative local adaptation. As expected, exons with the lowest p value differentiate the three populations in a PCA (Figure S4), and their average difference in allele frequency between populations is higher than in the rest of the exome (0.21–0.27 in outlier exons vs. 0.121–0.16 exome-wide for Far-West, 0.174-0.193 vs. 0.121–0.146 exome-wide for West, and 0.296-0.251 vs. 0.16–0.146 exome-wide for Central). However, these SNPs are not strongly differentiated, suggesting that there is either limited positive selection or that it is not monogenic. Only 17 exons have a GRoSS p value <0.01 after multiple testing correction (Benjamini-Hochberg): 11 in Far-West, four in West, and two in Central (Data S1A–S1D). Among them, perhaps the most interesting is the transcription-factor-coding gene *PRDM10* (adjusted GRoSS p value = 0.0020), which includes a missense variant (P > T) with a frequency of 0.86 in Far-West and absence in the other populations. This gene shows signatures of positive selection in the ancestor of modern humans after the split from Neanderthals.[31,32] Whether this and other candidates (in Data S1) are true targets of positive selection, and more generally whether genetic adaptation has contributed to the genetic differentiation of these bonobo populations, will require the analysis of additional samples, ideally of known geographic origin and including the eastern side of the bonobo range.

Conclusions

Our results reveal the presence of strong population substructure in bonobos, indicating at least three distinct populations in the west and central part of their range. The genetic differentiation and estimated divergence time between the Central and Western populations is of comparable depth to those between the closest chimpanzee subspecies, Central and Eastern, which diverged more than 100 kya,[11] and lower than that among the other chimpanzee subspecies.[11] Bonobos live exclusively in the DRC and, although they have been studied for some time, thus far little behavioral diversity has been observed among groups, with only some differences reported in drumming, grooming, and hunting among central bonobo communities—the best-studied ones through long-term data collection.[6,7] It is notable that the three bonobo populations are not only genetically differentiated but that they also have substantially different evolutionary histories and have experienced different levels of inbreeding. The Far-West populations show evidence of particularly long-term low N_e and genetic isolation (with sporadic gene flow), making them potentially vulnerable to future ecological changes and anthropogenic pressures such as human encroachment. Of note, without sampling wild individuals of known geographic origin, which is possible only with non-invasive samples, we are unable to discriminate between a model of clinal differentiation and one of discrete population divergence between the three groups. Further, our sampling does not appear to cover the eastern population (TL2 in Figure 2), which previous mtDNA studies identified as potentially highly differentiated, with a possibility of genetic exchange with other populations.[13,33] Thus, our work stresses the importance of performing genomic studies of (non-invasive) samples of wild bonobos across their full geographical range, especially the eastern end of the range, in order to identify all genetically differentiated bonobo populations, infer their evolutionary history, and establish their potential connections. Although information on eastern populations is unlikely to substantially modify our inferences for the Central and Western populations, it will almost certainly reveal even larger levels of population differentiation in bonobos, and it is essential to achieve a much-needed comprehensive understanding of the whole species. For now, the presence of genetically differentiated bonobo populations, and the high genetic isolation of some of these populations, should be taken into consideration in the planning of conservation efforts.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to, and will be fulfilled by, the lead contact, Aida M. Andrés (a.andres@ucl.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All the data used for the analyses in the study were previously published and are listed in the key resources table under the section “deposited data.”
- This paper does not report original code.
- All the analyses are described in detail in STAR Methods and any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

C.d.F. and A.M.A. were supported by intramural funds from the Max Planck Institute for Evolutionary Anthropology (Leipzig, Germany). C.d.F. was also supported by the European Research Council (ERC) grant agreement no. 694707 to Svante Pääbo. A.M.A. was additionally funded by the Wellcome Trust ISSF3 UCL award 204841/Z/16/Z. M.K. was supported by “La Caixa” Foundation (ID 100010434), fellowship code LCF/BQ/PR19/11700002, the Vienna Science and Technology Fund (WWTF) (10.47379/VRG20001), and the Austrian Science Fund (FWF) (FW547002). T.M.-B. was supported by the European Research Council (ERC) grant agreement no. 864203. We would like to thank Svante Pääbo for access to samples and discussions, Mimi Arandjelovic and Martin Surbeck for their valuable comments, Roger Mundry for the R-programming discussions, Hjalmar Kühl for sharing the data of the bonobo’s habitat, and Hiroyuki Takemoto for providing the location of the sampled individuals analyzed for mtDNA. We are grateful to the Life Science Computer Cluster of the University of Vienna.

AUTHOR CONTRIBUTIONS

A.M.A. and C.d.F. conceived the project. S.H., C.d.F., G.P., J.R.M., R.L., M.K., and P.F. analyzed data, with support from A.M.A., I.G., T.M.-B., and C.H. A.M.A., T.M.-B., and M.K. provided funding. S.H., C.d.F., and A.M.A. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data Preparation
  - Population substructure and differentiation
  - mtDNA analyses
  - Demographic inference
  - IBS, IBD and ROH
  - Positive selection
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.09.043.

Received: April 8, 2024
Revised: July 7, 2024
Accepted: September 17, 2024
Published: October 15, 2024

REFERENCES

1. Groves, C. (2001). Primate Taxonomy (Smithsonian Books).
2. Fischer, A., Prüfer, K., Good, J.M., Halbwax, M., Wiebe, V., André, C., Atencia, R., Mugisha, L., Ptak, S.E., and Pääbo, S. (2011). Bonobos fall within the genomic variation of Chimpanzees. PLoS One *6*, e21605. https://doi.org/10.1371/journal.pone.0021605.
3. IUCN; ICCN (2012). Bonobo (*Pan paniscus*): Conservation Strategy 2012–2022 (IUCN and ICCN).
4. Whiten, A. (2020). Does culture shape hunting behavior in bonobos? eLife *9*, e62104. https://doi.org/10.7554/eLife.62104.
5. Furuichi, T. (2011). Female contributions to the peaceful nature of bonobo society. Evol. Anthropol. *20*, 131–142. https://doi.org/10.1002/evan.20308.
6. Hohmann, G., and Fruth, B. (2003). Culture in bonobos? Between-species and within-species variation in behavior. Curr. Anthropol. *44*, 563–571. https://doi.org/10.1086/377649.
7. Samuni, L., Wegdell, F., and Surbeck, M. (2020). Behavioural diversity of bonobo prey preference as a potential cultural trait. eLife *9*, e59191. https://doi.org/10.7554/eLife.59191.
8. Eriksson, J., Hohmann, G., Boesch, C., and Vigilant, L. (2004). Rivers influence the population genetic structure of bonobos (Pan paniscus). Mol. Ecol. *13*, 3425–3435. https://doi.org/10.1111/j.1365-294X.2004.02332.x.
9. Eriksson, J., Siedel, H., Lukas, D., Kayser, M., Erler, A., Hashimoto, C., Hohmann, G., Boesch, C., and Vigilant, L. (2006). Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (Pan paniscus). Mol. Ecol. *15*, 939–949. https://doi.org/10.1111/j.1365-294X.2006.02845.x.
10. 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.
11. De Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V.C.V.C.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.

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

12. 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.

13. Kawamoto, Y., Takemoto, H., Higuchi, S., Sakamaki, T., Hart, J.A., Hart, T.B., Tokuyama, N., Reinartz, G.E., Guislain, P., Dupain, J., et al. (2013). Genetic structure of wild bonobo populations: diversity of mitochondrial DNA and geographical distribution. PLoS One *8*, e59660. https://doi.org/10.1371/journal.pone.0059660.

14. Teixeira, J.C., de Filippo, C., Weihmann, A., Meneu, J.R., Racimo, F., Dannemann, M., Nickel, B., Fischer, A., Halbwax, M., Andre, C., et al. (2015). Long-term balancing selection in LAD1 maintains a missense trans-species polymorphism in humans, chimpanzees, and bonobos. Mol. Biol. Evol. *32*, 1186–1196. https://doi.org/10.1093/molbev/msv007.

15. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.

16. Hickey, J.R., Nackoney, J., Nibbelink, N.P., Blake, S., Bonyenge, A., Coxe, S., Dupain, J., Emetshu, M., Furuichi, T., Grossmann, F., et al. (2013). Human proximity and habitat fragmentation are key drivers of the range-wide bonobo distribution. Biodivers. Conserv. *22*, 3085–3104. https://doi.org/10.1007/s10531-013-0572-7.

17. Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., and Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. *43*, 1031–1034. https://doi.org/10.1038/ng.937.

18. Kuhlwilm, M., Gronau, I., Hubisz, M.J.M.J., de Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H.A.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.

19. Beichman, A.C., Huerta-Sanchez, E., and Lohmueller, K.E. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. Annu. Rev. Ecol. Evol. Syst. *49*, 433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431.

20. Chavez, D.E., Gronau, I., Hains, T., Dikow, R.B., Frandsen, P.B., Figueiró, H.V., Garcez, F.S., Tchaicka, L., de Paula, R.C., Rodrigues, F.H.G., et al. (2022). Comparative genomics uncovers the evolutionary history, demography, and molecular adaptations of South American canids. Proc. Natl. Acad. Sci. USA *119*, e2205986119. https://doi.org/10.1073/pnas.2205986119.

21. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496. https://doi.org/10.1038/nature10231.

22. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925. https://doi.org/10.1038/ng.3015.

23. Hershkovitz, I., Marder, O., Ayalon, A., Bar-Matthews, M., Yasur, G., Boaretto, E., Caracuta, V., Alex, B., Frumkin, A., Goder-Goldberger, M., et al. (2015). Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans. Nature *520*, 216–219. https://doi.org/10.1038/nature14134.

24. Pawar, H., Rymbekova, A., Cuadros-Espinoza, S., Huang, X., de Manuel, M., van der Valk, T., Lobon, I., Alvarez-Estape, M., Haber, M., Dolgova, O., et al. (2023). Ghost admixture in eastern gorillas. Nat. Ecol. Evol. *7*, 1503–1514. https://doi.org/10.1038/s41559-023-02145-2.

25. 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.

26. Cagan, A., Theunert, C., Laayouni, H., Santpere, G., Pybus, M., Casals, F., Prüfer, K., Navarro, A., Marques-Bonet, T., Bertranpetit, J., and Andrés, A.M. (2016). Natural selection in the great apes. Mol. Biol. Evol. *33*, 3268–3283. https://doi.org/10.1093/molbev/msw215.

27. Pawar, H., Ostridge, H.J., Schmidt, J.M., and Andrés, A.M. (2022). Genetic adaptations to SIV across chimpanzee populations. PLoS Genet. *18*, e1010337.

28. Schmidt, J.M., de Manuel, M., Marques-Bonet, T., Castellano, S., and Andrés, A.M. (2019). The impact of genetic adaptation on chimpanzee subspecies differentiation. PLoS Genet. *15*, e1008485.

29. Ostridge, H.J., Fontsere, C., Lizano, E., Soto, D.C., Schmidt, J.M., Saxena, V., Alvarez-Estape, M., Barratt, C.D., Gratton, P., Bocksberger, G., et al. (2024). Local genetic adaptation to habitat in wild chimpanzees. Preprint at bioRxiv. https://doi.org/10.1101/2024.07.09.601734.

30. Refoyo-Martínez, A., da Fonseca, R.R., Halldórsdóttir, K., Árnason, E., Mailund, T., and Racimo, F. (2019). Identifying loci under positive selection in complex population histories. Genome Res. *29*, 1506–1520. https://doi.org/10.1101/gr.246777.118.

31. Peyrégne, S., Boyle, M.J., Dannemann, M., and Prüfer, K. (2017). Detecting ancient positive selection in humans using extended lineage sorting. Genome Res. *27*, 1563–1572. https://doi.org/10.1101/gr.219493.116.

32. Kuhlwilm, M., and Boeckx, C. (2019). A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. Sci. Rep. *9*, 8463. https://doi.org/10.1038/s41598-019-44877-x.

33. Takemoto, H., Kawamoto, Y., Higuchi, S., Makinose, E., Hart, J.A., Hart, T.B., Sakamaki, T., Tokuyama, N., Reinartz, G.E., Guislain, P., et al. (2017). The mitochondrial ancestor of bonobos and the origin of their major haplogroups. PLoS One *12*, e0174851. https://doi.org/10.1371/journal.pone.0174851.

34. Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

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

36. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.

37. Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics *23*, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233.

38. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment, software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

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

40. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.

41. 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.

42. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution (N.Y) *38*, 1358–1370. https://doi.org/10.2307/2408641.

43. Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. *16*, 111–120.

44. Michalakis, Y., and Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics *142*, 1061–1064. https://doi.org/10.1093/genetics/142.3.1061.

45. Schlebusch, C.M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A.R., Vicente, M., Steyn, M., Soodyall, H., et al.

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

(2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science *358*, 652–655.

46. Besenbacher, S., Hvilsom, C., Marques-Bonet, T., Mailund, T., and Schierup, M.H. (2019). Direct estimation of mutations in great apes reconciles phylogenetic dating. Nat. Ecol. Evol. *3*, 286–292. https://doi.org/10.1038/s41559-018-0778-x.

47. Jombart, T., and Ahmed, I. (2011). adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. Bioinformatics *27*, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.

48. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
whole-exome high-coverage (∼20X) Illumina sequencing data of 20 bonobos (*Pan paniscus*), 20 central chimpanzees (*Pan troglodytes troglodytes*) and 20 humans (*Homo sapiens*)	Teixeira et al.[14]	https://doi.org/10.1093/molbev/msv007
mtDNA sequences of the 20 bonobos	Fischer et al.[2]	https://doi.org/10.1371/journal.pone.0021605
10 high-coverage (∼27X) bonobo whole-genomes	Prado-Marinez et al.[10]	https://doi.org/10.1038/nature12228
D-loop of the mtDNA of 136 wild bonobo individuals	Kawamoto et al.[13]	https://doi.org/10.1371/journal.pone.0059660
Software and algorithms		
function ‘glPca’	R-package “adegenet”[34]	https://adegenet.r-forge.r-project.org/
function ‘nj’	R-package “ape”[35]	https://emmanuelparadis.github.io/
ADMIXTURE	ADMIXTURE[36]	https://dalexander.github.io/admixture/publications.html
CLUMPP	CLUMPP[37]	https://academic.oup.com/bioinformatics/article/23/14/1801/188285
mafft’ v7 (Katoh and Standley, 2013[38]	mafft’ v7[38]	https://mafft.cbrc.jp/alignment/software/
function ‘cmdscale’	A basic function in R “Stats”	https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cmdscale
function ‘upgma’	R-package “phangorn”[39]	https://www.rdocumentation.org/packages/phangorn/versions/2.11.1
G-PhoCS	G-PhoCS[17]	https://github.com/gphocs-dev/G-PhoCS
KING	KING[40]	https://www.kingrelatedness.com/manual.shtml
GroSS	GroSS[30]	https://github.com/FerRacimo/GroSS

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

As this study analyzed previously published datasets, there is nothing to declare.

METHOD DETAILS

Data Preparation

We analyzed whole-exome high-coverage (∼20X) Illumina sequencing data of 20 bonobos (*Pan paniscus*), 20 central chimpanzees (*Pan troglodytes troglodytes*) and 20 humans (*Homo sapiens*) that we published previously.[14] Bonobo and chimpanzee blood samples were collected in African sanctuaries (Lola ya bonobo sanctuary in Kinshasa, DRC; and Tchimpounga sanctuary, Jane Goodall Institute, Republic of Congo, respectively) and human samples belong to the Yoruba population from HapMap.[14] The mtDNA sequences of the 20 bonobos were previously published.[2] We also used 10 high-coverage (∼27X) bonobo whole-genomes which are from wild-born individuals residing in European zoos.[10] The protocols on how exomes and whole-genomes were mapped and prepared are available in these previous publications.[10,14] The exomes were used for PCA, ADMIXTURE, NJ, F_{ST}, and GRoSS, which require polymorphism information, while the genomes were used for G-PhoCS, which require putatively neutral loci in individual genomes.

Population substructure and differentiation

PCA were performed using the function ‘glPca’ from the R-package ‘adegenet’[34] and run for all individuals together and separately per species. NJ trees were generated using the function ‘nj’ from the R-package ‘ape’[35] also constructed for all individuals together, and separately per species, with a distance matrix of pairwise nucleotide differences such that the distance between a heterozygote and any homozygote is 1, and the distance between the two different homozygotes is 2. This was used to calculate the number of pairwise differences between individuals, which was then divided by the total number of base-pairs that passed the filters in all species (25,781,213) to have a comparative measurement. The number of variant sites across all autosomal sites (SNPs and/or fixed differences) is 228,488 for chimpanzees, 86,250 for bonobos, and 106,832 for Yoruba humans. ADMIXTURE[36] was run on a subset

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

of SNPs that minimizes linkage disequilibrium (LD) by removing high-LD SNPs with plink[41] with the following steps: 1) create a window of 200 SNPs; 2) calculate LD (as r^2) between each pair of SNPs in the window; 3) if $r^2 > 0.5$ remove one of a pair of SNPs; 4) shift the window 20 SNPs and repeat the procedure. ADMIXTURE was run from K=1 to K=8, each with 10 replicates using the following command line: "admixture -s time –cv INPUT.ped k". We run the cross-validation procedure ("–cv" flag) as described in Alexander et al.[36] to determine the number of K that best fits the data. The software CLUMPP[37] was used to condense the 10 admixture runs per K to identify modes where different runs have similar outcomes (>90%) by selecting pairs of replicates having a symmetric similarity coefficient G' > 0.9.

To measure pairwise population differentiation, we calculated the average F_{ST} across all sites that are polymorphic in at least one of the two populations, using a custom script implementing the formula of Weir and Cockerham's F_{ST}.[42] The pairwise comparisons are: 1) among bonobo groups using our exome data; 2) among all chimpanzee subspecies using the GAGDP dataset; 3) among six human populations from Africa (Yoruba and Luhya), Europe (Toscani and Finns) and Asia (Han Chinese and Japanese) using the 1000 Genomes data.[15] Heterozygosity is calculated for each individual simply as the number of heterozygous sites over the total number of base pairs.

mtDNA analyses

Kawamoto et al.[13] analyzed the D-loop of the mtDNA of 136 bonobo individuals to generate a phylogenetic tree of samples of known geographic origin. Together with the published mtDNA data[2] of our 20 bonobos, in total 156 mtDNA sequences were aligned with the software 'mafft' v7,[38] and filtered to remove positions with indels and missing data, retaining 1,101bp. We calculated pairwise sequence differences using the Kimura 2-parameters model[43] and performed MDS on these distances using the R-function 'cmdscale'. MDS is a dimensional reduction, similar to PCA albeit using different methodologies, which allows us to visualize the information contained in a matrix of pairwise distances with some degree of loss function called "stress". In order to visualize samples with identical sequences/haplotypes, we randomly added noise (or jittered) as 10% of the standard deviation for each dimension of the MDS, which are 0.0100 and 0.0045, for Dimensions 1 and 2, respectively. We calculated Φ_{ST} among 10 bonobo groups according to the formula of Michalakis & Excoffier[44] with a custom script, which then was used for building an UPGMA tree. Negative values were set to 0. For plotting, we used the function 'upgma' of the R-package 'phangorn'[39] with default parameters. To build NJ-trees, as for the autosome, we used the function 'nj' from R-package 'ape'[35].

Demographic inference

To infer the demographic history of the three populations, we used the Generalised Phylogenetic Coalescent Sampler (G-PhoCS[17]), a Bayesian sampling method that summarizes the information over local genealogies at short, putatively neutral loci in approximate linkage equilibrium. G-PhoCS infers demographic parameters, such as divergence times, effective population sizes and migration rates, given a pre-specified population phylogeny. We used the UPGMA tree from their mtDNA haplotypes (Figure 2A), which is consistent with the NJ tree topology from autosomal data (Figure 1B), for the tree topology of three bonobo populations and a western chimpanzee (Figure S3B). G-PhoCS can produce reliable parameters using one full genome per population, so we used bonobo whole genomes from Prado-Martinez et al.,[10] one genome per bonobo population, which were selected based on their PCA clustering with the exomes (Results and Figure S3A). Our analysis follows best-practice procedures previously defined,[17,18,45] including filters and parameters. We used eight quality filters downloaded from the UCSC genome annotation database for hg19 (http://hgdownload.cse.ucsc.edu/goldenpath/hg19/database/, the last date of access 29/04/2019) to remove known genic regions (refGene, knownGene), simple and complex repeat regions (simpleRepeat, genomicSuperDups), CpG islands (cpgIslandExt), repeat masker (rmsk), conserved regions across 46 placental species (phastConsElements46wayPlacental), and synteny net between the assemblies hg19 and PanTro4 (netPanTro4). We computed a set of 1000 bp long loci that avoided the genomic filters and separated by at least 10 kb. This resulted in 72,607 loci overall. We then partitioned this set into three subsets of 24,202 or 24,203 alternating loci, and used the first two subsets in our analysis (subset 1 and 2). Finally, we removed loci with more than 20% missing data for one of the analyzed individuals. As a result, each analysis considered 23,439 and 23,424 independent loci sets that are separated by at least 30 kb. These properties were shown to minimize the impact of recombination within loci and maximize independence between loci.[17] Only fragments with less than 20% missing data within each individual were analyzed. The numbers of fragments used in the G-PhoCS run are 23,439 for Set 1 and 23,424 for Set 2. We ran G-PhoCS with and without migration (Figure S3B), in both cases with 1,000,000 MCMC iterations in each run. We ran 10 replicate runs using each loci set and in each migration scenario, which results in 20 replicates per migration scenario. To estimate split time (Tau) and N_e (Theta), we discarded the first 100,000 iterations as burn-ins, taking one estimate every 1,000 iterations in order to minimize the issue of auto-correlation, which yielded 18,000 MCMC samples per parameter in each migration scenario. We then calculated mean values as point estimates and 95% Bayesian credible intervals as plausible ranges. To convert Tau and Theta to generations and N_e, we used a mutation rate of 0.43 x 10-9 per site per generation.[46] We further used 29 years as an average generation time to have the split times in calendar years. For migration between populations, we use 'migration probability', which ranges between 0 and 1, and calculated as $P = 1 - e^{\text{-rate}}$. This is a total rate, which is a product of the migration rate and time of migration, summing the rates in both directions.

IBS, IBD and ROH

Identity-by-state (IBS) can be observed at a given locus, for any given pair of individuals with genotype information, with three possible outcomes: the individuals have two different alleles (IBS0) or they share one (IBS1) or two (IBS2) alleles. Two individuals

Please cite this article in press as: Han et al., Deep genetic substructure within bonobos, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.09.043

who share 1 or 2 alleles IBS at a given locus may have inherited them from a recent common ancestor, in which case these alleles are identical-by-descent (IBD). We inferred IBD using a software KING[40] to infer the length and the number of IBD segments. Importantly, KING design allows robust inference of IBD segments in the presence of population substructure, as is the case in the populations used in this study. We ran KING with the "–ibdseg" and "–related" options, and other parameters were set to default values. IBD regions tend to be short between pairs of individuals derived from a given population that are not closely related, primarily because their last common ancestor was many generations ago; they tend to be long among closely related individuals. Using the set of informative nucleotide positions for the three species, we computed the number and length of the IBD regions (defined here as regions between IBS0 alleles) among all the individuals. Runs of homozygosity (ROH), which are continuous segments depleted of heterozygous positions likely because the two chromosomes, are derived from the same recent ancestor. As exome data are composed mostly of the genic part of the genome and cannot be used to infer with precision the ends of ROHs, we restricted the analysis to the comparison among individuals. We then compared the average ROH value in each bonobo group with the value of central chimpanzees and the value of Yoruba humans sequenced in the identical way.

Positive selection

To identify potential targets of differential adaptation among populations, we applied GroSS.[30] This method uses allele frequency data to identify genomic regions putatively under positive selection, and the branch where selection took place. Here, the three bonobo populations can be represented by a simple three-branch tree (Figure 3), where Far-West and West split from the common ancestor with Central. We note that the method does not distinguish between selection on the common branch of Far-West and West, or on the Central branch. For each individual, we removed sites with less than 5 reads, more than 99 reads, or genotype quality below 20, resulting in 1,078,889 segregating sites, and ran GroSS (https://github.com/FerRacimo/GroSS) using R version 3.4.2 (Data S1). We processed the output in the R environment, averaging the raw p-values for each exon with more than 3 SNPs (39,133 exons). Few exons reach nominal significance across the genome (at a p-value of < 0.01: Far-West: 36; West: 24; Central: 23). We assessed the most significant exons in a PCA (using an adjusted p-value cutoff of < 0.05) for each of the three branches (Far-West: 93 exons; West: 74 exons; Central: 75 exons), using the R-package 'adegenet'[47] (Figure S4). We inferred functional consequences of the genetic variants in these exons using the Ensembl Variant Effect Predictor.[48]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using R. A Mantel test based on Spearman Rank correlations was performed on pairwise nucleotide distances (Results). G-PhoCS and GRoSS with their respective statistical frameworks were applied according to the manuals.