ARTICLE

Ancient gene flow from early modern humans into Eastern Neanderthals

Martin Kuhlwilm[1]*, Ilan Gronau[2]*, Melissa J. Hubisz[3], Cesare de Filippo[1], Javier Prado-Martinez[4], Martin Kircher[1,5], Qiaomei Fu[1,6,7], Hernán A. Burbano[1,8], Carles Lalueza-Fox[4], Marco de la Rasilla[9], Antonio Rosas[10], Pavao Rudan[11], Dejana Brajkovic[12], Željko Kucan[11], Ivan Gušic[11], Tomas Marques-Bonet[4,13,14], Aida M. Andrés[1], Bence Viola[15,16], Svante Pääbo[1], Matthias Meyer[1], Adam Siepel[3,17] & Sergi Castellano[1]

It has been shown that Neanderthals contributed genetically to modern humans outside Africa 47,000–65,000 years ago. Here we analyse the genomes of a Neanderthal and a Denisovan from the Altai Mountains in Siberia together with the sequences of chromosome 21 of two Neanderthals from Spain and Croatia. We find that a population that diverged early from other modern humans in Africa contributed genetically to the ancestors of Neanderthals from the Altai Mountains roughly 100,000 years ago. By contrast, we do not detect such a genetic contribution in the Denisovan or the two European Neanderthals. We conclude that in addition to later interbreeding events, the ancestors of Neanderthals from the Altai Mountains and early modern humans met and interbred, possibly in the Near East, many thousands of years earlier than previously thought.

Based on the fossil record, Neanderthals diverged from modern humans at least 430,000 years ago[1], and the analysis of a Neanderthal genome from a cave in the Altai Mountains in Siberia suggests they diverged 550,000–765,000 years ago[2]. The analysis of a Denisovan genome from the same cave in the Altai Mountains further suggests that Neanderthals and Denisovans diverged 381,000–473,000 years ago[2]. This divergence was followed by admixture among archaic and modern human populations, including gene flow from Neanderthals into modern humans outside Africa[2–5], Denisovan gene flow into the ancestors of present-day humans in Oceania and mainland Asia[6,7], gene flow into the Denisovans from Neanderthals[2] and, possibly, gene flow into the Denisovans from an unknown archaic group that diverged from the other lineages more than one million years ago[2]. Genetic evidence of gene flow from modern humans into Neanderthals or Denisovans, however, remains elusive.

Divergence and heterozygosity in the archaic genomes

The Altai Neanderthal genome shares 5.4% more derived alleles with present-day Africans than does the Denisovan genome. This excess is particularly pronounced for derived alleles found at >0.9 frequency in Africans (Extended Data Table 1). These observations have been interpreted as evidence of gene flow from an unknown and more deeply diverged archaic hominin into the Denisovan lineage[2]. Here we examine whether gene flow from modern humans into the ancestors of the Altai Neanderthal may also have occurred.

Noting that regions in the Denisovan genome introgressed from a deeply divergent archaic hominin should have unusually high divergence to present-day Africans, and that regions of the Altai Neanderthal genome introgressed from modern humans should have unusually

low divergence to them, we examined the divergence of these archaic genomes to 504 African genomes[8] in 15,881 sequence windows of 100 kb (Supplementary Information section 9). Archaic alleles brought into Africa by Eurasians about 3,000 years ago[9,10] were excluded from these windows by using only derived alleles at >0.9 frequency in the combined African genomes. In the absence of information about the phase of the alleles in the two archaic genomes, we calculated their divergence to Africans using the archaic alleles in each window that give the minimum number of differences, to allow introgressed segments from modern humans to be more easily identified, if they exist. Noting also that introgressed regions in the Denisovan or Altai Neanderthal genome should have unusually high divergence to the other archaic genome, we calculated the divergence between the archaic genomes in the same windows by using the alleles that give the maximum number of differences.

We find that windows of the Denisovan genome with high divergence to Africans also have a high divergence to the Altai Neanderthal, whereas windows in the Altai Neanderthal genome with high divergence to Africans do not tend to have a high divergence to the Denisovan (Fig. 1a), consistent with gene flow from a deeply diverged hominin into the Denisovan ancestors. On the other hand, we find that windows of the Altai Neanderthal genome with low divergence to Africans have higher divergence to the Denisovan than Denisovan windows with low divergence to Africans (Fig. 1a). These windows in the Altai Neanderthal genome have higher heterozygosity than in the Denisovan genome (Fig. 1b), and 40.7% of their heterozygous sites share a derived allele with Africans, whereas 24.2% do so in the Denisovan. These observations raise the possibility of gene flow from modern humans into Neanderthals.

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [2]Efi Arazi School of Computer Science, Herzliya Interdisciplinary Center (IDC), Herzliya 46150, Israel. [3]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14850, USA. [4]Institute of Evolutionary Biology (UPF-CSIC), 08003 Barcelona, Spain. [5]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [6]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [7]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. [8]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany. [9]Área de Prehistoria, Departamento de Historia, Universidad de Oviedo, 33011 Oviedo, Spain. [10]Departamento de Paleobiología, Museo Nacional de Ciencias Naturales, CSIC, 28006 Madrid, Spain. [11]Anthropology Center of the Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia. [12]Croatian Academy of Sciences and Arts, Institute for Quaternary Paleontology and Geology, 10000 Zagreb, Croatia. [13]Catalan Institution of Research and Advanced Studies (ICREA), 08010 Barcelona, Spain. [14]Centro Nacional de Análisis Genómico (CRG-CNAG), 08028 Barcelona, Spain. [15]Department of Anthropology, University of Toronto, Toronto, Ontario M5S 2S2, Canada. [16]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [17]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.
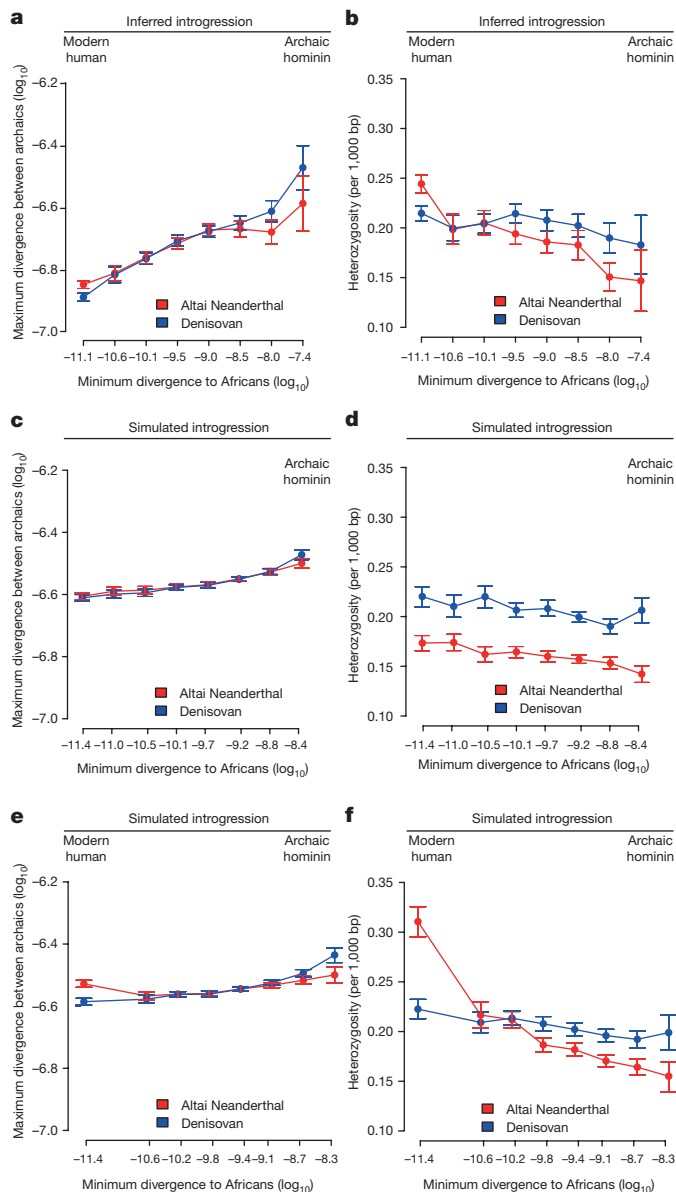*These authors contributed equally to this work.

Figure 1 | Divergence and heterozygosity in the Altai Neanderthal and Denisovan genomes. a, The maximum divergence between windows in the two archaic genomes versus their minimum divergence to Africans. Error bars represent the 95% confidence intervals from 1,000 bootstrap replicates. Regions previously described as inbred in the Altai Neanderthal genome[2] were excluded. **b**, Heterozygosity (per 1,000 bp) in windows of each archaic genome versus their minimum divergence to Africans. **c, d**, Simulation of a model with gene flow into the Denisovan lineage from both the Altai Neanderthal (0.65%, 50,000 years ago) and an unknown archaic hominin (1%, 200,000 years ago) that diverged from other hominins 1.5 million years ago. The constant mutation rate used makes the slope of the simulated curves less steep than in the actual genomes, where mutation rate varies among windows. **e, f**, Simulation of a model that also includes modern human gene flow into the Altai Neanderthal lineage (3.55%, 100,000 years ago).

Model–based inferences of gene flow

We assessed the possibility of modern human gene flow into the Altai Neanderthal lineage using the Generalized Phylogenetic Coalescent Sampler (G-PhoCS)[11], a Bayesian method for inferring divergence times, effective population sizes and rates of gene flow from genome sequences. We applied G-PhoCS in five separate analyses, each considering the Altai Neanderthal and Denisovan genomes and two present-day human genomes from an African, European

or Asian population (Supplementary Information section 8). We modelled gene flow among modern and archaic populations, including gene flow from an unknown deeply divergent archaic population, while accounting for the uncertainty in the ages of the archaic individuals.

The inferred demographic model confirms and provides quantitative estimates of previously inferred gene flow events among modern and archaic humans[2,3] (Extended Data Fig. 1). These include Neanderthal gene flow into modern humans outside Africa (3.3–5.8%) and gene flow from an unknown archaic hominin into the ancestors of Denisovans (0.0–0.5%). Interestingly, we also detect a signal of gene flow from modern humans into the ancestors of the Altai Neanderthal (1.0–7.1%). The precise source of this gene flow is unclear, but it appears to come from a population that either split from the ancestors of all present-day Africans or from one of the early African lineages, as significant admixture rates are estimated from San as well as Yoruba individuals. This introgression thus occurred in the opposite direction from the previously reported gene flow from Neanderthals to modern humans outside Africa[2,3,12].

Simulation of modern human gene flow

We used simulations to test if G-PhoCS correctly infers modern human gene flow into the Altai Neanderthal lineage (Extended Data Figs 2 and 3) and whether the patterns of divergence and heterozygosity observed in the Altai Neanderthal genome are expected from our inferred demographic model (Extended Data Fig. 1). Using parameters compatible with this model, we simulated windows of 100 kb for a model with gene flow into the Denisovan lineage from both the Altai Neanderthal and a deeply divergent archaic hominin[2], and a model including these admixture events together with modern human gene flow into the Altai Neanderthal lineage. Both models reproduced the observed patterns in windows most divergent to Africans (Fig. 1c and d), but only the model with modern human gene flow into the ancestors of the Altai Neanderthal reproduced the observed divergence and heterozygosity patterns in windows of the Altai Neanderthal least divergent to Africans (Fig. 1e and f).

Present-day human contamination among the DNA fragments from the Altai Neanderthal and Denisovan is around 1% (Table 1). After genotype calling, which is unaffected by low levels of error, these genomes should be largely free from contamination[2,7]. Even so, substituting gene flow from modern humans for present-day human contamination as high as 5% in the genotypes of the Altai Neanderthal fails to explain the observed sequence patterns (Extended Data Fig. 4).

Estimated ages of the introgressed haplotypes

The majority of haplotypes shared between present-day humans and an archaic genome should result from incomplete lineage sorting in the population ancestral to them and, thus, be old and short. However, if modern human introgression into the Altai Neanderthal lineage occurred after its separation from the Denisovan lineage we would expect a fraction of these shared haplotypes to be younger and longer in the Altai Neanderthal than in the Denisovan genome.

We examined these shared haplotypes making use of ARGweaver[13], a new computational method for sampling full genealogies and corresponding recombination events (ancestral recombination graphs) consistent with a collection of genome sequences (Supplementary Information section 10). We applied this method to six African genomes from three different populations (San, Mbuti, and Yoruba) and the two archaic genomes, and estimated the ages of haplotypes for which one archaic genome coalesces within the subtree of the African genomes more recently than it coalesces with the other archaic genome (Fig. 2a, inset). When we compare the age distribution of such 'African' haplotypes (≥50 kb), we find that the Altai Neanderthal genome has more young 'African' haplotypes (Fig. 2a, left) than the Denisovan genome (*P* < 0.01; fraction of MCMC replicates). The majority of

Table 1 | The archaic individuals analysed in this work

	Altai Neanderthal[2]	Denisovan[7]	El Sidrón Neanderthal	Vindija Neanderthal
Age (years old)	>50,000	>50,000	~49,000	~44,000
mtDNA contamination (%)	0.78	0.35	0.40	1.08
Nuclear contamination (%)	0.80	0.22	0.000023	1.12
Genome				
Average coverage	52.7-fold	30.9-fold	–	–
Heterozygosity (per kb)	0.19	0.22	–	–
Chromosome 21				
DNA enrichment	–	–	320-fold	120-fold
Average coverage	53.7-fold	31.1-fold	14.1-fold	35.9-fold
Heterozygosity (per kb)	0.13	0.21	0.24	0.26
Cumulative length of homozygous segments (Mb) 10–100 kb	9.68	22.60	20.50	20.50
>100 kb	19	4.80	5.10	5.10

Radiocarbon dates (uncalibrated), mean contamination estimates for the DNA fragments sequenced and summary statistics for the genomes and chromosome 21 sequences. mtDNA, mitochondrial DNA.

these young haplotypes are estimated to coalesce with the African genomes 100,000–230,000 years ago, suggesting that they entered into the ancestors of the Altai Neanderthal well before the reported gene flow from Neanderthals into modern humans outside Africa 47,000–65,000 years ago[12]. Both the cumulative and average length of the young 'African' haplotypes is longer in the Altai Neanderthal genome than in the Denisovan genome.

The introgression from a deeply divergent archaic population into the Denisovan lineage is a potential confounding factor in this analysis. However, this introgression event should affect older haplotypes in the Denisovan genome, rather than the young haplotypes examined above. Indeed, we find that the number of haplotypes in one archaic genome that coalesce outside Africans and the other archaic genome (Fig. 2b, inset) is higher in the Denisovan than in the Altai Neanderthal (Fig. 2b, right). Furthermore, the young 'African' haplotypes in the Altai Neanderthal genome do not significantly overlap with the older haplotypes in the Denisovan genome and in

simulations ARGweaver only infers them under a model with modern human gene flow into the Altai Neanderthal lineage (Extended Data Fig. 5).

Inference of gene flow in European Neanderthals

To investigate possible differences among Neanderthal populations with respect to introgression from modern humans, we designed oligonucleotide probes[14] based on the human reference sequence of chromosome 21, and used them to capture[15] this chromosome in a Neanderthal from Spain (El Sidrón Cave) and a Neanderthal from Croatia (Vindija Cave). We estimated their present-day human contamination to be around 1% (Table 1).

We find that the chromosome 21 of the Altai Neanderthal shares more derived alleles with Africans than the chromosome 21 of El Sidrón (3.5% more) and Vindija (4.9% more) Neanderthals, with the European Neanderthals sharing more derived alleles with Africans than the chromosome 21 of the Denisovan (9.8% more for El Sidrón, 8.8% more for Vindija). A comparison of the distribution of haplotype ages is not possible with the European Neanderthals, owing to insufficient amounts of data, but we compared the cumulative length of haplotypes coalescing within the African subtree for each Neanderthal lineage. This length is significantly greater for the Altai Neanderthal than for the European Neanderthals ($P < 0.01$; fraction of MCMC replicates), consistent with introgression from modern humans primarily into this Neanderthal lineage.

When we refine our estimates of gene flow by adding the chromosome 21 sequences of the European Neanderthals to our genome-wide data, G-PhoCS infers significant rates of gene flow from Neanderthals into modern humans outside Africa only for El Sidrón and Vindija Neanderthals (0.3–2.6%) (Fig. 3a), suggesting that Neanderthals from Europe are more closely related than the Altai Neanderthal to the population that interbred with modern humans outside Africa 47,000–65,000 years ago[12]. Conversely, significant rates of gene flow from modern humans into Neanderthals are inferred only into the ancestors of the Altai Neanderthal (0.1–2.1%) (Extended Data Figs 6 and 7). This suggests that modern human introgression into Neanderthals occurred mainly after the divergence of the Altai Neanderthal from El Sidrón and Vindija lineages 110,000 (68,000–167,000) years ago (Fig. 3b). However, it is possible that the lack of complete genomes from the European Neanderthals currently precludes the identification of modern human gene flow into them.

To explore the source of the modern human gene flow among the African populations, we simulated three scenarios in which the source of the gene flow into the Altai Neanderthal lineage was alternately an unknown population diverging from the ancestors of all
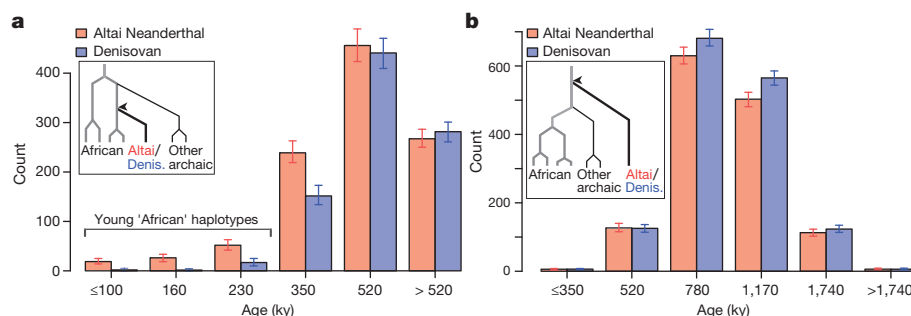

Figure 2 | Distinguishing between two scenarios of introgression into archaic humans. a, The age distribution of 'African' haplotypes (≥50 kb) in the Altai Neanderthal and the Denisovan genomes as inferred by ARGweaver. Error bars represent the 95% credible intervals from 302 Markov chain Monte Carlo (MCMC) replicates. An 'African' haplotype coalesces within the African subtree before coalescing with the other archaic individual (inset), and its age is inferred as that coalescent time (arrowhead). The majority of the young 'African' haplotypes in the Altai

Neanderthal genome are estimated to coalesce 100,000–230,000 years ago, with just a few estimated to coalesce less than 100,000 years ago (Supplementary Information section 10). **b**, The age distribution of 'deep ancestral' haplotypes (≥50 kb) in the Altai Neanderthal and Denisovan genomes. A 'deep ancestral' haplotype coalesces above the African subtree and the other archaic lineage (inset), and its age is inferred as that coalescent time (arrowhead). ky, thousand years.
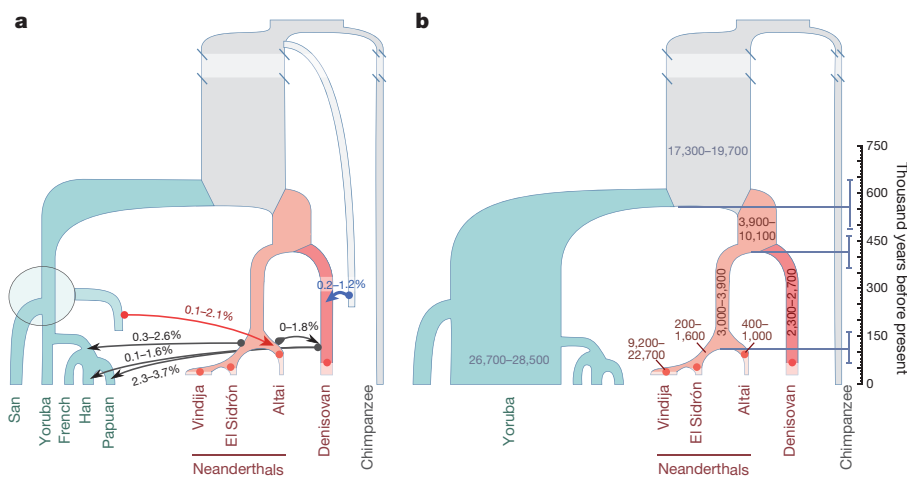
Figure 3 | Refined demography of archaic and modern humans. a, Total migration rates of six gene flow events inferred by G-PhoCS. The ranges correspond to 95% Bayesian credible intervals aggregated across runs. Five gene flow events have been previously reported, including gene flow from an unknown archaic group into Denisovans (blue arrow). In addition, we infer gene flow from a population related to modern humans into a population ancestral to the Altai Neanderthal (red arrow). It appears to come from a population that either split from the ancestors of present-day Africans or separated fairly early in the history of African populations (shaded circle). **b,** Effective population sizes and divergence times inferred by G-PhoCS. The ranges correspond to 95% Bayesian credible intervals aggregated across runs. The horizontal bars indicate posterior mean estimates for divergence times. Archaic samples (dots) are located at their estimated ages.

present-day Africans, of the San or of Yoruba lineage (Supplementary Information section 8). The G-PhoCS estimates from these three models are all similar and consistent with those in Fig. 3, and thus we cannot distinguish among them. However, it is clear that the source of the gene flow is a population equally related to present-day Africans and non-Africans (Extended Data Fig. 3). We conclude that the introgressing population diverged from other modern human populations before or shortly after the split between the ancestors of San and other Africans (Fig. 3a), which occurred approximately 200,000 years ago[11]. In agreement, the San, Mbuti and Yoruba genomes contribute equally to the young 'African' haplotypes in the Altai Neanderthal genome (Supplementary Information section 10).

Introgressed segments in the Altai Neanderthal
To shed light on possible functional implications of modern human gene flow into Neanderthals, we identified 163 putatively introgressed segments (≥50 kb) in the Altai Neanderthal genome (Supplementary Information section 9). These segments have no clear affinity to any present-day African population (Extended Data Fig. 8), and they overlap with 225 genes. Seven segments exceed 200 kb (Table 2) and the longest one (309 kb) overlaps with a region suspected to have been under positive selection in modern humans[3]. This region has a transcription factor gene (*NR5A2*) involved in liver development[16]. One segment of 150 kb is located within the *FOXP2* gene (Table 2), which encodes a transcription factor that may be relevant for language acquisition[17].

The number of putatively introgressed segments in the Altai Neanderthal decreases in regions of the genome under strong purifying selection (measured via background selection at linked sites[18]), and it is lower in the X chromosome compared to the autosomes. Because purifying selection purges deleterious alleles and the efficacy of purifying selection is higher on the X chromosome[19], this may indicate that modern human and Neanderthal[20] alleles were often not tolerated in each other's genetic background.

Population size in Neanderthals and Denisovans
Our demographic model suggests a long-term decline in the effective population size of Neanderthals and Denisovans since their divergence from the ancestors of present-day humans 484,000–640,000 years ago. However, the population ancestral to the Vindija Neanderthal appears to have expanded (Fig. 3b). In addition, the length distribution of homozygous stretches in the European Neanderthals resembles that

of the Denisovan, who lacks a signal of recent inbreeding[7], and not that of the Altai Neanderthal, whose parents were related at the level of half-siblings[2] (Table 1). Still, the European Neanderthals and the Denisovan exhibit signs of a history of mating in small populations[21], with a larger cumulative length of homozygous segments of 10–100 kb than present-day humans and great apes (Fig. 4). In agreement with purifying selection being less efficient in small populations, regulatory and conserved[22] regions in Neanderthals have a larger proportion of putatively deleterious alleles than present-day humans (Extended Data Fig. 9), as shown previously for their protein-coding genes[23].

Discussion
Our integrated demographic analysis of multiple archaic and present-day human genomes suggests a scenario of long-term decline in the populations of Neanderthals and Denisovans, with the consistently small Altai Neanderthal population perhaps reflecting a long period of isolation in the Altai Mountains. In addition, we provide evidence

Table 2 | Introgressed segments from modern humans into the Altai Neanderthal

Genomic region	SNPs	Sequence length (bp)	Genetic length (cM)	Genes in the region
Chr1: 199,707,795–200,016,460	161	308,665	0.047	*NR5A2; RNU6-609P; RNU6-716P; RNU6-778P*
Chr13: 49,532,446–49,790,867	103	258,421	0.040	*COX7CP1; FNDC3A; OGFOD1P1; RAD17P2; RNU6-60P; RNY3P2*
Chr2: 88,815,371–89,061,977	116	246,606	0.023	*EIF2AK3; RPIA; TEX37*
Chr3: 89,790,776–90,031,537	70	240,761	0.017	–
Chr3: 30,590,736–30,816,806	100	226,070	0.547	*GADL1; TGFBR2*
Chr6: 42,492,777–42,713,223	67	220,446	0.088	*ATP6V0CP3; PRPH2; RNU6-890P; TBCC; UBR2*
Chr8: 93,809,505–94,011,334	122	201,829	0.070	*IRF5P1; TRIQK*
Chr7: 113,813,987–113,963,584	37	149,597	0.055	*FOXP2*

The seven segments (≥200 kb) in the Altai Neanderthal genome that are enriched in heterozygous sites with derived alleles at high-frequency in Africans. These sites are homozygous ancestral in the Denisovan. The segment overlapping the *FOXP2* gene is also shown.
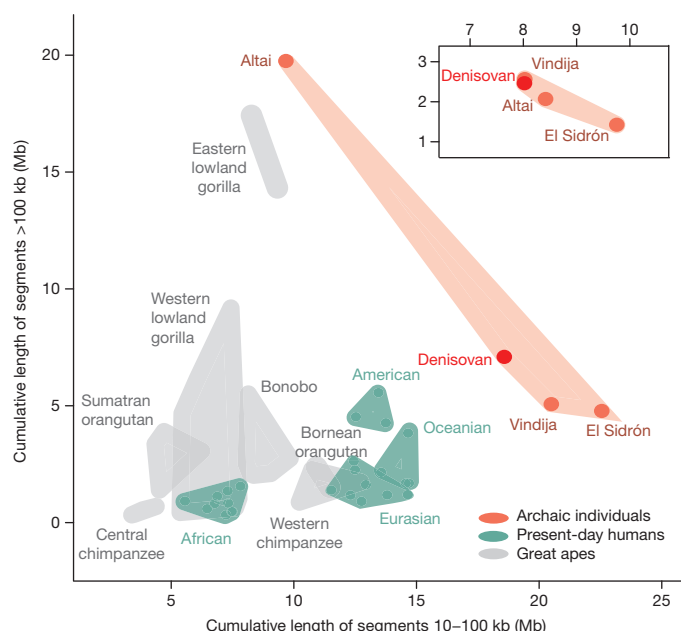
Figure 4 | Homozygous segments on chromosome 21. The range of the cumulative length (Mb) of homozygous segments is shown as the surface of a polygon, with individuals at the extremes of each group's range serving as vertices. Dots represent human individuals, archaic or otherwise, whereas great apes are not depicted individually. The Altai Neanderthal clusters with the other archaic individuals (inset) when recently inbred genomic regions are excluded.

for modern human introgression into the ancestors of this population of Neanderthals, and no such evidence in the European Neanderthals. These modern humans may represent a population that diverged early from other modern humans in Africa and later met the ancestors of the Altai Neanderthal. The finding of 'African' haplotypes as young as 100,000 years old in the Altai Neanderthal genome is consistent with interbreeding around that age.

Hublin[24] has proposed that Neanderthals expanded eastward from Europe during an interglacial period about 125,000 years ago (Oxygen Isotope Stage 5e). The presence of modern humans (at Skhul and Qafzeh) and Neanderthals (at Tabun) in the Levant as early as 120,000 years ago[25,26] provides one place where gene flow from early modern humans into Neanderthals could have occurred. Another place is Southern Arabia and the area around the Persian Gulf, where modern humans may have also settled early[27] and Neanderthals are likely to have been present[28]. The recent demonstration that modern humans may have been in China as early as 120,000 years ago[29] also suggests that modern humans migrated early out of Africa. Thus, early modern humans may have had the opportunity to admix with archaic hominins before the migration of the modern human ancestors of present-day non-Africans.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Arsuaga, J. L. et al. Neandertal roots: Cranial and chronological evidence from Sima de los Huesos. Science 344, 1358–1363 (2014).
2. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43–49 (2014).
3. Green, R. E. et al. A draft sequence of the Neandertal genome. Science 328, 710–722 (2010).
4. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514, 445–449 (2014).
5. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature (2015).
6. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468, 1053–1060 (2010).
7. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 222–226 (2012).
8. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526, 68–74 (2015).
9. Pickrell, J. K. et al. Ancient west Eurasian ancestry in southern and eastern Africa. Proc. Natl Acad. Sci. USA 111, 2632–2637 (2014).
10. Llorente, M. G. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science 350, 820–822 (2015).
11. Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. Nature Genet. 43, 1031–1034 (2011).
12. Sankararaman, S., Patterson, N., Li, H., Paabo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8, e1002947 (2012).
13. Rasmussen, M. D., Hubisz, M. J., Gronau, I. & Siepel, A. Genome-wide inference of ancestral recombination graphs. PLoS Genet. 10, e1004342 (2014).
14. Burbano, H. A. et al. Targeted investigation of the Neandertal genome by array-based sequence capture. Science 328, 723–725 (2010).
15. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl Acad. Sci. USA 110, 2223–2227 (2013).
16. Rausa, F. M., Galarneau, L., Bélanger, L. & Costa, R. H. The nuclear receptor fetoprotein transcription factor is coexpressed with its target gene HNF-3β in the developing murine liver intestine and pancreas. Mech. Dev. 89, 185–188 (1999).
17. Enard, W. FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. Curr. Opin. Neurobiol. 21, 415–424 (2011).
18. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5, e1000471 (2009).
19. Veeramah, K. R., Gutenkunst, R. N., Woerner, A. E., Watkins, J. C. & Hammer, M. F. Evidence for increased levels of positive and negative selection on the X chromosome versus autosomes in humans. Mol. Biol. Evol. 31, 2267–2282 (2014).
20. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507, 354–357 (2014).
21. Pemberton, T. J. et al. Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. 91, 275–292 (2012).
22. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050 (2005).
23. Castellano, S. et al. Patterns of coding variation in the complete exomes of three Neandertals. Proc. Natl Acad. Sciences USA 111, 6666–6671 (2014).
24. Hublin, J. J. in Neandertals and Modern Humans in Western Asia (eds Akazawa, T., Aoki, K. & Bar-Yosef, O.) (Kluwer Academic Publishers, 1998).
25. Mercier, N. H. V., Bar-Yosef, O., Vandermeersch B., Stringer, C. & Joron, J.-L. Thermoluminescence date for the Mousterian burial site of Es-Skhul, Mt. Carmel. J. Archaeol. Sci. 20, 169–174 (1993).
26. Grün, R. et al. U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. J. Hum. Evol. 49, 316–334 (2005).
27. Armitage, S. J. et al. The southern route "Out of Africa": evidence for an early expansion of modern humans into Arabia. Science 331, 453–456 (2011).
28. Rose, J. I. A. & Marks, A. E. "Out of Arabia" and the Middle–Upper Palaeolithic transition in the southern Levant. Quartär 61, 49–85 (2014).
29. Liu, W. et al. The earliest unequivocally modern humans in southern China. Nature 526, 696–699 (2015).

Author Contributions M.M. and Q.F. performed experiments; M.Ku., I.Gr., M.J.H., C.d.F., J.P.-M., M.Ki, Q.F., H.A.B., T.M.-B., A.M.A., S.P., M.M., A.S. and S.C. analysed genetic data; C.L.-F., M.d.l.R., A.R., P.R., D.B., Ž.,K., I.Gu. and B.V. analysed anthropological data; M.Ku., I.Gr., M.J.H., B.V., S.P., A.S. and S.C. wrote the manuscript.

Author Information Sequence data are available in the European Nucleotide Archive (ENA) under accession number PRJEB11828. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S. (asiepel@cshl.edu) or S.C. (sergi.castellano@eva.mpg.de).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

DNA extraction and library preparation. We prepared DNA extracts from two Neanderthal bones, SD1253 from El Sidrón Cave and Vi33.15 from Vindija Cave, as described in Rohland et al.[30] (Supplementary Table 1), and prepared DNA sequencing libraries containing a special 4 base pair (bp) clean-room tag sequence to avoid contamination in later steps[31,32]. During library preparation, we used a uracil-DNA-glycosylase (UDG) and endonuclease VIII mix to remove uracils resulting from cytosine deaminations[33].

Chromosome 21 capture experiment. We used a strategy previously described[15] that uses oligonucleotides synthesized on arrays to construct amplified probe libraries. We produced a probe library with a tile density of 3 nucleotides across the 29.8 Mb of non-repetitive sequences in chromosome 21 (GRCh37/hg19), with biotinylated probes similar to those described by Gnirke et al.[34]. We used this probe library, as previously described[23], to generate libraries from El Sidrón and Vindija Neanderthals. All libraries were subjected to a second round of amplification, followed by two rounds of hybridization capture. Capture eluates were amplified, barcoded with two indexes[32], pooled, and sequenced on the Genome Analyzer IIx (Illumina).

Contamination estimates. Estimates of present-day human mtDNA contamination in El Sidrón and Vindija libraries were previously reported in Castellano et al.[23]. These contamination estimates were calculated using diagnostic positions at which archaic mitochondrial genomes differed from sequences in a panel of 311 present-day human mitochondrial genomes. Nuclear DNA contamination estimates were calculated using a previously described maximum likelihood approach[7] that co-estimates the contamination and sequence error in the autosomes.

Computational correction of cytosine deaminations. Sequences may carry residual cytosine deaminations in the first positions of the 5′ end and in the last positions of the 3′ end in spite of the UDG treatment[33] (Supplementary Fig. 1). These bases are read as thymine and adenosine, respectively. As similarly described for the Altai Neanderthal genome[2], we decreased the quality to 2 of any 'T' base occurring within the first five bases or any 'A' base within the last five positions in El Sidrón and Vindija sequences.

Variation discovery. We called Neanderthal genotypes with GATK[35] and applied a previously described set of filters[23] (Supplementary Information section 3) to obtain high-quality sites for subsequent analyses. We obtained such calls for 17,014,623 and 20,582,399 sites for El Sidrón and Vindija chromosome 21, respectively. Genotypes in the Altai Neanderthal, Denisovan and present-day human genomes were similarly obtained (Supplementary Table 6), and a combined file for all individuals was created and annotated as in Meyer et al.[7]. Because multiple contaminated DNA fragments are needed for a contaminated genotype to be called, the proportion of contaminated genotypes is likely to be smaller than the reported contamination of 1% among DNA fragments.

Capture bias. In order to understand capture bias, we captured the chromosome 21 of the Altai Neanderthal to an average coverage of 46.9-fold. We then downsampled these sequences to assess capture bias at a wider range of average coverage from 8.1-fold to 35.7-fold, and did the same for the Altai Neanderthal shotgun sequences. The mean reference allele frequency is shifted from 0.52 in the shotgun sequences to 0.54–0.55, similar to the observed frequencies in the other archaic captured individuals (Supplementary Fig. 4). The mutation spectra after filtering do not change with coverage (Supplementary Fig. 2), and differences in allele frequency at heterozygous sites in the shotgun sequences are small (Supplementary Fig. 5). We observed that 3.8–5.2% of heterozygous sites in the shotgun sequences of chromosome 21 in the Altai Neanderthal are homozygous in the capture experiment at coverage from 14-fold to 46.9-fold (Supplementary Table 7). However, 22.3–45.4% of these heterozygous sites are filtered out, mainly due to low coverage in the capture experiment. The same is true for sites that are heterozygous in the capture experiment at 46.9-fold coverage, but homozygous (4.9–5.9%) or missing (14.1–21.6%) in the shotgun data at 15.1–53.7-fold coverage. In addition, the distribution of homozygosity stretches does not differ between the capture and the shotgun sequences (Supplementary Fig. 9). We conclude that capture bias does not distort our results in a systematic way.

Sequence patterns. Our analysis of the divergence of the archaic genomes to Africans and to each other sought to uncover the patterns that distinguish modern human gene flow into the Altai Neanderthal lineage from archaic gene flow into the Denisovan lineage. To do this, we analysed 15,881 sequence windows of 100 kb in length across the genomes of the two archaic individuals. These windows were required to have high-quality genotypes (as described in Supplementary Information section 3) in at least 50% of its length in both archaic genomes. Because the phase of the archaic alleles is unknown, the divergence of the archaic genomes to Africans was calculated using the archaic alleles in each window that

give the minimum number of differences to derived alleles at >0.9 frequency in 504 individuals from five African present-day populations (Yoruba, Mende, Luhya, Gambian, and Esan)[8]. Using the minimum divergence to Africans allows introgressed segments from modern humans to be more easily identified. In contrast, the divergence between the archaic genomes was calculated using the archaic alleles in each window that give the maximum number of differences. Using the maximum divergence between the archaic windows allows introgressed segments in either of the two archaic individuals to be more easily identified. Derived alleles were determined using the inferred ancestral base in the EPO six-primate alignments[36] and the minimum and maximum number of differences in a sequence window was divided by its number of high-quality genotypes. Regions of the genome described as inbred in the Altai Neanderthal[2] were excluded from this analysis. These are 103 regions >2.5 cM depleted in heterozygous sites. In this way, heterozygosity in Fig. 1b could be calculated from the same 15,881 sequence windows of 100 kb in Fig. 1a.

We used the program ms[37] to simulate 15,881 sequence windows of 100 kb in length, using parameters that are consistent with the G-PhoCS estimates (Supplementary Information section 8). We simulated scenarios with and without modern human gene flow into the Altai Neanderthal lineage (Supplementary Information section 9). The mutation rate of 0.5×10^{-9} mutations per bp and year[4,38] and an average generation time of 29 years[39] (as assumed in the G-PhoCS inferences) were also used. The number of chromosomes simulated were 1,008 for the Africans, two for the Neanderthal, two for the Denisovan, one for the unknown archaic, and one for the chimpanzee.

Alignments at neutral loci. Multiple sequence alignments were obtained for our main demography inference using G-PhoCS. Following the guidelines established in previous studies[11,40], we extracted multiple sequence alignments of the Altai Neanderthal, the Denisovan, and multiple present-day humans at 13,753 loci, 1 kb long, selected to minimize influence of direct selection, linkage between loci, and missing data. Among these, 2,960 loci were selected from chromosome 21, for which sequence data was available from El Sidrón and Vindija Neanderthals.

Demography inference. Our demography inference is based on five main G-PhoCS runs, each one containing the Altai Neanderthal, the Denisovan, the chimpanzee outgroup (panTro2), and two present day humans from a particular population. We considered populations from Africa (Yoruba and San), Europe (French), East Asia (China), and Oceania (Papuan). In five additional runs we added sequences from chromosome 21 of El Sidrón and Vindija Neanderthals. To account for the fact that different individuals lived at different times, we modified the algorithm to sample the times of the archaic individuals as four additional free parameters (Supplementary Information section 8). To validate the robustness of our estimates, we conducted additional inferences using subsets of the archaic individuals, different subsets of the loci, and allowing for gene flow from an unsampled (unknown) divergent human group, and explicitly modelling the source population of modern introgression into the ancestors of the Altai Neanderthal as an unsampled population branching off from the modern human population.

G-PhoCS setup. In each G-PhoCS run, we ran the Markov chain Monte Carlo (MCMC) sampler for 100,000 burn-in iterations and 200,000 subsequent sampling iterations, and checked manually for convergence of the Markov Chain. The samples were used to estimate a posterior mean and 95% Bayesian credible interval for each demographic parameter. For parameters common to the five runs with different present-day humans, we combined the five parameter traces to obtain aggregated estimates. Estimates of population divergence time and effective population size were calibrated by assuming an average mutation rate of 0.5×10^{-9} per base pair per year[4,38] and an average generation time of 29 years[39]. Estimates under different assumptions on mutation rate and generation time are obtained by simple scaling of the reported estimates. Gene flow is measured using the total migration rate, which is the estimated per-generation rate times the number of generations that migration is allowed in the model.

Simulations. To validate the G-PhoCS inferences we simulated, using ms[37], 10,000 loci of 1 kb of length for the Altai Neanderthal, Denisovan, three present-day humans from San, Yoruba, and European populations and the chimpanzee outgroup. Demographic parameters were set according to the ones inferred on the genomic sequences, with parameters describing divergence times of modern populations and growth of the European population taken from recent studies[11,41]. For these individuals, sequences were simulated under different scenarios for modern human introgression into the Altai Neanderthal population: (1) no introgression; (2) introgression from a population that diverged from present-day humans before the San divergence; (3) introgression from a population that diverged from the population ancestral only to Yoruba and Europeans; (4) introgression from a population that diverged from the population ancestral only to the San; and (5) introgression from a population ancestral only to Europeans. These five scenarios also included simulated gene flow from the Altai Neanderthal and an unsampled archaic population into the Denisovan population. G-PhoCS was run under

each scenario three times (one for each present-day individual) with the same settings used in the analysis of the actual genomes.

ARGweaver analysis. ARGweaver was run using the Altai Neanderthal and Denisovan genomes, six modern human genomes (two Yoruba, two San, and two Mbuti; Supplementary Table 2), and the chimpanzee reference genome (panTro4). Filters were applied to mask regions with uncertain genotype calls. The genome was divided into roughly 5 Mb blocks with 1 Mb overlap between adjacent blocks. A new method to integrate over genotype phase was used on the archaic and present-day human genomes (Supplementary Information section 10). Other settings, such as the recombination and mutation rate map, and the population size ($n = 11,534$), were the same as previously reported[13]. ARGweaver was run for 5,000 MCMC iterations, with an ancestral recombination graph sampled every 20 iterations starting at iteration 2,000. ARGweaver was run similarly with El Sidrón and Vindija chromosome 21 included. 'African' and 'deep ancestral' haplotypes were determined in each sampled ancestral recombination graph using only a single lineage from each archaic genome to avoid differences in power between them due to different levels of heterozygosity and inbreeding.
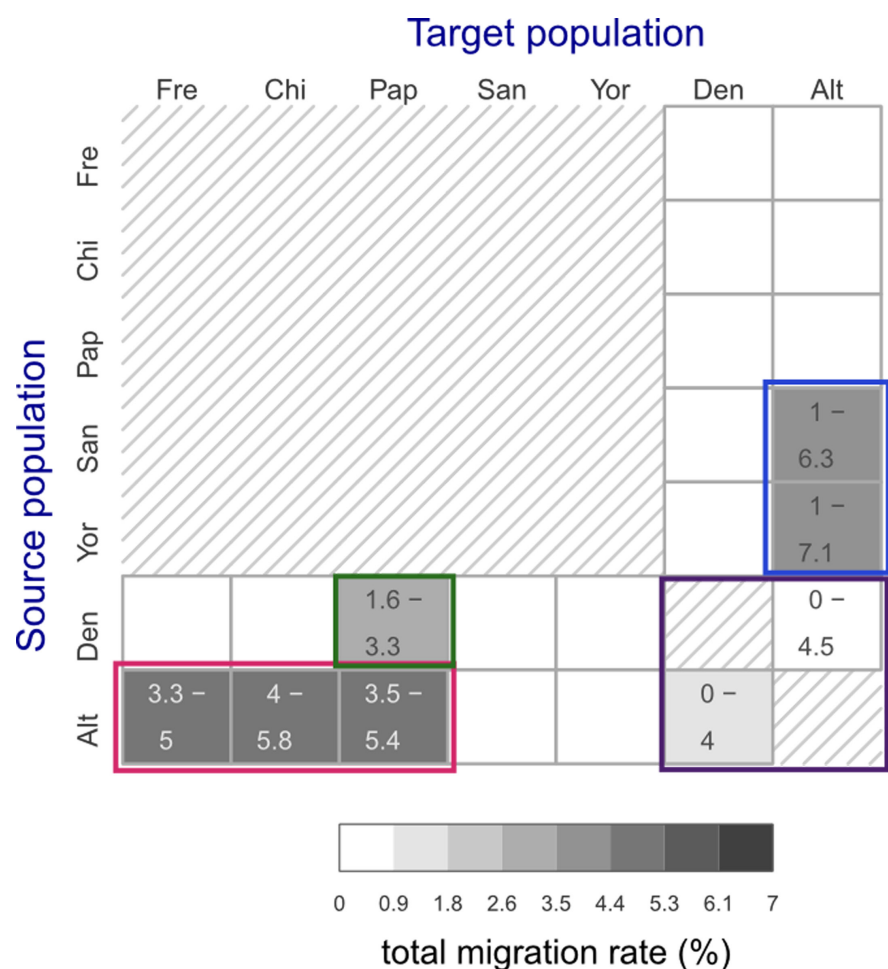
Screen for introgressed segments. A screen for modern human introgressed segments was performed using the frequency in Africans of derived alleles in sites that are heterozygous in one archaic genome (Altai Neanderthal or Denisovan) and homozygous ancestral in the other archaic genome. This allows us to identify segments that carry an archaic haplotype on three chromosomes, and a human haplotype only on one chromosome. Derived alleles were determined using the inferred ancestral base in the EPO six-primate alignments[36]. Genotypes and allele frequencies for the African individuals were obtained from the 1000 Genomes project[8]. We fitted the African derived allele frequencies along each of the archaic genomes using a locally weighted polynomial regression (loess function in *R*), and selected those genomic segments containing at least 10 sites where the fitted curve to the derived African allele frequencies consistently stayed over a frequency of 0.25 across 25 kb. Segments containing incompatible sites, that is, sites that were derived and shared in both archaic individuals, were removed. In the Altai Neanderthal, the average heterozygosity of the putatively introgressed segments is 4.9-fold higher than in random genome regions (Supplementary Information section 9).

Homozygosity segments. Homozygous segments were defined as maximal regions between two heterozygous positions of length between 10 and 100 kb or larger than 100 kb. To compare the hominin samples with great apes, we masked regions for which no data on great apes were available[42] in addition to the filters described in Supplementary Information section 3.

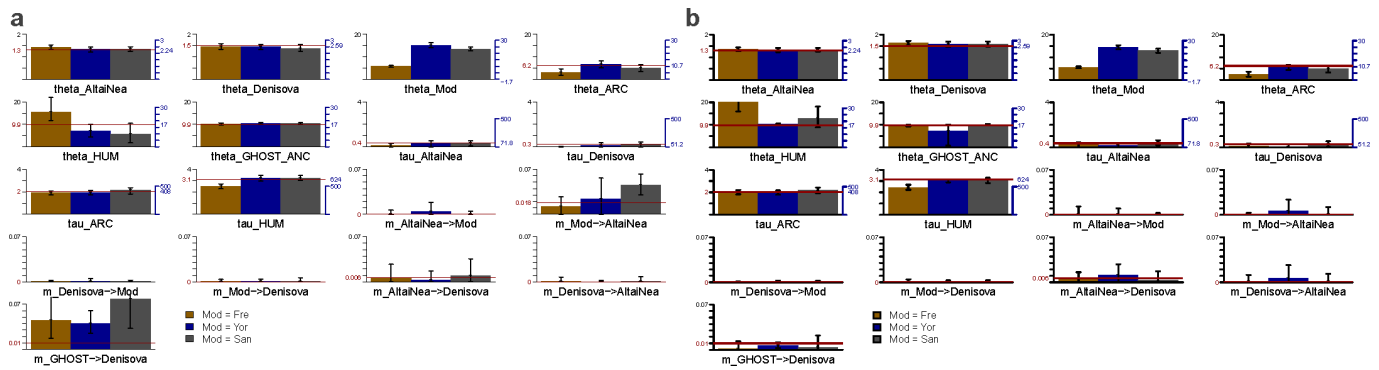**Prediction of functional consequences.** We tested the functional consequences of the derived alleles using conservation scores from PhastCons[22]. We calculated the fractions of mutations in deleterious sites for the different human groups (Supplementary Information section 7). We used annotations of transcripts from ENSEMBL[43] to define coding regions, untranslated regions, and 5,000 bases upstream of transcription start sites and downstream of transcription end sites. We used those as well as conserved transcription factor binding sites[44] and conserved elements, and sampled randomly for each category the same number of bases in neutral sites to calculate the ratio of "functional" to "neutral" polymorphism.

30. Rohland, N. & Hofreiter, M. Comparison and optimization of ancient DNA extraction. *Biotechniques* **42,** 343–352 (2007).
31. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protocols* **2010,** http://dx.doi.org/10.1101/pdb.prot5448 (2010).
32. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40,** e3 (2012).
33. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38,** e87 (2010).
34. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnol.* **27,** 182–189 (2009).
35. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
36. Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.* **18,** 1814–1828 (2008).
37. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18,** 337–338 (2002).
38. Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328,** 636–639 (2010).
39. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128,** 415–423 (2005).
40. Freedman, A. H. *et al.* Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10,** e1004016 (2014).
41. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc. Natl Acad. Sci. USA* **108,** 11983–11988 (2011).
42. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499,** 471–475 (2013).
43. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21,** 3439–3440 (2005).
44. Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor binding sites. *Nature Genet.* **45,** 723–729 (2013).
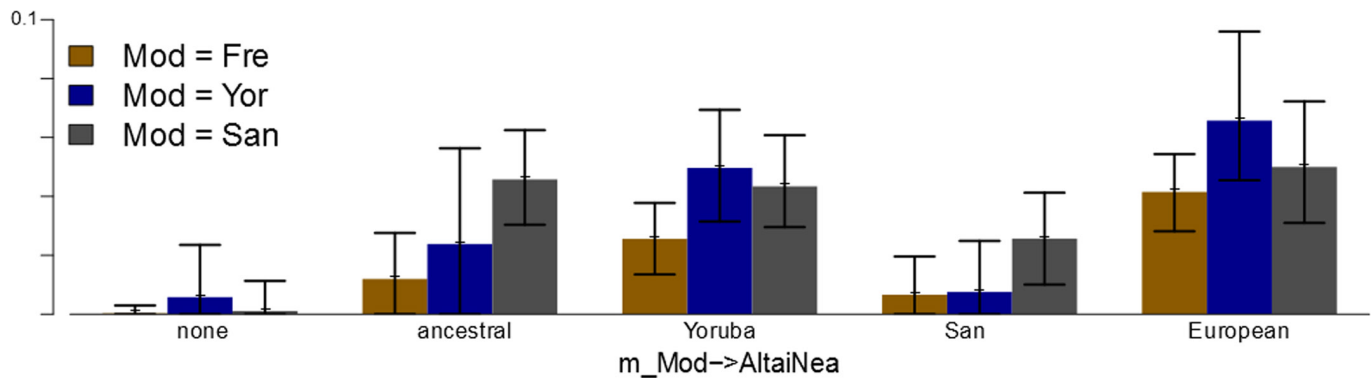
Target population

	Fre	Chi	Pap	San	Yor	Den	Alt
Fre							
Chi							
Pap							
San							1 – 6.3
Yor							1 – 7.1
Den			1.6 – 3.3				0 – 4.5
Alt	3.3 – 5	4 – 5.8	3.5 – 5.4			0 – 4	

Source population

total migration rate (%)

0   0.9   1.8   2.6   3.5   4.4   5.3   6.1   7

Extended Data Figure 1 | Migration rates in preliminary demographic inference. Total migration rates estimated for 22 directional migration bands in five separate preliminary G-PhoCS runs. Rows correspond to source populations and columns to the target populations. The 20 migration bands between modern and archaic populations were considered in five separate runs, each containing the four bands associated with a different modern human population (Supplementary Fig. 15A). The two migration bands between the Denisovan and the Altai Neanderthal populations were considered in all five runs, and the values shown here correspond to an aggregate of all five runs. The estimates are as shown in Supplementary Fig. 15B. Shade indicates the posterior mean total migration rate (legend), which approximates the probability that a lineage in the target population originated in the source population. The 95% Bayesian credible intervals from 2,000 MCMC replicates are indicated for migration bands whose upper credible interval bound is above 0.3%. We identified four clusters of migration bands, corresponding to what were likely at least four different cases of introgression between populations: (1) Neanderthals into non-African modern humans (red box), (2) Denisovans into Oceanians (green box), (3) between Neanderthals and Denisovans (magenta), and (4) modern humans into Neanderthals (blue box). Alt, Altai Neanderthal; Chi, Chinese; Den, Denisovan; Fre, French; Pap, Papuan; Yor, Yoruba.
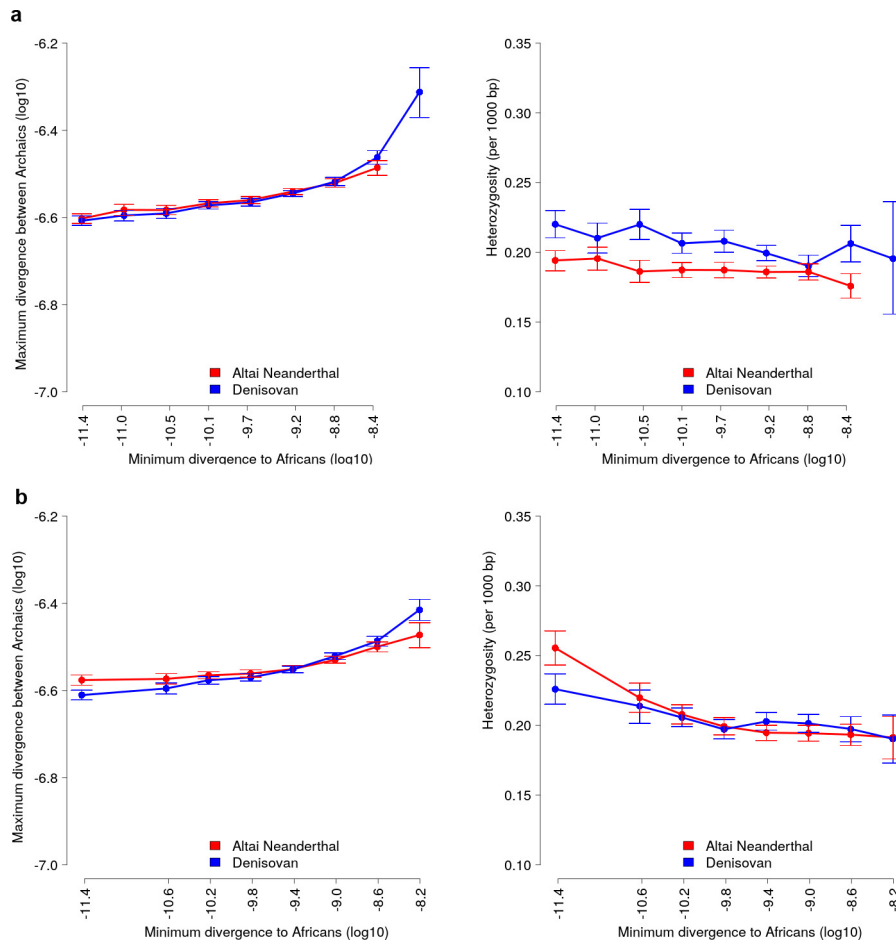
Extended Data Figure 2 | Demographic inference on simulated data.
Simulated data were generated under the demographic model as inferred by G-PhoCS (Supplementary Table 13). Each simulated data set consisted of 10,000 loci of 1 kb length. We simulated the Altai Neanderthal, the Denisovan, and three modern human populations corresponding to the San, Yoruba, and French, with modern human demography consistent with recent studies (Supplementary information section 8). Three migration bands were simulated: (1) from the Altai Neanderthal to the Denisovan, (2) from a population that diverged from the ancestors of all present-day humans 300,000 years ago into the Altai Neanderthal, and (3) from a population that diverged from the ancestors of all modern and archaic humans roughly 2.6 million years ago into the Denisovan. **a**, Estimates of effective population sizes (theta, θ), population divergence times (tau, τ) and migration rates (m) from three G-PhoCS runs on

data simulated with gene flow from modern humans into the Altai Neanderthal lineage. Each run analyses an individual from a different present-day population, using the exact same setup used in our main analysis (Supplementary Fig. 15A). Parameters are typically estimated accurately, with 95% Bayesian credible intervals containing the values used in simulations (horizontal red lines). Rates of archaic gene flow into Denisovan appear to be somewhat overestimated, and differences between analyses of African and non-African populations are consistent with those observed in the data analysis (Supplementary Fig. 15B). **b**, Similar analysis done on data simulated without gene flow from modern humans into the Altai Neanderthal lineage. Accurate estimates are obtained for all model parameters, and no gene flow is inferred from modern humans into the ancestors of the Altai Neanderthal. Error bars represent the 95% Bayesian credible intervals from 2,000 MCMC replicates.
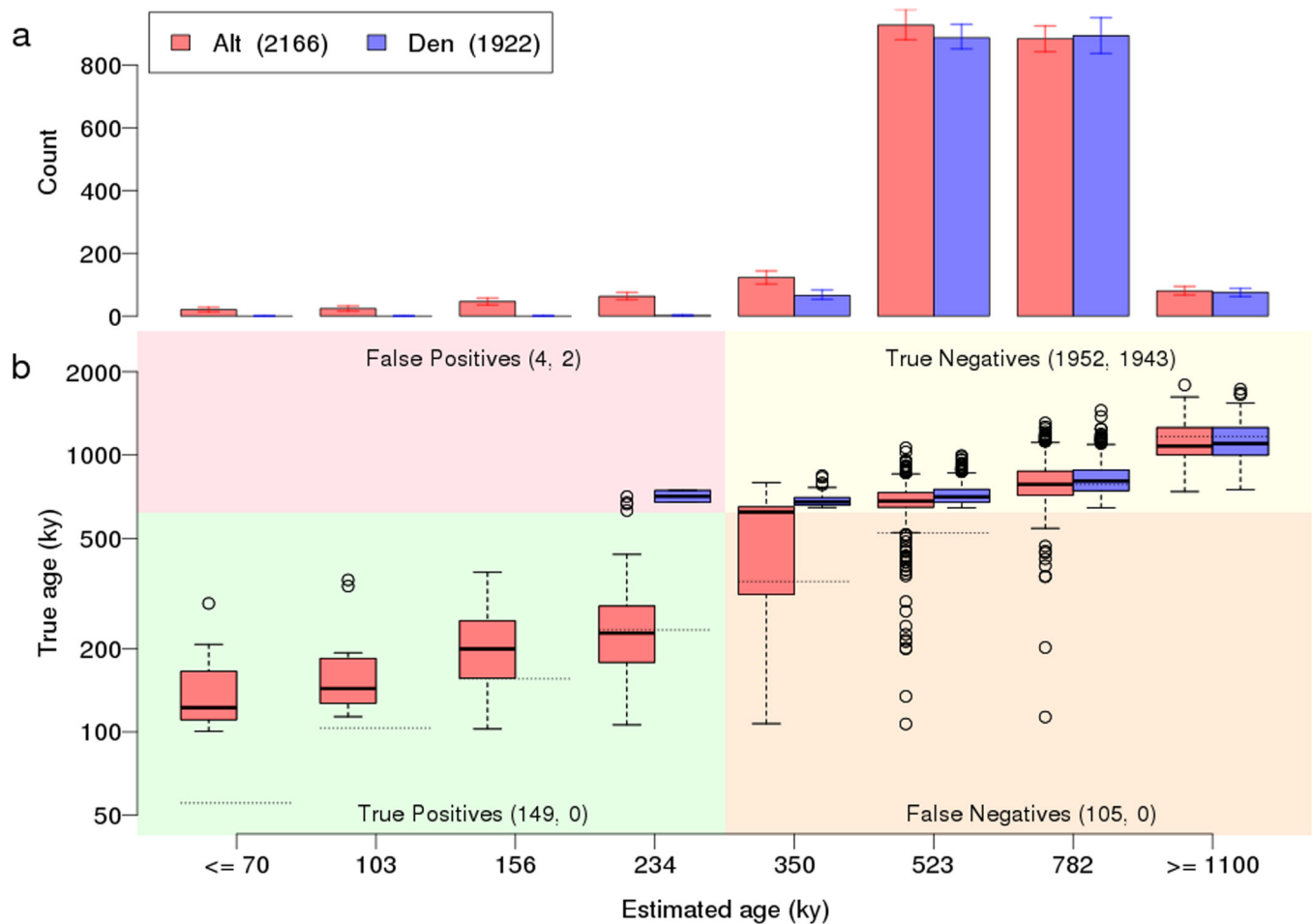
Extended Data Figure 3 | Simulation of different source populations for modern introgression into Neanderthals. Estimated rates of migration from the modern human population to the Altai Neanderthal population obtained from 15 G-PhoCS runs on five simulated data sets. All demographic parameters are set according to the values inferred by G-PhoCS in our data analysis (Supplementary Table 13), and the five data sets differ by the source population for migration: no migration (none), population ancestral to all present-day humans (ancestral), population ancestral to Yoruba and Europeans (Yoruba), San population (San), and European population (European). The first two sets are the ones analysed in Extended Data Fig. 2. Each data set is analysed three times, using different present-day samples: European (French), Yoruba, or San. Significant differences in estimates are observed between the data sets with and without gene flow and the data set with gene flow from a source population related to Europeans. Only minor differences were observed between values inferred for the three data sets with source population diverging from African populations (San, Yoruba, and ancestral). We conclude from this that the source population likely diverged from an African human population before the divergence of present-day Eurasians. The shaded circle in Fig. 3a represents this conclusion. Error bars represent the 95% Bayesian credible intervals from 2,000 MCMC replicates.
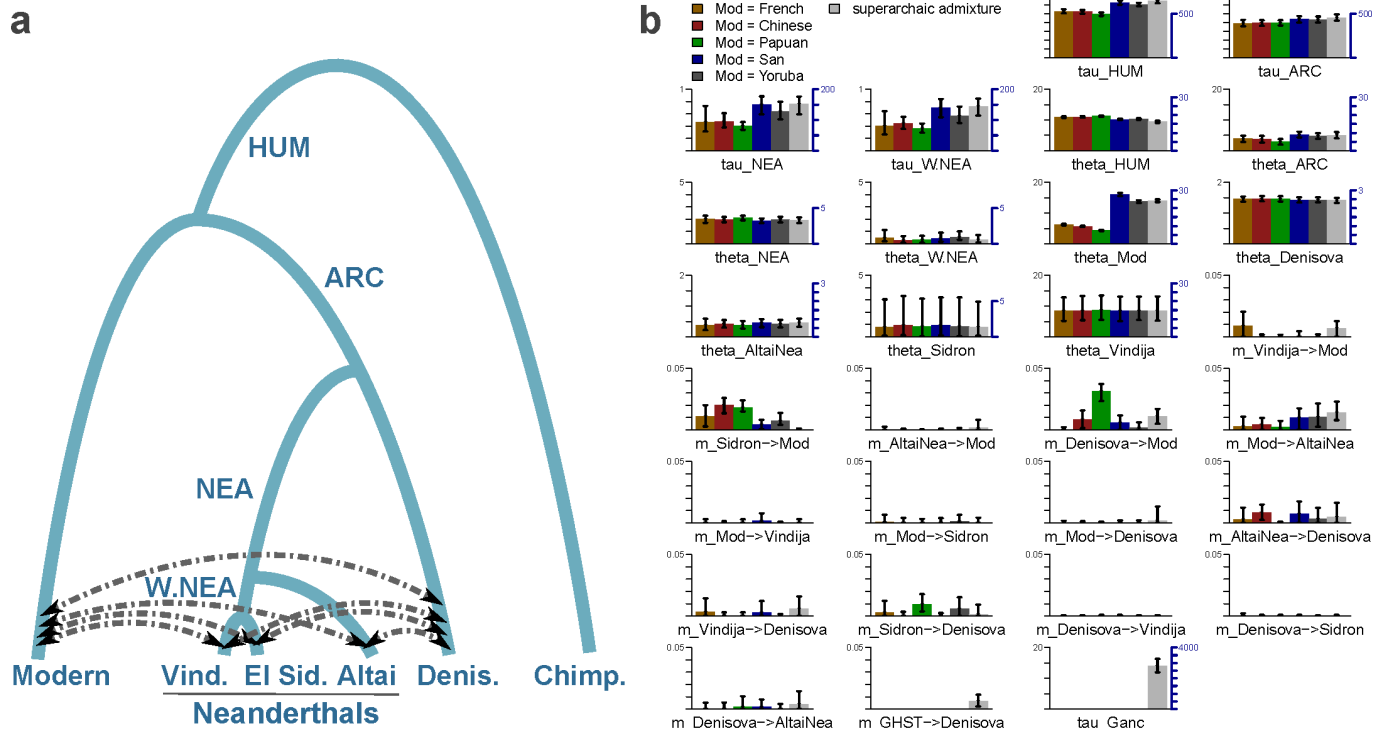
Extended Data Figure 4 | Simulation of present-day human contamination. Simulated windows of 100 kb for the Altai Neanderthal and Denisovan genomes with present-day human contamination of 5% at the genotype level. Windows are binned by their minimum divergence to Africans using derived alleles at >0.9 frequency in the simulated African population. The *x* and *y* axis as in Fig. 1. **a**, Gene flow from a deeply divergent archaic hominin into the Denisovan lineage (1%) and Altai Neanderthal gene flow into the Denisovan lineage (0.65%). **b**, Gene flow from a deeply divergent archaic hominin into the Denisovan lineage, Altai Neanderthal gene flow into the Denisovan lineage and modern human gene flow into the Altai Neanderthal lineage (1.8%). Error bars represent the 95% confidence intervals from 1,000 bootstrap replicates.

Extended Data Figure 5 | Haplotype ages inferred by ARGweaver on simulated data. a, Distribution of 'African' haplotype ages in sequences simulated with introgression into the Altai Neanderthal lineage from modern humans 100,000 years ago. 'African' haplotypes are identified as in Fig. 2. Error bars represent the 95% Bayesian credible intervals from 302 MCMC replicates. **b**, Distribution of true haplotype ages for each of the estimated ages. The horizontal dotted lines show the estimated age. The plot is divided into four quadrants; the lower half represents 'African' haplotypes having true ages between 100,000 and 620,000 years ago (the divergence time between archaic and present-day humans), which are necessarily due to post-divergence gene flow from modern humans.
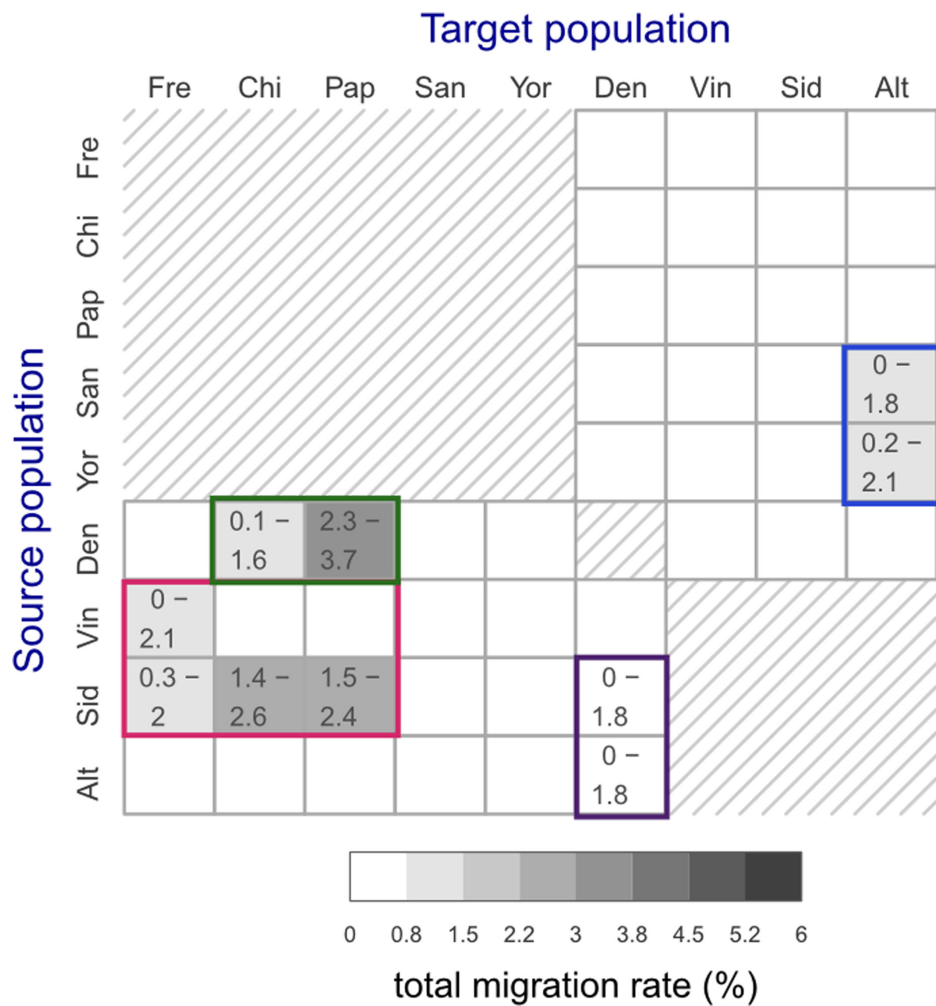
The left side of the plot represents regions that would be identified as introgressed based on a threshold of ≤ 234,000 years. The counts in each quadrant are for Altai Neanderthal (red) and Denisovan (blue), respectively. The counts for the Denisovan in the lower two quadrants are zero because there was no simulated migration from modern humans into the Denisovan lineage. Note that this is a somewhat nonstandard plot of true age versus estimated age; a more standard, reversed view is given in Supplementary Fig. 33 and demonstrates that the estimated ages are largely unbiased. Error bars as in the standard Tukey box plot (R boxplot function).

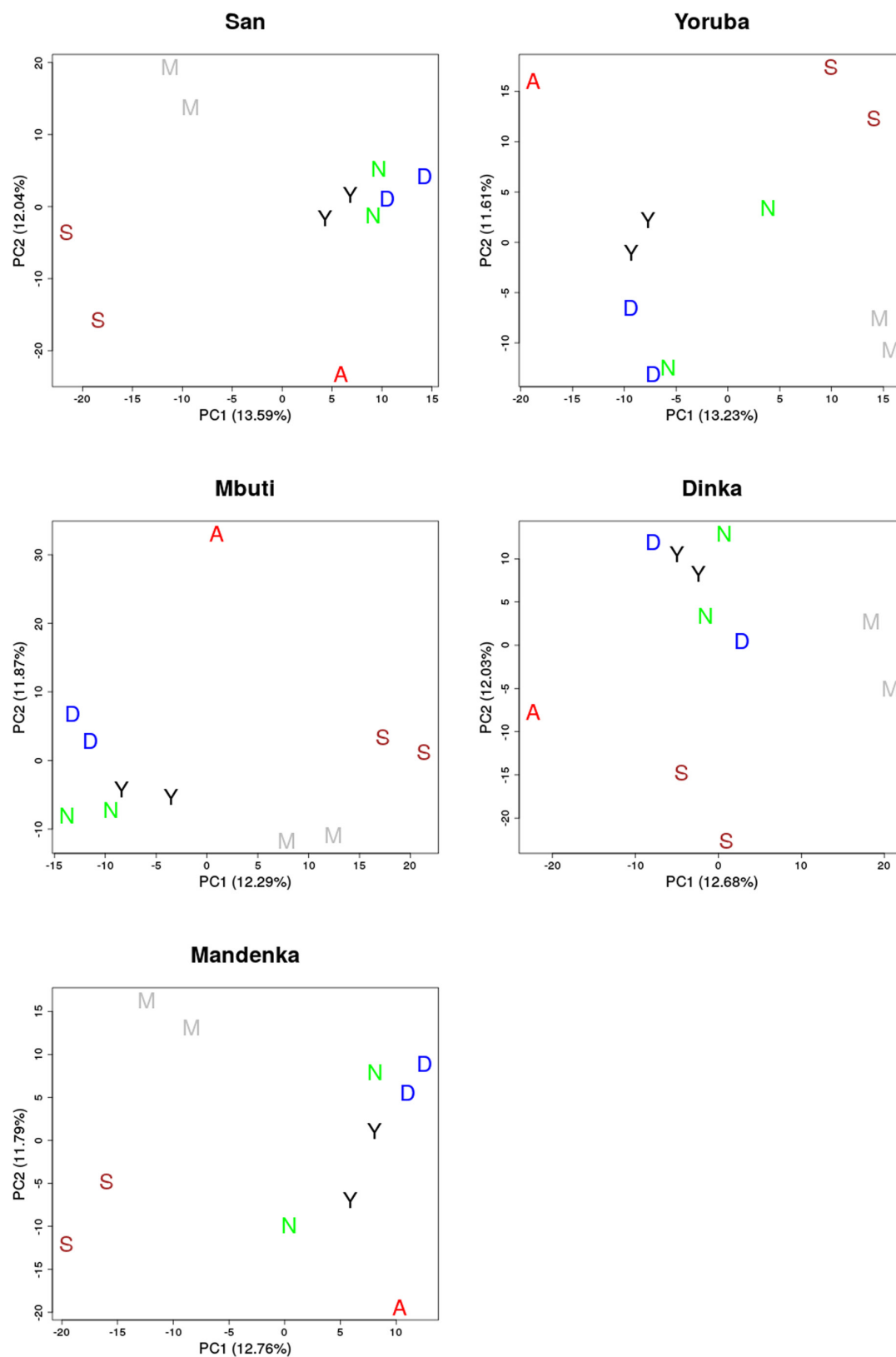**Extended Data Figure 6 | Main G-PhoCS demographic inference.** Summary of the main demographic inference using G-PhoCS in a model with four archaic populations and one modern human population. **a**, The population phylogeny assumed in each of the G-PhoCS runs. Labels on internal edges indicate names of the four ancestral populations: population ancestral to the two Western Neanderthals (W.NEA), population ancestral to all three Neanderthals (NEA), population ancestral to all four archaic individuals (ARC), and population ancestral to all human samples (HUM). We augmented the phylogeny with 14 directional migration bands (arrows) between all pairs of sampled populations except for the pairs of Neanderthal populations. In one of the runs we added an unknown 'ghost' population and a migration band from that population into the Denisovan population. **b**, Parameter estimates obtained by G-PhoCS in six separate runs analysing 13,754 neutral and loosely linked loci, substituting samples in the 'Modern' population with pairs of present-day humans from five different modern populations (Supplementary Table 11). The last run has gene flow from the 'ghost' population and uses two Yoruba individuals in the modern human population. Bar heights indicate posterior mean and error bars correspond to 95% Bayesian credible intervals. Estimates of divergence times (τ) and effective population sizes (θ) are given in raw form, scaled by number of mutations per 10 kb (left axis), and calibrated to absolute units, 1,000 years for time, and 1,000 individuals for effective population size, (right axis) assuming an average mutation rate of 0.5×10^{-9} mutations per year per bp and an average generation time of 29 years. For each of the 14 migration bands, we are showing the estimated total migration rates (m). See Supplementary Information section 8 for more information on parameter calibration and setup for G-PhoCS. A graphical summary of these estimates is given in Fig. 3.
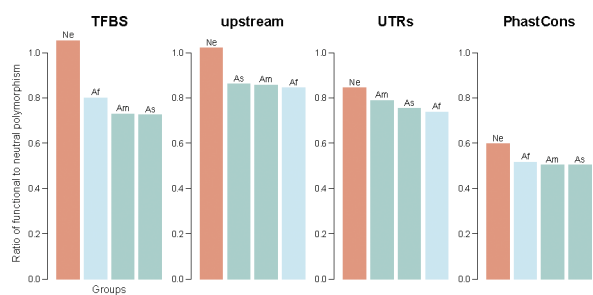
Extended Data Figure 7 | Migration rates in main demographic inference. Total migration rates estimated for 46 directional migration bands in five separate G-PhoCS runs. Rows correspond to source populations and columns to the target populations. The 40 migration bands between modern (present-day) and archaic populations were considered in five separate runs, each containing the eight bands associated with a different modern human population (Extended Data Fig. 6a). The six migration bands between the Denisovan population and the three Neanderthal populations were considered in all six runs, and the values shown here were estimated as an aggregate of all five runs. The estimates are as shown in Extended Data Fig. 6. Shade indicates the posterior mean total migration rate (legend), which approximates the probability that a lineage in the target population originated in the source population. The 95% Bayesian credible intervals from 2,000 MCMC replicates are indicated for migration bands whose upper credible interval bound is above 0.3%. We identified four clusters of migration bands, corresponding to what were likely at least four different cases of introgression between populations: (1) Western (European) Neanderthals into non-African modern humans (red box), (2) Denisovans into East Asian and Oceanians (green box), (3) Neanderthals into Denisovans (magenta), and (4) modern humans into Eastern Neanderthals (blue box). Directed arrows in Fig. 3a depict these introgression events. Sid, El Sidrón Neanderthal; Vin, Vindija Neanderthal.

Extended Data Figure 8 | Principal component analysis. The putatively introgressed segments in the Altai Neanderthal genome, defined by derived alleles in two individuals from the San, Yoruba, Mbuti, Dinka or Mandenka populations. The introgressed segments show no clear affinity to one present-day African population. A, Altai Neanderthal; S, San; M, Mbuti; Y, Yoruba; D, Dinka; N, Mandenka.

Extended Data Figure 9 | Natural selection in chromosome 21. Ratio of functional (putatively deleterious) to neutral polymorphism in archaic and present-day humans (Supplementary Information section 7). TFBS, transcription factor binding sites; upstream refers to 5 kb before the transcription start site of genes; UTRs, untranslated regions (in the mRNA). PhastCons ≥ 0.9 for a site to be used. Ne, Neanderthals; Af, Africans; As, Asians; Am, Americans.

Extended Data Table 1 | Shared derived alleles

African frequency	0≤f≤0.1	0.1<f ≤ 0.2	0.2<f ≤ 0.3	0.3<f ≤ 0.4	0.4<f ≤ 0.5	0.5<f ≤ 0.6	0.6<f ≤ 0.7	0.7<f ≤ 0.8	0.8<f ≤ 0.9	0.9<f≤ 1	**Fixed**
Altai (%)	66.86	5.01	3.89	3.42	3.11	2.79	2.71	2.55	2.68	**6.99**	**2.98**
Denisovan (%)	72.27	4.35	3.43	2.96	2.60	2.5	2.26	2.12	2.23	**5.30**	**1.90**

Percentage of derived alleles in one archaic genome (Altai Neanderthal or Denisovan) shared with African genomes in sites that are homozygous ancestral in the other archaic genome.
The percentages of shared derived alleles are binned by their African allele frequency. Fixed derived alleles in Africans are included in the '$0.9 < f \leq 1$' but also shown separately in the 'fixed' category.