Ghost admixture in eastern gorillas

Harvinder Pawar[1], Sebastian Cuadros[1], Marc de Manuel[1], Tom van der Valk[2,3], Irene Lobon[1], Marina Alvarez-Estape[1], Marc Haber[4], Olga Dolgova[5], Sojung Han[1], Qasim Ayub[6,9], Ruben Bautista[6], Joanna L. Kelley[8,9], Omar E. Cornejo[8,9], Oscar Lao[1], Aida M. Andrés[10], Katerina Guschanski[11,12,13], Benard Ssebide[14], Mike Cranfield[15], Chris Tyler-Smith[6], Yali Xue[6], Javier Prado-Martinez[1], Tomas Marques-Bonet[1,16,17,18]*, Martin Kuhlwilm[1,19,20]*

Affiliations:

[1] Institute of Evolutionary Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003 Barcelona, Spain

[2] Department of Bioinformatics and Genetics, Scilifelab, Swedish Museum of Natural History, Stockholm, Sweden

[3] Centre for Palaeogenetics, Svante Arrhenius väg 20C, SE-10691 Stockholm, Sweden

[4] Institute of Cancer and Genomic Sciences, University of Birmingham, Dubai, United Arab Emirates

[5] Integrative Genomics Lab, CIC bioGUNE - Centro de Investigación Cooperativa en Biociencias, Parque Científico Tecnológico de Bizkaia building 801A, 48160 Derio, Biscaya, Spain

[6] Wellcome Sanger Institute, Hinxton, Cambridge, UK

[7] Monash University Malaysia Genomics Facility, School of Science, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia

[8] School of Biological Sciences, Washington State University, Pullman, Washington, 99163 USA

[9] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Coastal Biology Building, 130 McAllister Way, Santa Cruz, CA 95060 USA

[10] UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London, Gower Street WC1E 6BT, UK

[11] Animal Ecology, Department of Ecology and Genetics, Uppsala University, 75236 Uppsala, Sweden

[12] Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom

[13] Science for Life Laboratory, 75237 Uppsala, Sweden

[14] Gorilla Doctors, Kampala, Uganda

[15] Gorilla Doctors, Karen C. Drayer Wildlife Health Center, One Health Institute, University of California Davis, School of Veterinary Medicine, One Shields Ave, Davis, CA 95616

[16] Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010 Barcelona, Spain

[17] CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain

[18] Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain

[19] Department of Evolutionary Anthropology, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria
[20] Human Evolution and Archaeological Sciences (HEAS), University of Vienna; Austria.

* These authors contributed equally.

Correspondence to: martin.kuhlwilm@univie.ac.at (M.K.), tomas.marques@upf.edu (T. M.-B.)

Abstract

Archaic admixture has had a significant impact on human evolution with multiple events across different clades, including from extinct hominins such as Neanderthals and Denisovans into modern humans. Within the great apes archaic admixture has been identified in chimpanzees and bonobos, but the possibility of such events has not been explored in other species. Here, we address this question using high-coverage whole genome sequences from all four extant gorilla subspecies, including six newly sequenced eastern gorillas from previously unsampled geographic regions. Using Approximate Bayesian Computation (ABC) with neural networks to model the demographic history of gorillas, we find a signature of admixture from an archaic 'ghost' lineage into the common ancestor of eastern gorillas, but not western gorillas. We infer that up to 3% of the genome of these individuals is introgressed from an archaic lineage that diverged more than 3 million years ago from the common ancestor of all extant gorillas. This introgression event took place before the split of mountain and eastern lowland gorillas, likely more than 40 thousand years ago, and may have influenced perception of bitter taste in eastern gorillas. When comparing the introgression landscapes of gorillas, humans and bonobos, we find a consistent depletion of introgressed fragments on the X chromosome across these species. However, depletion in protein-coding content is not detectable in eastern gorillas, possibly as a consequence of stronger genetic drift in this species.

Introduction

Gorillas are a member of the great apes, and form a sister clade to *Homo* (human) and *Pan* (chimpanzees and bonobos). Extant gorillas consist of four recognised subspecies, which cluster into two species, a western species of western lowland (*Gorilla gorilla gorilla*) and cross river (*Gorilla gorilla diehli)* gorillas and an eastern species of eastern lowland (*Gorilla beringei graueri*) and mountain gorillas (*Gorilla beringei beringei)* (Grubb et al. 2003). All gorilla subspecies are either endangered or critically endangered under IUCN criteria (Gray et al. 2013; Plumptre et al. 2016; Maisels et al. 2016).

The subspecies are distributed across western and eastern Africa in a non-continuous manner (Fig. 1A). The current geographic ranges of the different subspecies differ by size, continuity and ecology, impacting connectivity and population sizes (Fünfstück and Vigilant 2015). Western lowland gorillas are endemic to a largely continuous range of considerable size, whereas the other subspecies have much more fragmented distributions (Bergl and Vigilant 2007). Likewise, western lowland gorillas exhibit the highest genetic diversity of the

subspecies (Prado-Martinez et al. 2013; Xue et al. 2015; Fünfstück and Vigilant 2015), indicative of long-term high effective population sizes, while eastern gorilla effective population sizes are smaller (Thalmann et al. 2007). Mountain gorillas are currently isolated in two discrete areas, the Virunga Volcanoes Massif and the Bwindi Impenetrable National Park. The Bwindi National Park is located at a lower elevation than the Virunga Volcanoes, and as such exhibits warmer temperatures (Sarmiento et al. 1996; Robbins and Robbins 2018). Previous studies of the demographic history of gorillas did not incorporate information from all subspecies and were not conclusive, especially regarding the divergence time between the eastern and western clade (Kuhlwilm et al. 2016; McManus et al. 2015; Mailund et al. 2012; Becquet and Przeworski 2007; Thalmann et al. 2007). This might be due to gene flow from unsampled lineages, which is likely widespread but is often not sufficiently considered in evolutionary studies (Tricou et al. 2022; Pang and Zhang 2022). While uncovering such hidden introgression events in gorillas is not possible from ancient DNA from fossil remains, as has been performed in humans (Fontsere et al. 2019), it is possible to address such questions using genomic data from present-day individuals (Plagnol and Wall 2005; Vernot and Akey 2014; Green et al. 2010; Kuhlwilm et al. 2019).

To address this question, we use high-coverage whole genome sequences of 28 western and 21 eastern gorillas. In addition to previously published genomes (Prado-Martinez et al. 2013; Xue et al. 2015), we sequenced the genomes of five mountain gorillas from the Bwindi National Park, and one eastern lowland gorilla from the isolated population of Mount Tshiaberimu. These new genomes contribute to a more complete representation of the genomic diversity of eastern gorillas. Using this expanded dataset, representing all four known gorilla subspecies, we explored the demographic history of gorillas and specifically the hypothesis of ghost introgression, defined as gene flow from an unsampled archaic lineage. Given its significant impact in their sister taxa of *Pan* and *Homo* as well as many other species (Green et al. 2010; Kuhlwilm et al. 2019; Fontsere et al. 2019), such ghost introgression events may explain some of the uncertainties in previous demographic models for gorillas. Using an Approximate Bayesian Computation (ABC) approach, we find evidence for introgression from an extinct lineage into the common ancestor of eastern gorillas, and characterise some of the functional consequences of this introgressed genetic material.

Results

Eastern gorillas form two population clusters

We newly sequenced six eastern gorillas to high coverage (on average, 28.6X). After reprocessing the sequencing data from previous studies (Methods), we obtained a dataset of 49 individuals, with 27 western lowland gorillas, one Cross-river gorilla, 12 mountain gorillas and nine eastern lowland gorillas (Table S1). We performed a principal component analysis (PCA, Methods) to ascertain whether the newly sequenced individuals cluster with individuals from the same subspecies. The first PC separates western and eastern gorillas, as previously observed, and the second PC separates mountain gorillas from eastern lowland gorillas (Fig 1B). Since the new individual from the isolated Mount Tshiaberimu population clusters within the distribution of the other eastern lowland gorillas (Fig 1B), this individual is, as expected, considered a representative of this subspecies. The third PC reflects population stratification

within western lowland gorillas, whereas the fourth PC separates the eastern gorillas, with the two mountain gorilla populations from Virunga and Bwindi at the extremes (Fig 1C), explaining 3.2% of the variance. This is well in agreement with previous studies (Xue et al. 2015; Prado-Martinez et al. 2013).

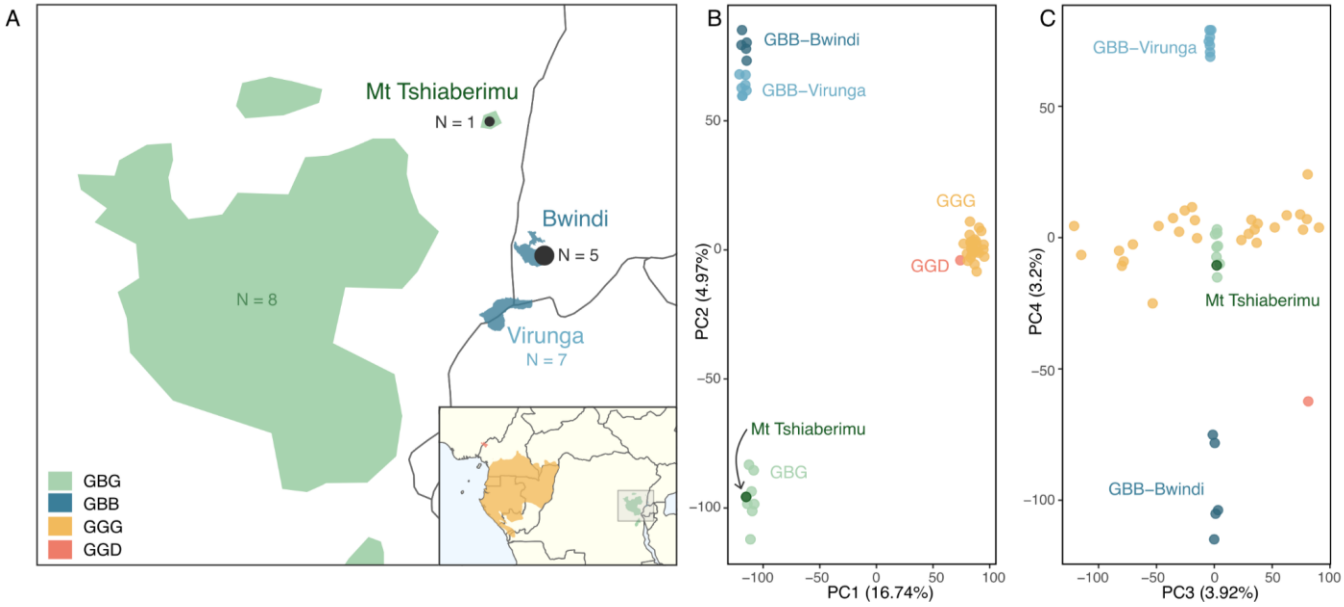

Fig 1: Gorilla samples used in this study. **A** Present geographic distribution of eastern gorillas, with that of the four gorilla subspecies shown in the inset. The newly sequenced samples are given in black, numbers of previously sequenced eastern gorillas are given in colour. GBG = *Gorilla beringei graueri* (Eastern lowland gorilla, n=9); GBB = *Gorilla beringei beringei* (Mountain gorilla, n=12); GGG = *Gorilla gorilla gorilla* (Western lowland gorilla, n=27); GGD = *Gorilla gorilla diehli* (Cross-river gorilla, n=1). **B** Principal Component Analysis (PCA) with principal components 1 and 2 shown. **C** PCA with principal components 3 and 4 shown.

Demographic modelling favours a ghost lineage in eastern gorillas

To infer a demographic model for the four extant gorilla subspecies, we used a neural-network based Approximate Bayesian Computation (ABC) modelling strategy using windowed summary statistics and extensive simulations (Methods), based on a previous implementation in the *Pan* clade (Kuhlwilm et al. 2019). A main improvement is the implementation of a broad range of informative summary statistics (Table S11), as is common practice for ABC studies in modern and ancient humans (Cooke and Nakagome 2018).

We first established a demographic null model of the four populations (Fig. S4A,S5), based on previous studies (Xue et al. 2015; van der Valk et al. 2019, 2018; Plumptre et al. 2016; Prado-Martinez et al. 2013; Tocheri et al. 2016; Roy et al. 2014). Notably, although none of these studies incorporated whole-genome data from all subspecies, our inferred parameters are largely coherent with previous work. Expanding upon this null demography (model A), we considered two more complex demographic models, in which we added the possibility of 'ghost' introgression into the common ancestor of eastern gorillas (model B) and western gorillas (model C). This allowed us to explicitly test whether the demographic null model was

improved by allowing for ghost admixture. We performed a formal comparison of these models (see Methods), to determine which fits the empirical data best. Model B, with archaic gene flow to the common eastern ancestor had the highest posterior model probability of 0.9973, compared to models A (0.0027) and C (0), and a substantially higher Bayes factor (374, vs. 0.0027 for model A and 0 for model C). In a cross-validation analysis, the model with archaic introgression into eastern gorillas was clearly distinguishable from the model without (Table S9). We conclude that a model with archaic introgression into the common eastern ancestor best explains the observed summary statistics in the empirical data.

We infer that eastern gorillas experienced bottlenecks and generally had lower effective population sizes than western gorillas, while mountain gorillas and eastern lowland gorillas experienced a particularly strong population decrease (Table S10), as described previously (Xue et al. 2015; van der Valk et al. 2019). We infer that the eastern subspecies split at 15 thousand years ago (kya) (14-16 kya, 95% credible interval (CI), Table S10). In agreement with previous studies (McManus et al. 2015), we see a population expansion in western lowland gorillas ~40 kya. Our null demographic model infers a large ancestral population size for western gorillas (Ne=98,135), in comparison to that of other gorilla populations, as well as a split between the western gorilla subspecies at ~454 kya (448-456 kya CI). Considering that not all summary statistics could be calculated for cross-river gorillas (where only one sample was available), and gene flow between western gorilla subspecies was not modelled, we caution that the confidence in this split time might be low. Finally, we infer that gorillas diverged into two species ~965 kya (729-1,104 kya CI), which is within the higher range of previous estimates (Thalmann et al. 2007; Scally et al. 2012; Kuhlwilm et al. 2016).

For simplicity, we modelled extant admixture as single migration pulses over one generation, finding a small contribution of gene flow from the common eastern ancestor to the western lowland gorillas of 0.80% (0.06% - 2.14% CI), as well as from western lowland gorillas to the common eastern ancestor of 0.27% (0.22% - 0.43% CI). We infer a contribution of 2.47% of gene flow from an archaic source into the common ancestor of eastern gorillas, with a narrow 95% CI of 2.38-2.49% (Fig 2B). We infer this ghost population diverged from the extant gorilla lineages ~3.4 million years ago (Mya) (2.98-3.8 Mya, 95% CI). We estimate the timing of this ghost gene flow to have occurred 38,281 years ago, although the CIs for this parameter are wide (22-108 kya, 95% CI) (Fig 2A,2B). By contrast, the posterior distributions for the archaic introgression proportion and the gorilla-ghost divergence time have narrow credible intervals, indicating a strong support with clear peaks for these parameters (Fig 2B). In contrast, our ABC analysis of model C does not confidently infer a contribution of a deeply divergent external lineage into the common ancestor of western gorillas. Instead, the best fit of this model suggests a 0.17% (0.09%-0.4%, 95% CI) contribution from an external lineage at ~1.1 Mya into the common ancestor of all extant gorillas (Table S10). This marginal contribution is inferred to originate from an external lineage which diverged from extant gorillas 1.9 Mya (1.5 - 3 Mya).

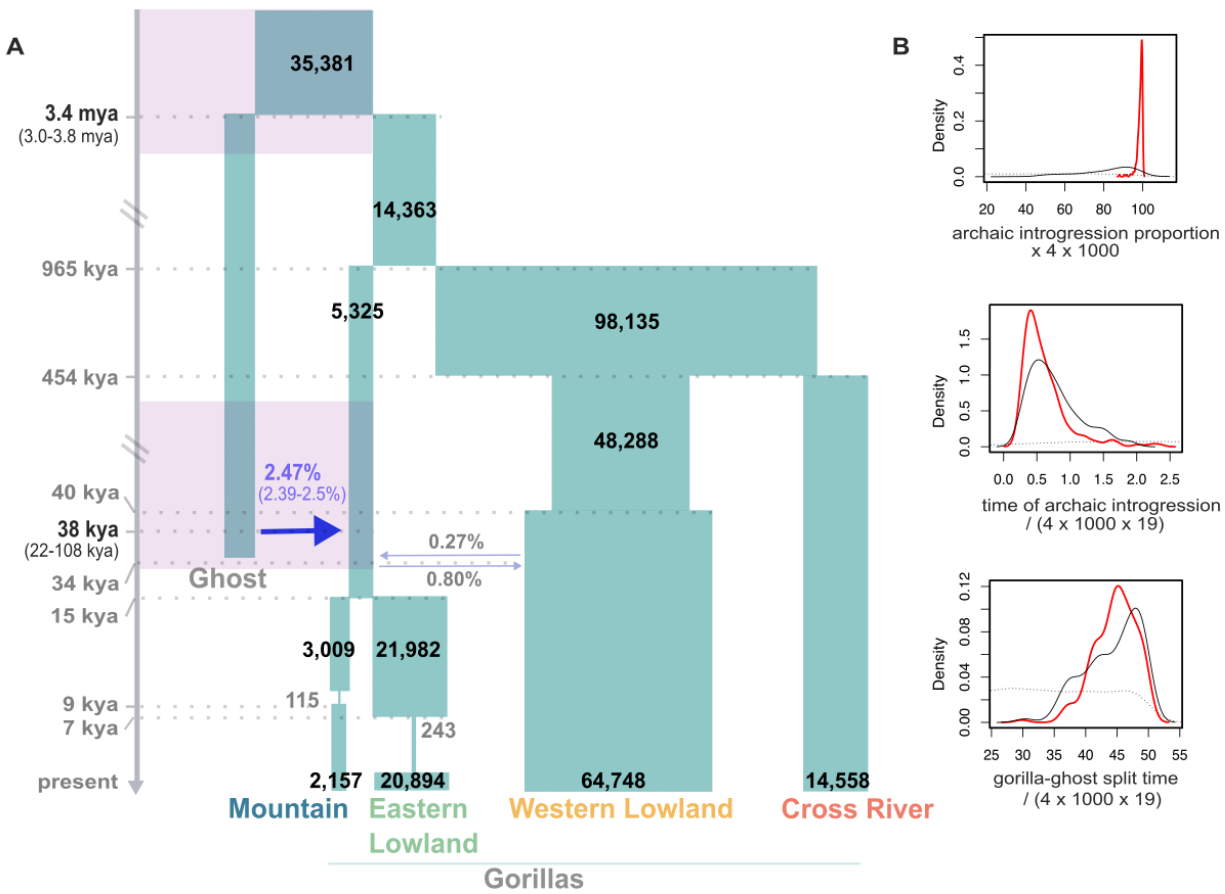

Fig 2: ABC-based demographic model. **A** Model of gorilla population history with archaic admixture from an unsampled 'ghost' lineage into the common ancestor of eastern gorillas. 95% credible intervals are shown for the archaic introgression proportion, timing of archaic introgression and archaic divergence (purple timeframes), inferred under ABC modelling. Numbers on blocks represent effective population sizes. kya = thousand years ago; mya = million years ago. **B** Posterior distributions for the archaic introgression proportion, time of archaic introgression, and gorilla-ghost split time. The dotted line indicates the prior distribution. The black line indicates the posterior inferred with a simple 'rejection' algorithm. The red line represents the posterior inferred with neural networks. Compared to the rejection algorithm, neural networks reduce the dimensionality of the summary statistics used and account for possible mismatch between the observed and simulated summary statistics (Csilléry et al. 2012).

The ghost introgression landscape in eastern gorillas

Having established that a model of ghost introgression into the common eastern ancestor provided the best fit to the empirical data, we aimed to identify the putative introgressed fragments in the genomes of eastern gorillas. To explore this landscape of ghost introgression, we implemented two complementary approaches: the S* statistic (Plagnol and Wall 2005; Vernot and Akey 2014) and the SkovHMM method, or hmmix (Skov et al. 2018). The S* statistic detects highly divergent windows relative to an outgroup, under a given demographic model, as introgressed sites (Plagnol and Wall 2005; Vernot and Akey 2014). Hence, the S*

approach depends on an accurate demographic model. By contrast, hmmix does not rely on a demographic model to identify putatively introgressed regions, but instead uses the density of private mutations in the ingroup to partition the genome into “internal” and “external” fractions, walking in small windows of 1,000 base pairs (bp) along the genomes (Skov et al. 2018).

We simulated the expected null distribution of S* scores for eastern gorillas with posterior parameter estimates from model A, *i.e.* a model without ghost introgression. This yields insights into the presence of any outlier windows in our empirical data using the 99% confidence interval for expected S* scores, given the mutation density (number of segregating sites) in each window (Fig. S9, see Methods). Indeed, at this threshold we identify an excess of S* outlier windows, suggestive of introgression from an external source into the common ancestor of eastern gorillas: Windows which fall outside the null expectation constitute, on average, 1.64% of eastern lowland genomes and 2.36% of mountain gorilla genomes, respectively (Table S12 - per individual).

Analogous to previous work (Kuhlwilm et al. 2019), we also employed hmmix to detect introgressed windows (Skov et al. 2018). Considering the strong support for ghost admixture into eastern gorillas, we again used western lowland gorillas as the outgroup and eastern gorillas as the ingroup. We find that 1.48-2.97% of the individual eastern gorilla genomes are inferred as external at a strict threshold for the mean probability of 0.95, with an estimated introgression time of 37-41 kya.

While we observe sharing of the putative introgressed regions across the eastern species, sharing is higher within each subspecies, which again is more pronounced in the mountain than in the eastern lowland gorillas (Fig 3A). This indicates that the majority of the putative introgressed regions are segregating rather than fixed. Pairwise nucleotide differences are elevated between eastern and western gorillas in putative introgressed regions in eastern gorillas, compared to random regions (Fig 3B). Likewise there is an excess of nucleotide differences between individuals of the eastern subspecies in the putative introgressed regions, indicative of an archaic origin of these regions (Fig 3B).

The overlap of the autosomal hmmix fragments and the S* outliers within each individual is, on average, 42% for eastern lowland gorillas and 51% for mountain gorillas (Table S13). For random genomic regions passing filtering criteria, the observed overlap is, on average, 6% for eastern lowland and 8% for mountain gorillas, suggesting that both methods detect to a large degree the same regions (Fig 3C, Table S16). We thus consider the regions in the intersect of the hmmix outliers and S* outliers as our high-confidence putative introgressed regions. The overlap between the two methods increases to 59% for eastern lowland and 68% for mountain gorillas, when using more lenient cutoffs for both methods, *i.e.* hmmix fragments of at least 40,000 base pairs (kbp) and 95% confidence interval S* outliers (Table S14). Mountain gorillas (with the exception of Turimaso) consistently exhibit higher proportions of overlapping base pairs of the two methods than the eastern lowland gorillas (Tables S13,S14).

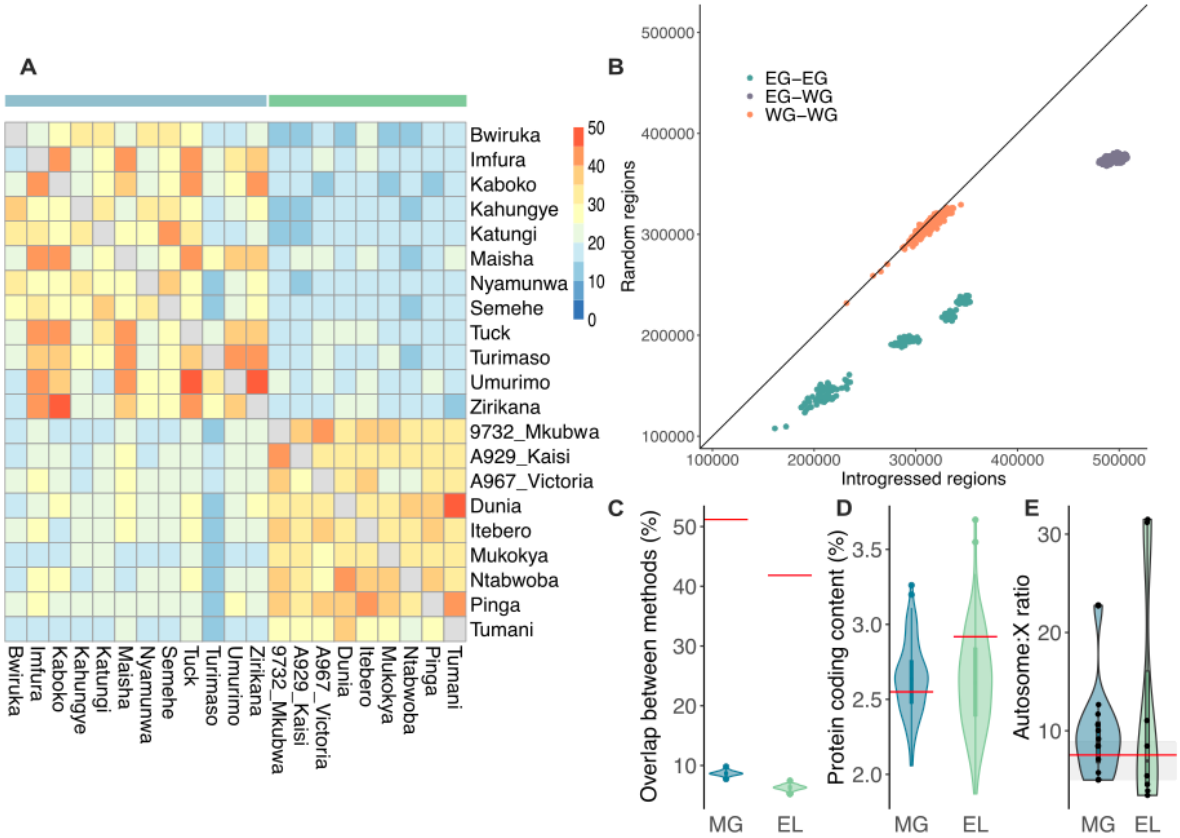

Fig 3: Characterization of introgressed fragments. **A** Sharing of putative introgressed regions across eastern gorillas for autosomal regions detected using the S* statistic and hmmix. **B** Pairwise nucleotide differences in introgressed regions (x axis) and in random regions (y axis). Colours indicate the comparison: among eastern gorillas (EG-EG, green), among western gorillas (WG-WG, orange), and between eastern and western gorillas (EG-WG, purple). **C** Percentage of overlapping base pairs in introgressed regions (red lines) and random regions (violin plots) for eastern gorillas. For details on the definition of random regions see Methods. **D** Percentage of protein coding content detected in introgressed regions (red lines) and random regions (violin plots) for eastern gorillas. **E** Autosome:X ratio of introgressed fragments inferred using hmmix for eastern gorillas (violin plots), with reference lines for the equivalent values for bonobos (red line) and humans (distribution as grey bar). In panels C-E, MG = mountain gorillas, EL = eastern lowlands.

The interaction of selection and introgression

In contrast to archaic introgressed regions identified in humans and bonobos, the putative introgressed regions in eastern gorillas are not significantly depleted in genic content compared to random genomic regions (Fig 3D). However, we find 127 Mb of autosomal segments longer than 5Mb that are depleted for introgressed fragments (Fig 4). Further, we observe a signal of depletion in archaic fragments on the X chromosome (Fig 3E), on a scale comparable to observations in modern humans (Sankararaman et al. 2014) and bonobos (Kuhlwilm et al. 2019).

Introgressed fragments can carry beneficial alleles and to explore signatures of adaptive introgression within eastern gorillas, we applied the method VolcanoFinder (Setter et al. 2020). VolcanoFinder scans the genome for a signal of a distorted local site frequency spectrum consistent with a selective sweep surrounding an introgressed allele. Outliers of the VolcanoFinder approach (95% composite likelihood ratio) within the putative introgressed regions identified above were considered putative targets of adaptive introgression. We identify seven candidate regions for adaptive introgression (Table S17), three of which are shared between eastern lowland and mountain gorillas. The region with the highest likelihood ratio in VolcanoFinder (chr12:11090005-11324172; max. LR=246.2) contains the bitter taste receptor *TAS2R14,* within which we find several protein-coding changes (Table S18).

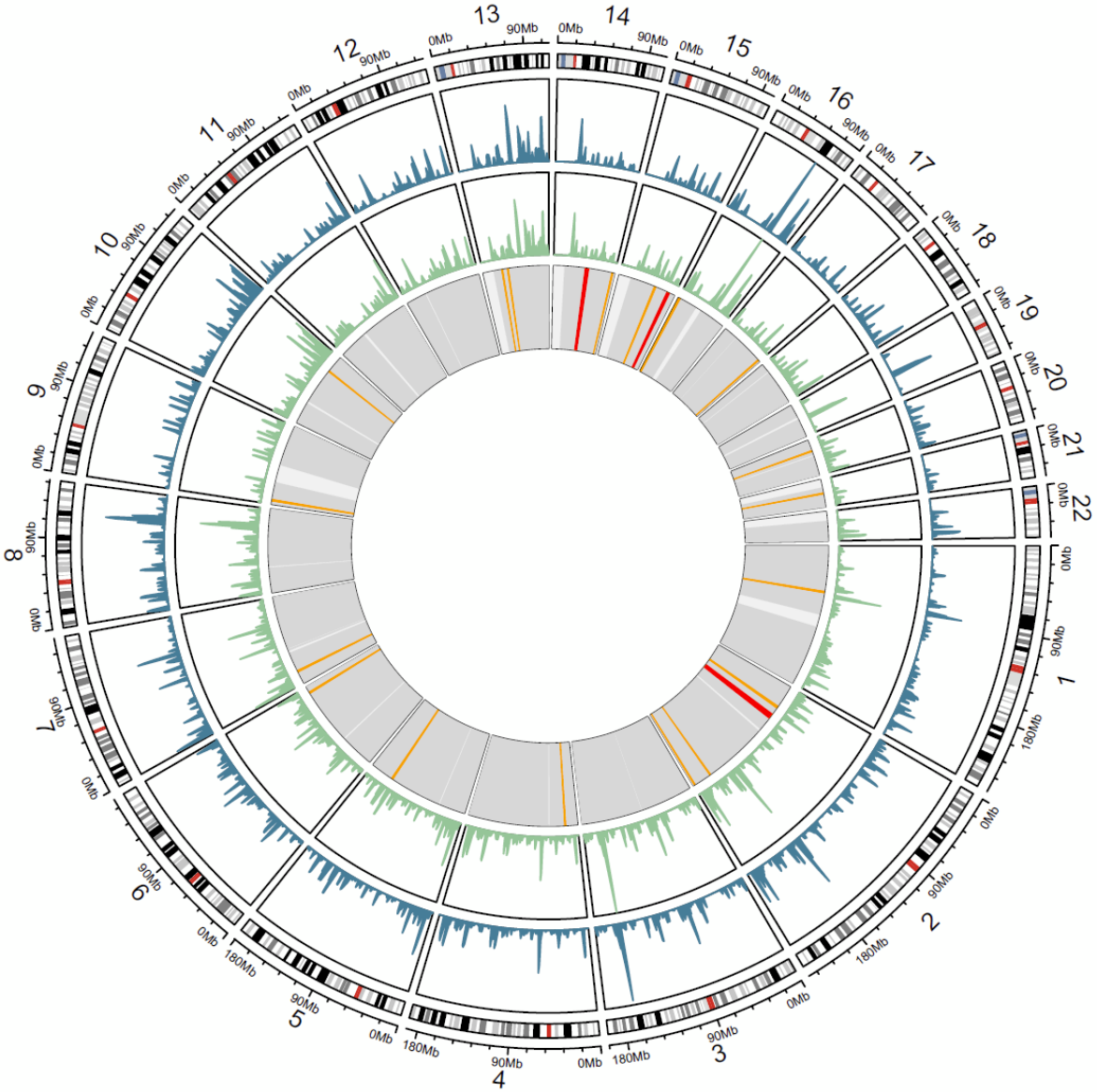

Fig 4: Distribution of introgressed fragments. Outer circle: karyogram of the autosomes based on the human genome (hg19). Second circle from outside: Introgression landscape in mountain gorillas (blue), as cumulative amount of introgressed material in sliding windows of 2 million base pairs (Mbp). Third circle from outside: Introgression landscape in eastern lowland gorillas (green) in sliding windows of 2 Mbp. Inner circle: long regions depleted of

introgression content are shown in orange (length >=5Mbp) and red (length >=8Mb). Grey: Genomic regions with sufficient data (>20% of 40 kbp windows passing threshold). White: Genomic regions without sufficient data.

Discussion

Here, we present the first demographic model inferred from representatives of all four extant gorilla subspecies, leveraging the most comprehensive dataset of gorilla genomes available to date, and a recent estimate for gorilla mutation rate (Besenbacher et al. 2019). The newly sequenced whole genomes of mountain gorillas from Bwindi National Park are genetically close to those from Virunga, but form a distinct cluster within their subspecies (Fig. 1B-C), confirming earlier results from microsatellite data (Baas et al. 2018). Eastern lowland gorillas, as represented in our dataset, seem to form a genetically less differentiated population, which includes the individual from Mount Tshiaberimu. Nonetheless, sample size remains a limitation, as high quality invasive samples are highly restricted for endangered species, given ethical and logistical constraints. A more fine-grained analysis of the evolutionary history and population structure of gorillas necessitates denser sampling, which most likely will only be possible through advances in the use of non-invasive samples. For example, a reconstruction of recent patterns of connectivity has been demonstrated from a large panel of faecal samples from chimpanzees (Fontsere et al. 2022). Furthermore, considering the rapid decline of great ape populations over the past centuries, more temporal sampling from historical specimens (van der Valk et al. 2018, 2019) has the potential to be highly informative on variation lost over time.

Previous estimates of demographic parameters varied greatly under different models, methods and input data (Mailund et al. 2012; Becquet and Przeworski 2007; Thalmann et al. 2007; McManus et al. 2015). The ABC approach presented herein leverages population-wise summary statistics. However, since only a single whole genome is currently available for cross-river gorillas, a subset of the statistics could not be obtained for this subspecies (Methods, Supplementary Table S11), and those calculated may be relatively less informative (for example, number of segregating sites). For all other populations, multiple individuals were included, yielding a better representation of their diversity in the summary statistics. As such, we have lower confidence in parameters involving cross-river gorillas, such as the relatively large divergence time inferred for the western lowland-cross-river split. This divergence time represents 47% of the inferred eastern-western species split, compared to 26% estimated in a previous study which also inferred a more recent species split time (McManus et al. 2015). We note this difference may be attributed to our inclusion of a greater number of western lowlands, known to have high levels of population structure (Thalmann et al. 2007; Anthony et al. 2007; Das et al. 2014). We also do not include gene flow between western lowlands and cross-river gorillas as a parameter in our modelling, which would reduce divergence estimates.

The inferred deep divergence time between the two species is at the upper end of previous estimates, and conservative for the detection of putatively introgressed windows under the null model, since larger S* scores would be expected to result from an increased number of segregating sites (Kuhlwilm et al. 2019). Indeed, even approximate demographic models with large divergence times may allow a detection of external gene flow into a target population

(Huang et al. 2022). Demographic modelling presented here finds the best model for gorilla demography to include archaic introgression from an unsampled 'ghost' lineage into the common ancestor of eastern gorillas. This accords with a growing literature on the prevalence of introgression from extinct lineages in humans (Green et al. 2010; Durvasula and Sankararaman 2020), bonobos (Kuhlwilm et al. 2019) and other species (Zhang et al. 2019; Rocha et al. 2022), as well as theoretical predictions and simulations showing an impact of admixture from unsampled lineages that is likely common rather than exceptional (Tricou et al. 2022; Pang and Zhang 2022). Using extensive simulations, we find strong support for a model including archaic admixture into eastern gorillas, compared to a null model without such ghost admixture, or a model of such an event in western gorillas. The latter may be rather considered similar to a model of deep substructure within gorillas, given the shallower times and small amounts of external gene flow inferred.

Our inference of 2.47% of ghost introgression is associated with high confidence as the posterior distribution is well differentiated from that of the prior (Fig 2B). This estimate agrees well with the estimates of genome-wide introgression proportions per individual inferred using the S* statistic and hmmix (Table S12). We likely underestimate the timing of archaic introgression, since shorter introgressed fragments are more likely to be missed, and another potential complication might be relatively high levels of homozygosity in eastern gorillas (Xue et al. 2015), leading to increased haplotype lengths. Our definition of putative introgressed regions as the overlap of outliers inferred with both the S* and hmmix methods (Fig 3C) is conservative and on the order expected for these methods, considering their relatively high false-positive rates (Huang, Kruisz, and Kuhlwilm 2022). Nonetheless, these methods are currently the only reliable tools available for detecting introgressed fragments in comparably small datasets of non-human species, without the availability of a source genome (Huang et al. 2022).

A higher degree of sharing of putative introgressed fragments is observed among mountain gorillas than in eastern lowlands (Fig 3A). This is consistent with smaller effective population sizes of these populations, increasing the impact of drift on introgressed genetic variation (Fontsere et al. 2019). High levels of genetic drift and reduced efficacy of natural selection likely also explain the absence of a detectable depletion of genic content in introgressed regions, in contrast to observations in introgressed material of humans and bonobos. However, we do find a number of "introgression deserts", *i.e.* regions depleted of introgressed material in the population (Fig 4), possibly as a result of purifying selection (Fontsere et al. 2019) shortly after the introgression took place. Furthermore, we observe a reduction of introgression on the X chromosome, as also seen in humans and bonobos (Sankararaman et al. 2014; Kuhlwilm et al. 2019; Skov et al. 2018). This is most likely a result of strong purifying selection against introgressed variation, as seen in humans and other species (Fontsere et al. 2019; Skov et al. 2022, 2018), possibly as a result of a combination with multiple factors (Chevy et al. 2022). Biased dispersal patterns (Harcourt et al. 1976) and high reproductive skew of gorilla males (Vigilant et al. 2015) might have led to a further reduction of the male-haploid X chromosome in introgressed material. Even though the observed patterns are likely a combination of these factors, we can currently not discern their respective contributions.

We note that our definition of adaptive introgressed targets is highly conservative, as the intersection of the outliers of three different methods S*, hmmix and VolcanoFinder as putative adaptive introgressed targets. However, in being conservative we aim to minimise the impact of potential false positives, which is a known caveat of the VolcanoFinder method (Zhang et

al. 2020; Setter et al. 2020). However, at present this is the only method available to localise signatures of adaptive introgression without a source genome. Interestingly, three candidate genes contain putative functional variants segregating in eastern gorillas and fixed ancestral in western gorillas. One of these genes is *TAS2R14,* which encodes a taste receptor implicated in perception of bitter tastes (Di Pizio and Niv 2015) and contains six missense variants. Eastern gorillas typically have more herbaceous diets than the frugivorous western gorillas (Robbins and Robbins 2018), as such taste receptors are plausible targets of adaptive introgression in eastern gorillas. Bitter taste receptors have been suggested as targets of recent positive selection in western lowland gorillas as well, including a region encompassing *TAS2R14* (McManus et al. 2015). It is possible that different mutations in the same region have been under selection in the different species. This could be interpreted in terms of the essential role of taste receptors to avoid toxicity. The gene *SEMA5A* contains a missense variant and a splice region variant; this gene has been associated with neural development, with implications in autism spectrum disorder (Carulli et al. 2021). However, the functional impact of these variants in gorillas demands further work in the future. Here, we do not find a contribution of adaptive introgression to altitude adaptation, a phenomenon observed in humans and other species (Fontsere et al. 2019; Witt and Huerta-Sánchez 2019). In mountain gorillas and eastern lowland gorillas at high altitude, this adaptation is likely driven by different mechanisms, such as the oral microbiome (Moraitou et al. 2022).

In conclusion, our work contributes improved resolution to our understanding of the evolutionary history of eastern gorillas. Across individuals, we recover a putative 16.4% of the autosomal genome of an extinct lineage (Table S15), adding to a growing literature revealing unsampled, now extinct lineages via analysis of variation present in present-day individuals.

Methods

Samples and Sequencing

Six eastern gorillas were sequenced as part of this study. Five Bwindi mountain gorillas were sampled after death by the Mountain Gorilla Veterinary Project. One Mount Tshiaberimu individual was sampled under anaesthesia. Convention on the Trade in Endangered Species of Wild Fauna and Flora (CITES) permits were obtained for all samples. Sequencing was performed on the Illumina HiSeq X platform. These six samples are publicly available in the European Nucleotide Archive (ENA) under the project number: PRJEB12821. Detailed information on all samples is provided in the Supplementary Materials (Table S1).

Data processing

We integrated the newly sequenced samples alongside previously published, high-coverage genomic data (Prado-Martinez et al. 2013; Xue et al. 2015). Raw sequencing reads were mapped to the human hg19 reference genome, as described previously (de Manuel et al. 2016). Given that the hg19 reference does not belong to any of the gorilla subspecies, equal mapping bias will be exerted across all gorillas in our dataset. The final dataset derives from 49 gorillas of known subspecies: 12 mountain (*Gorilla beringei beringei),* 9 eastern

lowland *(Gorilla beringei graueri),* 1 cross river *(Gorilla gorilla diehli)* and 27 western lowland *(Gorilla gorilla gorilla)* gorillas.

Processing of data to obtain genotypes followed procedures described in (Kuhlwilm et al. 2019). We used bcftools to retain genotypes with a coverage larger than 5-fold and lower than 101-fold, a mapping quality over 20, a proportion of MQ0 reads under 10%, and an allele balance of more than 0.1 at heterozygous positions; bedtools and jvarkit (Lindenbaum 2015) to filter the data by known repeats (RepeatMasker) and mappability (35 k-mer). Following a previous study (Kuhlwilm et al. 2019), we used the rhesus macaque reference genome (Mmul10) to infer ancestral allele states at each site and generate an ancestral binary genome, as implemented in the freezing-archer repository (https://github.com/ bvernot/freezing-archer). Only positions with genotype information in all individuals after filtering were used for calculating summary statistics for the demographic model and the S* analysis. For hmmix, missing data was allowed, genotypes were filtered for known repeats and mappability, and then an individual-based filtering was applied for sequencing coverage (depth 6-100), mapping quality (20) and retained only bi-allelic single nucleotide variants (SNVs).

Demographic modelling

Null demographic model

To infer a reliable null demographic model for the four extant gorilla subspecies, we performed Approximate Bayesian Computation (ABC) modelling using the R package abc (Csilléry et al. 2012) with neural networks, following a previously described strategy (Kuhlwilm et al. 2019). Previous demographic models did not include all of the four extant gorilla subspecies (McManus et al. 2015; Xue et al. 2015). We first attempted a merging of these models (Table S2), but in simulations this proved a poor fit to the empirical data in terms of the distributions of segregating sites, one of the main determinants of S* (Fig S2).

We used ms (Hudson 2002) to simulate data, and aimed to generate 35,700 coalescent simulation replicates, of which 35,543 were successful, whereby per iteration we generated 2,500 windows of length 40,000 base pairs (bp), randomly sampling from wide uniform priors informed by the literature (Table S3) (McManus et al. 2015; Xue et al. 2015; Besenbacher et al. 2019). We sampled local mutation rates from a normal distribution with mean of 1.235e-08, recombination rate from a negative binomial distribution with mean of 9.40e-09 and gamma of 0.5 and assumed a generation time of 19 years (Besenbacher et al. 2019). We scaled the mean mutation rate to 1.976 (1.235e-08 * window size of 40,000 * 4 * Ne of 1000) with a scaled standard deviation of 0.460408 (1.976*0.233). We also scaled mean recombination rate to 1.504 (9.40e-09 * 4 * Ne of 1000 * window size of 40,000). Per window and per population, we calculated the following summary statistics: mean and standard deviations of heterozygosity, nucleotide diversity (pi) and Tajima's D, as well as the number of population-wise fixed and segregating sites, the number of fixed sites per individual and pairwise F_{ST} (Table S11). These measures constitute the input summary statistics for all ABC analyses performed in this section. Given only one diploid sample is available for *G. gorilla diehli*, we did not use standard deviations of heterozygosity, nucleotide diversity and fixed sites per individual, as well as mean nucleotide diversity for this population.

We calculated the equivalent summary statistics normalised by data coverage for the empirical data, which had been pre-filtered by repeats, mappability and sufficiently informative

windows (>50% of sites with confident genotype calls in all individuals). We also filtered by sites fixed across all gorillas relative to the human reference genome. We accepted parameter values from the prior distribution if they generated summary statistics close to those of the empirical data. This was assessed using a tolerance of 0.005, logit transformation of all parameters and 100 neural networks in the ABC analysis.

Alternative demographic models

We performed parameter inference for two further demographic models, in which we allowed gene flow from a 'ghost' lineage into the common ancestor of B) eastern gorillas (*G. beringei beringei, G. beringei graueri*) and C) western gorillas (*G. gorilla gorilla*, *G. gorilla diehli)*. For each alternate demographic model, as above, we performed ABC analysis using 35,700 simulation replicates, whereby per iteration we generated 2,500 windows of length 40kbp. We fixed parameters with narrow CIs from model A, in order to reduce the complexity of these models.

To compare the three demographic models A) null demography, B) ghost gene flow into the eastern common ancestor, C) ghost gene flow into the western common ancestor, we simulated 10,000 replicates of 250 windows of 40kbp length, fixing the parameters as the weighted median posteriors for each model. To achieve an equal simulated timeframe (number of generations) in all models under comparison, we added a non-interacting ghost population to the null demography, with a divergence time between ghost and extant gorillas equal to that inferred under Model B above. To determine if the models could be differentiated from each other we performed cross-validation with the function cv4postpr (nval=1000, tol=0.05, method="neuralnet"). We calculated the posterior probabilities of each demographic model using the function postpr (tol=0.05, method="neuralnet"). The resulting confusion matrix is shown in Table S9.

Detecting introgressed fragments

Following (Vernot and Akey 2014; Vernot et al. 2016; Kuhlwilm et al. 2019) we calculated the S* statistic using a customized version of the package freezing-archer, accommodating non-human samples. We calculated the S* statistic genome-wide in 40kbp windows, sliding every 30kbp, using the following test (i = ingroup) and reference (o = outgroup) populations: 1) GBG (*G. beringei graueri*-i, *G. gorilla gorilla*-o) and 2) GBB (*G. beringei beringei*-i, *G. gorilla gorilla*-o).

Identifying outliers for the S* statistic requires a distribution of scores for local mutation densities (represented by numbers of segregating sites in the dataset) under a null demographic scenario. We used the weighted median posteriors for each parameter value from the above ABC analysis to generate simulated data, specifying the number of segregating sites in a stepwise manner (from 15 to 800 in steps of 5). For each stepwise segregating site (158 in total), we simulated 20,000 windows of length 40kbp, to which we applied the S* statistic for each of the scenarios (GBG, GBB). From this we obtained generalised additive models (gams) per scenario for three confidence intervals (95%, 99%, 99.5%) using the R package mgcv, following the procedures described in detail in (Vernot et al., 2016, Kuhlwilm et al., 2019). From these gams, we predicted the expected S* distributions under the null model without archaic introgression. Applying the gams to the empirical data we inferred whether any windows lay outside the expectation per scenario and per confidence

interval, assessing confidence intervals of 95%, 99% and 99.5%. As such, the threshold of significance is defined as the 95%, 99% or 99.5% confidence interval from the standard deviation for expected S* scores, given the mutation density (Vernot et al. 2016; Kuhlwilm et al. 2019).

In a complementary approach to the S* statistic, we applied hmmix (Skov et al. 2018). We obtained the input files for this method: weight files, local mutation rates and individual observations files using scripts provided with the repository for hmmix (https://github.com/LauritsSkov/Introgression-detection, as of 2018/08/02), as well as bcftools, bedtools, jvarkit and custom R scripts. The macaque allele (RheMac10 assembly) was used for polarisation of alleles. We then applied the method to the eastern gorillas using the following prior parameters: starting_probabilities = [0.98, 0.02], transitions = [[0.9995, 0.0005], [0.012, 0.988]], emissions = [0.05, 0.5]. We confirmed that using different parameters did not affect the results. We used a recombination rate of 9.40e-09 per generation and 19 years generation time with the median fragment length to estimate introgression time. Decoding, *i.e.* assigning internal and external states to specific genomic regions, was done with the script provided with the repository. Putative external fragments were filtered for posterior probabilities of 0.9 (lenient) or 0.95 (strict), and required to contain at least 5 private positions.

We note that only hmmix could be used to infer archaic introgressed fragments on the X chromosome, due to the lack of a gorilla demographic model for the sex chromosomes.

Exploring introgressed regions

To obtain a consensus set of putative introgressed regions, we overlapped the autosomal outlier regions inferred under the two methods within each eastern gorilla. For this overlap, we calculated the percentage of overlapping base pairs, considering in turn each S* confidence interval (95%, 99%, 99.5%), and with and without a 40kbp length cutoff for hmmix regions identified under the strict threshold. Imposing a 40kbp length cutoff retains 76.7% of the total strict hmmix regions (Tables S13,S14). We consider the intersect of the S* 99% outliers with the strict hmmix autosomal outliers, as our putative introgressed regions of high confidence. To determine whether the overlap obtained differed from random expectation we generated intersections of random regions, of equivalent distribution to the empirical data, for 100 iterations.

As a proxy for gene density we calculated the proportion of protein coding base pairs within these regions of high confidence. As above, we compared this to the proportion of protein coding base pairs within 100 iterations of random genomic regions, of equal length distribution as the putative introgressed regions within each eastern gorilla. We calculated pairwise nucleotide differences between individuals in the putative introgressed regions and in random genomic regions of equal length distribution and sufficient callable sites. This was conducted for three comparisons 1) among eastern gorillas 2) among western gorillas 3) between eastern and western gorillas.

We further explored our putative introgressed regions of high confidence using PCA (Fig S10). This was generated using the biallelic sites in our putative introgressed regions. For comparison, we also generated PCAs of one random set of random regions, with equal length distribution of random regions as the putative introgressed regions per eastern gorilla. The PCAs in Fig 1 were generated using biallelic SNPs of random genomic regions of equivalent

length distribution to the putative introgressed regions of GBB Bwiruka. This sample of random genomic regions is representative of the whole genome. All PCAs were generated with the R package adegenet (Jombart and Ahmed 2011). We generated phylogenetic trees of our putative introgressed regions and one random replicate (Fig S11), using the 'K80' model of nucleotide substitution, using the adegenet package (Jombart 2008). Haplotype networks were drawn using pegas (Paradis 2010).

We localised introgression deserts by screening 1Mb non-overlapping windows ('bins') spanning the genome. We filtered out bins overlapping centromeres and those at the end of each chromosome which were < 1Mb in size. Per bin we calculated the frequency of putative introgressed regions falling within the bin, for each eastern gorilla. We also calculated data coverage of the bins and filtered by mean callable proportion > 0.5. Deserts hence constitute bins where no eastern gorilla carried a putative introgressed region, and which had a reasonable number of callable sites.

Plots were created with ggplot2 (Wickham 2016), circlize (Gu et al. 2014) and pheatmap (Kolde et al. 2012). Genomic ranges were analysed with the GenomicRanges package (Lawrence et al. 2013).

Adaptive introgression

To explore signatures of adaptive introgression within eastern gorillas, we applied the genome-wide scan VolcanoFinder (Setter et al. 2020). To do so, we polarised the data to two outgroups. First, we polarised the human reference allele using the rhesus macaque allele, and subsequently polarised the gorilla genotypes by this polarised allele representing the ancestral state. To obtain the allele frequency input files per chromosome, we then filtered our data to only eastern gorilla genotypes at biallelic sites, and also filtered out sites with multiple ancestral alleles (where polarisation would be uncertain) and sites of reference homozygotes. The second input file required is an empirical unnormalised site frequency spectrum (SFS), which we generated by obtaining the unfolded SFS, normalising so all site categories sum to 1 and then filtering out the first category (the 0 entry). We called VolcanoFinder specifying "-big 1000, D = -1, P = 1, Model = 1". For computational efficiency, we performed the VolcanoFinder scan in blocks, whereby each chromosome was split into blocks of approximately equal numbers of base pairs. We placed a test site every 1,000 bp (-big 1000). We set D to -1, so VolcanoFinder iteratively tested a grid of values for genetic distance internally and selected the value that maximises the likelihood ratio (Setter et al. 2020). We set P to 1 as our input data was polarised. We used Model = 1, following procedures applied to human data (Setter et al. 2020), as well as non-human species (Moest et al. 2020; Liu et al. 2022).

We took the 95% outliers of composite likelihood ratio (CLR) scores calculated from VolcanoFinder and intersected these regions with our putative introgressed regions (identified above), to obtain putative adaptive introgressed targets. To explore potential functional consequences, we assessed which genes and which mutations fall within the putative adaptive introgressed targets, using the Variant Effect Predictor annotation (Version 83) (McLaren et al. 2016).

Acknowledgments

We thank Xin Huang and Aigerim Rymbekova for helpful discussions, and Derek Setter for valuable guidance in applying VolcanoFinder. We thank the Uganda Wildlife Authority for the Gorilla monitoring and research permission. This project has been funded by the Vienna Science and Technology Fund (WWTF) and the City of Vienna through project VRG20-001 to M.K.; the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 864203), BFU2017-86471-P (MINECO/FEDER, UE), "Unidad de Excelencia María de Maeztu", funded by the AEI (CEX 2018-000792-M), Howard Hughes International Early Career, NIH 1R01HG010898-01A1, Secretaria d'Universitats i Recerca and CERCA Program del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880) to T.M.-B.; H.P. was supported by a Formació de Personal Investigador fellowship from Generalitat de Catalunya (FI_B100131); M.A.-E. was supported by an FPI (Formación de Personal Investigador) PRE2018-083966 from Ministerio de Ciencia, Universidades e Investigación; C.T.-S. and Y.X. were funded by the Wellcome grant 098051; K.G. was supported by the Swedish Research Council grant (2020-03398); María de Maeztu Mobility Fellowship to J.L.K; O.D. was supported by a John Templeton Foundation grant (ID: 62178); A.A. received funding from UCL's Wellcome Trust ISSF3 award (204841/Z/16/Z); Q.A. is supported by strategic funding from Monash University (STG-000114).

References

Anthony, Nicola M., Mireille Johnson-Bawe, Kathryn Jeffery, Stephen L. Clifford, Kate A. Abernethy, Caroline E. Tutin, Sally A. Lahm, et al. 2007. "The Role of Pleistocene Refugia and Rivers in Shaping Gorilla Genetic Diversity in Central Africa." *Proceedings of the National Academy of Sciences of the United States of America* 104 (51): 20432–36.

Baas, Pauline, Tom van der Valk, Linda Vigilant, Urbain Ngobobo, Escobar Binyinyi, Radar Nishuli, Damien Caillaud, and Katerina Guschanski. 2018. "Population-Level Assessment of Genetic Diversity and Habitat Fragmentation in Critically Endangered Grauer's Gorillas." *American Journal of Physical Anthropology* 165 (3): 565–75.

Becquet, Celine, and Molly Przeworski. 2007. "A New Approach to Estimate Parameters of Speciation Models with Application to Apes." *Genome Research* 17 (10): 1505–19.

Bergl, Richard A., and Linda Vigilant. 2007. "Genetic Analysis Reveals Population Structure and Recent Migration within the Highly Fragmented Range of the Cross River Gorilla (Gorilla Gorilla Diehli)." *Molecular Ecology* 16 (3): 501–16.

Besenbacher, Søren, Christina Hvilsom, Tomas Marques-Bonet, Thomas Mailund, and Mikkel Heide Schierup. 2019. "Direct Estimation of Mutations in Great Apes Reconciles Phylogenetic Dating." *Nature Ecology & Evolution* 3 (2): 286–92.

Carulli, Daniela, Fred de Winter, and Joost Verhaagen. 2021. "Semaphorins in Adult Nervous System Plasticity and Disease." *Frontiers in Synaptic Neuroscience* 13 (May): 672891.

Chevy, Elizabeth T., Emilia Huerta-Sánchez, and Sohini Ramachandran. 2022. "Integrating Sex-Bias into Studies of Archaic Admixture on Chromosome X." *bioRxiv*. https://doi.org/10.1101/2022.08.30.505789.

Cooke, Niall P., and Shigeki Nakagome. 2018. “Fine-Tuning of Approximate Bayesian Computation for Human Population Genomics.” *Current Opinion in Genetics & Development* 53 (December): 60–69.
Csilléry, Katalin, Olivier François, and Michael G. B. Blum. 2012. “Abc: An R Package for Approximate Bayesian Computation (ABC).” *Methods in Ecology and Evolution*. https://doi.org/10.1111/j.2041-210x.2011.00179.x.
Das, Ranajit, Scott D. Hergenrother, Iván D. Soto-Calderón, J. Larry Dew, Nicola M. Anthony, and Michael I. Jensen-Seaman. 2014. “Complete Mitochondrial Genome Sequence of the Eastern Gorilla (Gorilla Beringei) and Implications for African Ape Biogeography.” *The Journal of Heredity* 105 (6): 752–61.
Di Pizio, Antonella, and Masha Y. Niv. 2015. “Promiscuity and Selectivity of Bitter Molecules and Their Receptors.” *Bioorganic & Medicinal Chemistry* 23 (14): 4082–91.
Durvasula, Arun, and Sriram Sankararaman. 2020. “Recovering Signals of Ghost Archaic Introgression in African Populations.” *Science Advances* 6 (7): eaax5097.
Fiona Maisels (Wildlife Conservation Society), Elizabeth A. Williamson (University of Stirling), and Richard Bergl. 2016. “IUCN Red List of Threatened Species: Gorilla Gorilla.” *IUCN Red List of Threatened Species*, April. https://www.iucnredlist.org/en.
Fontsere, Claudia, Martin Kuhlwilm, Carlos Morcillo-Suarez, Marina Alvarez-Estape, Jack D. Lester, Paolo Gratton, Joshua M. Schmidt, et al. 2022. “Population Dynamics and Genetic Connectivity in Recent Chimpanzee History.” *Cell Genomics* 2 (6): None.
Fontsere, Claudia, Marc de Manuel, Tomas Marques-Bonet, and Martin Kuhlwilm. 2019. “Admixture in Mammals and How to Understand Its Functional Implications: On the Abundance of Gene Flow in Mammalian Species, Its Impact on the Genome, and Roads into a Functional Understanding.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 41 (12): e1900123.
Fünfstück, Tillmann, and Linda Vigilant. 2015. “The Geographic Distribution of Genetic Diversity within Gorillas.” *American Journal of Primatology* 77 (9): 974–85.
Gray, Maryke, Justin Roy, Linda Vigilant, Katie Fawcett, Augustin Basabose, Mike Cranfield, Prosper Uwingeli, Innocent Mburanumwe, Edwin Kagoda, and Martha M. Robbins. 2013. “Genetic Census Reveals Increased but Uneven Growth of a Critically Endangered Mountain Gorilla Population.” *Biological Conservation* 158 (February): 230–38.
Green, Richard E., Johannes Krause, Adrian W. Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, et al. 2010. “A Draft Sequence of the Neandertal Genome.” *Science* 328 (5979): 710–22.
Grubb, Peter, Thomas M. Butynski, John F. Oates, Simon K. Bearder, Todd R. Disotell, Colin P. Groves, and Thomas T. Struhsaker. 2003. “Assessment of the Diversity of African Primates.” *International Journal of Primatology* 24 (6): 1301–57.
Gu, Zuguang, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors. 2014. “Circlize Implements and Enhances Circular Visualization in R.” *Bioinformatics* 30 (19): 2811–12.
Harcourt, A. H., K. S. Stewart, and D. Fossey. 1976. “Male Emigration and Female Transfer in Wild Mountain Gorilla.” *Nature* 263 (5574): 226–27.
Huang, Xin, Patricia Kruisz, and Martin Kuhlwilm. 2022. “Sstar: A Python Package for Detecting Archaic Introgression from Population Genetic Data with S.” *Molecular Biology and Evolution*, October. https://doi.org/10.1093/molbev/msac212.
Hudson, R. R. 2002. “Generating Samples under a Wright-Fisher Neutral Model of Genetic Variation.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/18.2.337.
Jombart, Thibaut. 2008. “Adegenet: A R Package for the Multivariate Analysis of Genetic Markers.” *Bioinformatics* 24 (11): 1403–5.
Jombart, Thibaut, and Ismaïl Ahmed. 2011. “Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data.” *Bioinformatics* 27 (21): 3070–71.
Kolde, Raivo. 2012. "Pheatmap: pretty heatmaps." *R package version* 1, no. 2: 726.
Kuhlwilm, Martin, Sojung Han, Vitor C. Sousa, Laurent Excoffier, and Tomas Marques-Bonet. 2019. “Ancient Admixture from an Extinct Ape Lineage into Bonobos.” *Nature*

Ecology & Evolution 3 (6): 957–65.
Kuhlwilm, Martin, Marc de Manuel, Alexander Nater, Maja P. Greminger, Michael Krützen, and Tomas Marques-Bonet. 2016. “Evolution and Demography of the Great Apes.” *Current Opinion in Genetics & Development* 41 (December): 124–29.
Lawrence, Michael, Wolfgang Huber, Hervé Pagès, Patrick Aboyoun, Marc Carlson, Robert Gentleman, Martin T. Morgan, and Vincent J. Carey. 2013. “Software for Computing and Annotating Genomic Ranges.” *PLoS Computational Biology* 9 (8): e1003118.
Lindenbaum, Pierre. 2015. “JVarkit: Java-Based Utilities for Bioinformatics,” May. https://doi.org/10.6084/m9.figshare.1425030.v1.
Liu, Shuyu, Lei Zhang, Yupeng Sang, Qiang Lai, Xinxin Zhang, Changfu Jia, Zhiqin Long, et al. 2022. “Demographic History and Natural Selection Shape Patterns of Deleterious Mutation Load and Barriers to Introgression across Populus Genome.” *Molecular Biology and Evolution* 39 (2). https://doi.org/10.1093/molbev/msac008.
Mailund, Thomas, Anders E. Halager, Michael Westergaard, Julien Y. Dutheil, Kasper Munch, Lars N. Andersen, Gerton Lunter, et al. 2012. “A New Isolation with Migration Model along Complete Genomes Infers Very Different Divergence Processes among Closely Related Great Ape Species.” *PLoS Genetics* 8 (12): e1003125.
Manuel, Marc de, Martin Kuhlwilm, Peter Frandsen, Vitor C. Sousa, Tariq Desai, Javier Prado-Martinez, Jessica Hernandez-Rodriguez, et al. 2016. “Chimpanzee Genomic Diversity Reveals Ancient Admixture with Bonobos.” *Science* 354 (6311): 477–81.
McLaren, William, Laurent Gil, Sarah E. Hunt, Harpreet Singh Riat, Graham R. S. Ritchie, Anja Thormann, Paul Flicek, and Fiona Cunningham. 2016. “The Ensembl Variant Effect Predictor.” *Genome Biology* 17 (1): 122.
McManus, Kimberly F., Joanna L. Kelley, Shiya Song, Krishna R. Veeramah, August E. Woerner, Laurie S. Stevison, Oliver A. Ryder, et al. 2015. “Inference of Gorilla Demographic and Selective History from Whole-Genome Sequence Data.” *Molecular Biology and Evolution* 32 (3): 600–612.
Moest, Markus, Steven M. Van Belleghem, Jennifer E. James, Camilo Salazar, Simon H. Martin, Sarah L. Barker, Gilson R. P. Moreira, et al. 2020. “Selective Sweeps on Novel and Introgressed Variation Shape Mimicry Loci in a Butterfly Adaptive Radiation.” *PLoS Biology* 18 (2): e3000597.
Moraitou, Markella, Adrian Forsythe, James A. Fellows Yates, Jaelle C. Brealey, Christina Warinner, and Katerina Guschanski. 2022. “Ecology, Not Host Phylogeny, Shapes the Oral Microbiome in Closely Related Species.” *Molecular Biology and Evolution*, December. https://doi.org/10.1093/molbev/msac263.
Pang, Xiao-Xu, and Da-Yong Zhang. 2022. “Impact of Ghost Introgression on Coalescent-Based Species Tree Inference and Estimation of Divergence Time.” *Systematic Biology*, July. https://doi.org/10.1093/sysbio/syac047.
Paradis, Emmanuel. 2010. “Pegas: An R Package for Population Genetics with an Integrated-Modular Approach.” *Bioinformatics* 26 (3): 419–20.
Plagnol, Vincent, and Jeff D. Wall. 2005. “Possible Ancestral Structure in Human Populations.” *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.0020105.eor.
Plumptre, Andrew J., Stuart Nixon, Deo K. Kujirakwinja, Ghislain Vieilledent, Rob Critchlow, Elizabeth A. Williamson, Radar Nishuli, Andrew E. Kirkby, and Jefferson S. Hall. 2016. “Catastrophic Decline of World’s Largest Primate: 80% Loss of Grauer's Gorilla (Gorilla Beringei Graueri) Population Justifies Critically Endangered Status.” *PloS One* 11 (10): e0162697.
Prado-Martinez, Javier, Peter H. Sudmant, Jeffrey M. Kidd, Heng Li, Joanna L. Kelley, Belen Lorente-Galdos, Krishna R. Veeramah, et al. 2013. “Great Ape Genetic Diversity and Population History.” *Nature* 499 (7459): 471–75.
Robbins, Martha M., and Andrew M. Robbins. 2018. “Variation in the Social Organization of Gorillas: Life History and Socioecological Perspectives.” *Evolutionary Anthropology* 27 (5): 218–33.
Rocha, Joana L., Pedro Vaz Pinto, Hans R. Siegismund, Matthias Meyer, Bettine Jansen van Vuuren, Luís Veríssimo, Nuno Ferrand, and Raquel Godinho. 2022. “African

Climate and Geomorphology Drive Evolution and Ghost Introgression in Sable Antelope." *Molecular Ecology* 31 (10): 2968–84.
Roy, Justin, Maryke Gray, Tara Stoinski, Martha M. Robbins, and Linda Vigilant. 2014. "Fine-Scale Genetic Structure Analyses Suggest Further Male than Female Dispersal in Mountain Gorillas." *BMC Ecology* 14 (July): 21.
Sankararaman, Sriram, Swapan Mallick, Michael Dannemann, Kay Prüfer, Janet Kelso, Svante Pääbo, Nick Patterson, and David Reich. 2014. "The Genomic Landscape of Neanderthal Ancestry in Present-Day Humans." *Nature* 507 (7492): 354–57.
Sarmiento, Esteban E., Thomas M. Butynski, and Jan Kalina. 1996. "Gorillas of Bwindi-Impenetrable Forest and the Virunga Volcanoes: Taxonomic Implications of Morphological and Ecological Differences." *American Journal of Primatology* 40 (1): 1–21.
Scally, Aylwyn, Julien Y. Dutheil, Ladeana W. Hillier, Gregory E. Jordan, Ian Goodhead, Javier Herrero, Asger Hobolth, et al. 2012. "Insights into Hominid Evolution from the Gorilla Genome Sequence." *Nature* 483 (7388): 169–75.
Setter, Derek, Sylvain Mousset, Xiaoheng Cheng, Rasmus Nielsen, Michael DeGiorgio, and Joachim Hermisson. 2020. "VolcanoFinder: Genomic Scans for Adaptive Introgression." *PLoS Genetics* 16 (6): e1008867.
Skov, Laurits, Ruoyun Hui, Vladimir Shchur, Asger Hobolth, Aylwyn Scally, Mikkel Heide Schierup, and Richard Durbin. 2018. "Detecting Archaic Introgression Using an Unadmixed Outgroup." *PLoS Genetics* 14 (9): e1007641.
Skov, L., M. Coll Macià, E. Lucotte, M. I. A. Cavassim, D. Castellano, M. H. Schierup, and K. Munch. 2022. "Extraordinary Selection on the Human X Chromosome Associated with Archaic Admixture." *bioRxiv*. https://doi.org/10.1101/2022.09.19.508556.
Thalmann, O., A. Fischer, F. Lankester, S. Pääbo, and L. Vigilant. 2007. "The Complex Evolutionary History of Gorillas: Insights from Genomic Data." *Molecular Biology and Evolution* 24 (1): 146–58.
Tocheri, Matthew W., René Dommain, Shannon C. McFarlin, Scott E. Burnett, D. Troy Case, Caley M. Orr, Neil T. Roach, et al. 2016. "The Evolutionary Origin and Population History of the Grauer Gorilla." *American Journal of Physical Anthropology* 159 (Suppl 61): S4–18.
Tricou, Théo, Eric Tannier, and Damien M. de Vienne. 2022. "Ghost Lineages Can Invalidate or Even Reverse Findings Regarding Gene Flow." *PLoS Biology* 20 (9): e3001776.
Valk, Tom van der, David Díez-Del-Molino, Tomas Marques-Bonet, Katerina Guschanski, and Love Dalén. 2019. "Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas." *Current Biology: CB* 29 (1): 165–70.e6.
Valk, Tom van der, Edson Sandoval-Castellanos, Damien Caillaud, Urbain Ngobobo, Escobar Binyinyi, Radar Nishuli, Tara Stoinski, et al. 2018. "Significant Loss of Mitochondrial Diversity within the Last Century due to Extinction of Peripheral Populations in Eastern Gorillas." *Scientific Reports* 8 (1): 6551.
Vernot, Benjamin, and Joshua M. Akey. 2014. "Resurrecting Surviving Neandertal Lineages from Modern Human Genomes." *Science* 343 (6174): 1017–21.
Vernot, Benjamin, Serena Tucci, Janet Kelso, Joshua G. Schraiber, Aaron B. Wolf, Rachel M. Gittelman, Michael Dannemann, et al. 2016. "Excavating Neandertal and Denisovan DNA from the Genomes of Melanesian Individuals." *Science* 352 (6282): 235–39.
Vigilant, Linda, Justin Roy, Brenda J. Bradley, Colin J. Stoneking, Martha M. Robbins, and Tara S. Stoinski. 2015. "Reproductive Competition and Inbreeding Avoidance in a Primate Species with Habitual Female Dispersal." *Behavioral Ecology and Sociobiology* 69 (7): 1163–72.
Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer.
Witt, Kelsey E., and Emilia Huerta-Sánchez. 2019. "Convergent Evolution in Human and Domesticate Adaptation to High-Altitude Environments." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 374 (1777): 20180235.
Xue, Yali, Javier Prado-Martinez, Peter H. Sudmant, Vagheesh Narasimhan, Qasim Ayub, Michal Szpak, Peter Frandsen, et al. 2015. "Mountain Gorilla Genomes Reveal the

Impact of Long-Term Population Decline and Inbreeding." *Science* 348 (6231): 242–45.
Zhang, Dezhi, Linfang Tang, Yalin Cheng, Yan Hao, Ying Xiong, Gang Song, Yanhua Qu, et al. 2019. "'Ghost Introgression' as a Cause of Deep Mitochondrial Divergence in a Bird Species Complex." *Molecular Biology and Evolution*, July. https://doi.org/10.1093/molbev/msz170.
Zhang, Xinjun, Bernard Kim, Kirk E. Lohmueller, and Emilia Huerta-Sánchez. 2020. "The Impact of Recessive Deleterious Variation on Signals of Adaptive Introgression in Human Populations." *Genetics* 215 (3): 799–812.