Report

Current Biology

The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans

Highlights

- Denisovan admixture into modern humans occurred after Neanderthal admixture
- There is more Denisovan ancestry in South Asians than expected from current models
- Denisovan ancestry has been subject to positive and negative selection after admixture
- Male infertility most likely occurred after modern human interbreeding with Denisovans

Authors

Sriram Sankararaman, Swapan Mallick, Nick Patterson, David Reich

Correspondence

sriram@cs.ucla.edu (S.S.), reich@genetics.med.harvard.edu (D.R.)

In Brief

Sankararaman et al. present a map of Denisovan and Neanderthal ancestry in 120 diverse populations and show that Denisovan admixture post-dated Neanderthal admixture. South Asians have more Denisovan ancestry than expected. There was selection both for and against archaic ancestry. Hybridization with Denisovans was probably associated with reduced male fertility.

Sankararaman et al., 2016, Current Biology *26*, 1–7
May 9, 2016
http://dx.doi.org/10.1016/j.cub.2016.03.037

CellPress

Please cite this article in press as: Sankararaman et al., The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans, Current Biology (2016), http://dx.doi.org/10.1016/j.cub.2016.03.037

Current Biology
Report

The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans

Sriram Sankararaman,[1,2,*] Swapan Mallick,[3,4,5] Nick Patterson,[4] and David Reich[3,4,5,*]
[1]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA
[3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[5]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: sriram@cs.ucla.edu (S.S.), reich@genetics.med.harvard.edu (D.R.)
http://dx.doi.org/10.1016/j.cub.2016.03.037

SUMMARY

Some present-day humans derive up to ~5% [1] of their ancestry from archaic Denisovans, an even larger proportion than the ~2% from Neanderthals [2]. We developed methods that can disambiguate the locations of segments of Denisovan and Neanderthal ancestry in present-day humans and applied them to 257 high-coverage genomes from 120 diverse populations, among which were 20 individual Oceanians with high Denisovan ancestry [3]. In Oceanians, the average size of Denisovan fragments is larger than Neanderthal fragments, implying a more recent average date of Denisovan admixture in the history of these populations ($p = 0.00004$). We document more Denisovan ancestry in South Asia than is expected based on existing models of history, reflecting a previously undocumented mixture related to archaic humans ($p = 0.0013$). Denisovan ancestry, just like Neanderthal ancestry, has been deleterious on a modern human genetic background, as reflected by its depletion near genes. Finally, the reduction of both archaic ancestries is especially pronounced on chromosome X and near genes more highly expressed in testes than other tissues ($p = 1.2 \times 10^{-7}$ to 3.2×10^{-7} for Denisovan and 2.2×10^{-3} to 2.9×10^{-3} for Neanderthal ancestry even after controlling for differences in level of selective constraint across gene classes). This suggests that reduced male fertility may be a general feature of mixtures of human populations diverged by >500,000 years.

RESULTS AND DISCUSSION

The Date of Denisovan Admixture into the Ancestors of Oceanians

In order to detect segments of archaic ancestry in modern humans, it is important to know the size scale of these ancestry segments, which in turn reflects the number of generations during which recombination has the chance to break up these segments and thus the date of admixture. The date of Neanderthal admixture has been estimated [3–5], but there has been no study that has estimated the date of Denisovan admixture.

To estimate the date of Denisovan admixture, we exploited the fact that sites at which Denisovans carry a derived allele not seen in both the Altai Neanderthal and a large panel of sub-Saharan Africans are highly likely to derive from Denisovan introgression [6]. We similarly identify sites likely to derive from Neanderthal introgression. Within each class, we fit an exponential function to the decay of linkage disequilibrium with genetic distance. The inverse of the decay constant translates to the average date of gene flow in generations (Supplemental Experimental Procedures, "Estimating the date of archaic gene flow into Oceanian populations"). In Oceanians, the rate of decay at sites informative of Denisovan ancestry is estimated to be slower than the rate at sites informative of Neanderthal ancestry ($p = 0.00004$ for the null hypothesis of the two dates being equal, based on a two-sided test using block jackknife) (Figure 1). We caution that the nominal date estimate of 1,000 ± 8 generations ago for Denisovan admixture and 1,121 ± 16 generations ago for Neanderthal admixture are likely to be biased. One source of bias arises from the fact that errors in the genetic map can produce systematic underestimates of dates inferred based on linkage disequilibrium; however, as both types of ancestry have been decaying in the same population, this will bias the inference of Denisovan and Neanderthal admixture dates in exactly the same way, and hence will not contribute to an artifactual inference of one date being more recent than the other. A second source of bias can arise due to extreme demographic events since admixture [4]. We carried out coalescent simulations to explore this possibility and found that a slower linkage disequilibrium decay provides unambiguous evidence of more recent admixture, even taking into account biases arising due to plausible differences in the demographic history of Denisovans and Neanderthals (Supplemental Experimental Procedures, "Simulations"; Table S1). Further, we verified that the observation of a more recent date of Denisovan admixture is unchanged when we fit a two-pulse model of admixture (we obtained nominal date estimates of 986 and 21,808 generations for Denisovan admixture and 1,197 and 90,262 generations for Neanderthal admixture). To obtain a rough estimate of the date of Denisovan admixture—cognizant of the fact that for Oceanians we do not have the information needed to fully correct for uncertainty in the genetic map—we calibrated to previous

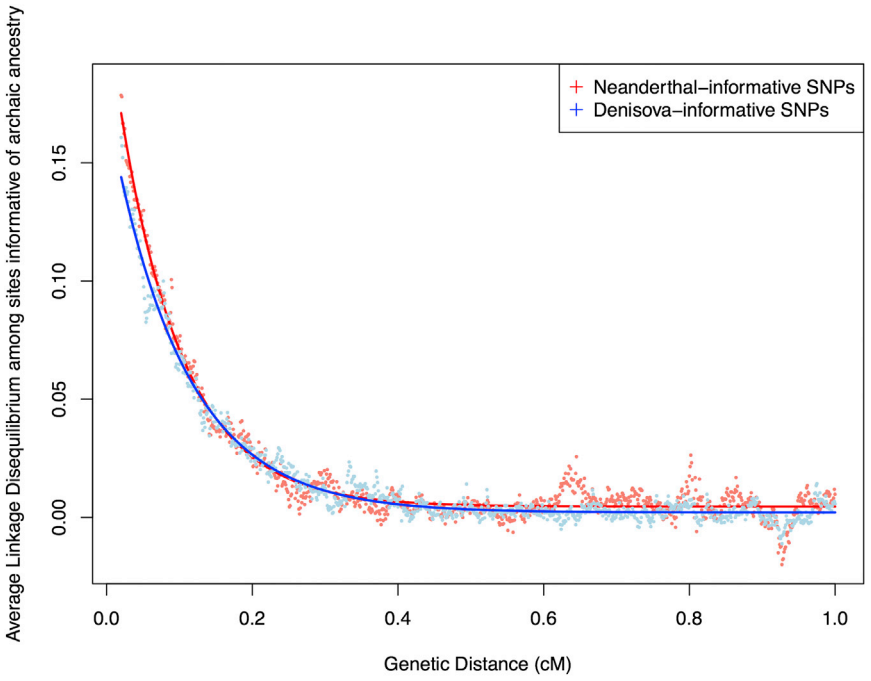


Figure 1. More Recent Date of Denisovan than Neanderthal Admixture

Average linkage disequilibrium (Lewontin's D) as a function of distance in Oceanians for SNPs informative of Neanderthal (red) and Denisovan (blue) ancestry. The Denisova decay is slower, implying a more recent date. See also Table S1.

estimates of the date of Neanderthal admixture, under the simplifying assumption that the date of Neanderthal admixture in the history of New Guineans is the same as the 50,000–60,000 years ago estimated for a radiocarbon-dated Upper Paleolithic Siberian [3]. Rescaling by 1,000/1,121, we estimate 44,000–54,000 years ago for Denisovan admixture.

Denisovan and Neanderthal Introgression Maps

To study the impact of the Denisovan and Neanderthal admixture events simultaneously, we developed methods that allow us to distinguish these two sources of archaic ancestry. We applied these methods to the Simons Genome Diversity Project (SGDP) dataset: 257 high-quality genomes from 120 non-African populations, including 20 Oceanian individuals from populations known to have high Denisovan admixture (unpublished data; Supplemental Experimental Procedures, "Data Processing").

For each individual, we inferred archaic ancestry segments across the autosomes (chromosomes 1–22) and chromosome X (our method did not allow us to test for archaic ancestry on chromosome Y because the archaic genomes are from females). Figure 2A plots the estimates of the proportion of confidently inferred Denisovan ancestry on a map, and Table 1 tabulates the results for six population pools (Table S2 tabulates the results for each population). Denisovan ancestry in Oceanians is greater than in other non-Africans [1] (Table 1). Both Neanderthal and Denisovan ancestry are greater in eastern non-Africans than in West Eurasians [6–10] (Supplemental Experimental Procedures, "Variation in the genome-wide proportions of archaic ancestry"; Table S3). We replicate previous findings of substantial Denisovan ancestry in New Guineans and Australians, as well as in populations that harbor admixtures of New Guinean ancestry [11]. However, we were surprised to detect a peak of Denisovan ancestry estimates in South Asians, both in the Himalayan region and in South and Central India (Figure 2A). The highest estimate is in Sherpas (0.10%), who have a Denisovan point estimate about one-tenth of that seen in Papuans (1.12%) (Table S3). Although this is notable in light of the likely Denisovan origin of the *EPAS1* allele that confers high-altitude adaptation in Tibetans [12, 13], *EPAS1* is not sufficient to explain the observation as Sherpas have the highest point estimate even without chromosome 2, on which *EPAS1* resides. To determine whether the peak of Denisovan ancestry in South Asia is significant, we tested whether the Denisovan ancestry proportion in diverse mainland Eurasians can be explained by differential proportions of non-West Eurasian ancestry (as it is already known that there is more Denisovan ancestry in East Eurasians than in West Eurasians [6]). For each Eurasian population X, we computed an allele frequency correlation statistic that is proportional to eastern non-African ancestry (Figure 2B; Supplemental Experimental Procedures, "Modeling the variation in Denisovan ancestry across populations"). We regressed the proportion of confidently inferred Denisovan ancestry against this statistic. Although the proportion of Denisovan ancestry in these populations is correlated with non-West Eurasian ancestry ($\rho_{Pearson} = 0.832$, block jackknife $p = 3.6 \times 10^{-10}$ for the correlation coefficient being non-zero), South Asian groups as a whole have significantly more Denisovan ancestry than expected (block jackknife Z score for residuals = 3.2, $p = 0.0013$ by a two-sided test for the null hypothesis that the Denisovan ancestry estimate in South Asians is predicted by their proportion of non-West Eurasian ancestry; Figure 2B; Supplemental Experimental Procedures, "Modeling the variation in Denisovan ancestry across populations"). The signal remains significant ($Z = 3.1$) when we remove from the analysis five populations that have ancestry very different from the majority of South Asians (Tibetan, Sherpa, Hazara, Kusunda, and Onge); however, the signals are non-significant for Central Asians ($Z = 1.2$) and Native Americans ($Z = 0.1$). Taken together, the evidence of Denisovan admixture in modern humans could in theory be explained by a single Denisovan introgression into modern humans, followed by dilution to different extents in Oceanians, South Asians, and East Asians by people with less Denisovan ancestry. If dilution does not explain these patterns, however, a minimum of three distinct Denisovan introgressions into the ancestors of modern humans must have occurred.

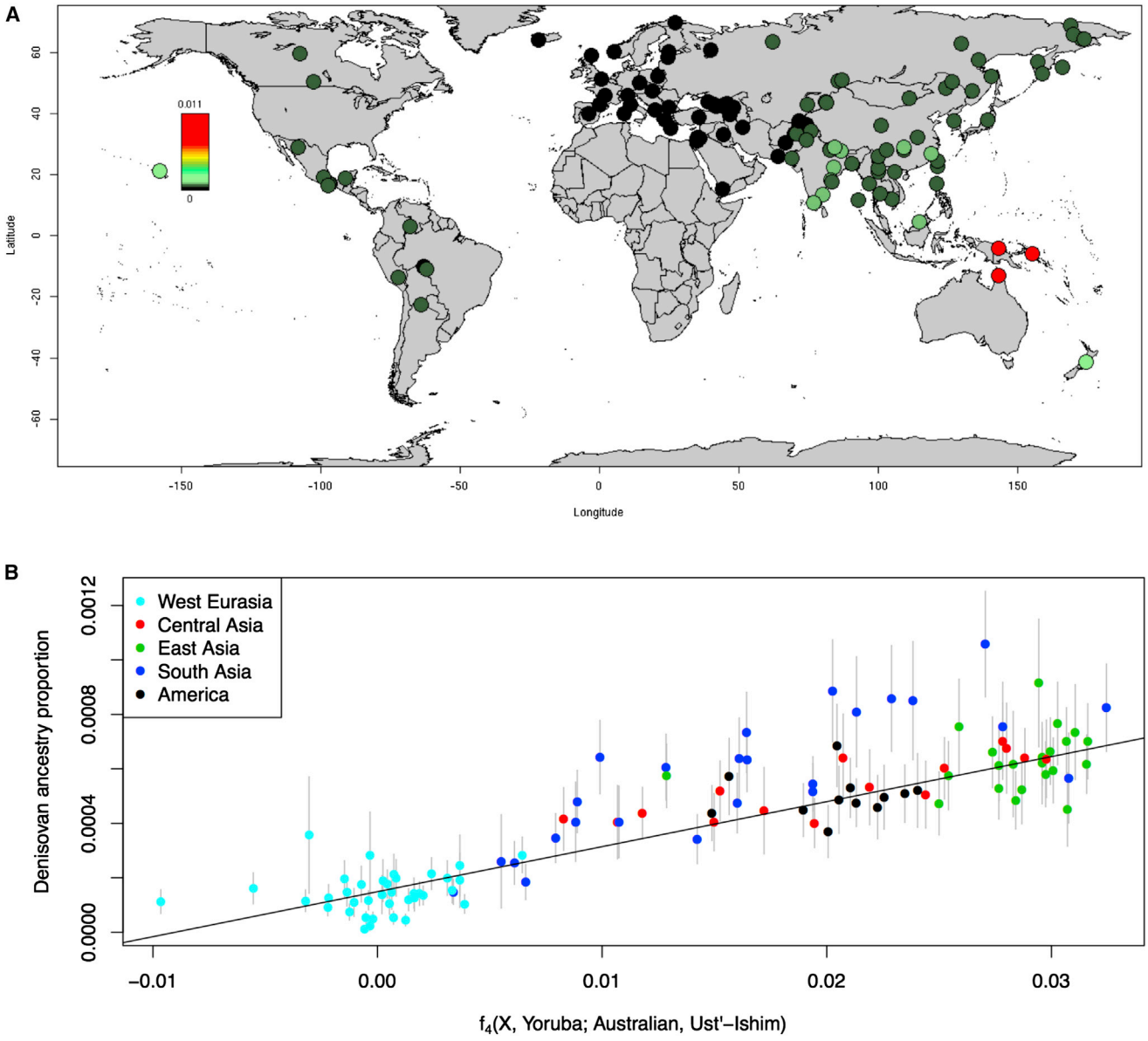


Figure 2. Variation in Denisovan Ancestry Proportion

(A) Proportion of the genome inferred to be Denisovan in ancestry in diverse non-Africans. The color scale is not linear to allow saturation of the high Denisova proportions in Oceania (bright red) and better visualization of the peak of Denisova proportion in South Asia.

(B) Proportion of the genome confidently inferred to be Denisovan in ancestry in mainland Eurasians plotted against the rate of allele sharing of each sample with non-West Eurasians as measured by an f_4 statistic. Error bars (1 SE) were obtained from a block jackknife. The Denisovan ancestry estimates in South Asians are systematically above expectation (fitted trend line) ($p = 0.0013$).

See also Table S3.

Tiling Path of Denisovan and Neanderthal Ancestry Inferred from Modern Genomes

The union of detected Denisovan haplotypes spans 257 Mb in Oceanians (Supplemental Experimental Procedures, "Coverage of archaic haplotypes"). The union of Neanderthal haplotypes spans 673 Mb over all non-Africans, which is smaller than the 1.1 Gb found in 1000 Genomes Project phase 1 data [14], most likely due to the fact that the total number of non-Africans genomes analyzed here is smaller. The positions of archaic ancestry are correlated across populations, with the strongest correlations at large spatial scales among the Neanderthal maps and weaker correlations between the Neanderthal and Denisovan maps (Figure 3B).

Regions with Elevated Proportions of Archaic Ancestry

We scanned all maps for windows with elevated proportions of archaic ancestry (average marginal probability $\geq$ 0.30 over a 100 kb window based on a published threshold [4]; Supplemental Experimental Procedures, "Genomic regions with elevated archaic ancestry"; Table S4). We identified 238

Table 1. Genome-wide Estimates of Archaic Ancestry

		Neanderthal Ancestry (%)		Denisovan Ancestry (%)	
Population	Individuals	Autosomes	X	Autosomes	X
America	29	1.37 ± 0.11	0.26 ± 0.18	0.05 ± 0.01	0.00 ± 0.00
Central Asia	27	1.40 ± 0.12	0.23 ± 0.18	0.05 ± 0.01	0.00 ± 0.00
East Asia	50	1.39 ± 0.11	0.32 ± 0.28	0.06 ± 0.02	0.00 ± 0.01
Oceania	26	1.54 ± 0.12	0.42 ± 0.36	0.85 ± 0.43	0.18 ± 0.17
South Asia	48	1.19 ± 0.11	0.40 ± 0.26	0.06 ± 0.03	0.01 ± 0.03
West Eurasia	77	1.06 ± 0.12	0.18 ± 0.19	0.02 ± 0.01	0.00 ± 0.00

We estimated the probability of Neanderthal and Denisovan ancestry for each phased genome in each population. We report the mean and SD of the proportion of confidently inferred archaic alleles (marginal probability >50%) across diploid individuals within each population. The highest point estimate of Neanderthal ancestry is in Oceania, and although this estimate is significantly higher than that in West Eurasia ($Z = 3.9$), consistent with previous reports [7, 8], it is not higher than that in East Asia ($Z = 0.7$). See also Table S2.

windows with elevated Neanderthal ancestry in a pool of all non-Africans and 48 with elevated Denisovan ancestry in Oceanians. Regions with elevated archaic ancestry may represent loci where archaic alleles have experienced positive selection, but a formal test is challenging due to the fact that archaic alleles, on average, do not evolve neutrally [14–17]. We also tested for sets of genes that have among the 5% highest archaic ancestry (hypergeometric test implemented in FUNC [18]; we report $p < 0.05$ after multiple testing correction; Supplemental Experimental Procedures, section S4). Genes involved in keratin filament formation related to skin and hair are enriched for Neanderthal ancestry, generalizing the results of previous analyses that were limited to Europeans and East Asians [14, 15]. Genes involved in phospholipid transporter activity related to fat metabolism and in trace-amine receptor activity related to detecting subtle scents are significantly enriched for Denisovan ancestry (Table S5).

Deserts of Archaic Ancestry

Some of the most striking features of the introgression maps are the archaic ancestry deserts: windows longer than 10 Mb at which the archaic ancestry proportion is <1/1000 (Figure 3A; Supplemental Experimental Procedures, "Analysis of genomic regions deficient in archaic ancestry"). We identified 18 Neanderthal ancestry deserts in a pool of all non-African individuals and 63 Denisovan deserts in Oceanians. Four windows (1:99–112 Mb, 3:78–90 Mb, 7:108–128 Mb, and 13:49–61 Mb) are both Neanderthal and Denisovan ancestry deserts. The desert on chromosome 7 contains the *FOXP2* gene, which has been hypothesized to have a role in enabling modern human speech and language [19] and has been identified as a desert in previous maps in Europeans and East Asians. Our finding that this region is also a desert of Denisovan ancestry strengthens the evidence that the modern human version of this gene may be critical for modern human biology [14, 15].

Archaic Ancestry Is Reduced in the Genomic Regions Most Constrained by Selection

We tested the relationship between archaic ancestry and regions of strong linked selection as measured by a B statistic [20]. Neanderthal ancestry decreases in proximity to functional elements in all populations ($\rho_{Spearman} = 0.25$–0.29; Figure 3C; Supplemental Experimental Procedures, "Correlation of archaic ancestry with B-statistics"; Table S6), as does Denisovan ancestry in Oceanians ($\rho_{Spearman} = 0.26$, Table S6), most likely reflecting greater selection against Neanderthal ancestry in low B statistic regions [14–17]. Power to detect archaic ancestry is elevated close to regions of linked selection due to a reduction in the rates of incomplete lineage sorting caused by the lower effective population size in these regions [14], so these results are not artifacts of reduced power. Thus, similar processes appear to have worked to remove Neanderthal and Denisovan ancestry near genes.

Archaic-Modern Admixture Was Most Likely Associated with Reduced Male Fertility

Our study provides new evidence in support of the hypothesis that reduced male fertility may be a common feature of admixture between human populations diverged by at least a half million years, a hypothesis that was previously suggested based on genetic patterns associated with the hybridization between Neanderthals and modern humans [14, 21]. We show that qualitatively similar signals are associated with Denisovan admixture.

One line of evidence for reduced fertility in male hybrids is that the proportion of archaic ancestry in modern humans is significantly reduced on chromosome X compared to the autosomes. This is suggestive of reduced male fertility as loci contributing to this phenotype are concentrated on chromosome X in hybrids of other species [22]. We confirm an extreme reduction of Neanderthal ancestry on chromosome X (16%–34% of the autosomes depending on the population) [14] and find a quantitatively similar reduction of Denisovan ancestry (21% of the autosomes in Oceanians) (Table 1).

The second line of evidence in support of the hypothesis of reduced fertility in hybrids is that there is a reduction of archaic ancestry in genes that are disproportionately expressed in testes, a known characteristic of male hybrid fertility [22]. To test for this signal in our data, we analyzed a set of genes having a significantly higher expression level in testes than any of 15 others tissues in an RNA sequencing dataset [23]. We detect a statistically significant depletion of Denisovan ($p = 1.21 \times 10^{-7}$ in Oceanians) and Neanderthal ($p = 2.1 \times 10^{-3}$ in Oceanians) ancestry in these genes relative to the genes in the other tissues (Table S7; Supplemental Experimental Procedures, "Association of Denisovan ancestry with tissue-specific expression"). We considered the possibility that these observations could be

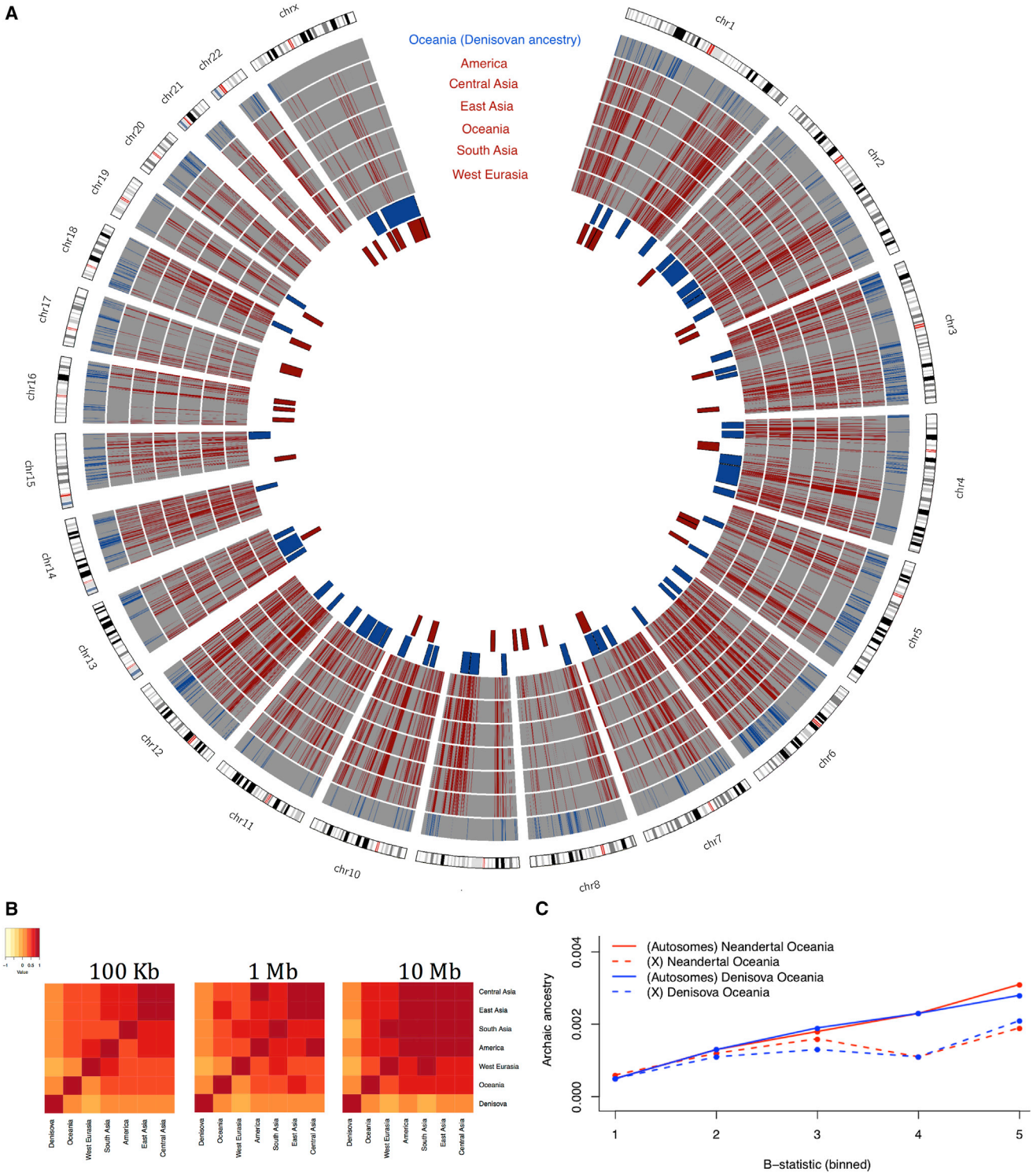


Figure 3. Fine-Scale Maps of Denisovan and Neanderthal Introgression

(A) Non-overlapping 100 kb windows that have non-zero inferred archaic ancestry in each of six populations (blue, Denisova; red, Neanderthal). In the innermost rings, we plot deserts (windows >10 Mb). See also Tables S4 and S5.

(B) Correlation of confidently inferred archaic ancestry (Neanderthal ancestry in six non-African populations and Denisovan ancestry in Oceanians) across populations in non-overlapping windows of size 100 kb, 1 Mb, and 10 Mb.

(C) We plot the median of the proportion of Denisovan and Neanderthal ancestry within quintiles of a B statistic measuring intensity of linked selection (low B indicates the regions most affected by linked selection). See also Tables S6 and S7.

explained by stronger linked selection at testes-expressed genes than at random places in the genome. However, when we correlate this pattern to B statistics (which are sensitive to linked selection [20]), we find that the gene sets that are disproportionally expressed in liver, heart, and skeletal muscle have even lower average B statistics than the genes most expressed in testes, and yet they do not show a depletion in archaic ancestry (Table S7). We also considered the possibility that the B statistic might not fully capture the degree of selective constraint at the genes disproportionately expressed in testes. However, when we use logistic regression to control for measures of selective constraint, we find that the significant reduction is observed not only when we control for B statistic at each gene ($p = 4.4 \times 10^{-7}$ for Denisovans; $p = 2.8 \times 10^{-3}$ for Neanderthals). It is also observed when we control for a direct estimate of the degree of selective constraint: the genetic diversity observed empirically at each gene in sub-Saharan Africans ($p = 3.2 \times 10^{-7}$ for Denisovans; $p = 2.9 \times 10^{-3}$ for Neanderthals; Supplemental Experimental Procedures, "Association of Denisovan ancestry with tissue-specific expression").

Conclusions

It has been suggested that the empirically observed reduction in Neanderthal ancestry in Europeans and East Asians near functionally important regions could be explained by a greater load of weakly deleterious alleles in Neanderthals due to the smaller population size of Neanderthals since separation, followed by purging of deleterious Neanderthal alleles in the mixed population [16, 17]. Since we have shown that similar patterns are associated with the Denisovan introgression event, it seems plausible that similar evolutionary forces operated to remove Denisovan ancestry segments. However, the model of a greater load of deleterious mutations in archaic humans cannot explain the observed reduction of both Neanderthal and Denisovan ancestry near genes that are disproportionately expressed in testes, suggesting that male hybrid sterility may have been associated with both introgressions. An important direction for future research is to understand the relative importance of purging of slightly deleterious alleles, as well as reduced fertility in hybrid males, in changing the content of genomes in the aftermath of the interbreeding that occurred between modern and archaic humans.

EXPERIMENTAL PROCEDURES

Inferring Segments of Archaic Ancestry

To determine the positions of archaic ancestry segments, we applied a machine-learning algorithm known as a conditional random field (CRF) [24]. The input data consists of the spatial distribution across the genome of derived alleles at sites informative about archaic ancestry (including the Neanderthal- and Denisova-informative sites). The CRF searches for runs of such alleles over the size scale expected for archaic introgression. The method we use is changed in important ways compared to the CRF previously implemented to solve the simpler problem of detecting Neanderthal ancestry (Supplemental Experimental Procedures, "An improved procedure for deconvolving Neanderthal and Denisovan ancestries") [14]. Not only do we use different classes of SNPs, but we also do not exploit the haplotype-based information used by the previously reported CRF, as we found that it leads to a bias in the inferred proportions of Denisovan ancestry in mainland Eurasians that have proportions of Denisovan ancestry of ~1/1000. This bias arises due to the previously described CRF being optimized for archaic admixture proportions of ~1/100; we found that this led erroneously to similar inferences of Denisovan ancestry in Han and French [6].

Assessing the Accuracy of Archaic Segment Inference

Discriminating between Neanderthal and Denisovan ancestral components in populations that have both, such as Oceanians, is challenging because Neanderthals and Denisovans are more similar to each other on average than either is to modern humans [1, 6]. To assess the accuracy of our inferences (Supplemental Experimental Procedures, "Empirical estimate of the accuracy of archaic ancestry estimates"; Figure S1), we devised a statistical procedure that uses previous genome-wide estimates of archaic ancestry to estimate the probability that the CRF infers Denisovan (or Neanderthal) ancestry when the true ancestry is Neanderthal, Denisovan, or modern human. For example, the inferred proportion of Neanderthal ancestry in African hunter-gatherers who most likely have negligible amounts allows us to estimate the rate of misclassification of modern human ancestry as Neanderthal. Similarly, the inferred proportion of Denisovan ancestry in West Eurasians who have negligible amounts allows us to estimate the rate of misclassification of non-Denisovan ancestry as Denisovan. This procedure enables us to estimate the false discovery rate (FDR) for an ancestry (the probability that segments assigned to a given ancestry are misclassified; Supplemental Experimental Procedures, "Empirical estimate of the accuracy of archaic ancestry estimates"; Figure S1). In Oceanians, at a threshold of 0.50 on the CRF probabilities, the FDR is 3% for Denisovan ancestry and 15% for Neanderthal ancestry. Our procedure also enables us to estimate the fraction of true archaic ancestry that we detect. We detect 24% of true Denisovan and 72% of true Neanderthal segments. The relatively low power to detect true Denisovan segments is likely to reflect the fact that the Siberian Denisovan genome that we use for detecting these segments is known to be deeply divergent from the Denisovan population that introgressed into the ancestors of Oceanians (much more divergent than the Altai Neanderthal genome is from the introgressing Neanderthal population) [6].

ACCESSION NUMBERS

The raw data analyzed for 277 of the samples analyzed here are available through the EBI European Nucleotide Archive under accession numbers EBI-ENA: PRJEB9586 and ERP010710. A version of the genotype data we analyzed that is small enough to download by FTP is available at http://reich.hms.harvard.edu/pub/datasets/sgdp/. The remaining 23 samples are only available to researchers who provide a signed letter affirming that they will abide by specific restrictions for using the samples, and they can be accessed by writing to D.R.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, one figure, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.03.037.

AUTHOR CONTRIBUTIONS

S.S. and S.M. performed analyses. N.P. and D.R. supervised the study. S.S. and D.R. wrote the manuscript.

ACKNOWLEDGMENTS

D.R. was supported by NIH grant GM100233 and by NSF grant HO BCS-1032255 and is a Howard Hughes Medical Institute investigator. S.S. was supported in part by NIH grant 5K99GM111744-02 and 4R00GM111744-03.

Received: February 26, 2016
Revised: March 10, 2016
Accepted: March 17, 2016
Published: March 28, 2016

REFERENCES

1. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.

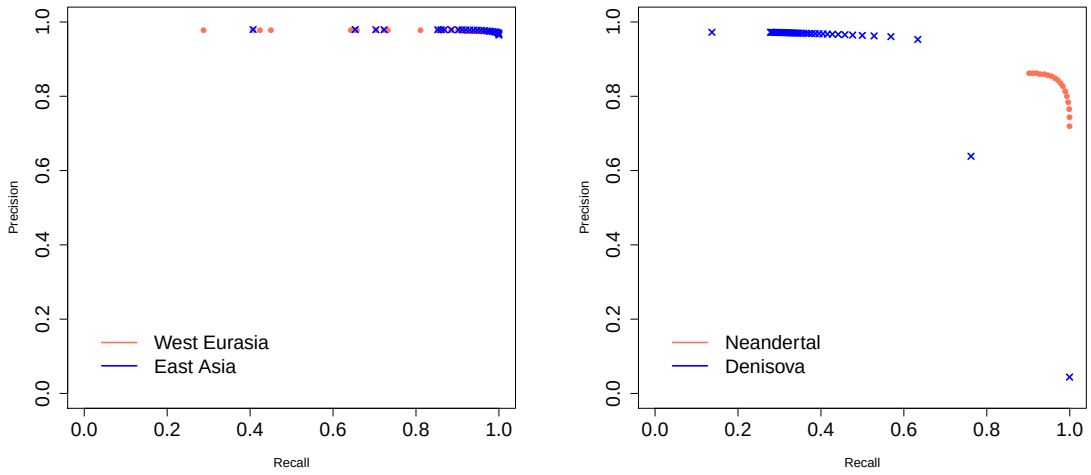
2. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.
3. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.
4. Sankararaman, S., Patterson, N., Li, H., Pääbo, S., and Reich, D. (2012). The date of interbreeding between Neandertals and modern humans. PLoS Genet. *8*, e1002947.
5. Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., et al. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science *346*, 1113–1118.
6. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.
7. Wall, J.D., Yang, M.A., Jay, F., Kim, S.K., Durand, E.Y., Stevison, L.S., Gignoux, C., Woerner, A., Hammer, M.F., and Slatkin, M. (2013). Higher levels of neanderthal ancestry in East Asians than in Europeans. Genetics *194*, 199–209.
8. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.
9. Vernot, B., and Akey, J.M. (2015). Complex history of admixture between modern humans and Neandertals. Am. J. Hum. Genet. *96*, 448–453.
10. Skoglund, P., and Jakobsson, M. (2011). Archaic human ancestry in East Asia. Proc. Natl. Acad. Sci. USA *108*, 18301–18306.
11. Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M.R., Pugach, I., Ko, A.M., Ko, Y.C., Jinam, T.A., Phipps, M.E., et al. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am. J. Hum. Genet. *89*, 516–528.
12. Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature *512*, 194–197.
13. Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D.B., Pritchard, J.K., Beall, C.M., and Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. Nat. Commun. *5*, 3281.
14. Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. Nature *507*, 354–357.
15. Vernot, B., and Akey, J.M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. Science *343*, 1017–1021.
16. Juric, I., Aeschbacher, S., and Coop, G. (2015). The strength of selection against Neanderthal introgression. BioRxiv, doi: http://dx.doi.org/10.1101/030148.
17. Harris, K., and Nielsen, R. (2015). The genetic cost of Neanderthal introgression. BioRxiv, doi: http://dx.doi.org/10.1101/030148.
18. Prüfer, K., Muetzel, B., Do, H.H., Weiss, G., Khaitovich, P., Rahm, E., Pääbo, S., Lachmann, M., and Enard, W. (2007). FUNC: a package for detecting significant associations between gene sets and ontological annotations. BMC Bioinformatics *8*, 41.
19. Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519–523.
20. McVicker, G., Gordon, D., Davis, C., and Green, P. (2009). Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. *5*, e1000471.
21. Currat, M., and Excoffier, L. (2011). Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. Proc. Natl. Acad. Sci. USA *108*, 15129–15134.
22. Presgraves, D.C. (2008). Sex chromosomes and speciation in Drosophila. Trends Genet. *24*, 336–343.
23. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.
24. Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, C.E. Brodley, and A.P. Danyluk, eds. (Morgan Kaufmann Publishers), pp. 282–289.

Current Biology, Volume *26*

Supplemental Information

The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans

Sriram Sankararaman, Swapan Mallick, Nick Patterson, and David Reich



(a) **(b)**

Figure S1: Empirical precision-recall curves for archaic local ancestry inference (related to main text Experimental Procedures). (a) Empirical precision-recall curve of the modified method for inferring Neanderthal local ancestry in West Eurasian and East Asian populations. (b) Empirical precision-recall curve of the modified method for inferring Neanderthal and Denisovan local ancestry in Oceanian (Australians, Papuans and Bougainville Islanders) populations. The method is a modification of the previously proposed CRF [S1] to improve the ability to deconvolve the contributions of Neanderthal and Denisovan ancestries (described in Section).

Denisovan date	Neanderthal date	Estimated Denisovan date	Estimated Neanderthal date	Neanderthal-Denisovan dates (Z-score)
1500	2000	1579.8±25.7	1924.1±40.5	7.2
2000	1500	2069.8±41.1	1555.9±30.8	-9.8
1800	2000	1881.0±33.0	1882.5±39.5	0
2000	1800	2106.1±43.0	1826.9±37.6	-4.8
1900	2000	1939.3±38.0	1924.4±45.9	-0.3
2000	1900	2018.5±39.2	1850.5±40.2	-3.1
2000	2200	1914.4±39.3	1943.1±34.6	0.5
2200	2000	2225.7±55.1	1773.9±30.37	-7.2

Table S1: Evaluation of Neanderthal and Denisovan admixture date estimates on simulated data (related to Figure 1). For every setting of the true Neanderthal and Denisovan dates, we show the Block Jackknife corrected point estimate and standard errors as well as the block Jackknife Z-score for the difference in the estimates. The top rows correspond to a simple demographic model while the bottom two rows correspond to data simulated under a demographic model based on the model used in [S2]

Populations	Neanderthal ancestry		Denisovan ancestry	
	A (%)	X (%)	A (%)	X (%)
Abkhasian	0.976	0.100	0.011	0.000
Adygei	1.126	0.119	0.020	0.000
Albanian	1.203	0.334	0.019	0.000
Aleut	1.357	0.368	0.044	0.000
Altaian	1.413	0.445	0.064	0.000
Ami	1.440	0.183	0.047	0.019
Armenian	1.077	0.121	0.013	0.000
Atayal	1.531	0.785	0.062	0.000
Australian	1.559	0.300	0.895	0.105
Balochi	1.070	1.046	0.026	0.000
Basque	1.100	0.098	0.011	0.000
BedouinB	0.858	0.386	0.007	0.000
Bengali	1.261	0.268	0.063	0.000
Bergamo	1.134	0.015	0.020	0.000
Bougainville	1.622	1.375	0.861	0.032
Brahmin	1.101	0.635	0.064	0.000
Brahui	1.099	0.261	0.018	0.000
Bulgarian	1.078	0.250	0.005	0.000
Burmese	1.334	0.427	0.057	0.000
Burusho	1.272	0.200	0.035	0.061
Cambodian	1.419	0.538	0.075	0.000
Chane	1.338	0.761	0.042	0.000
Chechen	1.019	0.000	0.025	0.000
Chipewyan	1.633	0.384	0.049	0.000
Chukchi	1.228	0.161	0.040	0.000
Cree	1.260	0.126	0.057	0.000
Crete	0.993	0.187	0.014	0.000
Czech	1.067	0.000	0.028	0.000
Dai	1.314	0.211	0.064	0.014
Daur	1.359	0.475	0.067	0.010
Druze	0.965	0.186	0.011	0.000
Dusun	1.438	0.312	0.086	0.000
English	1.085	0.210	0.015	0.000
Eskimo Chaplin	1.500	0.000	0.053	0.000
Eskimo Naukan	1.401	0.408	0.060	0.000
Eskimo Sireniki	1.491	0.265	0.051	0.000
Estonian	1.076	0.167	0.021	0.000
Even	1.411	0.229	0.064	0.000
Finnish	1.165	0.302	0.013	0.000
French	1.023	0.188	0.012	0.000
Georgian	1.134	0.000	0.012	0.000
Greek	0.975	0.579	0.005	0.000
Han	1.495	0.144	0.062	0.005
Hawaiian	1.342	0.184	0.117	0.000
Hazara	1.225	0.324	0.034	0.000
Hezhen	1.399	0.277	0.053	0.000
Hungarian	1.122	0.057	0.019	0.000
Icelandic	1.237	0.147	0.015	0.000
Igorot	1.399	0.503	0.048	0.000
Iranian	0.968	0.351	0.022	0.000
Iraqi Jew	0.926	0.231	0.020	0.000
Irula	1.199	0.212	0.089	0.000
Itelman	1.428	0.042	0.045	0.000
Japanese	1.308	0.444	0.058	0.000
Jordanian	0.810	0.282	0.005	0.000
Kalash	1.113	0.409	0.025	0.000
Kapu	1.069	0.705	0.055	0.000
Karitiana	1.374	0.120	0.037	0.000
Kashmiri Pandit	1.175	0.235	0.041	0.000
Kharia	1.133	0.380	0.085	0.000
Khonda Dora	1.207	0.157	0.086	0.000
Kinh	1.448	0.433	0.052	0.000
Korean	1.457	0.539	0.062	0.000
Kurumba	1.313	0.751	0.081	0.000
Kusunda	1.256	0.581	0.075	0.061
Kyrgyz	1.306	0.101	0.040	0.000
Lahu	1.358	0.075	0.061	0.000
Lezgin	1.125	0.338	0.014	0.019
Madiga	1.126	0.795	0.073	0.000
Makrani	1.041	0.141	0.015	0.000
Mala	1.127	0.527	0.052	0.000
Mansi	1.311	0.091	0.040	0.000
Maori	1.252	0.000	0.136	0.000
Mayan	1.386	0.183	0.069	0.000
Miao	1.341	0.151	0.073	0.000
Mixe	1.342	0.222	0.048	0.000
Mixtec	1.252	0.414	0.044	0.000
Mongola	1.389	0.346	0.068	0.000
Nahua	1.332	0.263	0.046	0.000
Naxi	1.371	0.106	0.070	0.000
North Ossetian	1.079	0.226	0.013	0.000
Norwegian	1.157	0.297	0.001	0.000
Onge	1.325	0.533	0.057	0.000
Orcadian	1.132	0.077	0.004	0.000
Oroqen	1.399	0.540	0.059	0.000
Palestinian	0.909	0.074	0.010	0.000
Papuan	1.596	0.366	1.123	0.269
Pathan	1.097	0.469	0.041	0.000
Piapoco	1.318	0.236	0.053	0.000
Pima	1.437	0.266	0.052	0.000
Polish	1.086	0.240	0.036	0.000
Punjabi	1.156	0.156	0.061	0.000
Quechua	1.361	0.333	0.045	0.000
Relli	1.190	0.572	0.064	0.019
Russian	1.148	0.243	0.018	0.000
Saami	1.363	0.000	0.028	0.000
Samaritan	0.888	0.000	0.002	0.000
Sardinian	1.133	0.200	0.009	0.000
She	1.468	0.224	0.077	0.000
Sherpa	1.395	0.250	0.106	0.000
Sindhi	1.174	0.188	0.048	0.022
Spanish	1.031	0.130	0.018	0.000
Surui	1.446	0.011	0.050	0.000
Tajik	1.064	0.068	0.016	0.000
Thai	1.458	0.584	0.048	0.000
Tibetan	1.389	0.169	0.082	0.010
Tlingit	1.261	0.211	0.042	0.000
Tu	1.466	0.232	0.045	0.000
Tubalar	1.391	0.261	0.052	0.000
Tujia	1.430	0.266	0.092	0.010
Turkish	1.024	0.226	0.014	0.000
Tuscan	1.151	0.131	0.016	0.000
Ulchi	1.508	0.177	0.064	0.000
Uygur	1.170	0.398	0.057	0.019
Xibo	1.437	0.438	0.066	0.000
Yadava	1.157	0.469	0.047	0.000
Yakut	1.525	0.155	0.070	0.000
Yemenite Jew	0.947	0.277	0.012	0.000
Yi	1.387	0.036	0.070	0.000
Zapotec	1.360	0.329	0.051	0.000

Table S2: Summary of proportion of the genome confidently inferred to be archaic in ancestry (related to Table 1). Archaic ancestry estimates refer to the fraction of SNPs which have a marginal probability of either Neanderthal or Denisovan ancestry > 0.50. The fraction of Neanderthal ancestry in individual i is estimated by the statistic $tia^{(n)}(i)$ while the fraction of Denisovan ancestry in individual i is estimated by $tia^{(d)}(i)$ (see Equation 1 in Section). We report the mean across individuals within each population and use a threshold of 0.50. A and X refer to estimates across the autosomes and X chromosome respectively.

Population	Region	*tia*	Z-score
Dai	East Asia	0.000642887	4.552016
Daur	East Asia	0.000665048	4.483456
Han	East Asia	0.000617295	4.39127
Japanese	East Asia	0.000580133	4.473346
Naxi	East Asia	0.000701322	4.644555
She	East Asia	0.000766465	4.230774
Xibo	East Asia	0.000660518	4.16324
Yi	East Asia	0.000701042	4.120488
Bengali	South Asia	0.000634354	4.110124
Sherpa	South Asia	0.00105782	4.737869
Tibetan	South Asia	0.000824393	4.136213
Eskimo_Naukan	Central Asia	0.000601988	4.080411
Even	Central Asia	0.000641474	4.61232
Australian	Oceania	0.00894954	9.019621
Bougainville	Oceania	0.0086141	8.616564
Hawaiian	Oceania	0.00117403	4.052288
Maori	Oceania	0.00136216	4.79091
Papuan	Oceania	0.0112295	10.79779

Table S3: Populations with a higher proportion of the genome confidently inferred to be Denisovan compared to French (related to Figure 2). We report populations where the difference in the confidently inferred proportion of Denisovan ancestry (*tia*) is statistically significant (Z-score$>$ 4).

Table S4: Regions of elevated archaic ancestry proportion in American populations (related to Figure 3A). We report 100 kb non-overlapping windows with average marginal probability of archaic ancestry: $la \geq 0.30$). We report windows of elevated Neanderthal ancestry in Americans, Central Asians, East Asians, Oceanians, South Asians, and West Eurasians as well as windows of elevated Denisovan ancestry in Oceanians.
This table is provided as an Excel file.

GO term	America	Central Asia	East Asia	South Asia	West Eurasia	Oceania	Denisova
cellular response to cadmium ion		0.001	0.001	0.001			
cellular response to inorganic substance		0.001	0.001	0.012			
cellular response to metal ion		0.001	0.001	0.002			
cellular response to zinc ion		0.001	0.001	0.001			
chemokine-mediated signaling pathway				0.004			
cytokine production involved in inflammatory response					0.048		
glycosphingolipid metabolic process						0.002	
positive regulation of keratinocyte proliferation	0.012						
regulation of cytokine production involved in inflammatory response					0.023		
response to cadmium ion			0.017	0.001			
extracellular region					0.026		
intermediate filament	0.001	0.001					
intermediate filament cytoskeleton	0.001	0.012					
intracellular						0.005	
intracellular membrane-bounded organelle						0.015	
intracellular organelle						0.012	
intracellular part						0.008	
invadopodium						0.023	
keratin filament	0.001	0.001	0.001	0.001	0.036		
membrane-bounded organelle						0.023	
organelle						0.015	
cadmium ion binding			0.032	0.03			
C-C chemokine receptor activity				0.001			
chemokine receptor activity				0.001			
cytokine receptor activity				0.001			
G-protein coupled chemoattractant receptor activity				0.001			
phospholipid transporter activity							0.044
trace-amine receptor activity							0.001

Table S5: Gene Ontology categories with elevated archaic ancestry (related to Figure 3A). We list GO-categories that have significantly elevated archaic ancestry (FWER p-value < 0.05) grouped by biological process, cellular component and molecular function. We list categories that are significantly enriched for Neanderthal ancestry in each of six non-African groups as well as categories that are enriched for Denisovan ancestry in Oceanians (Papuans, Australians and Bougainville Islanders).

	Population		la		$ta_{0.25}$		$ta_{0.9}$
		ρ (se)	$-log_{10}(pval)$	ρ (se)	$-log_{10}(pval)$	ρ (se)	$-log_{10}(pval)$
Autosomes	East Asians	0.255 (0.0196)	38.063	0.0408 (0.0193)	1.461	-0.0294 (0.0181)	0.980
X	East Asians	0.379 (0.0688)	7.436	0.159 (0.0931)	1.061	0.146 (0.0933)	0.926
Autosomes	Oceanians	0.29 (0.0196)	48.752	0.0157 (0.0202)	0.361	-0.0234 (0.0192)	0.648
X	Oceanians	0.284 (0.0964)	2.499	-0.0193 (0.156)	0.045	-0.113 (0.143)	0.369
Autosomes	West Eurasians	0.252 (0.0169)	49.162	0.0448 (0.0162)	2.246	-0.0306 (0.0148)	1.419
X	West Eurasians	0.38 (0.079)	5.827	0.178 (0.0517)	3.239	0.138 (0.0509)	2.167
Autosomes	Oceanians	0.263 (0.018)	47.764	-0.029 (0.017)	1.053	-0.0647 (0.014)	5.405
X	Oceanians	0.333 (0.0906)	3.630	0.125 (0.193)	0.286	0.125 (0.173)	0.326

Table S6: Relationship between archaic ancestry and B-statistic (related to Figure 3C). On top, relationship between Neanderthal ancestry and B-statistic for West Eurasians, East Asians and Oceanians (Australians, Papuans and Bougainville Islanders). On bottom, relationship between Denisovan ancestry and B-statistic in Oceanians (Australians, Papuans and Bougainville Islanders) on the autosomes and the X chromosome. ρ refers to Spearman's correlation coefficient, la, $ta_{0.9}$ and $ta_{0.25}$ refer to different summaries of archaic ancestry. We show results on autosomes and X chromosome.

Tissue	Uncorrected			B-statistic						Heterozygosity					
	A+X	A	X	A+X		A		X		A+X		A		X	
				Mean $\times 10^{-4}$	p	Mean $\times 10^{-4}$	p	Mean $\times 10^{-4}$	p	Mean $\times 10^{-4}$	p	Mean $\times 10^{-4}$	p	Mean $\times 10^{-4}$	p
Adipose	0.8	0.75	0.98	724	0.67	728	0.61	738	0.92	11	0.73	11	0.69	10	0.89
Adrenal	0.74	0.69	1	736	0.56	732	0.53	NA	NA	12	0.66	11	0.61	NA	NA
Blood	0.93	0.92	0.34	722	0.77	722	0.75	679	0.5	9.7	0.92	9.6	0.91	10	0.5
Brain	1	1	0.2	685	1	692	1	685	0.22	9.1	1	9.1	1	10	0.14
Breast	0.53	0.47	0.96	728	0.35	733	0.32	644	0.78	12	0.43	12	0.38	9.3	0.86
Colon	0.28	0.28	0.8	695	0.23	702	0.22	718	0.5	9.9	0.24	9.9	0.23	10	0.5
Heart	0.95	0.94	0.76	668	0.96	668	0.95	725	0.5	9.6	0.95	9.5	0.94	9.6	0.49
Kidney	0.71	0.7	0.64	677	0.7	677	0.67	667	0.5	10	0.67	10	0.66	9.4	0.5
Liver	0.39	0.24	0.98	662	0.51	660	0.36	673	0.96	9.9	0.36	9.9	0.22	9.6	0.96
Lung	0.82	0.77	0.91	723	0.65	715	0.6	659	0.8	12	0.74	12	0.69	9.6	0.75
Lymph	0.91	0.91	0.9	675	0.9	665	0.88	753	0.5	9.9	0.89	9.6	0.88	11	0.5
Ovary	0.13	0.12	0.9	755	0.05	736	0.049	642	0.5	10	0.1	9.8	0.095	8.4	0.5
Prostate	0.64	0.59	0.95	754	0.41	756	0.38	778	0.5	10	0.58	11	0.53	9.4	0.5
Skeletal muscle	0.56	0.44	0.95	600	0.91	600	0.84	672	0.94	9	0.56	8.9	0.44	9.7	0.89
Testes	**1.2e-07**	**6e-06**	0.46	673	**4.4e-07**	672	**1.7e-05**	682	0.046	9.2	**3.2e-07**	9.4	**9.6e-06**	9.6	0.3
Thyroid	0.73	0.7	0.9	724	0.59	718	0.56	657	0.5	10	0.68	10	0.65	11	0.5
Adipose	0.26	0.17	0.93	724	0.18	728	0.11	738	0.84	11	0.19	11	0.12	10	0.79
Adrenal	0.85	0.86	1	736	0.77	732	0.78	NA	NA	12	0.78	11	0.8	NA	NA
Blood	0.058	0.037	0.91	722	0.029	722	0.015	679	0.88	9.7	0.056	9.6	0.037	10	0.87
Brain	0.97	0.94	0.98	685	0.97	692	0.92	685	0.98	9.1	0.98	9.1	0.95	10	0.98
Breast	0.65	0.68	0.62	728	0.54	733	0.56	644	0.37	12	0.52	12	0.56	9.3	0.34
Colon	0.91	0.87	0.97	695	0.89	702	0.83	718	0.91	9.9	0.89	9.9	0.84	10	0.89
Heart	0.83	0.88	0.3	668	0.82	668	0.88	725	0.49	9.6	0.81	9.5	0.87	9.6	0.49
Kidney	0.79	0.83	0.46	677	0.77	677	0.81	667	0.29	10	0.75	10	0.8	9.4	0.29
Liver	0.72	0.7	0.83	662	0.76	660	0.76	673	0.76	9.9	0.7	9.9	0.68	9.6	0.75
Lung	0.89	0.87	0.85	723	0.82	715	0.79	659	0.75	12	0.81	12	0.79	9.6	0.76
Lymph	0.96	0.97	0.62	675	0.95	665	0.97	753	0.49	9.9	0.95	9.6	0.97	11	0.49
Ovary	0.15	0.18	0.62	755	0.089	736	0.11	642	0.49	10	0.12	9.8	0.14	8.4	0.49
Prostate	0.24	0.26	0.78	754	0.15	756	0.17	778	0.49	10	0.19	11	0.22	9.4	0.49
Skeletal muscle	0.68	0.68	0.71	600	0.83	600	0.86	672	0.6	9	0.69	8.9	0.69	9.7	0.59
Testes	**2.2e-03**	6.8e-03	0.015	673	**2.8e-03**	672	8.8e-03	682	4e-03	9.2	**2.9e-03**	9.4	7.8e-03	9.6	6.7e-03
Thyroid	0.37	0.42	0.62	724	0.27	718	0.31	657	0.49	10	0.3	10	0.35	11	0.49
Adipose	0.088	0.053	0.93	724	0.025	728	0.011	738	0.84	11	0.039	11	0.024	10	0.8
Adrenal	0.6	0.51	1	736	0.38	732	0.32	0	0	12	0.4	11	0.34	0	0
Blood	0.93	0.84	0.87	722	0.63	722	0.43	679	0.86	9.7	0.92	9.6	0.83	10	0.82
Brain	1	1	0.98	685	1	692	0.99	685	0.98	9.1	1	9.1	1	10	0.98
Breast	0.85	0.82	0.85	728	0.66	733	0.64	644	0.71	12	0.67	12	0.67	9.3	0.68
Colon	0.27	0.15	0.98	695	0.18	702	0.082	718	0.93	9.9	0.22	9.9	0.12	10	0.91
Heart	0.6	0.55	0.38	668	0.66	668	0.61	725	0.24	9.6	0.57	9.5	0.53	9.6	0.24
Kidney	0.93	0.97	0.022	677	0.93	677	0.96	667	0.49	10	0.89	10	0.94	9.4	0.49
Liver	0.83	0.71	0.55	662	0.94	660	0.9	673	0.46	9.9	0.74	9.9	0.64	9.6	0.44
Lung	0.88	0.8	0.87	723	0.64	715	0.53	659	0.79	12	0.84	12	0.76	9.6	0.8
Lymph	0.63	0.67	0.39	675	0.62	665	0.63	753	0.49	9.9	0.57	9.6	0.62	11	0.49
Ovary	0.37	0.3	0.86	755	0.13	736	0.092	642	0.65	10	0.3	9.8	0.24	8.4	0.64
Prostate	0.83	0.79	0.62	754	0.55	756	0.52	778	0.49	10	0.76	11	0.73	9.4	0.49
Skeletal muscle	0.58	0.38	0.94	600	0.99	600	0.96	672	0.92	9	0.68	8.9	0.47	9.7	0.91
Testes	**3.6e-09**	**3.1e-05**	0.0086	673	**6.7e-09**	672	**7.7e-05**	682	**0.0011**	9.2	**1.9e-08**	9.4	**5e-05**	9.6	0.004
Thyroid	0.54	0.43	0.86	724	0.32	718	0.24	657	0.66	10	0.42	10	0.35	11	0.63

Table S7: Enrichment of tissue-expressed genes in regions of the genome depleted in Denisovan ancestry (top), Neanderthal ancestry (middle) in Oceanians populations and Neanderthal ancestry in mainland Eurasians (bottom) (related to Figure 3C). We compare tissue-expressed genes (defined as genes that are more significantly expressed in a given tissue compared to all other tissues) to all genes that are specific to at least one tissue. We report the one-sided P-value for Fisher's exact test for the genes on the autosomes, X chromosomes and for the combined set across autosomes and X chromosomes. Only testes-expressed genes remain statistically significantly enriched in regions with low Denisovan ancestry after correcting for 16 tests in each case (highlighted). We also repeated this analysis correcting for the B-statistic and for the local heterozygosity from a panel of Africans. We report the one-sided value of a test of the coefficient associated with a gene being present in a given tissue in a logistic regression of the depletion status of a genes that also included as a covariate the B-statistic or the local heterozygosity. Local heterozygosity for each gene is calculated across 76 African chromosomes, restricting to sites which pass filter level >=1, and to sites where at least half the samples have a valid genotyping call. Samples from panel A were excluded as the error rate is known to be higher. Only testes-expressed genes show a statistically significant enrichment in regions with low archaic ancestry. Oceanian populations refer to Papuans, Australians and Bougainville Islanders. A - autosomes, X - X chromosomes, A+X - combined autosomes and X chromosome, Mean - mean of the B-statistic or the heterozygosity across the class of genes examined, p - P-value for Fisher's exact test. We note that adipose-expressed and blood-expressed genes appear to be nominally depleted for Neanderthal ancestry in mainland Eurasians and Oceanians respectively though the corresponding P-values are not significant after multiple testing correction.

Supplemental Experimental Procedures

Estimating the date of archaic gene flow into Oceanian populations

As a first step towards understand the history of Denisovan gene flow into the Oceanian populations, we need to infer the date of this gene flow event (or more precisely, the date of last exchange of genes between the ancestral populations). To do this, we will measure the extent of admixture linkage disequilibrium (LD) (such a statistic was used to estimate Neanderthal gene flow in Europeans [S3]). A limiting factor in estimating accurate dates of admixture events that are more than thousands of years old is the accuracy of the genetic maps used. To estimate dates accurately, [S3] developed a procedure to correct the nominal LD-based dates using estimates of the error of genetic maps. In turn, the errors in a given genetic map were estimated by comparing the map to crossovers observed in a European pedigree [S4]. Alternately, the error in the map could be assessed within the statistical framework used to estimate map. However, this procedure limits the applicability of LD-based admixture date estimation as it requires access to both a population-specific map as well as an estimate of the error associated with the map.

Rather than attempt to estimate absolute dates (which requires us to characterize the errors in the genetic maps), we attempt to obtain a relative ordering of Neanderthal and Denisovan admixture events. Given that Oceanian populations have a history of gene flow from populations related to Neanderthals as well as Denisovans, we can ask if Denisovan gene flow event pre or post-dated the Neanderthal gene flow event (we use the term gene flow to refer to the date of last exchange of genes – it is possible and quite likely that there were multiple episodes of gene flow or a period of continuous gene flow between two populations). Since we are estimating the date of gene flow in the same population, it is meaningful to compare these dates.

Our procedure for dating gene flow in a target population begins by ascertaining a set of SNPs. For all pairs of ascertained SNPs at a given genetic distance x, we compute $C(x)$ defined to be the average of Lewontin's D in the target population. We then fit an exponential function to $C(x)$ as a function of x using ordinary least squares for x in the range of 0.02 cM to 1 cM and use the rate of decay as an estimate of the time of gene flow. To estimate standard errors of this estimate, we use a weighted Block Jackknife [S5] with 10 Mb blocks having a minimum of 100 SNPs.

To estimate the date of Neanderthal gene flow in a test population, we ascertain SNPs at which a single randomly chosen Neanderthal allele (from a diploid Neanderthal genome) is derived relative to the human-chimp ancestor, a single randomly chosen Denisovan allele is ancestral, all alleles in a panel of sub-Saharan Africans are ancestral and that are polymorphic in the test population. We term this ascertainment $nd10$. For the Neanderthal and Denisovan alleles, we use the diploid genotypes from the high-coverage Altai Neanderthal genome [S6] and Denisovan genome [S7] respectively. For the sub-Saharan Africans, we use a panel of 44 high-coverage genomes sequenced as part of the Simons Genome Diversity Project (SGDP) [S8] that we determined are closely related to the Yoruba relative to Altai Neanderthal (see for details on processing of SGDP data). More precisely, we included all populations such that the Z-score of the D-statistic $D(X, Yoruba; Neanderthal, Chimp)$ is less than 2, where X is one of n African populations sequenced in the SGDP. We computed this D-statistic restricting to transversions.

To estimate the date of Denisovan gene flow, we ascertain SNPs at which Denisova is derived and Neanderthal is ancestral and all sub-Saharan Africans are ancestral ($nd01$). For the genetic map, we used the combined Oxford LD-based map [S9].

Our test panel consists of individuals from Papua New Guinea, Aboriginal Australians and Bougainville islanders (16 Papuans, 2 Australian Aborigines and 2 Bougainville islanders).

We estimate the nominal time of gene flow λ in Oceanians as $\lambda = 1121 \pm 16$ for Neanderthal gene flow and $\lambda = 1000 \pm 8$ for Denisovan gene flow. Thus, the nominal date of last exchange of genes

between Denisovans and Oceanians postdates the corresponding date for Neanderthals and Oceanians (Block Jackknife two-sided P-value 4.3×10^{-5}). This date is consistent with a model in which Denisovan gene flow occurred after the divergence of these populations from other Eurasian populations.

It is plausible that there were multiple introgression events associated with either archaic so that a single pulse of admixture is not a good model. To test this, we fit a model that is a mixture of two exponentials. For Neanderthal gene flow, we estimate nominal admixture dates of $\lambda_1 = 1197, \lambda_2 = 90262$. For Denisovan gene flow, we estimate $\lambda_1 = 986, \lambda_2 = 21808$. Thus, our estimates are relatively insensitive to the assumption of one vs two pulses of admixture. Further, in both the Neanderthal and Denisovan gene flow events, the older date is substantially older (at least 20 times) and at least as old as the split times of the archaics from modern humans suggesting little evidence for additional older admixture events since the split of archaic and modern human populations.

Simulations

To test the robustness of our results, we performed coalescent-based simulations under a demographic model in which a present-day non-African population had gene flow from both Neanderthals and Denisovans.

We generated 3000 independent 1 Mb regions. We set the mutation rate to $1.2e-8$ and the recombination rate to $1.3e-8$. We simulated 100 Oceanian and African chromosomes and 1 Neanderthal and Denisovan chromosome. All effective population sizes were fixed at 10000. We set the Archaic-modern human split, Neanderthal-Denisovan split, and African-non-African split to 12000, 8000 and 2500 generations respectively. The Neanderthal and Denisovan mixture proportions were set to 2% and 4.5% respectively. We fixed the time of the older admixture event to 2000 generations and varied that of the more recent admixture event across 1500, 1800 and 1900 generations. For each parameter instantiation, we considered a setting where the Neanderthal admixture pre-dated the Denisovan admixture and vice-versa. Table S1 shows the estimated dates. We see that the estimates are unbiased when the difference between the admixture dates is at least 500 generations. As the difference decreases, estimates of the older dates in particular tend to be biased. This bias tends to affect the Neanderthal estimate more than the Denisovan estimate. This is likely an effect of the smaller Neanderthal admixture proportion that leads to a noisier LD decay signal. However, the relative order of dates is always consistent. We computed a block Jackknife difference for a difference in the two estimates. In the cases where the null of no difference was rejected, the direction of the difference is consistent with the direction of the difference of the two parameters. There appears to be less power to reject the null in the cases where the Neanderthal admixture is older than the Denisovan admixture. We note that the power of this test is expected to be higher in simulations than in real data due to the fact that we simulate independent 1 Mb long regions so that the simulations carry more independent loci than real data.

We performed an additional set of simulations using a more realistic demographic model. We do not have a detailed joint demographic model relating Oceanians, Neanderthals and Denisovans. Instead, we modified the demographic model of non-Africans and Neanderthals used in [S2] that is, in turn, based on a demographic model fit by [S10]. We used the East Asian demographic parameters as a proxy for the Oceanian populations. We added both Neanderthal and Denisovan populations to this model. Neanderthal and Denisovan admixture proportions were set to 2% and 4.5% respectively. The split time of the two archaics was set to 8000 generations with their effective population sizes set to 2500 as was done in [S2]. We also modeled the observation [S6] that the divergence of the introgressing and sequenced Denisovans is larger than that of Neanderthals by setting these split times to 5600 and 2800 generations for Denisovans and Neanderthals respectively. We considered a model where the Neanderthal and Denisovan admixture dates are 2000 and 2200 generations respectively as well as one where the Denisovan admixture occurred earlier. Table S1 again shows that our estimates detect a statistically significant difference in the correct direction when the Denisovan admixture pre-dates Neanderthal admixture. However, in the opposite setting, the difference is no longer significant though the difference of the point estimates has

the same sign as that of the true parameter values.

These simulations indicate that our estimate of the relative dates of archaic admixture is robust although the absolute estimates themselves are quite sensitive (both to demographic parameters as well as to errors in the genetic map that we have not considered here but have been shown to affect these statistics previously [S3]).

Maps of archaic ancestry in diverse present-day humans

To infer maps of Neanderthal and Denisovan ancestry, we first applied a Conditional Random Field that had previously been developed to infer Neanderthal ancestry in Eurasian populations [S1]. The CRF used in [S1] was designed to infer archaic ancestry in populations that have a single dominant archaic ancestry component. While the inference from this application are reasonable for populations with a single dominant archaic component, we propose a modified method in Section that we show has improved accuracies for populations that have both Neanderthal and Denisovan ancestries.

To infer Neanderthal (Denisovan) ancestry, we applied the CRF using the high-coverage Altai Neanderthal [S6] (the high-coverage Denisovan genome [S7]) as an archaic reference (essentially, performing two two-way classifications). Inferences in the CRF require us to estimate model parameters. We fixed the model parameters to the values estimated in [S1].

We applied the CRF to present-day human genomes from diverse populations that were sequenced as part of the Simons Genome Diversity Project (SGDP) combined with genomes from the panel A individuals sequenced in an earlier study [S6]. The sequencing reads for the panel A individuals were processed using the same pipeline as the SGDP. We grouped the individuals according to five continental populations: West Eurasians, East Asians, Oceanians, South Asians, Americans and Central Asians. Of particular interest in this dataset are the populations that harbor a substantial fraction of Denisovan ancestry. To study these populations, we considered a subset of the Oceanian populations – Australians, Papuans and Bougainville Islanders, that consists of 16 individuals from Papua New Guinea, 2 from Bougainville Islands and 2 Australian Aborigines.

We used 43 African genomes from 17 populations as a reference panel of modern humans assumed to carry no archaic ancestry. These genomes were chosen from populations that are similar to the west African Yoruba (YRI) in their relationship to the Altai Neanderthal, *i.e.*, we chose populations X for which the Z-score associated with the D-statistic, $D(X, YRI; Altai\ Neanderthal, Chimpanzee)$ is less than two (where the standard error of the D-statistic is estimated using a weighted block jackknife with 5 cM blocks [S11]).

To infer Denisovan ancestry, we constructed two reference panels: one panel consists of the Denisovan genome [S7] sequenced to 31-fold coverage while the other consists of the 43 African genomes [S6]. To infer Neanderthal ancestry, one of the reference panels consists of the Altai Neanderthal genome [S6] sequenced to 52-fold coverage while the other consists of the 43 African genomes. For each haplotype $i \in \{1, \ldots, I\}$ in the target population and SNP $s \in \{1, \ldots, S\}$, we apply the CRF to estimate $\gamma_{i,s}^{(n)}$ and $\gamma_{i,s}^{(d)}$– the marginal probabilities of Neanderthal and Denisovan ancestry at SNP s of haplotype i.

Data Processing

Genotypes were called using the procedure described in [S8]. Briefly, BWA-MEM [S12] alignments were used as input for single-sample genotype calls using a reference-bias-free modification of the Unified Genotyper from the Genome Analysis Toolkit (GATK) [S13]. Sites which were found to be both polymorphic in at least one sample compared with chimpanzee and which pass filters (at filter level 1) were compiled (62.6M sites). At these discovered positions, genotype calls for samples were compiled at filter level 0 (the lower filter level is justified as the sites are known to be polymorphic in at least one sample).

We applied previously described filters to the Altai Neanderthal genome [S6] and the Denisovan genome [S7]. These filters restrict to sites that are non-repetitive, uniquely mappable and are not outliers with respect to coverage. Due to the high-coverage of these genomes, we work directly with the genotypes called from the ancient DNA reads (using GATK [S13]). This reduces the effects of genotyping errors as well as contamination (which is estimated to be at 1% at the read-level and hence substantially smaller at the genotype level for either of the genomes).

We restricted to SNPs that are biallelic across chimpanzee, ancient and modern genome sequences. We also filtered sites at which more than half the African reference genotypes are missing as well as sites where the Neanderthal or Denisovan genotype is missing.

The CRF requires phased genomes as input. We simultaneously phased all the genotypes in SGDP and panel A using SHAPEIT with default parameters [S14]. The ancestral allele at each site was determined from the 1000 Genomes ancestral sequence. Genetic distances were obtained from the combined LD map [S9] lifted over to hg19 coordinates. For the X chromosome, we obtained a sex-averaged map by scaling the X chromosome LD-based map by $\frac{2}{3}$.

Genome-wide analysis of archaic ancestry

For each individual i and archaic ancestry $a \in \{n, d\}$, we estimated the proportion of the genome that is confidently inferred to harbor archaic ancestry, $tia^{(a)}(i)$, to be the fraction of SNPs for which the marginal probability $\gamma_{i,s}^{(a)} > 0.90$.

$$tia^{(a)}(i) = \frac{1}{|H(i)|} \sum_{h \in H(i)} \frac{\sum_{s=1}^{S} \mathbf{1}\{\gamma_{i,s}^{(a)} > 0.90\}}{S} \tag{1}$$

We will drop the superscript when the archaic ancestry being referred to is clear from context. Here $H(i)$ indexes the haplotypes that belong to individual i. The above equation also holds for estimating Neanderthal ancestry on the X chromosome. In the case of the X chromosome, we average over both chromosomes for females only.

Empirical estimate of the accuracy of archaic ancestry estimates

We can estimate the accuracy of the archaic ancestry estimates on the SGDP data under several assumptions. The basic idea is as follows: the inferred Neanderthal ancestry in a target population can be modeled as arising from a process that classifies true Neanderthal, Denisovan or modern human alleles as Neanderthal. Given previous estimates of the Neanderthal and Denisovan ancestry in the target population, we can estimate these classification probabilities. These classification probabilities, in turn, provide information on the accuracy of the inference. For example, the proportion of Neanderthal ancestry inferred in a population like the African hunter-gatherers gives us the probability that a modern human allele is classified as Neanderthal assuming that the African hunter-gatherers have neither Neanderthal nor Denisovan ancestries. The procedure that we describe integrates out the uncertainty in the true Neanderthal and Denisovan ancestries across populations to estimate classification probabilities which in turn can be converted into estimates of precision and recall.

In these analyses, we will consider African hunter-gatherers (Khomani San), West Eurasians, East Asians and Oceanians. For SNPs that are assigned a marginal probability $\geq t$, let $p_{i,j}(t), i, j \in \{n, d, m\}$ denote the probability of that an allele of ancestry i was assigned ancestry j, where the ancestries $\{n, d, m\}$ refer to Neanderthal, Denisovan and modern human ancestries respectively. We assume that these probabilities are constant across the populations analyzed. This assumption holds if the Neanderthal and Denisovan ancestries in these populations are derived from similar ancestral populations and if the demographic histories of these populations do not affect the accuracy of the CRF. The first assumption is reasonable given the close relatedness of existing Neanderthal genomes obtained from a

wide range of spatial and temporal separation [S6]. This assumption might also be violated if the archaic admixtures occurred at different times across populations. For Neanderthal ancestry, current estimates strongly suggest that most of the Neanderthal ancestry in non-African populations traces its origin to a shared admixture event (eastern non-Africans have about 25% more Neanderthal ancestry than west Eurasians [S6; 15]). Further, we have shown previously that the precision of our method changes by about 1% when the time of Neanderthal admixture varies across more than 1000 generations [S1] so that we expect these probabilities to be relatively robust to variation in the time of admixture. Another reason to expect that the assumption of constant probabilities might not hold is that the target populations differ in their recent demographic histories. Nevertheless, our method analyzes single haploid genomes in each of these populations and hence, should be robust to these differences. Further, since we are analyzing non-African genomes relative to archaic and African genomes, genomes from distinct non-African populations should show similar relationships to the African and archaic genomes. Non-African populations that have substantial African-related gene flow might violate this assumption. The African hunter-gatherer might appear to also violate these assumptions given that they might share recent ancestry with the African reference genomes.

Let $f_{i,k}, i \in \{n, m, d\}, k \in \{san, we, ea, me\}$ denote the true proportions of ancestry i in population k. For a given threshold t, we observe $\tilde{f}_{n,k}(t), \tilde{f}_{d,k}(t)$, the fraction of sites with marginal probability of Neanderthal and Denisovan ancestry of at least t in population k. We can then find $p_{i,j}(t)$ by solving the following optimization problem.

$$\begin{aligned} \{p^*_{i,a}(t)\} \quad = \quad & argmin \sum_{a\in\{n,d\},k\in\{san,we,me,ea\}} \left(\tilde{f}_{a,k}(t) - \sum_{i\in\{d,m,n\}} f_{i,k}p_{i,a}(t)\right)^2 \\ & 0 \le p_{i,a}(t) \le 1 \end{aligned}$$

The precision (proportion of archaic ancestry called at threshold t that is true) for the estimates of archaic ancestry a in target population k is given by $p^*_{a,a}(t)$ while the recall (proportion of true archaic ancestry that is called at threshold t) is given by $\frac{f_{a,a}p^*_{a,a}(t)}{\sum_{i\in\{d,m,n\}} f_{i,k}p^*_{i,a}(t)}$. By computing these estimates for all values of $t \in [0, 1]$, we can estimate a precision-recall curve for each archaic ancestry estimate in a given population.

Since the true values of mixture proportions $f_{i,k}$ are not known, we sampled 100 times from the range of plausible values estimated for these quantities and averaged our precision and recall estimates over these samples. Specifically, we assumed $f_{n,san} = f_{d,san} = 0$, $f_{n,ea} = f_{n,me} \sim \mathcal{N}(0.0189, (0.0013)^2)$, $f_{n,we} = rf_{n,ea}, r \sim \mathcal{N}(0.76, (0.06)^2)$, $f_{d,we} = 0$,$f_{d,ea} \sim$ Unif$(0, 0.002)$ and $f_{d,me} \sim$ Unif$(0.03, 0.06)$. In words, we assume that the African hunter-gatherers have no archaic ancestry, West Eurasians have no Denisovan ancestry, East Asians and Oceanians have the same proportion of Neanderthal ancestry, Oceanians have substantial Denisovan ancestry while East Asians have a small fraction and that West Eurasians have slightly less Neanderthal ancestry than East Asians (consistent with previous studies). The use of the normal distribution for Neanderthal ancestry in West Eurasians and East Asians is motivated by the fact that these estimates are endowed with formal standard errors. We use a uniform distribution for the other estimates.

This procedure reveals that in populations such as West Eurasians and East Asians, which are well-modeled as a two-way admixture between modern and archaic humans, the CRF attains reasonable recall at high precision (attaining recalls $> 50\%$ at precisions $> 95\%$) . However, as the Oceanians have substantial Neanderthal and Denisovan ancestries, the precision tends to be substantially lower (less than 80% for any recall). Further insight into the error processes can be obtained by inspecting the classification probabilities $p^*_{\cdot,\cdot}(t)$. For example, at a probability threshold $t = 0.90$, the probability of classifying an allele of modern human ancestry as either Neanderthal or Denisovan ($p^*_{m,n}(0.90)$ and $p^*_{m,d}(0.90)$) is of the order of 10^{-4} while the probability of classifying a Neanderthal allele as Denisovan

or vice-versa ($p^*_{n,d}(0.90)$ or $p^*_{d,n}(0.90)$) is of the order of 0.10 which is of the same order of magnitude as the probability that a Neanderthal (or Denisovan) allele is classified correctly ($p^*_{n,n}(0.90)$ or $p^*_{d,d}(0.90)$).

An improved procedure for deconvolving Neanderthal and Denisovan ancestries

We considered a modified procedure to improve the accuracy of archaic ancestry inference.

Firstly, we modified the reference panels. To infer Denisovan ancestry, we constructed two reference panels: one panel consists of the Denisovan genome while the other consists of 43 African genomes pooled with the Neanderthal genome. Analogously, to infer Neanderthal ancestry, one of the reference panels consists of the Neanderthal genome while the other consists of African and Denisovan genomes. A second modification we made is to set the model parameter associated with haplotypic feature to zero because we discovered a small bias induced by this parameter for populations with proportions of archaic ancestry that are of the order of $\frac{1}{1000}$. The bias arises because the CRF was trained to estimate ancestries of the order of 1/100 leading to a specific weighting of the haplotype parameter relative to the SNP parameters. This weighting is not appropriate when the true admixture fraction is substantially different. As a result, the method infers similar proportion of Denisovan ancestry in French and Han Chinese ($\approx 0.6\%$) in contradiction to [S1]. This modification allows the CRF to be applied to study Denisovan ancestries in mainland Eurasia.

We estimated the empirical accuracy of this modified procedure as described in Section . At a nominal probability threshold of 0.90, the CRF now attains a recall of around 47% at a precision of 97% for Neanderthal ancestry inference in West Eurasians and East Asians (Figure S1a). In Oceanians, it attains a recall of around 43% at a precision of 83% for Neanderthal ancestry and a recall of 14% at a precision of 97% for Denisovan ancestry (Figure S1b).

Given the relatively high precision of these estimates as well as the profile of the precision-recall curves that suggest that the precision remains high for lower probability thresholds, we chose a marginal probability threshold of 0.50 to call a SNP as archaic. At this threshold, the CRF has a recall of around 72% at precisions of around 97%, 97% and 85% respectively for Neanderthal ancestry in West Eurasians, East Asians and Oceanians whereas for Denisovan ancestry in Oceanians, the recall is around 24% at a precision of 97%. Our power to detect Denisovan ancestry in Oceanian populations is still lower than the power to detect Neanderthal ancestry. A likely explanation for this reduced power is the substantially larger divergence of the sequenced and introgressing genomes for the Denisovans compared to the Neanderthals [S6]. While the quantitative estimates of accuracy obtained in this framework depend on several assumptions about the distributions of archaic ancestries, the qualitative conclusion is that discriminating between distinct archaic ancestral components in a population such as Oceanians is challenging.

We can again obtain additional insights into the accuracy of our estimates by inspecting the classification probabilities $p^*_{\cdot,\cdot}$. The probability of classifying a modern human allele as archaic is of the order of 10^{-4} or smaller, in the modified procedure (specifically, we estimate $p^*_{m,n}(0.50) = 2 \times 10^{-4}$,$p^*_{m,d}(0.50) = 4 \times 10^{-5}$). On the other hand, the probability of classifying one archaic allele as another is of the order of 0.01 (specifically, $p^*_{d,n}(0.50) = 0.05$,$p^*_{n,d}(0.50) = 0.01$), reduced relative to the method analyzed in Section . As a result, there is an increased probability that an allele classified as Denisovan is truly Denisovan, particularly in populations that have substantial Neanderthal ancestry. These observations hold across a wide range of thresholds on probability t (including at $t = 0.25$, $t = 0.50$ and $t = 0.90$). Further, $p^*_{d,n} > p^*_{n,d}$ at these values of t, *i.e.*, our procedure is more likely to misclassify a Denisovan allele as Neanderthal than vice-versa. Thus, these classification probabilities tell us why the power or recall for Denisovan ancestry inference is lower than that for Neanderthal ancestry inference for the same probability threshold. This result is consistent with the larger divergence of sequenced and introgressing genomes for Denisovans relative to Neanderthals. Further, these probabilities also provide insight into why the

precision for Neanderthal ancestry inference in Oceanians is lower than that for Neanderthal ancestry inference in other non-Africans as well as for Denisovan ancestry inference in Oceanians. This observation is likely due to the higher rate of misclassification of Denisovan alleles as Neanderthal compared to the reverse process in combination with the higher proportion of Denisovan ancestry in Oceanians.

Variation in the genome-wide proportions of archaic ancestry

To formally test for differences in archaic ancestry, we tested for a difference in the *tia* statistic (Equation 1) across pairs of populations. Specifically, for a reference population r and a target population t, we computed $\delta(r,t) = tia(t) - tia(r)$. We assessed statistical significance using a block jackknife with 10 Mb blocks [S11].

Apart from Oceania, several populations in East, South, and Central Asia have higher values of *tia* for Denisovan ancestry (Table S3). For example, while the French have a mean *tia* of 0.01%, the Han have a mean *tia* of 0.06% (Z-score of 4.35). These proportions are rather small: the East Asian proportion of the genome called as Denisovan is about 5.8% of the corresponding proportion for Oceanians. Among the populations with elevated *tia* compared to French are the Tibetans and the Sherpa. The increased ancestry in the Sherpa and Tibetan populations is interesting in light of the evidence for Denisovan introgression at the EPAS1 locus that contributed to high-altitude adaptation in these populations [S16]. One possible explanation is that the increased ancestry is caused by Denisovan introgression at the EPAS1 locus. To test this explanation, we computed the *tia* statistic after removing chromosome 2 that contains the EPAS1 locus. We find that the *tia* statistic is highly concordant whether or not we include chromosome 2 ($\rho = 0.968$). Sherpa remains the population with the highest *tia* in both analyses while the ranks of the Tibetans are 7 and 5 depending on whether we include or exclude chromosome 2. Finally, we do not detect statistically significant increases (Z-score > 4 correcting for the multiple hypotheses tested) in Denisovan ancestry relative to Han Chinese within mainland Eurasians. We study this variation in more detail in Section .

Modeling the variation in Denisovan ancestry across populations

To understand how variation in Denisovan ancestry might be related to known population relationships, we tried to model the proportion of the genome inferred to be Denisovan in a given mainland Eurasian population as a linear function of its proximity to non-West Eurasians. Specifically, given that East Asians have higher Denisovan ancestry relative to West Eurasians, we asked if the Denisovan ancestry proportion in other mainland Eurasians can be explained by differential proportions of non-West Eurasian ancestries in these populations. For each mainland Eurasian population X, we computed the f_4-statistic $f_4(X, Yoruba; Australian, Ust' - Ishim)$ which measures the drift shared by population X with East Eurasians since their split from the ancestors of West Eurasians. We then regressed an estimate of Denisovan ancestry against this f_4 statistic measured on West Eurasian and East Asian populations, *i.e.*, we learned the parameters of the regression on West Eurasians and East Asians. We then used this regression to predict the mean Denisovan ancestry in the other Eurasian populations. Figure 2(B) shows the Denisovan ancestry inferred by the CRF versus the Denisovan ancestry expected under the model. In American, central and south Asian (that includes populations such as the Sherpa) populations, the proportion of Denisovan ancestry is positively correlated with the f_4 statistic (Pearson's correlation $\rho_{Pearson} = 0.832$, $Z = 6.27$).

We also observe that American, central and South Asian populations have systematically higher proportions of Denisovan ancestry than predicted by the linear model – the mean of the residuals is 1.36. Testing this model which involves a test of the residuals having mean zero presents two challenges: i) the analyzed populations are not independent observations as they share drift to various degrees, and ii) the estimates of both ancestries and the f_4 statistics are noisy. To test the model, we computed block jackknife standard errors for the mean of the residuals as well as the f_4 statistics. We deleted a

10 Mb block of the genome, in turn, and then computed Jackknife estimates of the Denisovan ancestry proportion and the f_4 statistics. We then ran the estimation procedure on the Jackknife estimates and computed the mean of the residuals on the American, central and South Asian populations. The Z-score for the mean of the residuals is 2.84 using the ancestry estimates from the CRF. An additional concern is that a handful (four) of the West Eurasian and East Asian populations used for parameter estimation appear to be outliers to the linear model (absolute value of standardized residuals>2). We reran the inference after excluding these outliers and found that the results became more significant. We estimated Z-scores of 3.74 for the CRF.

To further narrow down this signal of increased Denisovan ancestry, we inspected the residuals for each population. For the ancestries estimated by the CRF, none of the residuals is individually significant. However, when we ranked the populations according to their residuals, we find that the south Asian populations are ranked highest. We reran the block jackknife testing the mean of the residuals in south Asians, central Asians and Americans. The Z-scores are 3.20, 1.21 and 0.13 for south Asians, central Asians and Americans respectively with the CRF estimates.

Coverage of archaic haplotypes

We defined archaic haplotypes as runs of consecutive alleles along a haploid genome with marginal probability of archaic ancestry > 0.50. We merged the inferred archaic haplotypes (Neanderthal haplotypes inferred across all non-Africans, Denisovan haplotypes inferred across all Oceanians). We reconstructed 2235 Neanderthal contigs that cover a total length of 673 Mb and 967 Denisova contigs with a total length of 257 Mb.

Genomic regions with elevated archaic ancestry

We screened for non-overlapping 100 Kb windows with elevated proportions of archaic ancestry as estimated by $la^{(a)}(w) = \frac{\sum_{s \in S(w)} \sum_{i=1}^{I} \gamma_{i,s}^{(a)}}{I|\{j \in S(w)\}|}$. Here I is the number of haploid genomes, $S(w)$ refers to the set of SNPs that belong to window w, $a \in \{n, d\}$ refers to either Neanderthal or Denisovan ancestries, and $\gamma_{i,s}^{(a)}$ refers to the marginal probability of archaic ancestry a at SNP s in individual i. We selected the windows with this statistic exceeding 0.30 and merged consecutive windows.

We identified a number of windows with elevated proportions of Neanderthal ancestry – 88, 27, 37, 116, 2 and 11 in American, Central Asian, East Asian, Oceanian, South Asian and West Eurasian populations respectively. Further, we identified 48 windows with elevated proportions of Denisovan ancestry in Oceanians(Table S4). Our scan recovered previously identified loci such as BNC2 in West Eurasians [S1; 17] as well as POU2F3 [S1]/TMEM136 [S17].

GO analysis

We tested whether specific sets of genes have significantly elevated frequencies of archaic ancestry. To do this, we classified CCDS genes as having high archaic ancestry ancestry if the gene ranked in the top 5% of genes ranked according to the average of the marginal probability of archaic ancestry across all SNPs within the gene and all individuals in the population. For Neandertal ancestry in mainland Eurasians, we used the method from [S1] as it has greater power for populations with a single dominant archaic ancestry. For Neanderthal and Denisovan ancestry in Oceanians, we used the modified method proposed here. We then tested for an enrichment of Gene Ontology categories [S18] using the hypergeometric test implemented in FUNC [S19]. We report categories that are signficant at the 0.05 level after multiple testing correction using 1000 permutations.

Analysis of genomic regions deficient in archaic ancestry

We searched for large regions that are deficient in Neanderthal and Denisovan ancestry in the different populations in SGDP, restricting our analysis to the Oceanian population (Australians, Papuans and Bougainville Islanders) in the case of Denisovan ancestry.

As described previously [S1], to assess the existence of regions deficient in archaic ancestry in a robust manner, we measured the fraction of archaic ancestry $ta_t^{(a)}(w)$ that exceeds a threshold t for archaic ancestry $a, a \in \{n, d\}$, averaged across all SNPs and individuals within window w:

$$ta_t^{(a)}(w) = \frac{\sum_{i=1}^{I}\sum_{s=1}^{S}\mathbf{1}\{\gamma_{i,s}^{(a)} > t\}}{I|\{s \in w\}|} \tag{2}$$

Here $t \in [0, 1]$ is a threshold. We chose $t = 0.25$ to reduce our false negative rate and chose to examine large windows ($w = 10$ Mb) that overlap each other by 1 Mb. We excluded all windows that overlap (over any part of their length) the centromeres or the telomeres. We further restricted our analysis to windows in which the number of SNPs that pass filters is at least 1000 and over which the genetic length ≥ 2 cM. We declared a window as a desert if $ta_t^{(a)}(w) < \frac{1}{1000}$ and merged overlapping deserts.

We identified a number of regions that are deserts for archaic ancestries in different populations of the SGDP. Of particular interest are regions that are deserts for both Neanderthal and Denisovan ancestries across all populations. We identified four windows longer than 10 Mb ($1 : 99 - 112$, $3 : 78 - 90$, $7 : 108 - 128$, and $13 : 49 - 61$ Mb) that are deserts for both Neanderthal and Denisovan ancestries across all populations. The locus on chromosome 7 is particularly interesting as it contains the FOXP2 gene [S17]. This observation is interesting because previous attempts that identified deserts of Neanderthal ancestry [S17; 1] could not rule out the possibility that these deserts were the result of demographic events [S1]. The observation of deserts that are shared across distinct introgression events might suggest that these regions are resistant to introgression because of their importance for the modern human phenotype and represent selection against the introgressing alleles. To test the null hypothesis that the Neanderthal and Denisovan deserts are independently located along the genome, for each chromosome, we randomly placed the Neanderthal deserts (avoiding centromeres and telomeres since the original deserts were chosen to be non-overlapping with these features) and then counted the length of intersection of these deserts to the Denisovan desert. The P-value we report is the proportion of random datasets for which the overlap length is longer than that observed in data. We obtain a permutation p-value 0.67. We also observe two shared deserts larger than 10 Mb on chromosome X ($X : 62 - 78$, $X : 109 - 143$ Mb).

Correlation of archaic ancestry with B-statistics

To interrogate the effects of selection against introgressing archaic alleles, we analyzed the proportion of archaic ancestry in a genomic region as a function of the B-statistic, a measure of background selection [S20].

B-statistics were lifted over to hg19 coordinates. We then annotated each of the SNPs that we analyzed with the B-statistic of the genomic region in which the SNP falls. In our first analysis, we partitioned SNPs into quintiles based on their B-statistic annotation. At each SNP, we considered several estimates of the archaic ancestry : $la^{(a)}$ which computes the average over the marginal probability of archaic ancestry assigned to each individual haplotype, $ta_{0.9}^{(a)}$ which computes the average fraction of alleles across individuals that attain a marginal probability of ≥ 0.90 and $ta_{0.25}^{(a)}$ that computes the analogous statistic for a threshold of 0.25. Under a model where the archaic alleles are not under purifying selection, the power to detect archaic ancestry increases with decreasing B-statistic [S1] so that we expect the summaries of archaic ancestry to increase with decreasing B. On the other hand, under a model where the archaic alleles are subject to purifying selection, these statistics are expect to decrease with decreasing B.

We estimated Spearman's correlation coefficient ρ between Neanderthal ancestry and B-statistic in West Eurasians, East Asians and Oceanians (Table S6). We performed a block jackknife in 10 Mb windows to estimate the standard error of ρ. We see a statistically significant correlation between B-statistic and different summaries of the Neanderthal ancestry. The significance is strongest for the *la* statistic on the autosomes. While the *ta* statistics are not always significant, this trend is expected given the reduction in power to detect archaic ancestry with increasing value of B [S1] and that increasing the threshold is bound to exacerbate the difference in power across B quintiles. We also see an analogous, though weaker, trend on the X as would be expected given the reduced number of observations on the X.

Next, we estimated ρ for the Denisovan ancestry in Oceanians and observed an analogous positive correlation of *la* with B consistent with the effect of purifying selection on Denisovan introgressed alleles (Table S6).

Association of Denisovan ancestry with tissue-specific expression

We analyzed the Illumina BodyMap 2.0 data for genes that are highly expressed in each of 16 tissues. We used the definition of tissue-expressed genes in [S1] as genes that are significantly highly expressed in a given tissue than in any of the other tissues using the DESeq package [S21]. We defined a gene as being depleted in Denisovan (Neanderthal) ancestry when all sites across all Oceanian (Melanesian, Australian and Bougainville) individuals in the gene are assigned a marginal probability ≤ 0.10. We tested whether there is a statistically significant enrichment of tissue-expressed genes in genes with depleted Denisovan (Neanderthal) ancestry. We also tested for enrichment of tissue-expressed genes in genes depleted for Neanderthal ancestry in mainland Eurasians defined as a gene that is depleted in each of West Eurasian, East Asian, South Asian, American and Central Asian maps.

As a check for whether testes-specific genes might be depleted in archaic ancestry due to differences in the strength of purifying selection, we compared the B-statistics across testes-expressed genes to other tissue-expressed genes. For each gene, we computed an average B-statistic [S20]. Testes-expressed genes had a slightly reduced B-statistic on average compared to other tissue-expressed genes (0.673 ± 0.007 vs 0.684 ± 0.004) but the reduction is not statistically significant (Mann-Whitney one-sided test P-value $= 0.07$). Other tissue-expressed gene sets such as liver, heart and skeletal muscle have lower average B-statistics than testes but do not show a statistically significant depletion in archaic ancestry. We further investigate the influence of the B-statistic by performing a logistic regression of the depletion status of each tissue-expressed gene using as covariates the specific tissue in which it is expressed as well as the B-statistic. Table S7 shows that only testes-expressed genes are enriched in regions of low archaic ancestry.

It is plausible that the B-statistic is not strongly correlated with selective constraint, particularly in testes-expressed genes. To further investigate this possibility, we estimated the local heterozygosity at each gene. To reduce the potential for interaction between the local heterozygosity and the accuracy of our method for archaic ancestry inference, we estimated the local heterozygosity for each gene is calculated across 76 African chromosomes in the SGDP, restricting to sites which pass filter level $>= 1$, and to sites where at least half the samples have a valid genotyping call. Samples from panel A were excluded as the error rate is known to be higher. We then repeated the analysis carried out using B-statistics but now replacing B-statistics with local heterozygosity. Table S7 shows that, under this model, testes-expressed genes remain the only set of genes that are enriched in regions of low archaic ancestry.

Supplemental References

S1. S. Sankararaman, S. Mallick, M. Danneman, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, and D. Reich. The landscape of Neandertal ancestry in present-day humans. *Nature*, 2014.

S2. Qiaomei Fu, Heng Li, Priya Moorjani, Flora Jay, Sergey M Slepchenko, Aleksei A Bondarev, Philip LF Johnson, Ayinuer Aximu-Petri, Kay Prüfer, Cesare de Filippo, et al. Genome sequence of a 45,000-year-old modern human from western siberia. *Nature*, 514(7523):445–449, 2014.

S3. S. Sankararaman, N. Patterson, H. Li, S. Pääbo, and D. Reich. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.*, 8(10):e1002947, 2012.

S4. G. Coop, X. Wen, C. Ober, J. K. Pritchard, and M. Przeworski. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science*, 319:1395–1398, Mar 2008.

S5. Hans R Kunsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989.

S6. K. Prüfer, F. Racimo, N. Patterson, Flora. Jay, S. Sankararaman, and S. Sawyer. The complete genome sequence of a neandertal from the altai mountains. *submitted*, 2013.

S7. Matthias Meyer, Martin Kircher, Marie-Theres Gansauge, Heng Li, Fernando Racimo, Swapan Mallick, Joshua G. Schraiber, Flora Jay, Kay Prüfer, Cesare de Filippo, Peter H. Sudmant, Can Alkan, Qiaomei Fu, Ron Do, Nadin Rohland, Arti Tandon, Michael Siebauer, Richard E. Green, Katarzyna Bryc, Adrian W. Briggs, Udo Stenzel, Jesse Dabney, Jay Shendure, Jacob Kitzman, Michael F. Hammer, Michael V. Shunkov, Anatoli P. Derevianko, Nick Patterson, Aida M. Andrés, Evan E. Eichler, Montgomery Slatkin, David Reich, Janet Kelso, and Svante Pääbo. A high-coverage genome sequence from an archaic denisovan individual. *Science*, 2012.

S8. Swapan Mallick, Heng Li Li, Mark Lipson, Iain Mathieson, Melissa Gymrek, Fernando Racimo, Mengyao Zhao, and Niru Chennagiri. The landscape of human genome diversity. *in review*, 2015.

S9. Simon Myers, Leonardo Bottolo, Colin Freeman, Gil McVean, and Peter Donnelly. A fine-scale map of recombination rates and hotspots across the human genome. *Science*, 310(5746):321–324, 2005.

S10. Simon Gravel, Brenna M Henn, Ryan N Gutenkunst, Amit R Indap, Gabor T Marth, Andrew G Clark, Fuli Yu, Richard A Gibbs, Carlos D Bustamante, David L Altshuler, et al. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences*, 108(29):11983–11988, 2011.

S11. F Busing, E Meijer, and R Leeden. Delete-m jackknife for unequal m. *Statistics and Computing*, 9:3–8, 1999.

S12. H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv e-prints*, March 2013.

S13. Mark A DePristo, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, Christopher Hartl, Anthony A Philippakis, Guillermo Del Angel, Manuel A Rivas, Matt Hanna, et al. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature genetics*, 43(5):491–498, 2011.

S14. Olivier Delaneau, Jonathan Marchini, and Jean-François Zagury. A linear complexity phasing method for thousands of genomes. *Nature methods*, 9(2):179–181, 2012.

S15. J. D. Wall, M. A. Yang, F. Jay, S. K. Kim, E. Y. Durand, L. S. Stevison, C. Gignoux, A. Woerner, M. F. Hammer, and M. Slatkin. Higher Levels of Neanderthal Ancestry in East Asians Than in Europeans. *Genetics*, Feb 2013.

S16. Emilia Huerta-Sánchez, Xin Jin, Zhuoma Bianba, Benjamin M Peter, Nicolas Vinckenbosch, Yu Liang, Xin Yi, Mingze He, Mehmet Somel, Peixiang Ni, et al. Altitude adaptation in tibetans caused by introgression of denisovan-like dna. *Nature*, 512(7513):194–197, 2014.

S17. Benjamin Vernot and Joshua M Akey. Resurrecting surviving neandertal lineages from modern human genomes. *Science*, 343(6174):1017–1021, 2014.

S18. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, 25(1):25–29, May 2000.

S19. K. Prufer, B. Muetzel, H. H. Do, G. Weiss, P. Khaitovich, E. Rahm, S. Paabo, M. Lachmann, and W. Enard. FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics*, 8:41, 2007.

S20. Graham McVicker, David Gordon, Colleen Davis, and Phil Green. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet*, 5(5):e1000471, 05 2009.

S21. Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome Biol*, 11(10):R106, 2010.