PNAS RESEARCH ARTICLE | ANTHROPOLOGY

Estimating bonobo (*Pan paniscus*) and chimpanzee (*Pan troglodytes*) evolutionary history from nucleotide site patterns

Colin M. Brand[a,1,2], Frances J. White[a], Alan R. Rogers[b], and Timothy H. Webster[b,1]

Edited by Marcus Feldman, Stanford University, Stanford, CA; received January 16, 2022; accepted March 16, 2022

Admixture appears increasingly ubiquitous in the evolutionary history of various taxa, including humans. Such gene flow likely also occurred among our closest living relatives: bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*). However, our understanding of their evolutionary history has been limited by studies that do not consider all *Pan* lineages or do not analyze all lineages simultaneously, resulting in conflicting demographic models. Here, we investigate this gap in knowledge using nucleotide site patterns calculated from whole-genome sequences from the autosomes of 71 bonobos and chimpanzees, representing all five extant *Pan* lineages. We estimated demographic parameters and compared all previously proposed demographic models for this clade. We further considered sex bias in *Pan* evolutionary history by analyzing the site patterns from the X chromosome. We show that 1) 21% of autosomal DNA in eastern chimpanzees derives from western chimpanzee introgression and that 2) all four chimpanzee lineages share a common ancestor about 987,000 y ago, much earlier than previous estimates. In addition, we suggest that 3) there was male reproductive skew throughout *Pan* evolutionary history and find evidence of 4) male-biased dispersal from western to eastern chimpanzees. Collectively, these results offer insight into bonobo and chimpanzee evolutionary history and suggest considerable differences between current and historic chimpanzee biogeography.

admixture | Congo River | demography | introgression | male reproductive skew

Significance

There is genomic evidence of widespread admixture in deep time between many closely related species, including humans. Our closest living relatives, bonobos and chimpanzees, may also exhibit such patterns. However, assessing the exact degree of interbreeding remains challenging because previous studies have resulted in multiple inconsistent demographic models. We use an approach that addresses these gaps by analyzing all lineages, simultaneously estimating parameters, and comparing previously models. We find evidence of considerable introgression from western into eastern chimpanzees. We also show more breeding females than males and evidence of male-biased dispersal in western chimpanzees. These findings highlight the extent of admixture in bonobo and chimpanzee evolutionary history and are consistent with substantial differences between past and present chimpanzee biogeography.

It is increasingly apparent that gene flow between populations after divergence is common not only in plants but among animals as well (1–3). While some introgressed sequences may be maladaptive, others may also be neutral or even advantageous (2, 3). Despite the difficulty in detecting admixture, work using whole-genome sequencing data points to its near ubiquity in the evolutionary history of large mammals, including bears (4), elephants (5), and hominins (6–13). This central role in hominins suggests that this is likely also true for other nonhuman primates (14).

Our closest living relatives, bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*), have long been studied for genomic signatures of admixture. Early analyses for admixture in *Pan* from autosomal loci yielded inconsistent results and did not include data from Nigeria–Cameroon chimpanzees (15–19) (Fig. 1). Analysis of whole-genome sequences from all five *Pan* lineages replicated some but not all of these earlier results and also suggested additional introgression events (20, 21), including from an extinct *Pan* lineage into bonobos (22) (Fig. 1). Disagreement among previously proposed models may stem from 1) the exclusion of Nigeria–Cameroon chimpanzees in earlier studies because such data were not available and 2) failing to simultaneously consider all lineages when estimating parameters. Further, more recent studies have 3) not comprehensively compared previous models. A more accurate picture of bonobo and chimpanzee evolutionary history is central to understanding how evolutionary forces have shaped *Pan* genetic variation through time and may shed light on past biogeography.

Here, we address this gap in knowledge by applying a recently developed method, Legofit (23), to comprehensively compare previously proposed models for *Pan* evolutionary history and estimate demographic parameters, including 1) divergence times, 2) effective population sizes, and 3) the timing and degree of introgression. This approach employs site pattern frequencies to infer deep population history by simultaneously estimating all model parameters. There are a few advantages to this approach compared with other commonly used methods for demography. First, within-population variation is ignored, and recent changes in population size, therefore, cannot affect analyses (23). This results in fewer parameters that must be estimated. Second, the uncertainty introduced by statistical identifiability (i.e., when more than one model fits the data well) that is commonly encountered when ascertaining complex demographies can be incorporated into CIs via model averaging (23). Third, simultaneous estimation of all demographic parameters may reduce bias that has been described for other

Author contributions: C.M.B., F.J.W., A.R.R., and T.H.W. designed research; C.M.B. and T.H.W. performed research; C.M.B., A.R.R., and T.H.W. analyzed data; and C.M.B., F.J.W., A.R.R., and T.H.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: colin.brand@ucsf.edu or timothy.h.webster@utah.edu.

[2]Present addresses: Bakar Computational Health Sciences Institute, University of California, San Francisco, CA 94158 and Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94158.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2200858119/-/DCSupplemental.

Published April 22, 2022.

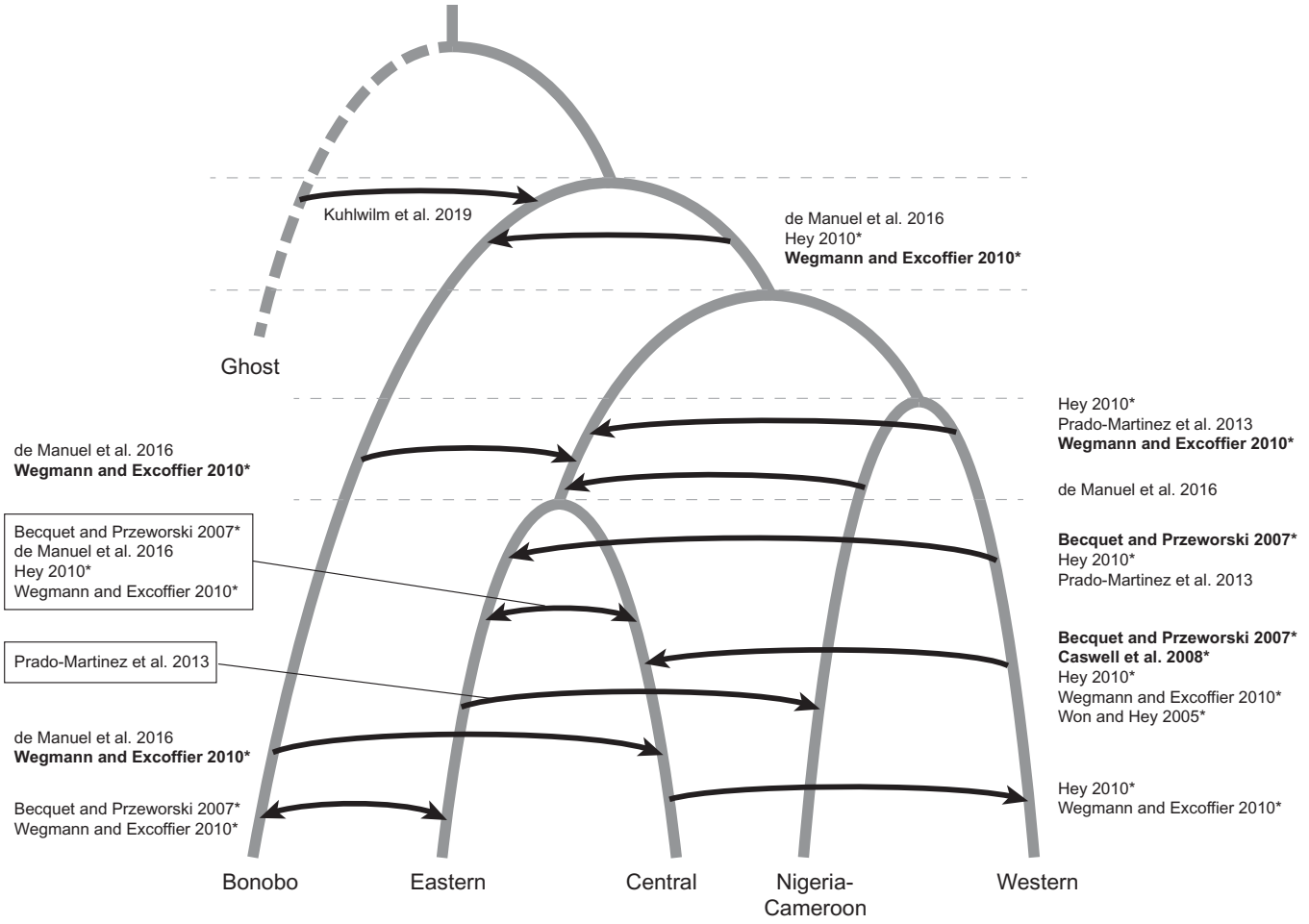

Fig. 1. Previous evidence of admixture in *Pan*. Branch lengths are not to scale. Events only reflect the lineages involved and do not indicate the order/timing of those events between lineage divergences (the dashed horizontal lines). Bolded studies examined gene flow in both directions rather than from one population to another. *These studies did not consider Nigeria–Cameroon chimpanzees in their analyses.

introgression methods (24, 25). In addition to the autosomes, we also consider the demographic history of the X chromosome in a separate analysis and compare it with the autosomes to assess the potential sex biases in *Pan* evolutionary history.

Results

Introgression from Western into Eastern Chimpanzees Best Explains *Pan* Nucleotide Site Patterns. Legofit aligned 2,366,070,805 loci across all six lineages and determined the ancestral allele for 52,809,700 sites. These sites were used to determine site pattern frequencies in the data and generate 50 bootstrap replicates (Fig. 2). A locus exhibits a given site pattern if a sample of random nucleotides drawn from the lineages reflected in that site pattern carries the derived allele and the other lineages carry the ancestral allele (23). For example, the site pattern cn refers to the case in which the derived allele is present in random nucleotides drawn from central and Nigeria–Cameroon chimpanzees but is absent from those drawn from the other populations (bonobos, eastern chimpanzees, and western chimpanzees). As expected from the ∼1.88-Ma divergence between bonobos and chimpanzees and the absence of multiple extant bonobo lineages, the singleton site pattern from bonobos (b) was the most common. The next most common patterns were singleton site patterns from each of the four chimpanzee subspecies (e, c, n, and w) and the site pattern shared across all chimpanzees (ecnw). Among the remaining nonsingleton patterns, the most frequent site patterns were sites unique to both Nigeria–Cameroon and western chimpanzees (nw) and sites unique to both eastern and central chimpanzees (ec). These results are consistent with previously suggested clustering of the chimpanzee subspecies; however, the divergence between Nigeria–Cameroon and western chimpanzees appeared younger than the divergence between eastern and central chimpanzees. This pattern could also be explained by a large effective population size in eastern and central chimpanzees. While previous evidence largely supports this second hypothesis (20, 21), we considered both possibilities. Hereafter, model names that end with "two" reflect a younger divergence between eastern and central chimpanzees. Each of the remaining site patterns accounted for <2% of the total distribution.

We ranked an initial set of models by their bootstrap estimate of predictive error (bepe) (26, 27), calculated by fitting each model to every bootstrap replicate, generating site pattern frequencies for the bootstraps, and calculating the mean squared difference between the bootstrap site pattern frequencies and those in the observed data (23). This initial set encompassed all possible subsets of the α-, β-, γ-, δ-, ε-, and ζ-introgression events ($n = 64$) and where the divergence between Nigeria–Cameroon and western chimpanzees was younger than the divergence between eastern and central chimpanzees (Fig. 3*A*). All models containing γ or introgression from eastern into Nigeria–Cameroon chimpanzees (21) were ranked lowest. We also previously considered introgression

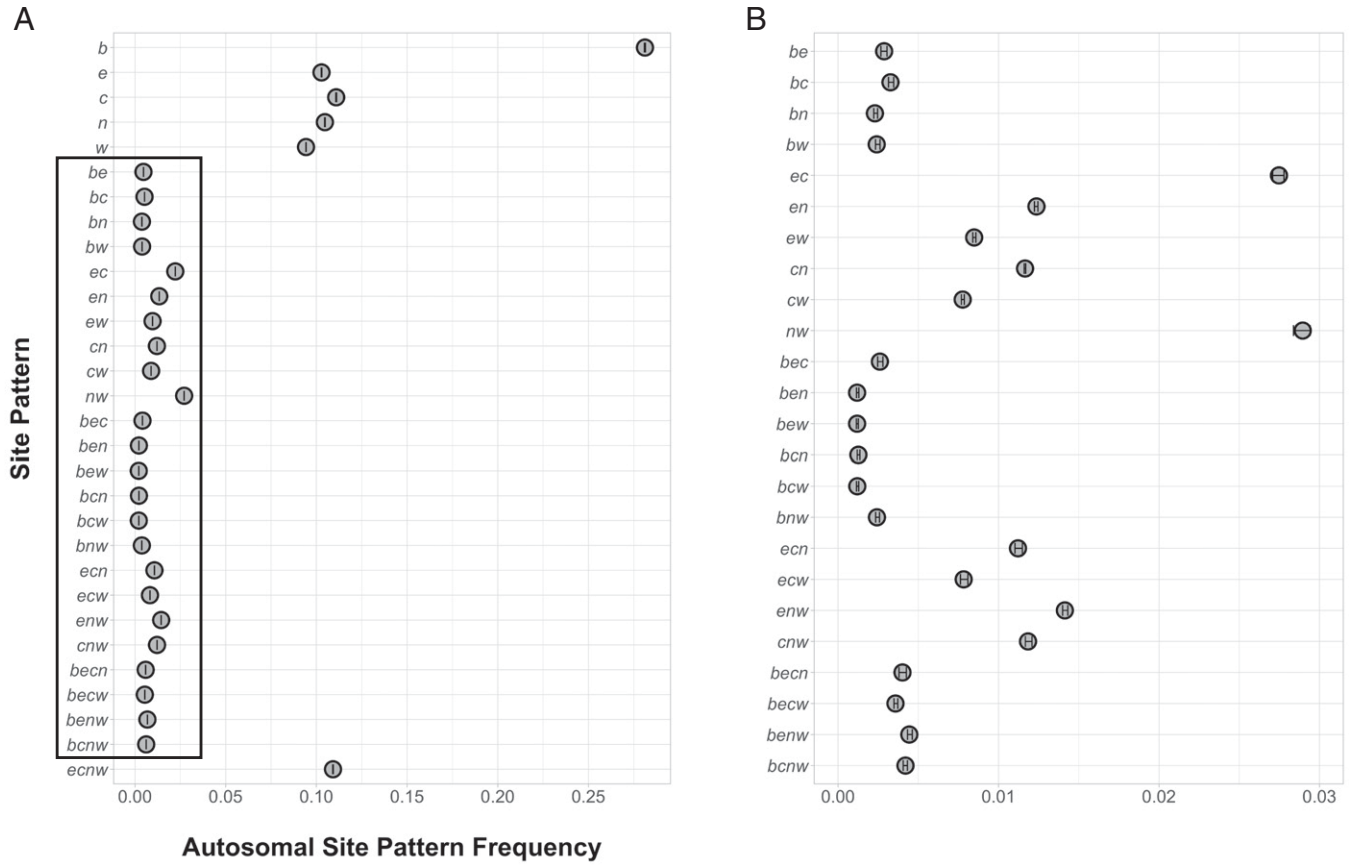

Fig. 2. Observed autosomal site patterns. *A* shows the overall distribution of site patterns, and *B* zooms in on the region encompassed by the black box in *A*. b indicates *P. paniscus*, e indicates *P. t. schweinfurthii*, c indicates *P. t. troglodytes*, n indicates *P. t. ellioti*, and w indicates *P. t. verus*. Points represent the point estimate, and horizontal error bars represent the 95% CIs per site pattern.

from the ancestor of eastern and central chimpanzees into Nigeria–Cameroon chimpanzees (20) and vice versa, but models containing this event were also consistently ranked lowest. Therefore, we excluded γ when we evaluated models ($n = 32$) where the divergence between eastern and central chimpanzee was younger than the divergence between Nigeria–Cameroon and western chimpanzees (Fig. 3*B*).

We also considered whether introgression from bonobos into eastern chimpanzees (η) as well as from western into the ancestor of eastern and central chimpanzees (θ) for models that allowed for this scenario (Fig. 3*B*) would improve model fit of our best-fit models. We added these events separately and together to the top five models (Dataset S2).

Of these 107 models, we found a single model that best fit the observed site patterns: βε2 (Table 1). This model includes two episodes of introgression: 1) from bonobos into the ancestor of eastern and central chimpanzees and 2) from western into eastern chimpanzees. This model exhibited small residuals (Fig. 4) and had the smallest bepe value compared with any other model (1.162×10^{-5}). We also considered the weight for each model or booma, which describes the proportion of times that the model had the lowest bepe value compared with all other models (23, 28). With the observed data and 50 bootstrap replicates, all models were simultaneously evaluated 51 times. βε2 exhibited the lowest bepe each time, resulting in a booma of one for that model and zero for the other models (Dataset S2). As βε2 was superior to all others, model averaging was not invoked.

Point estimates and CIs for the βε2-model parameters are provided in Table 1. This model estimated the age for the most recent common ancestor of all chimpanzees to be 987,000 y old, while the divergence between western and Nigerian–Cameroon chimpanzees dates to 114,000 y old, and the divergence between eastern and central chimpanzees was dated to 33,000 y old. The model also estimated effective population size to vary considerably over time with ∼38,000 individuals at the time of bonobo–chimpanzee divergence and 16,000 chimpanzees immediately prior to the divergence of the chimpanzee common ancestor. The effective population size of both subsequent lineages increased before diverging. The first introgression event in this model occurred from bonobos into the ancestor of eastern and central chimpanzees ∼510,000 y old; however, the estimated admixture proportion was extremely small: 0.006. Given how small this value is, it is possible that the parameter is actually zero, and we consider this event to be possible but tentative. We estimate that the second introgression event occurred 16,000 y old and suggest that ∼21% of the eastern chimpanzee genome derives from western chimpanzees.

After simulating data using this model, calculating site patterns, and fitting these site patterns to the model, we found minimal bias in our parameter estimates for admixture and the effective population size of older events (Fig. 5). The effective population size for the individual lineages appears to be underestimated despite the high estimates for these parameters. It is possible that these values are artifacts of our approach. Both divergence times 1) between eastern and central chimpanzees and 2) between Nigeria–Cameroon and western chimpanzees appear to be underestimated as well. This also appears to be true for the timing of the introgression events themselves. However, our estimated age for the ancestor of all chimpanzees agreed with the simulated data.

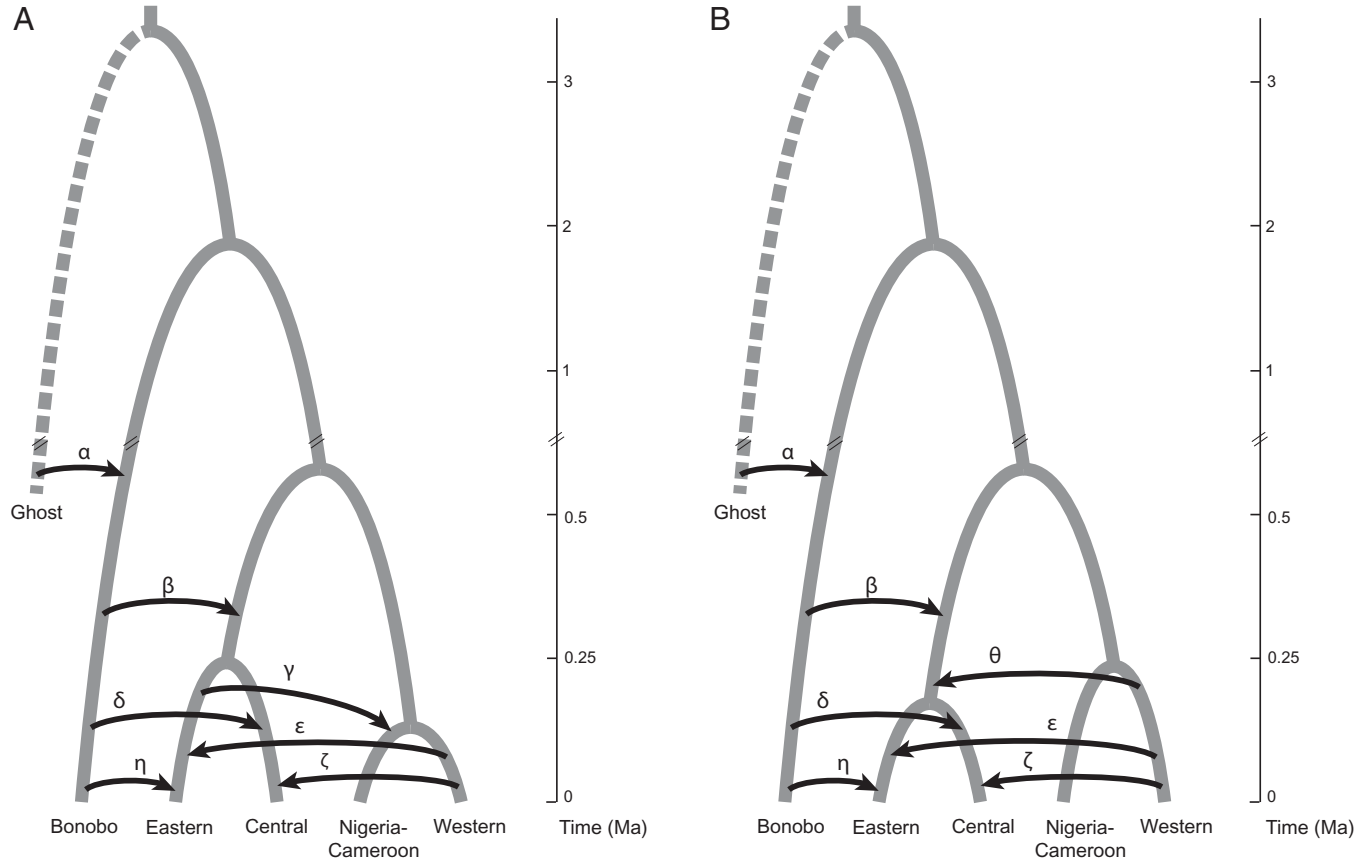

Fig. 3. Introgression events considered in this analysis. In *A*, the divergence time between eastern and central chimpanzees is older than the divergence between Nigeria–Cameroon and western chimpanzees based on the observed site patterns. *B* shows a model where these ages are reversed. All divergence times are estimated from refs. 20 and 22. We initially considered all possible subsets of events α, β, γ, δ, ε, and ζ in *A*. We then considered all possible subsets for those events except for γ in *B* as this event consistently resulted in a poor fit. Finally, we considered whether the addition of η, θ, or η and θ improved model fit for the top five models, only including these events when the topology of the model allowed such inclusion.

X-Chromosome Site Patterns Are Also Explained by the Autosomal Introgression Events and Potentially Additional Introgression. Next, we considered the added complexity of sex bias in *Pan* evolutionary history. We calculated site patterns for the X chromosome (Fig. 6) and three autosomes that are similarly sized to the X chromosome: chromosomes 5, 7, and 8. Site patterns could be determined for 1,932,892 loci on the X chromosome or 3.7% of the loci used in the autosomal analyses and exhibited similar patterns to chromosomes 5, 7, and 8 (Fig. 7). First, we fit these site patterns to the best autosomal model, βε2. All four chromosomes had nearly identical residuals, with the X chromosome exhibiting larger CIs for most of the site patterns (Fig. 7). Despite this fit, we proceeded to evaluate all our previous autosomal models ($n = 96$) as the X chromosome may have a different evolutionary history than the autosomes. As with the autosomal analysis, we considered whether bonobo introgression into eastern chimpanzees (η) and western chimpanzee introgression into the ancestor of eastern and central chimpanzees (θ) improved the fit of our top five X-chromosome models individually and together. We found large agreement in the rank of models evaluated in both the autosomes and X chromosome ($\rho = 0.956$, $n = 102$, $P < 0.001$). Three models exhibited low bepe values: βεη2, βε2, and αβδεηθ2 (Dataset S3). These models were comparable such that the top model was not superior to the others. This resulted in two models that were differentially weighted: βεη2 = 0.863, βε2 = 0, and αβδεηθ2 = 0.137 (Dataset S3).

Table 2 summarizes model averaged parameters for the X-chromosome top model set. The most effective population size and time parameters were congruent with the best autosomal model. The X-chromosome model also included a 0.07 admixture proportion for the α-introgression event (introgression of an extinct *Pan* lineage into bonobos), and it yielded a population size of 178,530 and a 2.03-My-old ancestor for bonobos, chimpanzees, and this extinct *Pan* lineage. In addition to α, the averaged model included introgression from bonobos into central chimpanzees (δ), bonobos into eastern chimpanzees (η), and western chimpanzees into the ancestor of eastern and central chimpanzees (θ). Admixture proportions for these first two events ranged from small (δ = 0.019) to negligible (η = 0.008), and reversing the direction of η did not improve model fit. However, the admixture proportion of the older chimpanzee introgression event, θ, was comparable with that of the western into eastern chimpanzee event: 0.186. Similar to the best autosomal model, the population size estimates for introgressing lineages were unreasonably large.

***Pan* Evolutionary History Is Characterized by Male Reproductive Skew and Male-Biased Dispersal from Western to Eastern Chimpanzees.** While the site patterns from the X chromosome support a slightly different evolutionary history than the autosomes, this history includes both introgression events estimated for the autosomes. Indeed, the top-ranked autosomal model was ranked second for the X chromosome based on its bepe value, and a model with a single additional introgression event comprises the majority of the weight on the averaged X-chromosome model. Therefore, we decided

Table 1. Best-fit autosomal model parameter estimates

Parameter	Point estimate	Lower bound	Upper bound
Admixture			
ε	0.21402	0.20873	0.217315
β	0.00603518	0.00539112	0.00721139
Population size			
b	1,107,330	96,696.0	120,117.5
e	146,408	137,731	165,923.5
w	463,900	416,495.5	50,125.0
ec	145,941.5	142,575.5	149,140.5
nw	80,913.5	79,786.5	82,440
ecnw	15,829.6	15,741.8	15,942.35
becnw	37,676.6	37,527.3	37,955.6
Time			
ε	16,730.125	6,707.8	30,155
β	510,447.5	500,817.5	524,242.5
ec	33,460.25	13,415.575	60,310
nw	114,075.5	101,862.25	121,390.5
ecnw	987,437.5	985,207.5	990,092.5

ε is introgression from *P. t. verus* into *P. t. schweinfurthii*. β is introgression from *P. paniscus* into the ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*. ec is the ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*. nw is the ancestor of *P. t. ellioti* and *P. t. verus*. ecnw is the ancestor of all *P. troglodytes* lineages. becnw is the ancestor of *P. paniscus* and *P. troglodytes*. Admixture is reported as the admixture proportion, population sizes are reported as the number of diploid individuals, and time is reported in years.

to estimate historic differences in female and male effective population sizes that reflect reproductive skew and female–male asymmetry in migration rates during introgression events using the best autosomal model on the X chromosome and a similarly sized autosome: chromosome 7. We used the estimated effective population size and divergence time parameters from the βε2-autosomal model as initial parameter values for both chromosomes and constrained these parameters by a parameter: s1. Constraining these parameters ensures that parameters scale up or down together but that their ratios do not change. We also input the estimated admixture proportions from βε2 as initial values for admixture proportion and constrained both introgression events with a second parameter: s2. Both parameters (s1 and s2) were allowed to range from 0 to 10. We then separately fit these models to the site patterns calculated from chromosome 7 and the X chromosome to estimate the value of both parameters in each chromosome. The chromosome 7 parameters agreed with our predictions. We found that the chromosome 7 effective population size and time parameters scaled closely to those from all autosomes, s1 = 0.990964, as did the admixture proportions, s2 = 1.01159. All else equal, the hemizygosity of the X chromosome in males should result in population size and time parameters from X-chromosome site patterns that scale by 0.75, while sex biases will drive that value up or down depending on both the measure and direction of sex bias (29). We estimated s1 to be 0.92512, indicating more breeding females than breeding males or male reproductive skew. Further, s2 was 0.317951, which suggests male-biased dispersal from western to eastern chimpanzees.

Discussion

Various approaches have been used to estimate demographic parameters for *Pan* evolutionary history. Some of these estimates vary between studies, and there is no clear agreement on the degree and distribution of interbreeding in this genus (Fig. 1). Further, some gene flow is hard to reconcile with present *Pan* biogeography. Bonobos and chimpanzees can hybridize in captivity (30), but wild populations are completely separated by

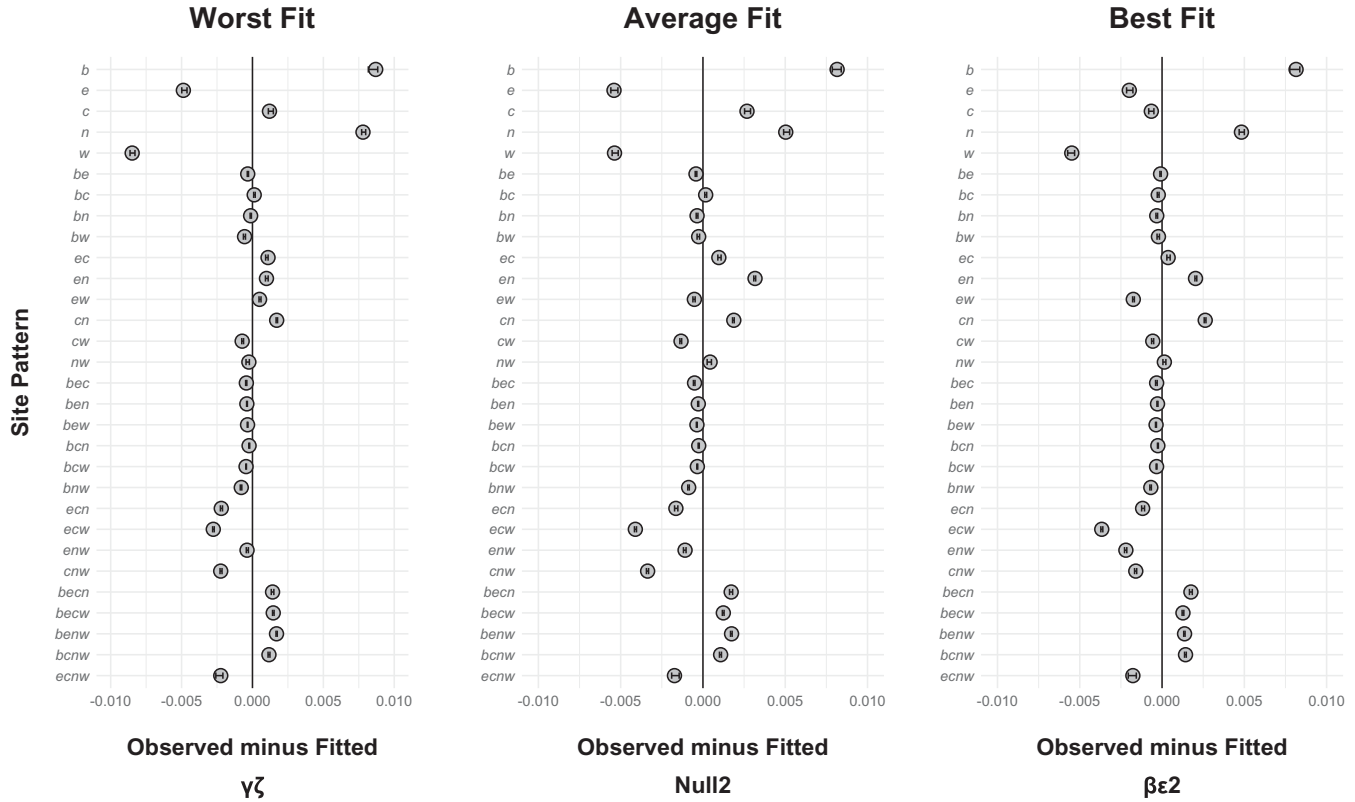

Fig. 4. Fitted autosomal model residuals. We display the worst-fit model, an average model, and the best-fit model. Points represent the point estimate, and horizontal error bars represent the 95% CIs per site pattern.

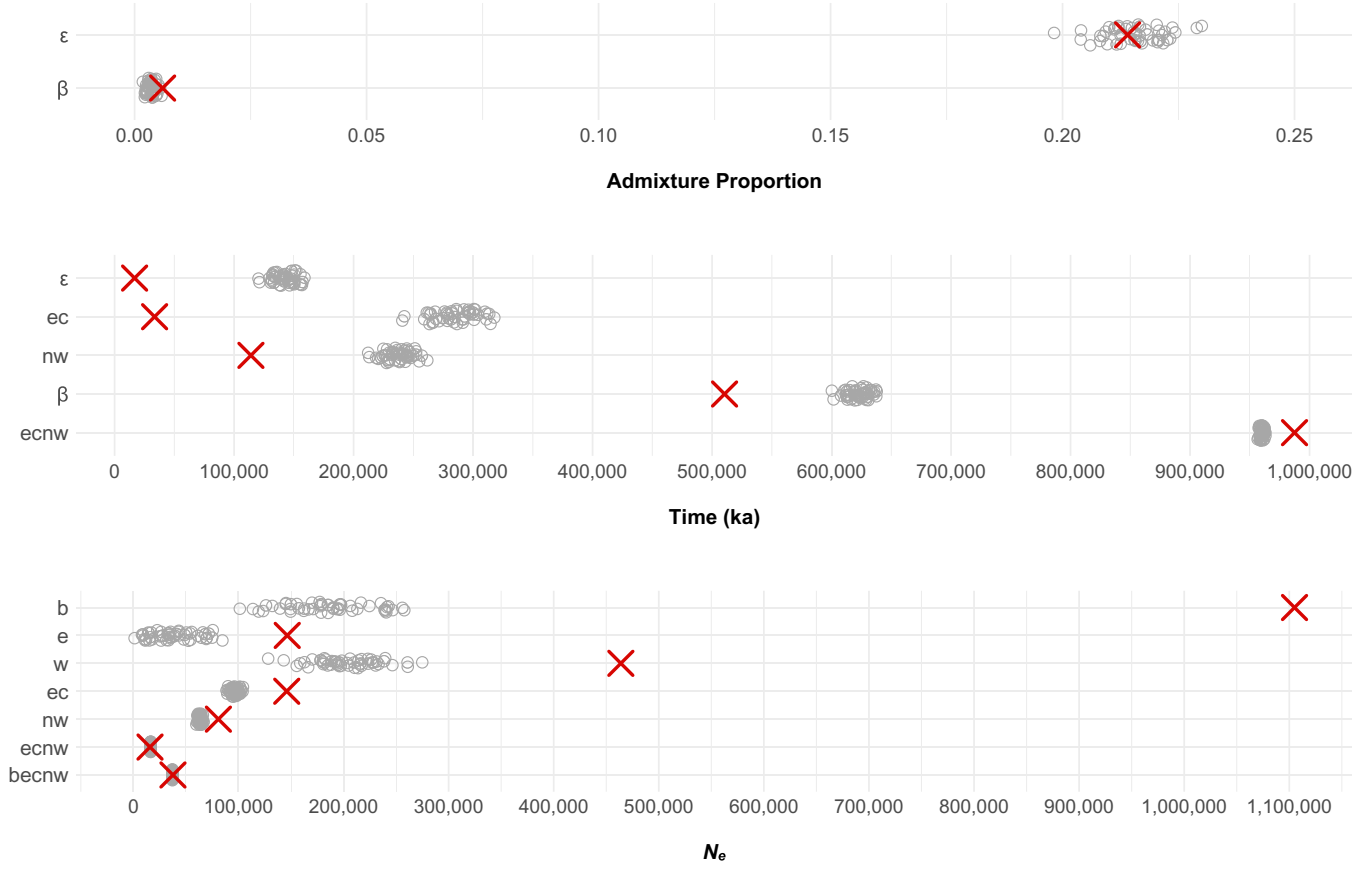

Fig. 5. Parameter estimate bias. The red crosses represent point estimates for the parameters from the βε2-model. Open gray circles represent the 50 values estimated by legofit using site patterns generated from data simulated with the βε2-model parameters using msprime. If the simulated data are less than the point estimate, the point estimate is underestimated, while if the simulated data are more than the point estimate, the point estimate is overestimated. ε indicates introgression from *P. t. verus* into *P. t. schweinfurthii*, β indicates introgression from *P. paniscus* into the ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*, ec indicates the ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*, nw indicates the ancestor of *P. t. ellioti* and *P. t. verus*, ecnw indicates the ancestor of all *P. troglodytes* lineages, and becnw indicates the ancestor of *P. paniscus* and *P. troglodytes*.

the Congo River, which may be difficult to traverse. The Congo River appears to be considerably older than previously thought, up from 1 to 2 Ma to 34 Ma (31–34). Such riverine barriers also separate three of the four chimpanzee subspecies, while western chimpanzees occur west of a large forest–savannah mosaic known as the Dahomey Gap (35). Chimpanzees have been characterized as poor swimmers (36) and as being afraid of water (37), yet some populations enter bodies of water to forage (38, 39) and thermoregulate in hot, dry habitats (40). Interestingly, bonobos are not characterized as having this aversion to water (37) and are known to routinely forage in swamps (41, 42). It is quite possible and likely that these rivers have experienced variation in discharge, which may facilitate gene flow between bonobos and chimpanzees as well as geographically proximate chimpanzee subspecies. Accurate demographic estimates, particularly those concerning gene flow, are, therefore, critical to inferring past biogeography from more than paleoenvironmental data alone.

In this study, we not only estimate demographic parameters but also, comprehensively compare previous models. We find that a model (βε2) containing two introgression events best fits *Pan* autosomal nucleotide site patterns: 1) bonobo introgression into the ancestor of eastern and central chimpanzees and 2) western chimpanzee introgression into eastern chimpanzees. The admixture proportion of this first event, bonobos into chimpanzees, was estimated to be 0.006. With a value this small, this event may not have even occurred, which would mean an even simpler evolutionary history. Despite the uncertainty surrounding this event, the other event appears to have involved substantial admixture, and both events have important implications for *Pan* biogeography.

We estimate that ∼21% of eastern chimpanzee DNA is derived from western chimpanzees. Our simulations from the best-fitting model that generated this parameter indicate that this admixture proportion is unlikely to be biased. This is also true for the tentative introgression from bonobos to the ancestor of eastern and central chimpanzees, which has been previously described (16, 20). This event would have likely been possible given variation in the discharge of the Congo River, and such contact could have happened at many points along that river and its tributaries. Indeed, sections of the Congo River near Kisangani appear to be strong candidates for such a location based on current hydrology (33). Introgression from bonobos into the ancestor of eastern and central chimpanzee may be further evidenced by putatively adaptive alleles introgressed from bonobos into chimpanzees (43).

Introgression from western into eastern chimpanzees, which presently occupy the ends of the species' geographic range, is perplexing when examined based on current biogeography. However, such an event is easily explained by differences in the current and historic range of these taxa. Variation in suitable chimpanzee habitat, including that for eastern and western chimpanzees, is well described for the past 120,000 y, particularly forest refugia (e.g., ref. 44). Contact between these lineages

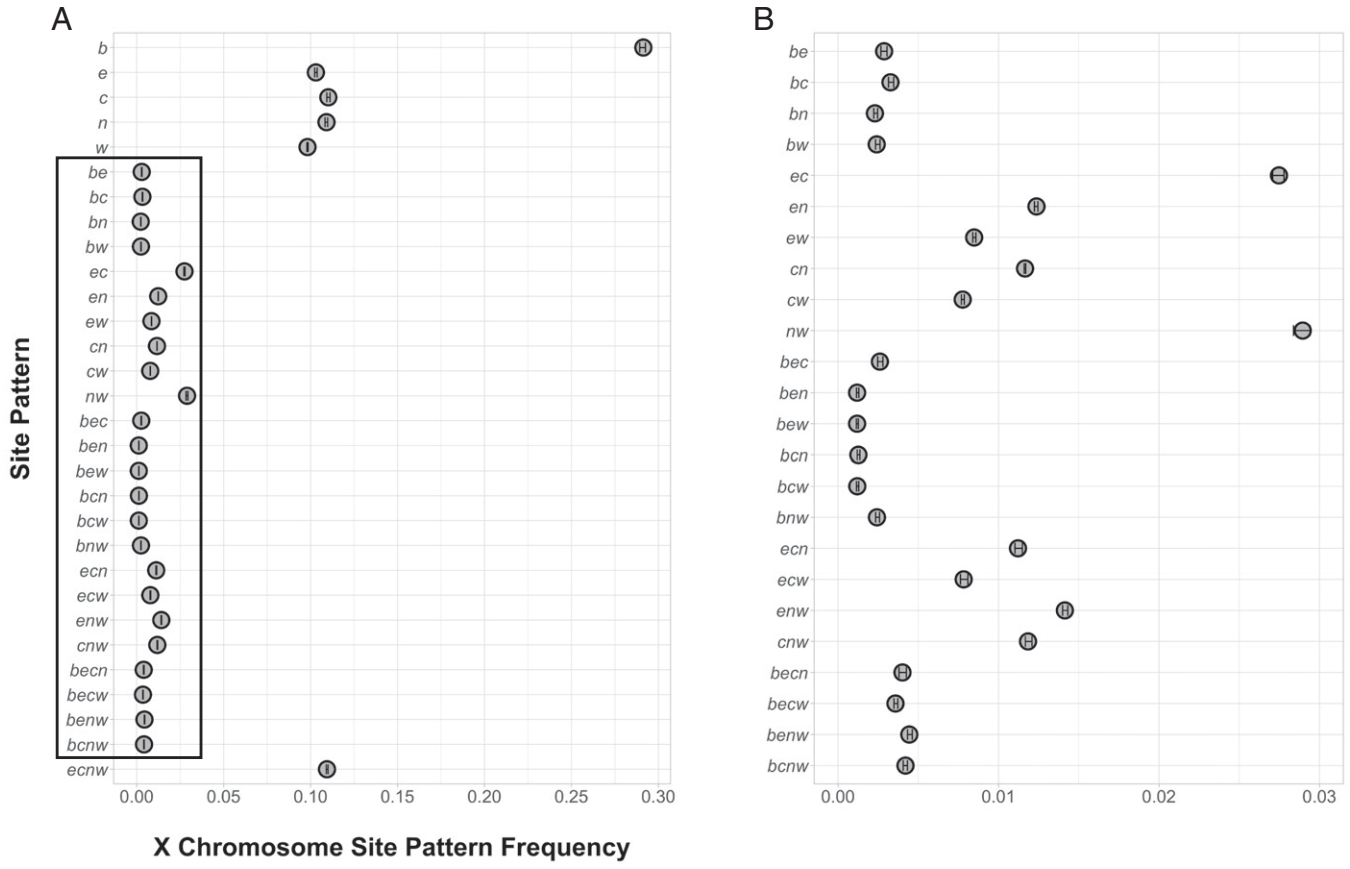

Fig. 6. Observed X-chromosome site patterns. *A* shows the overall distribution of site patterns, and *B* zooms in on the region encompassed by the black box in *A*. b indicates *P. paniscus*, e indicates *P. t. schweinfurthii*, c indicates *P. t. troglodytes*, n indicates *P. t. ellioti*, and w indicates *P. t. verus*. Points represent the point estimate, and horizontal error bars represent the 95% CIs per site pattern.

would require a connection through or north of the Dahomey Gap and may have occurred northeast of the current Nigeria–Cameroon chimpanzee range, possibly in Cameroon, the Central African Republic, Chad, or Nigeria. The exact history of the Dahomey Gap is only partially understood, but multiple primate species occur on both sides of the gap in the Upper Guinean and Congolian (or Lower Guinean) rainforests (45). Further, paleoenvironmental data suggest the Dahomey Gap has been subject to fluctuating periods of forest cover since at least 1.05 Ma (46). Some of the periods may have resulted in

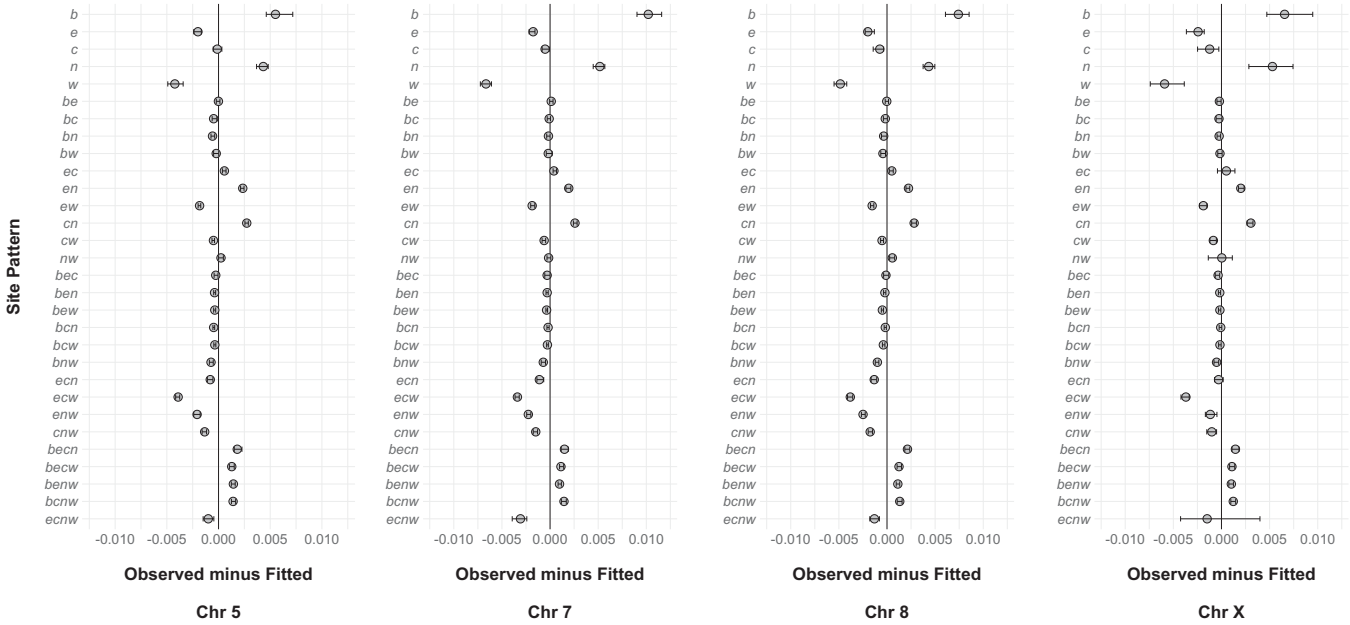

Fig. 7. Fitted individual autosomes and X-chromosome model residuals. All models were fit to the best autosomal model: βε2. Points represent the point estimate, and horizontal error bars represent the 95% CIs per site pattern.

Table 2. Model-averaged X-chromosome parameter estimates

Parameter	Point estimate	Lower bound	Upper bound
Admixture			
θ	0.1855	0.160988	0.195045
η	0.00809596	0.00386893	0.00946833
ε	0.19156471	0.17784714	0.19995339
δ	0.0195319	0.0160513	0.0207716
β	0.01404684	0.01148188	0.01557468
α	0.0751312	0.0672275	0.0779001
Population size			
b	949,621.275	686,457.745	1,857,385.88
e	207,168.235	104,257.035	551,706.039
c	136,3985	824,800	1,821,115
w	656,344.314	519,809.059	718,870.588
ec	67,434.7039	65,122.2147	70,568.3647
nw	64,631.3951	61,803.4206	70,613.1628
ecnw	17,200.048	16,921.348	17,733.0431
becnw	21,536.8834	20,948.4798	22,202.7255
gbecnw	178,530	171,603.5	195,779.5
Time			
η	4,673.59916	4,057.65167	5,464.66843
ε	9,347.19841	8,115.30333	10,929.3004
δ	13,5640.25	117,190.25	146,826
θ	333,120	313,685	342,647.5
β	428,807.5	424,840.147	431,972.255
α	1,193,287.5	1,185,135	1,197,285
ec	37,311.6515	32,315.5771	41,372.7843
nw	108,982.775	73,002.2863	132,691.059
ecnw	811,812.5	802,022.598	818,832.255
gbecnw	2,037,285	1,984,597.5	2,098,690

θ is introgression from *P. t. verus* into the ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*. η is introgression from *P. paniscus* to *P. t. schweinfurthii*. ε is introgression from *P. t. verus* into *P. t. schweinfurthii*. δ is introgression from *P. paniscus* into *P. t. troglodytes*. β is introgression from *P. paniscus* into the ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*. α is introgression from an extinct *Pan* lineage into *P. paniscus*. ec is the ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*. nw is the ancestor of *P. t. ellioti* and *P. t. verus*. ecnw is the ancestor of all *P. troglodytes* lineages. becnw is the ancestor of *P. paniscus* and *P. troglodytes*. gbecnw is the ancestor of extant *Pan* lineages and extinct *Pan* lineages. Admixture is reported as the admixture proportion, population sizes are reported as the number of diploid individuals, and time is reported in years.

substantial forest expansion (47, 48), enough to potentially allow for the introgression event supported by this study. Further, it is clear that chimpanzees occurred outside their present range at least once in the deep past based on the recovery of fossil chimpanzee teeth from the Kapthurin Formation in Kenya dated to ∼500,000 y ago (49).

The timing of western chimpanzee introgression into eastern chimpanzees is unclear. Our estimates for 1) divergence between eastern and central chimpanzees and 2) divergence between Nigeria–Cameroon and western chimpanzees are much more recent than expected (17, 18, 20, 21), and our assessment of parameter bias suggests that this might even be overestimated. This would point to a very recent divergence for eastern and central chimpanzees, <30,000 y ago, implying that the introgression from western chimpanzees occurred within the past few thousand years. While possible, it seems more plausible that these lineages diverged around the times proposed by other studies, ∼100,000 and 250,000 y ago. Admixture following divergence, as evidenced by broad time parameter CIs, may lead Legofit to infer a more recent point estimate. A more tractable approach to dating the western into eastern introgression event would involve the identification of putatively introgressed loci or haplotypes and assessing their age.

Our estimate of the ancestor of all four extant chimpanzee lineages (∼987,000 y ago) appears to be robust and is consistent with expectations from simulations of the best-fitting autosomal model. However, our estimate is hundreds of thousands of years older than other estimates [e.g., 544,000 to 633,000 y ago (20)]. We note that this estimate is largely consistent across the 107 autosomal models evaluated. Further, the estimate for this parameter from X-chromosome site patterns (∼812,000 y ago) is similar.

Our estimates for population size largely support previous findings (20, 21). Following divergence, the common ancestor of all chimpanzees experienced a period of decline. This was followed by substantial increases in both the ancestor of Nigeria–Cameroon and western chimpanzees and particularly, the ancestor of eastern and central chimpanzees. The estimated N_e for each lineage suggests that each subspecies experienced a population decline after divergence with their common ancestor. However, our population size estimates for lineages at the time of introgression are puzzling. We found a large effective population size for bonobos, eastern chimpanzees, and western chimpanzees. This may represent an instance of statistical identifiability where parameters are correlated, resulting in a broader CI (23). Indeed, some of these parameters are tightly correlated with each other and the β-admixture proportion (*SI Appendix*, Fig. S1). Such high parameter values could also be explained by geographic population structure (50). Bonobo population structure has been inferred from craniodental morphology (51), malarial infection (52), and mitochondrial haplotypes (53, 54). The geographic origins for the bonobos used in this analysis are unknown, and the eastern and western chimpanzees used here also span a large geographic range that could introduce population structure into the sample (21). Population structure would result in increased effective population size estimates and warrants further study. Another potential explanation for large effective population sizes is gene flow between bonobos and an extinct sister lineage—a different lineage than the previously proposed ghost lineage. Finally, as the admixture proportion for the introgression event from bonobos into the ancestor of eastern and central chimpanzees is near zero and may actually be zero, our data likely cannot provide information on the bonobo effective population size for that time in their evolutionary history.

While our best model for the demographic history of the X chromosome was similar to that of the autosomes, there were some interesting differences. The best model (βεη2) had an additional introgression event from bonobos to eastern chimpanzees. Further, a more complex model (αβδεηθ2) was weighted to ∼0.13 and included an additional three introgression events, one of which, western chimpanzees into the ancestor of eastern and central chimpanzees, exhibited a substantial admixture proportion: ∼0.18. This suggests that the X chromosome may capture additional facets of *Pan* evolutionary history due its size, inheritance patterns, and hemizygosity in males. However, this contrasts with patterns of reduced Denisovan and Neanderthal ancestry in human X chromosomes, regions described as "introgression deserts" (7, 8). The support for ghost admixture in our X-chromosome model is particularly perplexing because the original proposal for this event found that the bonobo X chromosome was largely devoid of ghost *Pan* ancestry (22). This result warrants further investigation, but one major difference between the studies is the correction for sex-chromosome mismapping in this study. Specifically, regions of homology on the sex chromosomes lead to read mismapping and downstream technical artifacts in variant calls (55). Thus, our correction might have increased our power for

detecting this ghost admixture on the X chromosome. We also draw attention to the reduced power that is inherent to studying the X chromosome; however, we feel that statistical power is not an issue here as the results of the X chromosome closely match similarly sized autosomes.

Despite exhibiting an equal sex ratio among adult individuals, sex-biased reproduction in *Pan* is well described (56–60). Indeed, extended periods of sexual receptivity in bonobos as compared with chimpanzees have prompted the hypothesis that male competition for mating opportunities is reduced in bonobos compared with chimpanzees (61), resulting in lower male reproductive skew. However, the bonobo communities studied to date at LuiKotale and Wamba exhibit higher reproductive skew than all but one eastern chimpanzee community (62–64). This pattern in bonobos has been previously suggested from genomic data (65). While we cannot speak to differences between bonobos and chimpanzees, we found that the estimated population size and time parameters did not scale as expected compared with a similarly sized autosome, providing evidence of male reproductive skew throughout *Pan* evolutionary history. Our analyses were only able to identify a single pattern for the clade, so future studies should investigate possible lineage-specific differences. Comparison of admixture proportions between the X chromosome and chromosome 7 also suggests sex bias in the introgression events described in this study, such that western chimpanzees exhibited male-biased dispersal . This scenario is intriguing because western chimpanzee males are slightly larger than eastern chimpanzee males (66) and therefore, may be more likely to win in disputes over females. We caution that reduced estimates of admixture on the X chromosome compared with the autosomes may also be the result of purifying selection, which is expected to be more efficient on the X chromosome due to its hemizygosity in males as proposed for the absence of archaic ancestry in humans (7, 8).

There are several important considerations for this analysis. First, Legofit is unable to estimate subsequent introgression between recently diverged lineages; therefore, other introgression events may have occurred that we cannot directly model using this approach. However, if introgression occurred shortly after a lineage diverged, we would expect the CI of this time parameter to be quite large. This may be the case for the ancestors of both eastern and central chimpanzees whose lower- and upper-bound estimates span a considerable time period. Yet, this interval is small for the divergence of Nigeria–Cameroon and western chimpanzees (range: ∼20,000 y) and even smaller for the common ancestor of all chimpanzees (range: ∼5,000 y). As we set the divergence time between bonobos and chimpanzees as the fixed parameter in this analysis, we do not have a resulting CI to infer subsequent introgression as estimated by other studies. Therefore, in addition to introgression from bonobos into the ancestor of eastern and central chimpanzees and from western into eastern chimpanzees, admixture between bonobos and chimpanzees and between eastern and central chimpanzees may also have occurred after their respective splits.

As noted above, the parameters estimated from this analysis were generated by setting one fixed parameter (the bonobo–chimpanzee divergence date or Tbecnw) to set the molecular clock. The point estimate used in this analysis was the median of a range from ref. 20. Thus, if the true divergence date is different from that used here, our parameter estimates would change as well. Additional genomic data from bonobos and chimpanzees may yield more accurate estimates of this critical parameter. The ordering of events may influence parameter estimates beyond the timing of each introgression event as well as model fit, although this seems unlikely. We also did not allow for an introgression event to occur multiple times (e.g., multiple pulses of introgression between two lineages). A better approach for determining multiple events is estimating the age of introgressed regions (67–69). Differently aged haplotypes in the same lineage would point to multiple events (20), and we encourage further study of this in future research.

Nucleotide site patterns in bonobos and chimpanzees confirm multiple aspects of their evolutionary history while offering insights into others. We find support for one introgression event from western into eastern chimpanzees. However, the biogeography of this event remains difficult to explain without invoking differences in the range of these subspecies over the course of the Late Pleistocene compared with the present. Collectively, the best-fit demographic model is simpler than more recently proposed models. Finally, our results point to a deeper divergence time for common chimpanzees. Additional genomic and paleoenvironmental data would be immensely informative in deciphering the evolutionary history of our closest living relatives and may provide insight into the evolution of other taxa in this region during this time period, including humans.

Materials and Methods

Genomic Data. We used raw short read data from all five extant *Pan* lineages from the Great Ape Genome Project (21). These data come from high-coverage genomes from 13 bonobos (*Pan paniscus*), 18 central chimpanzees (*Pan troglodytes troglodytes*), 19 eastern chimpanzees (*Pan troglodytes schweinfurthii*), 10 Nigeria–Cameroon chimpanzees (*Pan troglodytes ellioti*), and 11 western chimpanzees (*Pan troglodytes verus*). For an outgroup, we also used short read data from a high-coverage human female, HG00513, collected as part of the 1000 Genomes Project (70).

Read Mapping and Variant Calling. We used genotypes generated in ref. 71. Briefly, these data were reassembled using 1) sex-specific reference genome versions for mapping generated with XYalign (55) and 2) a contamination filter during variant calling with GATK4 (72). The use of male- and female-specific versions of the reference genome improves variant calling on the X chromosome (55), a critical step for our analyses of sex bias. The contamination filter was necessary because multiple samples in this dataset suffer from contamination from other samples both within and across taxa (21). All quality control, read mapping, and variant calling steps are described in detail in ref. 71 and contained in an automated Snakemake (73) available on GitHub (https://github.com/thw17/Pan_reassembly). The repository also contains a Conda environment with all software versions and origins, most of which are available through Bioconda (74).

Variant Filtration. We excluded unlocalized scaffolds ($n = 4$), unplaced contigs ($n = 4{,}316$), and the mitochondrial genome from these analyses. We used bcftools (75) to perform further variant filtering and provide the command line inputs in parentheses. We first normalized variants by joining biallelic sites and merging indels (insertions and deletions) and single nucleotide polymorphisms (SNPs) into a single record ("norm -m +any") using the panTro6 FASTA. We also only included sites ("-v snps") that were biallelic ("-m2 -M2") and at least 5 base-pairs from an indel ("-g 5"). On a per sample basis within each site, we marked genotypes where sample read depth was less than 10 and/or genotype quality was less than 30 as uncalled ("-S . -i FMT/DP $\geq$ 10 && FMT/GT $\geq$ 30"). To ensure that missing data did not bias our results, we further excluded any sites where less than ∼80% of individuals ($n = 56$) were confidently genotyped ("AN $\geq$ 112"). We also removed any positions that were monomorphic for either the reference or the alternate allele ("AC $>$ 0 && AC $\neq$ AN"). While lack of or low coverage at a locus is problematic, loci with excessive coverage are also of concern. These sites may yield false heterozygotes that are usually the result of copy number variation or paralogous sequences (76). As our data exhibit a high degree of interindividual and interchromosomal variation in mean coverage (71), we applied Li's (76) recommendation for a maximum depth filter ($d + 4\sqrt{d}$, where d is mean depth) to the mean chromosomal coverage of the individual in our

sample (*Pan* or *Homo*) with the highest coverage and excluded any loci that exceeded this value ("filter -e FMT/DP > d + 4$\sqrt{}$d") (*SI Appendix*, Table S1). These filtrations steps yielded between 2,413,791,600 and 2,493,198,004 variants for our downstream analyses (Dataset S1). After filtration, we generated reference allele frequency (RAF) files for each population that denote the chromosome, the site, the reference allele, the alternate allele, and the frequency of the reference allele.

Autosomal Analysis. We used Legofit (13, 23, 77) to estimate demographic history in the five extant lineages of bonobos and chimpanzees. We first used the "sitepat" function, version 1.87, on autosomal data to 1) call ancestral alleles; 2) tabulate site patterns from the RAF files, including singletons; and 3) generate 50 bootstrap replicates using a moving blocks bootstrap. Ancestral alleles were called using the human genome as an outgroup. We used the allele frequencies within the sample from each population to calculate the probability that a random haploid subsample would exhibit each site pattern. Site pattern labels reflect the samples that exhibit the derived allele: b = bonobo, e = eastern chimpanzee, c = central chimpanzee, n = Nigeria–Cameroon chimpanzee, and w = western chimpanzee. For example, the site pattern cn refers to the case in which the derived allele is present in random nucleotides drawn from c and n but is absent from those drawn from the other populations (b, e, and w).

After visualizing the frequency of the observed site patterns (Fig. 2) and examining those for 1) eastern and central chimpanzees (site pattern ec) and 2) Nigeria–Cameroon and western chimpanzees (site pattern nw), we decided to construct two sets of demographic models. In one, the divergence between eastern and central chimpanzees was older than the divergence between Nigeria–Cameroon and western chimpanzees, as possibly suggested by the site patterns. The other set of models considered the reverse, and these models are noted by ending in two. Next, we constructed various demographic models based on previously proposed introgression events.

Legofit cannot easily identify introgression between sister lineages (e.g., from eastern to central chimpanzees after their divergence or vice versa), so we do not consider those events in this analysis. We prioritized events from whole-genome studies and initially considered all possible subsets of six unidirectional events: α, β, γ, δ, ε, and ζ ($n = 64$) (Fig. 3*A*). α denotes introgression from a ghost *Pan* lineage into bonobos (22). β denotes introgression from bonobos into the ancestor of eastern and central chimpanzees (16, 20). γ denotes introgression from Nigeria–Cameroon into eastern chimpanzees (21). δ denotes introgression from bonobos into central chimpanzees (16, 20). ε denotes introgression from western into eastern chimpanzees (17, 18, 21). ζ denotes introgression from western into central chimpanzees (15–19). γ consistently resulted in poorly fit models as did introgression from Nigeria–Cameroon chimpanzees into the ancestor of eastern and central chimpanzees. Therefore, we excluded this event from the second set of models ($n = 32$) in which the eastern and central chimpanzees diverged from each other after Nigeria–Cameroon and western chimpanzees diverged from each other (Fig. 3*B*). We also considered whether adding an introgression event from bonobos into eastern chimpanzees (η) and from western chimpanzees into the ancestor of eastern and central chimpanzees (θ) individually and together improved model fit for the five best-fitting models (see below). We only considered θ for models that allowed for such an event (i.e., models from Fig. 3*B*).

Legofit requires at least one "fixed" parameter to set the molecular clock, so we chose to set the divergence time between bonobos and chimpanzee to the median value as estimated from de Manuel et al. (20). This value (1.88 Ma) was input in generation units (75,200) based on a generation time of 25 y (78). While each of the remaining nodes was initially set with the median estimate from de Manuel et al. (20) (ref. 22 has models that included α), we designated these parameters to be "free," which prompts Legofit to generate parameter estimates. We also estimated population size by setting these parameters to be free. We used initial values that ranged from 20,000 diploid individuals for the oldest event to 5,000 individuals for the most recent events. Introgression events were set to be "constrained" parameters, where a parameter is a function of another parameter. Designating parameters as such is useful for reducing the number of free parameters. Events were initially set to occur halfway between one or two divergence parameters or another introgression event. We ordered the timing of the introgression events such that models with multiple introgression events were ordered from oldest to youngest by their Greek letter designation, except for models that included θ, which we placed after β (Fig. 3). The order of the more recent chimpanzee introgression events (γ through η) is arbitrary and is not based on other results given the discordant findings of previous studies. However, given that these events are more recent and only impact one lineage potentially twice (eastern chimpanzees), we reasoned that event order would not robustly impact model fit. Indeed, this was observed for early analyses (79). We did not consider multiple occurrences of the same introgression event. Initial effective population sizes were set to decrease through time such that population sizes decreased upon each divergence. Initial admixture proportions were set to 0.01 for all introgression events (20), except for ghost admixture into bonobos, which was initially set at the median value (0.027) from Kuhlwilm et al. (22).

Legofit can be run using one of two algorithms: deterministic and stochastic (77). We employed the deterministic algorithm in all models as it is faster and more precise for all but the most complex models (77). We ran the "legofit" function in Legofit, version 2.3.2-3-gd31699a, per the demographic model on our real data and each of the 50 bootstrap replicates. Legofit estimates parameters for each model by maximizing the composite likelihood via the legofit function. Full likelihood is not maximized because information on linkage disequilibrium is not considered. Legofit employs differential evolution (DE) to maximize composite likelihood. We conducted this in several stages following Rogers et al. (13). In stage 1, points in the DE swarm were scattered widely across parameter space. As some legofit jobs may converge on different local maxima of the composite likelihood surface, each of the legofit jobs wrote its own swarm of points to a state file. In stage 2, each legofit job initialized its DE swarm by reading all of the state files produced in stage 1, enabling legofit to choose among local optima discovered in stage 1. At this point, we used the "pclgo" function to reexpress free variables as principal components. Some free parameters may be tightly correlated, and this can result in broader CIs because there are fewer dimensions than parameters. This issue can be addressed by reducing the dimension of the parameter space. Our early analyses used a value of 0.001 ("–tol 0.001") such that principal components were only retained if they explained >0.001% of the variance. However, as the exclusion of dimensions may introduce bias, we retained the full dimension. Reexpression of dimensions as principal components can also improve model fit because it allows legofit to operate on uncorrelated dimensions (77). This step produces a new model file (.lgo file). We then repeated stages 1 and 2 as stages 3 and 4 using the new .lgo file.

Models were compared by calculating the bepe for each model using the "bepe" function in Legofit (23). We also determined whether the top model was superior to all others by using the Legofit program "booma." Briefly, booma calculates weights based on whether each model has the lowest bepe value for the real data and each of the bootstrap replicates (23).

We tested for potential bias in the parameter estimates by generating simulations using msprime (80) and fitted those simulated data to the model that best fit the observed site patterns. We used parameter point estimates from that model, the previous fixed time parameter (75,200 generations or 1.88 Ma for the divergence between bonobos and chimpanzees), and median effective population sizes from Prado-Martinez et al. (21) for lineages where we did not have an estimate for N_e from our model. We simulated 1×10^4 chromosomes, each 2×10^6 basepairs in length, and used a mutation rate of 1.5×10^{-8} (81) and a recombination rate of 1.2×10^{-8} (82). This was repeated to generate a total of 50 simulated datasets to which we fit the model using all four stages of the deterministic approach described above. We then visually compared the model's point estimates with these simulated bootstraps to assess parameter bias.

X-Chromosome Analysis. We used the same methods as described above to generate site patterns for the X chromosome using the sitepat function, version 2.3.2-3-gd31699a. We used the same methods and criteria to filter variants. We further excluded any variants from the first pseudoautosomal region (PAR1), the first 2.7 Mb of the X chromosome, because it differs from the rest of the X chromosome in recombination rate, mutation rate, and effective size (83). Because PAR2 and the X-transposed region (XTR) are regions of homology between the X and Y in humans (84, 85), we used a human female sample with no Y chromosome as our outgroup to prevent potential biases causes by this homology (55). We also generated site patterns for three autosomes of comparable length to the X chromosome: chromosomes 5, 7, and 8. Given that the X chromosome

may have a different evolutionary history than the autosomes, we fit each of the previous autosomal models ($n = 96$) to the X-chromosome site patterns. As with our autosomal models, we considered whether η and Θ improved model fit individually and together for the five best-fitting X-chromosome models. If the autosomes and X chromosome exhibited the same evolutionary history by sharing the best-fitting model, we could assess sex bias in *Pan* demography.

We fit the best autosomal model to chromosome 7 and the X chromosome using the same deterministic approach described above. In these analyses, all population size and time parameters were fit using a model of form b = s1a, where b is the parameter in the chromosome 7 or X-chromosome analysis, a is the fitted value of that parameter in the autosomal analysis, and s1 is a free parameter that is shared by all population size and time parameters. This ensures that all population size parameters move up and down together, but their ratios do not change. We also included a second parameter, s2, for admixture proportion parameters. We used the same fixed time parameter for bonobo and chimpanzee divergence (1.88 Ma). s1 was given an initial value of 1 for chromosome 7 and 0.75 for the X chromosome. If sex-biased processes are absent from *Pan* evolutionary history, the effective population size inferred from the X chromosome should be 0.75 that inferred from the autosomes (29). Thus, departures from 0.75 suggest that female and male effective population sizes were previously unequal. A larger number of breeding males than females should produce $s1 < 0.75$, whereas $s1 > 0.75$ indicates fewer breeding males than females. s2 was also set to range from 0 to 10 and given an initial value of 1. We expected $s2 > 1$ if a greater number of females emigrated during introgression than males, while $s2 < 1$ would suggest a greater number of emigrating males.

We compared the agreement of model ranks between the autosomes and X chromosomes by ranking each set of models by increasing bepe. We then removed models not assessed for both the autosomes and X chromosome. We ran Spearman's correlation on these ranks in R, version 3.6.3 (86), using the base R stats package.

Data Availability. The archived version of the models and code used in this analysis has been deposited in Zenodo (DOI: 10.5281/zenodo.6331044). A non-archived version has been deposited in GitHub (https://github.com/brandcm/Pan_Demography). Many figures were generated in R using ggplot2, version 3.3.3 (87). Correlations between the estimated parameters for the best-fit model were visualized in R using corrplot, version 0.90 (88). Code and input data for generating these figures are available in the repositories described above as a Jupyter notebook. This notebook also includes the model rank correlation analysis. Some figures were further edited using Inkscape, version 1.1.1 (89). Previously published data were also used for this work (20, 21, 70). The raw *Pan* data underlying this article were previously published (20, 21) and are publicly available from the Sequence Read Archive (accession nos. PRJNA189439 and SRP018689) and the European Nucleotide Archive (accession no. PRJEB15086). The human sample, Biosample ID SAME123526, is also publicly available from The International Genome Sample Resource website (https://www.internationalgenome.org/data). The pipeline used to generate the data is available in GitHub (https://github.com/thw17/Pan_reassembly).

ACKNOWLEDGMENTS. The ideas and analyses presented here benefited from helpful discussion with many individuals. We thank current and former members of the Sterner and Ting laboratories: Tanner Anderson, Savanah Bird, Diana Christie, Elisabeth Goldman, Enrique Gomez, Claire Goodfellow, Alex Hickmott, Jenneca McCarter, Sam Queeno, Noah Simons, Kirstin Sterner, Jessica Stone, Nelson Ting, and Hannah Wellman. We also thank members of the Capra laboratory for their feedback: Tony Capra, Shiron Drusinsky, Sarah Fong, Erin Gilbertson, Bian Li, Evonne McArthur, Souhrid Mukherjee, David Rinker, and Keila Velázquez-Arcelay. This work was funded by NSF Grant BCS 1945782. We are grateful for computational resources and support from the University of Utah Center for High Performance Computing and access to the University of Oregon computing cluster, Talapas.

Author affiliations: [a]Department of Anthropology, University of Oregon, Eugene, OR 97403; and [b]Department of Anthropology, University of Utah, Salt Lake City, UT 84112

1. J. Mallet, Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237 (2005).
2. E. J. Baack, L. H. Rieseberg, A genomic view of introgression and hybrid speciation. *Curr. Opin. Genet. Dev.* **17**, 513–518 (2007).
3. P. W. Hedrick, Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–4618 (2013).
4. J. A. Cahill *et al.*, Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol. Ecol.* **24**, 1205–1217 (2015).
5. E. Palkopoulou *et al.*, A comprehensive genomic history of extinct and living elephants. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2566–E2574 (2018).
6. J. D. Wall, M. F. Hammer, Archaic admixture in the human genome. *Curr. Opin. Genet. Dev.* **16**, 606–610 (2006).
7. S. Sankararaman, S. Mallick, N. Patterson, D. Reich, The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr. Biol.* **26**, 1241–1247 (2016).
8. B. Vernot *et al.*, Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016).
9. S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, J. M. Akey, Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* **173**, 53–61 (2018).
10. V. Slon *et al.*, The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* **561**, 113–116 (2018).
11. G. S. Jacobs *et al.*, Multiple deeply divergent Denisovan ancestries in Papuans. *Cell* **177**, 1010–1021 (2019).
12. F. A. Villanea, J. G. Schraiber, Multiple episodes of interbreeding between Neanderthal and modern humans. *Nat. Ecol. Evol.* **3**, 39–44 (2019).
13. A. R. Rogers, N. S. Harris, A. A. Achenbach, Neanderthal-Denisovan ancestors interbred with a distantly related hominin. *Sci. Adv.* **6**, eaay5483 (2020).
14. J. Tung, L. B. Barreiro, The contribution of admixture to primate evolution. *Curr. Opin. Genet. Dev.* **47**, 61–68 (2017).
15. J. L. Caswell *et al.*, Analysis of chimpanzee history based on genome sequence alignments. *PLoS Genet.* **4**, e1000057 (2008).
16. D. Wegmann, L. Excoffier, Bayesian inference of the demographic history of chimpanzees. *Mol. Biol. Evol.* **27**, 1425–1435 (2010).
17. C. Becquet, M. Przeworski, A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* **17**, 1505–1519 (2007).
18. J. Hey, The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* **27**, 921–933 (2010).
19. Y.-J. Won, J. Hey, Divergence population genetics of chimpanzees. *Mol. Biol. Evol.* **22**, 297–307 (2005).
20. M. de Manuel *et al.*, Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* **354**, 477–481 (2016).
21. J. Prado-Martinez *et al.*, Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013).
22. M. Kuhlwilm, S. Han, V. C. Sousa, L. Excoffier, T. Marques-Bonet, Ancient admixture from an extinct ape lineage into bonobos. *Nat. Ecol. Evol.* **3**, 957–965 (2019).
23. A. R. Rogers, Legofit: Estimating population history from genetic data. *BMC Bioinformatics* **20**, 526 (2019).
24. A. R. Rogers, R. J. Bohlender, Bias in estimators of archaic admixture. *Theor. Popul. Biol.* **100C**, 63–78 (2015).
25. M. Petr, S. Pääbo, J. Kelso, B. Vernot, Limits of long-term selection against Neandertal introgression. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1639–1644 (2019).
26. B. Efron, Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Am. Stat. Assoc.* **78**, 316–331 (1983).
27. B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, 1993).
28. S. T. Buckland, K. P. Burnham, N. H. Augustin, Model selection: An integral part of inference. *Biometrics* **53**, 603–618 (1997).
29. T. H. Webster, M. A. Wilson Sayres, Genomic signatures of sex-biased demography: Progress and prospects. *Curr. Opin. Genet. Dev.* **41**, 62–71 (2016).
30. H. Vervaecke, L. Van Elsacker, Hybrids between common chimpanzees (*Pan troglodytes*) and pygmy chimpanzees (*Pan paniscus*) in captivity. *Mammalia* **56**, 667–669 (1992).
31. L. Beadle, *The Inland Waters of Tropical Africa: An Introduction to Tropical Limnology* (Longman, 1981).
32. J. A. Myers Thompson, A model of the biogeographical journey from Proto-pan to *Pan paniscus*. *Primates* **44**, 191–197 (2003).
33. H. Takemoto, Y. Kawamoto, T. Furuichi, How did bonobos come to range south of the congo river? Reconsideration of the divergence of *Pan paniscus* from other *Pan* populations. *Evol. Anthropol.* **24**, 170–184 (2015).
34. A. D. Horn, The taxonomic status of the bonobo chimpanzee. *Am. J. Phys. Anth.* **51**, 273–281 (1979).
35. J. D. Lester *et al.*, Recent genetic connectivity and clinal variation in chimpanzees. *Commun. Biol.* **4**, 283 (2021).
36. S. Angus, Water-contact behavior of chimpanzees. *Folia Primatol. (Basel)* **14**, 51–58 (1971).
37. T. Kano, *The Last Ape: Pygmy Chimpanzee Behavior and Ecology* (Stanford University Press, 1992).
38. D. P. Watts, K. B. Potts, J. S. Lwanga, J. C. Mitani, Diet of chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo, Kibale National Park, Uganda, 1. Diet composition and diversity. *Am. J. Primatol.* **74**, 114–129 (2012).
39. D. P. Watts, K. B. Potts, J. S. Lwanga, J. C. Mitani, Diet of chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo, Kibale National Park, Uganda, 2. Temporal variation and fallback foods. *Am. J. Primatol.* **74**, 130–144 (2012).
40. J. D. Pruetz, P. Bertolani, Chimpanzee (*Pan troglodytes verus*) behavioral responses to stresses associated with living in a savanna-mosaic environment: Implications for hominin adaptations to open habitats. *Paleoanthropology* **2009**, 252–262 (2009).
41. S. Uehara, Utilization patterns of a marsh grassland within the tropical rain forest by the bonobos (*Pan paniscus*) of Yalosidi, Republic of Zaire. *Primates* **31**, 311–322 (1990).
42. G. Hohmann, S. Ortmann, T. Remer, B. Fruth, Fishing for iodine: What aquatic foraging by bonobos tells us about human evolution. *BMC Zool.* **4**, 5 (2019).
43. J. Nye *et al.*, Selection in the introgressed regions of the chimpanzee genome. *Genome Biol. Evol.* **10**, 1132–1138 (2018).
44. C. D. Barratt *et al.*, Quantitative estimates of glacial refugia for chimpanzees (*Pan troglodytes*) since the Last Interglacial (120,000 BP). *Am. J. Primatol.* **83**, e23320 (2021).
45. A. H. Harcourt, M. A. Wood, Rivers as barriers to primate distributions in Africa. *Int. J. Primatol.* **33**, 168–183 (2012).
46. L. M. Dupont, B. Donner, R. Schneider, G. Wefer, Mid-Pleistocene environmental change in tropical Africa began as early as 1.05 Ma. *Geology* **29**, 195–198 (2001).
47. L. M. Dupont, M. Weinelt, Vegetation history of the savanna corridor between the Guinean and the Congolian rain forest during the last 150,000 years. *Veg. Hist. Archaeobot.* **5**, 273–292 (1996).

48. C. S. Miller, W. D. Gosling, Quaternary forest associations in lowland tropical West Africa. *Quat. Sci. Rev.* **84**, 7–25 (2014).
49. S. McBrearty, N. G. Jablonski, First fossil chimpanzee. *Nature* **437**, 105–108 (2005).
50. M. Nei, N. Takahata, Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.* **37**, 240–244 (1993).
51. V. Pilbrow, C. Groves, Evidence for divergence in populations of bonobos (*Pan paniscus*) in the Lomami-Lualaba and Kasai-Sankuru regions based on preliminary analysis of craniodental variation. *Int. J. Primatol.* **34**, 1244–1260 (2013).
52. W. Liu *et al.*, Wild bonobos host geographically restricted malaria parasites including a putative new *Laverania* species. *Nat. Commun.* **8**, 1–14 (2017).
53. Y. Kawamoto *et al.*, Genetic structure of wild bonobo populations: Diversity of mitochondrial DNA and geographical distribution. *PLoS One* **8**, e59660 (2013).
54. J. Eriksson, G. Hohmann, C. Boesch, L. Vigilant, Rivers influence the population genetic structure of bonobos (*Pan paniscus*). *Mol. Ecol.* **13**, 3425–3435 (2004).
55. T. H. Webster *et al.*, Identifying, understanding, and correcting technical artifacts on the sex chromosomes in next-generation sequencing data. *Gigascience* **8**, giz074 (2019).
56. U. Gerloff, B. Hartung, B. Fruth, G. Hohmann, D. Tautz, Intracommunity relationships, dispersal pattern and paternity success in a wild living community of Bonobos (*Pan paniscus*) determined from DNA analysis of faecal samples. *Proc. Biol. Sci.* **266**, 1189–1195 (1999).
57. C. Boesch, G. Kohou, H. Néné, L. Vigilant, Male competition and paternity in wild chimpanzees of the Taï forest. *Am. J. Phys. Anthropol.* **130**, 103–115 (2006).
58. E. Inoue, M. Inoue-Murayama, L. Vigilant, O. Takenaka, T. Nishida, Relatedness in wild chimpanzees: Influence of paternity, male philopatry, and demographic factors. *Am. J. Phys. Anthropol.* **137**, 256–262 (2008).
59. E. E. Wroblewski *et al.*, Male dominance rank and reproductive success in chimpanzees, *Pan troglodytes schweinfurthii*. *Anim. Behav.* **77**, 873–885 (2009).
60. N. E. Newton-Fisher, M. E. Thompson, V. Reynolds, C. Boesch, L. Vigilant, Paternity and social rank in wild chimpanzees (*Pan troglodytes*) from the Budongo Forest, Uganda. *Am. J. Phys. Anthropol.* **142**, 417–428 (2010).
61. B. Hare, V. Wobber, R. Wrangham, The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression. *Anim. Behav.* **83**, 573–585 (2012).
62. M. Surbeck, K. E. Langergraber, B. Fruth, L. Vigilant, G. Hohmann, Male reproductive skew is higher in bonobos than chimpanzees. *Curr. Biol.* **27**, R640–R641 (2017).
63. S. Ishizuka *et al.*, Paternity and kin structure among neighbouring groups in wild bonobos at Wamba. *R. Soc. Open Sci.* **5**, 171006 (2018).
64. M. S. McCarthy, J. D. Lester, M. Cibot, L. Vigilant, M. R. McLennan, Atypically high reproductive skew in a small wild chimpanzee community in a human-dominated landscape. *Folia Primatol. (Basel)* **91**, 688–696 (2020).
65. K. Prüfer *et al.*, The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012).
66. R. J. Smith, W. L. Jungers, Body mass in comparative primatology. *J. Hum. Evol.* **32**, 523–559 (1997).
67. V. Plagnol, J. D. Wall, Possible ancestral structure in human populations. *PLoS Genet.* **2**, e105 (2006).
68. S. Sankararaman *et al.*, The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
69. M. J. Hubisz, A. L. Williams, A. Siepel, Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. *PLoS Genet.* **16**, e1008895 (2020).
70. A. Auton *et al.*; 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
71. C. M. Brand, F. J. White, N. Ting, T. H. Webster, Soft sweeps predominate recent positive selection in bonobos (Pan paniscus) and chimpanzees (Pan troglodytes). bioRxiv [Preprint] (2021). https://doi.org/10.1101/2020.12.14.422788 (Accessed 8 January 2021).
72. R. Poplin *et al.*, Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv [Preprint] (2018). https://doi.org/10.1101/201178 (Accessed 10 July 2020).
73. J. Köster, S. Rahmann, Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
74. B. Grüning *et al.*; Bioconda Team, Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* **15**, 475–476 (2018).
75. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
76. H. Li, Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843–2851 (2014).
77. A. R. Rogers, An efficient algorithm for estimating population history from genetic data. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.01.23.427922 (Accessed 6 August 2020).
78. K. E. Langergraber *et al.*, 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 (2012).
79. C. M. Brand, *A Genomic Investigation of Bonobo (*Pan paniscus*) and Chimpanzee (*Pan troglodytes*) Divergence* (University of Oregon, 2021).
80. J. Kelleher, A. M. Etheridge, G. McVean, Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.* **12**, e1004842 (2016).
81. S. Besenbacher, C. Hvilsom, T. Marques-Bonet, T. Mailund, M. H. Schierup, Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat. Ecol. Evol.* **3**, 286–292 (2019).
82. L. S. Stevison *et al.*; Great Ape Genome Project, The time scale of recombination rate evolution in great apes. *Mol. Biol. Evol.* **33**, 928–945 (2016).
83. D. J. Cotter, S. M. Brotman, M. A. Wilson Sayres, Genetic diversity on the human X chromosome does not support a strict pseudoautosomal boundary. *Genetics* **203**, 485–492 (2016).
84. F. J. Charchar *et al.*, Complex events in the evolution of the human pseudoautosomal region 2 (PAR2). *Genome Res.* **13**, 281–286 (2003).
85. M. T. Ross *et al.*, The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
86. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
87. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
88. T. Wei, V. Simko, R package "corrplot": Visualization of a Correlation Matrix 2017 (Version 0.84, 2021).
89. Inkscape Project, Inkscape (Version 1.1.1, 2021).