www.sciencemag.org/cgi/content/full/science.aao1887/DC1

Supplementary Materials for

A high-coverage Neandertal genome from Vindija Cave in Croatia

Kay Prüfer,* Cesare de Filippo,† Steffi Grote,† Fabrizio Mafessoni,† Petra Korlević, Mateja Hajdinjak, Benjamin Vernot, Laurits Skov, Pinghsun Hsieh, Stéphane Peyrégne, David Reher, Charlotte Hopfe, Sarah Nagel, Tomislav Maricic, Qiaomei Fu, Christoph Theunert, Rebekah Rogers, Pontus Skoglund, Manjusha Chintalapati, Michael Dannemann, Bradley J. Nelson, Felix M. Key, Pavao Rudan, Željko Kućan, Ivan Gušić, Liubov V. Golovanova, Vladimir B. Doronichev, Nick Patterson, David Reich, Evan E. Eichler, Montgomery Slatkin, Mikkel H. Schierup, Aida Andrés, Janet Kelso, Matthias Meyer, Svante Pääbo*

*Corresponding author. Email: pruefer@eva.mpg.de (K.P.); paabo@eva.mpg.de (S.P.)
†These authors contributed equally to this work.

Published 5 October 2017 on *Science* First Release
DOI: 10.1126/science.aao1887

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:
(available at www.sciencemag.org/cgi/content/full/science.aao1887/DC1)

Contents

S1: Sampling, Radiocarbon Dating, DNA Extraction and Library Preparation

Petra Korlević*, Charlotte Hopfe, Sarah Nagel, Mateja Hajdinjak, Tomislav Maricic, Qiaomei Fu and Matthias Meyer

* To whom correspondence should be addressed (petra_korlevic@eva.mpg.de)

Sample selection

In efforts leading to the sequencing of the first Neandertal genome in 2010(*3*), 44 bone fragments from Vindija Cave, most of which were morphologically undiagnostic in regards to the species they originated from, were screened for the presence of Neandertal DNA. Of these fragments, 19 tested positive. Since the experiments in that study had been performed using a mixture of methods, including direct amplification of mtDNA fragments by PCR and enrichment of mtDNA fragments from DNA libraries using primer extension capture(*26*), we prepared new DNA extracts(*27*) from 20 to 200 mg of each specimen to determine their suitability for further genome sequencing efforts. From these extract double stranded libraries(*28*) were prepared, enriched for human mitochondrial DNA(*29*) and sequenced using Illumina technology. We identified several bone fragments that exhibited a high content of Neanderthal DNA (approximated based on mtDNA coverage) as well as low estimates of present-day mtDNA human contamination(*30*) (Table S1). Of these, one of the largest samples, Vindija 33.19, was selected for further experiments.

The Vindija 33.19 Neandertal is a morphologically undiagnostic bone shaft splinter discovered by Malez and co-workers in the G3 layer of Vindija cave (Croatia) during excavations in 1980(*31*). Direct radiocarbon dating of the bone was performed at the Oxford Radiocarbon Accelerator Unit specifically for this project (sample reference OxA 32,278), and the determined age was 45.3 ± 2.3 kBP (uncalibrated), with the calibrated date extending beyond 50 kBP (older than 45.5 kBP with 95% probability).

In addition, we included in this study a rib fragment of Mezmaiskaya 1, a partial Neandertal skeleton from a neonatal individual found in quadrant M-26 of layer 3 in Mezmaiskaya cave (Russia)(*32*). Its age has been estimated to 60-70 kBP(*33*).

Sampling, radiocarbon dating, DNA extraction and library preparation

After the initial screening, the Vindija 33.19 fragment was sampled further both for DNA extraction and radiocarbon dating. For radiocarbon dating, 567 mg of bone was sawed off using a sterile diamond disc and sent to the Oxford Radiocarbon Accelerator Unit. Since DNA preservation can be highly variable within one specimen, we sampled different areas of the bone fragment. Bone powder was removed using a sterile dentistry drill and DNA was extracted using a silica-based method(*17, 34*). The Mezmaiskaya 1 rib fragment was sampled and the DNA extracted as part of a previous project(*2*).

Single-stranded DNA libraries were prepared from varying volumes of input extract (10–15 µL out of 50 µL

extract volume) using the protocol from Gansauge and Meyer 2013(*18*) with modifications(*34*). A subset of the libraries was prepared using Uracil-DNA-glycosylase (UDG) and endonuclease VIII treatment to remove uracils in the interior of the molecules(*1*), while for other libraries this step was omitted to minimize fragmentation and loss of endogenous DNA. Libraries were double-indexed by amplification with pairs of sample-specific indexing primers(*28*) using AccuPrime Pfx DNA Polymerase(*35*). All libraries were amplified into PCR plateau and purified using the MinElute PCR purification kit (Qiagen). A one-cycle reamplification step with Herculase II Fusion DNA Polymerase was performed to remove heteroduplices(*1, 35*). Libraries were diluted based on Agilent 2100 Bioanalyzer concentration measurements, and sequenced using Illumina technology. Sequences were assigned to each library requiring a perfect match to the respective index combination, forward and reverse reads were overlap-merged, and the resulting sequences were mapped to the human reference genome (hg19) (for a detailed description see S2). Summary statistics, including predicted genomic coverage in the whole library, were calculated as described in Gansauge et al. 2017(*36*). An outline of prepared extracts and libraries for the initial quality control screening of Vindija 33.19 is shown in Table S2.

High coverage genome generation

DNA libraries for high throughput sequencing were prepared from varying volumes of input extract (2.5–15 µL out of 50 µL extract volume) using the same single-stranded library preparation approach as above. Following amplification and double-indexing, molecules <45 bp were removed from library B8744 by gel excision as described elsewhere(*1*), while all other libraries were sequenced without size fractionation. After heteroduplex removal, DNA concentrations in the libraries were determined on an Agilent 2100 Bioanalyzer DNA 1000 chip, and libraries were diluted for sequencing on an Illumina HiSeq 2500 platform in rapid or high output mode (see SI2). A summary of all libraries is shown in Table S3.

S2: Sequencing and Initial Data Processing

Barbara Hoeber, Antje Weihmann, Udo Stenzel, Svante Pääbo, Janet Kelso and Kay Prüfer*

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

We generated a total of 30x genomic coverage from nine Vindija 33.19 Neandertal sequencing libraries. Approximately 24% of this data was produced from a UDG treated library, removing most of the typical C to T errors present in ancient DNA, while the remaining ~76% of the data were not treated. In addition, 1.4x coverage was generated from two non-UDG treated libraries of the previously to 0.5x sequenced Mezmaiskaya 1 individual.

Sequencing and Base Calling

We sequenced 185 lanes of the nine Vindija 33.19 libraries and eight lanes of the two Mezmaiskaya 1 libraries on the Illumina HiSeq 2500 platform (Table S4 and S5). An indexed ΦX 174 library was added to each library before sequencing and the sequencer was run in 76 basepair paired-end mode with two seven basepair indices (*37*).

A total of 33 lanes, containing the Vindija 33.19 libraries A9401, A9402, A9403, A9404 and B8744, were processed with the standard Illumina base-caller Bustard. All remaining lanes were base-called with *Ibis* (version: FreeIbis 2dcf022bf40793c12d226c17d68583683e74e057) (*38*, *39*).

Read-Merging and Index Assignment

We used the program *leeHom* with the parameter "--ancientdna" to merge overlapping mate-pairs and trim adapter sequences (*40*). After merging, sequences were assigned to a library based on their index sequences using *deML* (*41*) with default parameters. We only retain those sequences for further analyses that were assigned to the correct library.

Alignment and Duplicate Removal

All reads were aligned to a modified human reference GRCh37 from the 1000 Genomes project, which includes the ΦX 174 genome (NC_001422.1), the human herpesvirus 4 type 1 (NC_007605), and the 1000 Genomes Phase 2 decoy sequences(*42*) and an updated mitochondrial sequence. As in previous analyses, we used *BWA* (*43*) (version: 0.5.10-evan.9-1-g44db244 (*44*)) with parameters: –n 0.01 –o 2 –l 16500 to increase sensitivity. Sequences shorter than 35 basepairs were removed and duplicates were merged into one sequence using *bam-rmdup* (version: 0.6.3) (*45*). In this process, unaligned merged sequences, half-aligned pairs and disjoint pairs were removed and only aligning properly-paired or merged sequences were retained for further analysis. Out of a total of 24 billion Vindija 33.19 sequences 2.4 billion aligned and 1.8 billion remain after duplicate removal.

Coverage Estimates

We estimated the nuclear coverage for each sequencing library by counting the number of bases (Q≥30) in the part of sequences that overlap confidently alignable regions of the human genome (map35_100 track: see Supplementary section 5b, page 35, ref. (*2*)) on all autosomes and dividing by the total length of alignable regions (Table S6). Summing over all data we estimated a genome-wide coverage of 30x for Vindija 33.19 and 1.4x for Mezmaiskaya 1. Chromosome X shows a similar coverage to that of the autosomes, demonstrating that both samples stem from female individuals.

GC Dependent Coverage

Previous high-coverage ancient genomes showed a bias towards higher coverage with lower GC content. We calculated GC content in windows of 51 bases and assigned the value for the middle position for each base in the genome. Coverage was calculated for each position using Vindija 33.19 sequences aligning with high confidence and excluding low quality bases (MQ ≥ 25, base-quality ≥ 30, map35_100 regions), and binned according to GC content (Fig.S1).

Length Distribution and Ancient DNA Substitution Patterns

The length distribution for sequences aligning on chromosome 21 from all Vindija 33.19 and Mezmaiskaya 1 libraries are shown in Figure S2. Except for Vindija library B8744 which underwent size-selection (see S1), all libraries show a mode close to our length cutoff of 35 base pairs. Mezmaiskaya sequences, with an average length of 47 base pairs, are shorter than Vindija 33.19 libraries sequences with an average length of 53 base pairs.

We also analyzed substitution patterns along sequences by counting substitutions to homozygous genotypes in the published Altai Neandertal VCFs files within the first 20 and last 20 bases from the sequence ends. Except for the UDG-treated library B8744 all Vindija 33.19 libraries show similar substitution patterns with C to T substitutions reaching ~40% at both sequence ends, falling to ~1% in the interior of sequences (Fig. S3). In contrast, B8744 shows elevated C to T substitutions mainly at the first position from the 5’ and the first two positions from the 3’ end. Mezmaiskaya 1 shows higher C to T substitution patterns at the 3’ end and in the interior of sequences compared to the untreated Vindija 33.19 libraries (Fig. S4).

Reprocessing of other Genomes

Previous alignments of the Altai Neandertal (*2*), Denisovan (*1*), Ust’Ishim (*46*) and Loschbour (*47*) high-coverage genomes and the Mezmaiskaya 1 (*2*), Vindija 33.16, Vindija33.25 and Vindija33.26 (*3*) low-coverage genomes did not include the 1000 Genomes decoy sequences as part of the reference. To make alignments comparable, we realigned these genomes to the same reference used for Vindija33.19 and the new Mezmaiskaya 1 data. For alignment, the data of the high-coverage genomes were used in the processing

described in the original publications, with the exception that steps that lower base-qualities or trim off sequence-ends were omitted (see Genotyping section for details on how the resulting error at sequence ends is counteracted). Low-coverage genomes were realigned using the processed data described in ref.(*2*).

S3: Genotyping

Kay Prüfer* and Cesare de Filippo

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

Ancient DNA contains specific substitutions due to the accumulation of miscoding lesions over time. While the molecules extracted for previous high-coverage ancient genome projects were enzyme-treated to remove most of these miscoding lesions, this step was omitted for most of the Vindija 33.19 and Mezmaiskaya 1 libraries in order to maximize sequence coverage. The resulting large fraction of erroneous substitutions leads to an excess of false heterozygous calls with standard software. To overcome this issue, we implemented a genotyping software, *snpAD*, that incorporates a position-dependent error-profile and estimates the expected probabilities for each genotype by maximum likelihood.

Applying *snpAD* to high-coverage Vindija 33.19 data reduces heterozygous calls to levels similar to previous modern and archaic genomes. Using a large run of homozygosity on the Altai Neandertal chromosome 21, we show that heterozygous calls are specifically reduced in the inbred region, where few to no heterozygous calls are expected. In order to generate a set of comparable calls, we run *snpAD* on the data of Vindija 33.19, Altai Neandertal, Denisova, Ust'Ishim, Loschbour and Mezmaiskaya 1.

Ancient DNA Damage Introduces Erroneous Heterozygous Calls

Previous ancient genomes, including the Altai Neandertal (*2*) and Denisovan (*1*) genomes, have been genotyped using the software package *GATK* (*48*). The data produced for these genomes have been UDG treated (*21*) so that most of the damage associated C to T changes were removed; appreciable fractions of C to T changes only remained at the first base and the last two bases in forward orientation due to a lower efficiency of the UDG enzyme in removing damaged bases at DNA ends (*1, 49*). To reduce the impact of these remaining C to T changes, the end-bases with elevated C to T changes were removed from sequences in the Denisovan data and reduced in base-quality scores in the Altai Neandertal data before genotyping. However, compared to these previous ancient genomes, most of the Vindija 33.19 data stems from libraries that have not been UDG treated and C to T changes remain at frequencies of ~2.5% even in the interior of sequences (Fig S5).

To test whether the high extent of damage associated error leads to false genotype calls, we ran *GATK-UnifiedGenotyper* (version: 1.2-58) on the Vindija 33.19 data on chromosome 21 and compared these genotypes to the published calls in the Altai Neandertal, Denisovan and a present-day African modern human

(Dinka "DNK07"). About 10-fold more heterozygous sites are called in Vindija as compared to the other two archaic individuals and about 5-fold more heterozygous sites than in the African genome. Figure S6a shows that this high heterozygosity in Vindija is primarily due to an excess of C/T and A/G heterozygotes. While heterozygotes are expected to show on average 50% of each allele, the calls for C/T heterozygotes show a highly skewed distribution; the majority of sequences at these sites show more than 80% C and less than 20% T (Fig. S6b). The sequences carrying T are almost always in forward orientation (Fig. S6c), as expected from the single stranded library preparation which preserves the strand orientation of ancient DNA damage (*1, 18*). The mirror images of these signals are observed for G/A heterozygotes.

We also tested whether masking T bases at the ends of sequences is sufficient to remove the bias in calls. For this, we masked T bases within the first and last 6 basepairs on each sequence, leading to 7.9% of all bases to be masked. GATK genotype calls on these masked data still showed an excess of damage-induced heterozygotes (C/T and G/A heterozygotes constituted 42% and 43% of all calls, respectively). Based on these results, we conclude that our previous procedures for calling genotypes cannot be applied to the Vindija 33.19 data.

Determinants of Sequence Quality and Filtering

In alignments of ancient DNA sequences, four sources of error contribute to a base being misrepresented as a different base:

(1) False alignment of unrelated or distantly related contaminating DNA,
(2) alignment of endogenous sequences to a paralogue region in the genome or alignment ambiguity near insertions/deletions,
(3) sequencing error, and
(4) ancient DNA damage.

In addition to restricting the differences in alignment, category (1) and (2) errors can be reduced by requiring a minimum sequence length and mapping quality, and by restricting the analysis to unique regions of the genome. For all samples, we apply a minimum mapping quality of 25 and a minimum sequence length of 35 bases. Only positions that do not overlap a 35mer that aligns with less than one mismatch to another position in the genome are considered (map35_100, SI5b ref.(*2*)). In addition, all sequences were aligned to the decoy containing reference to remove further ambiguous alignments due to regions not represented in the human reference genome (*50*) (see S2). All alignments were post-processed by the indel-realignment implemented in *GATK* (*48*)(version: 1.3.14).

The base-calling procedures used for the Vindija and Mezmaiskaya sequencing data give a base-quality as an estimated sequencing error probability for all sequences. In modern DNA studies these base-qualities are often used directly or in a recalibrated form when calling genotypes(*51*). However, the length of endogenous ancient DNA sequences is often shorter than the read length, and most bases are effectively observed twice by the two mate-pairs, allowing the reads to be merged and increasing confidence in calls. In our Vindija and Mezmaiskaya data, we saw that 90% and 96% of aligning sequences were shorter than the

sequencing read-length of 76 basepairs. Less than 1% of bases had a quality lower than 30 in both samples, demonstrating that most sequences have been merged. In order to test whether these high scores are reflected in the error rates, we compared sequences from chromosome 1 to the Altai Neandertal genotype calls (*2*), excluding sites where Altai is heterozygous or sites outside uniquely mappable regions. The differences to the Altai Neandertal genome are generally low when considering sites with a base-quality greater than 30 and are not strictly monotonically decreasing from this point on (Fig.S7), showing that bases with quality scores ≥ 30 do not consistently decrease in error rates with higher quality scores. Here, we have chosen a fixed base-quality cutoff of 30 to reduce error rates, and do not further incorporate base-qualities into genotyping.

In contrast to sequencing error, substitutions due to ancient DNA damage or amplification will not be reflected in lower quality scores or be reduced by the merging of mate-pairs and duplicates, since the original molecule in the extract, from which the library-molecule is derived, already contained the substitution. However, ancient DNA damage is not equally distributed along the sequences and lead to C->T exchanges particular at the ends of molecules (for single stranded library preparation, as used for all libraries here) (see S2). Strand-orientation and position within a sequence can be considered to arrive at better estimates of the probabilities of substitutions.

Genotype Calling Method

Our method follows a simple probabilistic model for calling genotypes in a single diploid individual (see e.g. (*51, 52*)). At each position in the genome, we observe bases $X = [X_1, \dots, X_n]$ in sequences covering this position. Assuming independence between the observations in different sequences, the probability to observe the bases X when the true genotype is $G \in \{\mathrm{AA}, \mathrm{CC}, \mathrm{GG}, \mathrm{TT}, \mathrm{AC}, \mathrm{AG}, \mathrm{AT}, \mathrm{CG}, \mathrm{CT}, \mathrm{GT}\}$ is $\prod_i P(X_i|G)$. The probability of each individual observation X_i given a genotype G can be further broken down into the individual probabilities for each allele H_1 and H_2 with: $G = H_1H_2$. By assuming that X_i has equal chance to be sampled from each allele we can calculate: $P(X_i|G) = \frac{P(X_i|H_1)+P(X_i|H_2)}{2}$. In this last equation $P(X_i|H)$ corresponds to the probability that X_i was observed when the true base was H. In our implementation these probabilities are supplied as parameters taking into account the position within the sequence and the orientation of the sequence alignment (see below for details on estimating these parameters).

In order to estimate the prior probabilities $P(G)$ for all genotypes, we maximize the likelihood $L(G|Data) = P(Data|G) \propto P(G|Data) * P(G)$. Treating sites as independent this likelihood is the product over all sites and $P(X|G) * P(G)$ at any individual site can be calculated as described above. In practice, we implement this step in C++ using the derivative-free optimization algorithm BOBYQA (*53*) as implemented in the library *nlopt (54)*. In order to reduce the number of free parameters for optimization, we do not estimate the individual probabilities for the homozygous states. Instead, the base composition over all sites is estimated by choosing the base B with the highest probability $P(X|B)$ according to our error model. Since the vast majority of sites are expected to be homozygous in our data, we can use this base composition $P_{[\mathrm{A,C,G,T}]}$ as a

proxy for the fraction of homozygous site of each type and calculate the homozygous priors as $P_{[\mathrm{AA,CC,GG,TT}]} = P_{[\mathrm{A,C,G,T}]} * (1 - P_{het})$ where P_{het} is the sum over the heterozygous priors.

With the prior probabilities, the likelihood of genotype Y can be calculated at each site as

$$L(G_Y|X) = \frac{P(G_Y|X) * P(G_Y)}{\sum_i P(G_i|X) * P(G_i)}$$

We call the best genotype among all and give the ratio of best to second best genotype $\log_{10}\left(\frac{L(G_{best}|X)}{L(G_{2nd-best}|X)}\right)$ as the GQ field in the VCF file and the phred-scaled ratio of the best genotypes likelihood to the sum over all others as the QUAL field. Individual likelihoods are listed in the PP field.

Estimating Error Profiles

In order to estimate the probabilities of base exchanges ("error profiles"), we compared Mezmaiskaya 1 and Vindija 33.19 sequences to sites from the published VCF files of the Altai Neandertal genome that are not heterozygous and pass the filtering criteria (published map35_100 filter with GC-based coverage cutoffs and simple repeats removed). Bases that passed quality criteria (length ≥ 35, MQ ≥ 25, base-Q ≥ 30) and fell within unambiguous regions (map35_100), were used to calculate frequencies of base exchanges in forward-orientation along sequences to the Altai reference separately for the 15 5'-most and 15 3'-most bases and averaged over the bases #16 and #17 from either side. The latter average over four bases is used as a proxy for the exchange rates in the interior of sequences, i.e. those bases that are not within 15 bases of either sequence end.

We also calculated error profiles for previously published high-coverage ancient genomes. Sequences for each genome were realigned to the human reference with decoy sequences (see S2). The published genotypes in VCF format, removing heterozygotes, were used as a reference to compare sequences and error profiles were calculated grouping libraries with similar treatment (see Table S7).

Estimating Priors

To test our implementation, we first used two simulated datasets: the first dataset used a heterozygosity of 8/10,000 (similar to heterozygosity in present-day modern humans) while the second simulated 2/10,000 (similar to heterozygosity in the Altai Neandertal). Both datasets simulated 5 million sites at 20x sequence coverage for each site and used the estimated error profile of the non-UDG treated Vindija data to add substitutions according to randomly drawn positions and strand orientation. The estimated genotype priors deviated by less than 0.6% from the simulated genotype probabilities in both tests[1].

The error profile for Vindija 33.19 was calculated by comparing Vindija sequences with the Altai Neandertal genotypes, while sequences from other high coverage ancient individuals were compared to their

[1] We note that such small-scale differences in the estimated prior will affect the called genotype at a given site only if two or more genotypes are nearly equally probable. With sufficient coverage, such events are expected to be rare.

own, previously published genotypes. Although the differences between Vindija and Altai Neandertals of approximately 3×10^{-4} differences per basepair is about an order of magnitude smaller than the error rate in the enzyme treated Vindija 33.19 library B8744, we tested the effect on estimated priors by simulating the same two datasets with a reduced error profile. This error profile was calculated by subtracting the rate of substitutions between the Altai Neandertal and Vindija 33.19 genotypes (see Genotypes section below). We then estimated priors using the error profile without subtracting divergence, thus giving an overestimate of error rates to the prior calculation. We found that priors deviated by a maximum of 2% from simulated genotype probabilities, showing that a slight misspecification of the error profile has minor consequences on the estimated priors for high-coverage data.

Genotype priors were estimated for all archaic genomes (see Table S8) using filters and error profiles discussed in the last sections. Figure S8 shows that the Denisovan, Altai Neandertal and Vindija Neandertal have comparable estimates of prior probabilities for each genotype. Mezmaiskaya shows reduced estimates for some types of heterozygotes, possibly due to a bias because of the much lower coverage in this individual. Loschbour and Ust'Ishim show higher prior probabilities, as expected for modern human individuals with a higher heterozygosity. The latter two individuals are males and accordingly the estimates for the haploid chromosome X are reduced to between 3.8-7.4% of that of the autosomal average. The Y-chromosome shows higher genotype probabilities in the range of 12-32% of the autosomal average, potentially due to the smaller number of sites and the repetitive structure of the Y chromosome.

Genotyping

Genotypes for all individuals were called using the estimated genotype priors together with the error profiles. In comparison to the *GATK* calls we do not observe a bias towards forward oriented sequences with low-frequency T at CT heterozygote sites or a bias towards reverse oriented sequences with low-frequency A at GA heterozygote sites (Fig.S9). The absence of this signal demonstrates that our approach reduces false heterozygous calls due to damage compared with *GATK*.

To gain further insight into the quality of called genotypes we compared the calls on the high-coverage Altai Neandertal individual between *snpAD* and *GATK* on chromosome 21. This chromosome contains a large region that is largely devoid of heterozygous sites and originates from recent inbreeding in the ancestry of the Altai Neandertal. Due to the close ancestry in the region we expected it to be almost completely devoid of heterozygous sites and called heterozygotes are expected to be largely due to error. When comparing the number of heterozygous calls in the inbred region and outside of the inbred region, we found that *snpAD* shows less calls in the inbred region as compared to the *GATK* calls (Table S9). This reduction in the inbred region is not due to a general reduction in called heterozygotes for *snpAD* since the number of calls outside the inbred region is similar between both genotypers. This suggests that the new method improves the accuracy of heterozygous calls also in UDG treated high-coverage ancient genomes.

General Filters on Genotypes

Similar to the Altai Neandertal genome (SI5b (*2*)), we used additional filters to further reduce the fraction of erroneous calls in all high-coverage individuals. We filtered sites with a coverage of less than 10x and removed sites that fall within the 2.5% extremes of the GC-corrected coverage distribution. Simple-repeats were removed according to Tandem Repeats Finder(*55*) tracks in the UCSC genome browser(*56*) for hg19. Indel differences to the human reference, called by *GATK*, were excluded. The filters significantly reduce the number of calls in the Altai inbred region on chromosome 21 (Table S9; Fisher exact test on the four values in the *snpAD* column: p-value < 2.2e16).

S4: Contamination Estimates

Cesare de Filippo and Kay Prüfer

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

We estimated the extent of modern human contamination among the Vindija 33.19 and Mezmaiskaya 1 sequences. The mitochondrial contamination was estimated to be 1.6% for Vindija 33.19 and 2.0% for Mezmaiskaya 1. Nuclear contamination point-estimates for Vindija 33.19 fall in the range of 0.2-0.3% per basepair and 0.6% contamination (based on 5 observations; 95% confidence interval (CI): 0.2-1.5%) is estimated after genotyping. The new Mezmaiskaya 1 data yield nuclear contamination ranging from 2 – 3%.

Mitochondrial Contamination

To estimate the fraction of modern human sequences among all mitochondrial sequences generated from Vindija 33.19 and Mezmaiskaya 1 libraries, we first determined sites at which the previously published Vindija 33.19 or the Mezmaiskaya 1 mitochondrial sequence differ from 311 modern human mitochondrial sequences. This yielded 76 informative positions for Mezmaiskaya 1 and 81 informative positions for Vindija 33.19. Following previous approaches (see ref.(*30*)), a contamination estimate can then be calculated by counting the sequences matching the Neandertal state as compared to the modern-human base. To avoid misclassification of sequences due to the high rate of deamination, we did not test sequences aligning in forward orientation when either the modern human or Neandertal base was "C" and sequences aligning in reverse direction when either base was "G". We also excluded bases with a base-quality smaller than 30. The Vindija 33.19 mitochondrial contamination was estimated to 1.6% (binomial CI: 1.45 – 1.71%) and Mezmaiskaya 1 contamination to 2.0% (CI:1.62-2.49%) (Table S10). When including the previously sequenced ~0.5x nuclear coverage data of Mezmaiskaya 1 (*2*), the contamination estimate drops to 1.5%. Note that this large difference is driven by the mitochondrial coverage in the previous data set being higher than in the data generated here.

Y-chromosomal Contamination

Since both Mezmaiskaya 1 and Vindija 33.19 are female individuals, sequences matching the human reference Y-chromosome can be used to estimate contamination in the Neandertals originating from modern human males. Therefore, we estimated the coverage on autosomes and chromosome Y for sequences aligning with a mapping quality of at least 25 and within mappable regions (map35_100% track as described in SI5 ref.(*2*)). Since the Y-chromosome is haploid, male autosomal contamination can be calculated as twice the coverage on Y divided by the coverage on the autosomes (see Table S11). Male contamination was estimated to be 0.74% and 2.25% for Vindija 33.19 and Mezmaiskaya 1, respectively. Contamination estimates for Mezmaiskaya 1 do not change substantially when including previous sequence data.

Maximum-Likelihood Estimate of Nuclear Contamination

We used a previously described maximum likelihood method to estimate contamination together with sequencing error (*1*, *2*). Briefly, the method uses reads covering positions at which humans show a fixed derived difference compared to the human-chimpanzee common ancestor. At these positions, the tested archaic genome can either be homozygous derived, heterozygous or homozygous ancestral. At homozygous derived sites, contamination cannot be detected and only error contributes to low-frequency ancestral alleles observed among reads. Contamination at heterozygous sites will increase the frequency of reads showing the derived allele (unless the base is misread due to error). At homozygous ancestral sites, both error and contamination lead to an increase in observed derived variants. By restricting the analysis to bases with a quality of at least 30 and reads with a mapping quality of at least 30 we use this model to estimate two error-rates and their relative proportion alongside population genetic parameters (fraction of sites heterozygous/homozygous) and contamination. Table S12 shows the estimated contamination rates.

Nuclear Contamination Estimate based on Neandertal-Ancestral Sites

Like mitochondrial contamination, nuclear contamination can also be estimated from sites at which the contaminating sequences are expected to differ from true endogenous sequences (see e.g.(*57*)). Here we consider autosomal sites where all modern humans show a fixed derived allele and the Altai Neandertal and Denisova genomes are homozygous for the ancestral allele (see supplementary section 18, ref.(*2*)). We found a total of 22,500 sites that fell into regions of unique alignability ("map35_100%", supplementary section 5b, ref.(*2*)). We further restricted this dataset to sites where a randomly sampled sequence from each of the three low-coverage Neandertals (Vindija33.15, Vindija33.25, Vindija33.26) (*3*) supported the ancestral state in Neandertals, leaving us with a total of 412 informative sites. Using these sites, we estimated contamination as the number of Vindija 33.19 or Mezmaiskaya 1 sequences matching the human derived states over the total number of sequences overlapping these diagnostic sites (Table S13). As for the mitochondrial contamination estimate, we avoid misclassification of sequences due to deamination by restricting our analysis to sequences aligning in reverse orientation when either the ancestral or derived state is C and to sequences aligning in forward orientation when either the ancestral or derived state is G.

We also estimated contamination in our genotype calls based on the list of diagnostic Neandertal sites. Contamination is calculated as the number of chromosomes matching the human allele over the total number of chromosomes. In other words, we count twice the homozygous and once the heterozygous genotypes. This approach yielded an estimate of 0.63% (5 heterozygous sites among 398 informative sites; 95% CI from binomial distribution 0.20%-1.46%). This estimate does not differ significantly from the point estimate when considering individual sequences of 0.29 (Fisher's exact test p=0.18).

S5: Inbreeding and Heterozygosity

Cesare de Filippo* and Kay Prüfer*

* To whom correspondence should be addressed (cesare_filippo@eva.mpg.de, pruefer@eva.mpg.de)

We estimate the autosomal diversity as the number of heterozygous sites over the total number of sites in the archaic genomes, and compare it to that of 15 modern humans from five continents (Africa, Europe, Asia, Oceania and America) recently published (*22*). The Vindija Neandertal carries on average 1.62 heterozygous sites every 10,000 nucleotides, similar to the Altai Neandertal (1.58) but slightly lower than Denisova (1.83). Modern humans are more diverse with heterozygosity ranging from 4.97 to 9.42 per 10,000 nucleotides.

Archaic and some modern human genomes also show long regions (>2.5cM) that are very low in heterozygosity, a sign of inbreeding. Among the three archaic genomes, the Altai Neandertal genome is estimated to contain the largest fraction of bases in regions of homozygosity (641 Mb), followed by the Vindija Neandertal (306 Mb) and Denisova (104 Mb). As described previously, the Altai Neandertal also shows an excess of large regions longer than 10 cM, totaling 380Mb, most likely due to very recent inbreeding (*2*). Only a small number of such large regions are observed in Vindija (36.8 Mb) and Denisova (14.5 Mb). After removing putative regions of inbreeding, the heterozygosity is more similar among the archaic genomes (1.78 in Vindija, 1.99 in Altai, 1.89 in Denisova).

Data and filtering

For both the detection of inbred regions and estimates of heterozygosity, we apply the following set of filters to the archaic genomes by considering only those sites that are:

1) in regions of unique alignability ("map35_100%" track described in SI5b in (*2*)),
2) not falling in simple repeats as annotated by the Tandem Repeat Finder (*55, 58*),
3) not indels
4) passed the GC-corrected coverage filters (see S2).

For the inbreeding analyses, we further filter sites within five base-pairs of indels, sites with a genotype quality lower than 40 (QUAL values in *snpAD* output, see S3), and heterozygous sites that show imbalance between the two alleles (i.e. more than 70% of reads show one allele).

We note that while the latter three filters aim at increasing the confidence in individual called variants, they are not necessarily unbiased in how well they are able to correct errors at heterozygous compared to homozygous sites. For the estimates of heterozygosity, we therefore consider only sites that passed the above four filters in all samples and do not apply further filters on genotype quality, allelic imbalance and a larger distance to indels.

For comparison, we choose three female individuals per continent from the Simons Genome Diversity Project (SGDP) dataset (*22*), apply the alignability and Tandem Repeat Finder filters, removed indels and sites with genotype quality lower than 1 (see further for more details).

Detection of inbreeding

To detect regions of inbreeding, we apply the same method as in (*2*). Briefly, we define a strict run of homozygosity (ROH) as a stretch of sequence longer than 50,000 base-pairs that does not contain any heterozygous site and has at least 50% of sequences after filtering the data. Sequencing errors as well as true mutations can create some heterozygous sites that interrupt runs of homozygosity. For this reason, regions of homozygosity are further collected into homozygous by decent (HBD) tracts by shifting a window of 1 Mb over the genome in steps of 100 kb and merging consecutive and overlapping 1-Mb windows that contain a minimum proportion π of sites in ROH[2]. In order to determine the adequate parameter π, we find the largest π for which HBD tracts do not show significant clustering along chromosomes. To ascertain the significance of the clustering of HBD tracts, for each chromosome we calculate the p-value as the proportion of times the minimum observed distance between/among tracts was shorter than or equal to that generated by 1000 reshuffling of the tracts. Finally, we select the largest value of π where the percentage of chromosomes showing significant clustering is lower than 5% (Figure S10 and Table S14); for samples where there is insufficient number of HBD tracts, we chose the minimum value ($\pi = 0.8$). Physical length was converted to genetic length by using an average recombination rate of 1.3 cM/Mb. Using the African-American recombination map (*59*) did not change our results qualitatively for the comparisons between samples, even though individual tract lengths changed (see Figure S13 and S14).

We note that while our procedure is identical to the one used before, we observe some differences in the results. Upon further investigation, we can attribute these differences to the use of decoy sequences in the alignment and a different genotype calling method (see S3). For the Altai Neandertal, we find that 90% of the HBD tracts overlap with the previous calls and that the length distribution and total length of HBD tracts is comparable between old and new processing and old and new genotyper (Figure S11). No significant differences were found in the distribution of HBD chunks among the different processing (all Mann-Whitney U-test p-values > 0.223) and the total length of the chunks (Figure S11).

Comparison between the new and old release of modern human data.

We use individuals from the SGDP dataset (*22*) to call HBD tracts and estimate heterozygosity in comparison to ancient samples. To ensure that results are comparable to our previous processing, we compare our previous results for Karitiana "SS6004476" to those using the SGDP processed sequence data for the same individual. As for the ancient genome, we also reprocessed the individual by realigning it to the human reference including decoy and calling genotypes using the GATK-*UnifiedGenotyper (60)*. We found that the overall autosomal

[2] The parameter π was called "scan-p parameter" in 2. K. Prüfer *et al.*, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49 (2014).

heterozygosity estimates with the minimal set of filters are very similar: 5.4 and 5.3 heterozygous in 10,000 sites. We identified HBD tracts using a cutoff of genotype quality of 40 for the reprocessed data and 1 for the SGDP processing (*22*). There was no significance difference among the distributions (all Mann-Whitney U-test p-values > 0.542, see Figure S12).

Inbreeding in ancient and modern human genomes.

Figure S13 shows the total length of the putative HBD tracts > 2.5cM for archaic and modern human samples. The Vindija Neandertal, show moderate levels of inbreeding (total 348 cM) that are half of those in the Altai Neandertal (748 cM), higher than Denisova (125 cM), and lower than or comparable to some modern human individuals: Karitiana (459 cM) and Pima (273 cM) from America. Long HBD tracts >10cM, which are indicative of a particularly close relationship between the parents of the individual, are present in the Altai Neandertal in agreement with previous results (*2*), but constitute a small proportion of only 1.3% in Vindija. However, when only considering HBD tracts that are between 2.5 and 10 cM in length, the Vindija Neandertal yields a similar total length of HBD tracts as the Altai Neandertal (Figure S13-14). None of the modern human genomes shows a similar high fraction of the genome in HBD tracts 2.5-10 cM in length. Only the archaic genomes and the modern human American Karitiana and Pima show HBD tracts longer than 10 cM.

Heterozygosity

We calculate heterozygosity as the number of heterozygous sites over the total number of sites (approximatively 793 Mb on the autosomes) that passed the minimal set of filters in all genomes considered: three archaic and 15 modern humans from the SGDP dataset (*22*). The results reported in Table S14 and Figure S15 show that the heterozygosity of Vindija is similar to that of Altai and Denisova, and approximatively between 3.2 and 5.3 times lower than the heterozygosity of modern humans. The heterozygosity estimate of 1.6×10^{-5} for Altai and Vindija translates into an effective population size $N_e \approx 2{,}800$ and the estimate of 1.8×10^{-5} for Denisova into $N_e \approx 3{,}100$, assuming a mutation rate of $\mu = 1.45 \times 10^{-8}$ per generation.

Together with the average nucleotide sequence divergence between Vindija and Altai (Div_{VA}), shown in Table S15, we can estimate $F_{ST} = \frac{Div_{VA} - H}{Div_{VA}}$, where H is the Vindija or Altai heterozygosity estimate. We find that $F_{ST} > 0.4$ independent of the heterozygosity estimate used. This value is larger than reported modern human values between Europe and Asia ($F_{ST} \approx 0.1$) and between sub-Saharan Africa and non-Africans ($F_{ST} < 0.2$) (*61*). However, we caution that the Neandertal F_{ST} estimates may be inflated by recent inbreeding and the different times at which the Altai and Vindija Neandertal were sampled.

S6: Relationship to other Vindija Samples

Kay Prüfer* and Fabrizio Mafessoni

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

We compared the three low-coverage genomes from the Vindija samples 33.16, 33.25 and 33.26, and the chromosome 21 capture data from Vindija 33.15 to the high-coverage Vindija 33.19 genome. High-confidence calls of heterozygous sites in Vindija 33.19 and Vindija 33.15 match in 99% of the cases, suggesting that these two bones carry the same genome. In contrast, the three low-coverage Neandertals show no evidence of sharing the same genome sequence as the high-coverage samples.

Mitochondrial Sequences

Previous results have shown that the Neandertal mitochondrial sequence from Vindija 33.16 (*30*) is likely identical to the incomplete mitochondrial sequence of Vindija 33.26 (*3*). The Vindija 33.19 mitochondrial sequence (*19*) and the Vindija 33.15 mitochondrial sequence (*14*) are 100% identical to the 33.16 sequence. In contrast, the third sample with a low-coverage genome, Vindija 33.25, shows a mitochondrial sequence that differs at 10 positions from these four samples (*62*). Thus, according to the mitochondrial sequence, Vindija 33.15, Vindija 33.16, Vindija 33.19 and Vindija 33.26 could originate from the same individual while Vindija 33.25 is unquestionably from a different individual.

We note that even if Vindija samples with identical mitochondria do not stem from the same individual, the matching mitochondria still indicates a close maternal ancestry. Using a mitochondrial mutation rate of $\mu = 2.67 \times 10^{-8}$ per basepair per year (*46*) and assuming an exponential distribution with rate μl for the probability that a mutation occurs in the mitochondrial sequence of length l, we estimate that identical mitochondrial sequences are separated by less than 6,773 years (95% probability). When using the lower confidence interval for the mitochondrial mutation rate from ref. (*46*) ($\mu = 2.16 \times 10^{-8}$), the separation is estimated to be within 8,371 years with 95% probability.

Comparison between Vindija 33.19 and Vindija 33.15 Genotypes

We used the recently published chromosome 21 capture data from Vindija 33.15 (~33x coverage) and El Sidron 1253 (~12x coverage) (*13*) to call genotypes as described in S3. Briefly, all data were first realigned to the decoy-containing human reference and indel-realigned using *GATK*. The error profile was calculated by comparison to homozygous Altai Neandertal calls and genotype priors were estimated by maximum likelihood for sites within alignable regions and using sequences with confident mapping ($MQ \geq 25$) and bases with confident calls ($BQ \geq 30$). The estimated priors are similar to the priors estimated for the Vindija 33.19 chromosome 21. Genotypes were called and sites were excluded unless they fell within the GC-corrected

central 95% coverage distribution, had a coverage of at least 10 sequences, and did not overlap with tandem repeats.

In order to test whether Vindija 33.19 and Vindija 33.15 are from the same individual, we counted how often high-confidence (QUAL ≥ 40) heterozygous calls in both genotyping datasets match. Approximately 99% of heterozygous calls are identical (see Table S16), suggesting that Vindija 33.15 and Vindija 33.19 contain identical genomes.

Comparison to Low-Coverage Genomes

Green et al. (2010) (*3*) compared positions between two low-coverage genomes, where one sample yielded at least two sequences and the other at least one sequence. Based on the differences within sequences of the same sample in comparison with the differences to the other sample they concluded that all three samples carry different nuclear genomes. Here we modified the approach to test whether any of the three low-coverage genomes shows sufficient similarity to the high-coverage Vindija 33.19 genome to indicate identical genome sequences. For this, we use high-confident Vindija 33.19 heterozygous sites (QUAL ≥ 40). When these heterozygous sites overlap two sequences from a low-coverage genome (with basequality≥30 and MQ≥25) the site is counted as matching if the two sequences show the two alleles of the heterozygous site and as non-matching otherwise. If the genomes of the low-coverage sample and the high-coverage sample were the same, and in the absence of sequencing error, we would expect that the two sequences match with a probability of 0.5. None of three low-coverage Vindijas nor Mezmaiskaya 1 shows sufficiently high matching rates to indicate that the genomes are identical (Table S17; Fig. S16). As a positive control, we also included a 0.4x coverage subsample from Vindija 33.19 (library A9368, non-UDG treated) and observed signals close to 0.5. However, the confidence intervals for the estimate do not include 0.5 (Fig. S16). The factors that likely contribute most to this difference are (i) errors in the low-coverage sample or the genotype calls which reduce the sharing of heterozygotes, (ii) reference-bias in the alignment of sequences which reduces the chance to observe both alleles, and (iii) human contaminating sequences.

S7: Population Size Inferences, Split Times and Branch Shortening

Fabrizio Mafessoni* and Kay Prüfer*

To whom correspondence should be addressed (fabrizio_mafessoni@eva.mpg.de, pruefer@eva.mpg.de)

We estimated the effective population size over time using PSMC for the Vindija33.19 Neandertal, Altai Neandertal and Denisovan genomes. All three genomes show a similar demographic history characterized by a strong reduction in Ne after the estimated split with human populations.

The age of the archaic samples is expected to lead to a shortening in lineage length compared to present-day human genomes, since new mutations ceased to be accumulated when the archaic individuals died. We found that all archaic lineages show branch-shortening. The Vindija sample is estimated to be about 50 ky old (assuming $\mu = 0.5 \times 10^{-9}$/bp/year) and thus 60-70ky younger than the Altai, and ~20 ky younger than the Denisova finger bone, under the assumption of an unchanging mutation rate on all lineages.

Based on the F(A|B) measure of divergence together with the estimated population sizes and the branch shortening, we estimate that the split between Vindija and Altai occurred around 140 ky before present (assuming a 13Mya human-chimpanzee split), and that Mezmaiskaya 1 split from Vindija around 80-100kya. Split estimates between Neandertals and Denisovan, and from modern humans are compatible with previous estimates.

Population history

In order to estimate the demographic history of the Vindija individual compared to those of the other ancient genomes we applied the Pairwise Sequential Markovian Coalescent (PSMC) method (*63*) to the genotype calls (applying the recommended filters for coverage, tandem repeats and indels as descibed in S3). Since PSMC inferences rely on the distribution of heterozygous sites across the genome, errors that lead to false heterozygous calls are expected to have a disruptive effect. To guard against false calls, previous studies have restricted their analysis to regions of the genome that show little similarity to other regions in the genome (*2*, *47*). Here, we test two such mapability filters: map35_50, which retains sites in the genome where at least 50% of all overlapping 35mers do not align elsewhere in the genome allowing for up to one mismatch, and map35_100, which requires all overlapping 35mers to align uniquely with up to one mismatch.

We find that archaic genomes show a higher heterozygosity in regions that are specific to the less restrictive map35_50 filter compared to the map35_100 filter (Table S18). In addition, PSMC estimates of effective population size over time consistently show a higher effective population size for recent times with map35_50 compared to map35_100, consistent with more interspersed erroneous heterozygous calls in the less restrictive mapability filter. Part of this signal is removed when filtering for higher genotype quality. However, this filter is not sufficient to remove the entire signal when using the looser map35_50 filter, while

a combination of genotype quality and map35_100 filter yield estimates without high effective population size for recent times (FigS17). Based on these observations, we deem the map35_100 filter more effective in removing false heterozygous calls.

The two mapability tracks show also differences in the estimated effective population sizes at older times, with the more conservative map35_100 filter yielding consistently smaller effective population sizes compared to the less stringent map35_50 filter (Fig.S18). These differences are observed for the Altai Neandertal with two different genotypers (snpAD and GATK) and independent of whether decoy sequences were used in the alignment (FigS18b). Present-day human sequences also exhibit a consistent difference between the two mapability filters (FigS18a), showing that the effect is not specific to ancient DNA.

One possible explanation for the consistently smaller estimated effective population sizes may be a smaller mutation rate in the regions retained by the more stringent mapability filter. To estimate the difference in mutation rate between the two filters we use a quantity that is expected to be proportional to the mutation rate and calculated the rate of nucleotide sequence differences within map35_50 and within map35_100 regions between the reference human genome (hg19) and the chimpanzee genome (pantro2) using the whole genome alignments from UCSC. The estimated rate of difference is 0.0123 bp^{-1} with map35_100 and 0.0126 bp^{-1} with map35_50, showing that the estimated mutation rate difference of 2-3% between the two tracks is too small to account for the differences in the estimated effective population sizes.

Since neither processing nor mutation rate differences can account for the different results, we next sought to test whether similar effects are observed in simulated data. For this, we ran the coalescent simulator *ms* (*64*) using the published estimate of effective population sizes over time for the Altai Neandertal to generate sequences of identical sizes to those of the human reference chromosomes. We then applied the map35_50 and map35_100 filter to these simulated sequences. Surprisingly, the filtering recapitulates the observed differences between the two mapability tracks (Fig S18c). Since the true effective population size is known for the simulation, we are also able to determine that even the more permissive mapability filter leads to an underestimate of effective population sizes. We conclude that the fragmentation of data when filtering for mapability leads to a bias in PSMC estimates.

In order to correct for this bias we devised a simulation-based approach that involves the following steps:

1. Demographies are estimated using PSMC with the filtered data.
2. The estimated demography is re-scaled by multiplying θ with factors ranging from 0.9 to 2 (step-size 0.025).
3. The rescaled demographies are simulated using *ms* and the resulting sequences are filtered identically to the real data.
4. PSMC is estimated on all filtered simulations.
5. The scaling parameter for the simulation with the least-squares-fit to the real data is used to correct the estimate of the real data.

The procedure aims at identifying a simulated scenario that, when filtered, would give rise to an estimated PSMC demography that closely matches the one estimated from the filtered real data (Fig.S19). The least-

squares fit in step 5 compares values over 10,000 time points following the density of an exponential distribution over time to match PSMCs resolution in time. The least-squares distances are then linearly interpolated in order to increase the resolution in estimating the rescaling factor. This approach estimates rescaling factors of 1.351515, 1.30303 and 1.29697 for Altai, Vindija33.19 and Denisova, respectively, that we apply to the demographies to correct for the bias introduced by filtering. The resulting estimates for the Altai Neandertal and Denisovan individual are similar to those previously reported (Fig S20). Command lines to generate simulations using the corrected demographies are listed in Figure S21.

The PSMC estimates after correction reveal similar demographic histories for both Neanderthals and the Denisova, who share a recent history of very low effective population sizes.

Branch shortening

When comparing the PSMC estimates for the three archaic individuals, we notice that the curve of the Altai individual is shifted leftward compared to both Vindija and Denisova. This observation is consistent with this individual being older, hence having a shorter history of low recent Ne. To estimate the age difference between the two samples, we aligned the Altai and the Vindija PSMC curves, similarly to previous approaches (*46*), and estimated a difference in time of 62 kya (Fig.S22). This age difference is compatible with previous estimates of the age of the Altai sample of over 100ky.

In order to gain further insight into the age differences between the archaic humans, we count the number of derived changes on each lineage since the common ancestor of humans with the great apes. Since derived variants accumulate over time, a smaller number of derived changes indicates a shorter lineage, i.e. an older age for the sample material.

We infer the ancestral state for a given base in the human genome by requiring that the human alignments of the reference genomes of chimpanzee, bonobo, gorilla, orangutan and macaque, are present and identical at that nucleotide. A difference to this ancestral state in any of the three archaic humans, and the three ancient modern humans Ust'-Ishim (*46*), Loschbour and Stuttgart (*47*) are counted. In order to arrive at a date estimate, we also require a present-day modern human genome as a calibration point. For this, we use the Mbuti individual (S-Mbuti-2) from the SGDP dataset (*22*) with a quality-value≥ 0. With the number of derived changes in Mbuti, n_M, and the derived changes in an archaic or ancient modern human, n_A, we calculate the relative age of the individual as $(n_M - n_A)/n_M$, and arrive at an estimate in years by multiplying this quantity with an assumed divergence time to the common ancestor of humans and chimpanzee of 13 million years. We note that this divergence time corresponds to a mutation rate of 0.5×10^{-9} per base pair per year (1.3% human-chimpanzee sequence divergence divided by 2×13 million years). This mutation rate is in better agreement with directly observed generation times (*65, 66*) and per generation mutation rates from sequencing of parent-offspring trios in humans and chimpanzees (*67, 68*) than the previously used mutation rate estimate of 1×10^{-9}. This higher mutation rate would make our estimates of branch shortening and split times half as large, i.e. they can be brought into concordance with the higher mutation rate by dividing them by 2.

In order to compute the variance in the dating estimates we calculated branch shortening in windows of 5Mb and computed bootstrap confidence intervals as 1.96 × standard error around the estimated dates. As reported previously, branch shortening estimates are strongly affected by the quality and filtering of the data (*1*). In order to minimize biases due to the different quality of the genomes we restricted our analyses to regions retained after applying the basic filters described in S3. We then tested four different filtering strategies in addition to the basic filtering: no additional filters, genotypes with QUAL ≥ 60, intergenic regions (excluding exons and the surrounding 5kb using the Refseq Genes track downloaded from the UCSC browser), and counting only transversions as derived changes (Table S19, Fig. S23). The estimates indicate that the Vindija33.19 sample is between 60 and 70ky younger than the Altai individual, consistent with the PSMC curve alignment estimates of the age difference, and between 20 and 30kya younger than the Denisovan individual.

When comparing the estimated age of Vindija 33.19 with that of Ust'Ishim, we find consistently younger point-estimates for the age of Ust'-Ishim compared to Vindija. The calibration of the radiocarbon date of Vindija yields only a minimum age 45.5kya (see S1), while Ust'Ishim was radiocarbon dated to 45.0kya, calibrated. The branch-shortening estimates thus suggest that the Vindija sample is older than the Ust'Ishim sample.

The analysis of branch shortening also shows an apparent excess of changes in Loschbour and Stuttgart when considering all sites, most likely due to the lower quality in contrast to a present-day sample. However, further filtering removes this apparent excess, and estimates using only transversions are closer to published radiocarbon age-estimates of ca. 8000 years for Loschbour and ca. 7000 years for Stuttgart. Figure S24 shows branch shortening estimates separately for different types of substitutions; Loschour and Stuttgart show an excess of substitutions for three types of transitions (A->G,G->A,C->T) in addition to other types of substitutions.

Since estimates based on transversion are closest to the reported radiocarbon dates for Ust'Ishim, Loschbour and Stuttgart, we use age-estimates based on transversions to transform split-time estimates between two ancient samples into estimates before present. Figure S25 shows the PSMC estimates corrected for branch shortening.

Population split dates

In previous ancient genome papers (*1-3*) population split times have been estimated by calculating F(A|B), i.e. the proportion of sites in which an haploid genome from an individual A is derived in respect to a common ancestor, given that a reference genome B is heterozygous at the same site. This statistic is expected to decrease with time, since new mutations emerge on the B lineage that generate heterozygotes with derived alleles that are not shared with A; shared derived sites that were heterozygous in B are also lost due to drift on the B lineage. The decrease thus depends on the specific demography of the reference lineage B, while demography on lineage A does not influence results. We use coalescent simulations generated with *ms* and based on our corrected PSMC estimates (see Fig. S21) to estimate the expected F(A|B) values given a specific split time

(Fig. S26). The observed F(A|B) values are then fitted against these calibration curves in order to obtain the time of population. In order to avoid recurrent mutations, F(A|B) is calculated using only transversions (*2*).

Table S20 shows the split-time estimates among the three archaic humans and to a present-day Mbuti individual. The estimated split times between the archaic humans and modern humans range from 522-634kya and overlap the previously published range of 550-765kya (*2*). The range of point estimates for the Neandertal-Denisovan splits (392-438kya) fall within the range of previous estimates (381-473kya).

As expected, Vindija33.19 is more closely related to the Altai Neandertal than it is to the Denisovan, and the last common ancestor of both Neandertals was estimated to have lived 130-145kya. Due to the fact that the F(A|B) measure yields split times relative to the length of the B lineage, we are also able to compare the two estimates F(Altai|Vindija) and F(Vindija|Altai) to test whether the split time estimates reflect the relative difference in branch length. The split times are ~80 kya and ~20kya before branch shortening, consistent with a relative age difference of ~60ky between the two Neandertals. We also compared F(Neandertal|Denisova) and F(Denisova|Neandertal) to estimate the relative age of Denisova to the Vindija and Altai Neandertals. We find that F(Vindija|Denisova) < F(Denisova|Vindija), consistent with an older age of Denisova compared to Vindija, and that F(Altai|Denisova) > F(Denisova|Altai), consistent with an older age of Altai compared to Denisova. The age differences are ~50ky and ~5ky for Vindija-Denisova and Denisova-Altai, respectively, and do not agree well with the branch shortening estimates.

We also calculated the split times between Mezmaskaya1 on the one hand, Vindija33.19, Altai, Denisova and Mbuti. Since neither demography nor heterozygous sites can be estimated reliably for the lower coverage Mezmaiskaya genome, we calculated the F(A|B) statistics using only the high coverage archaic genomes as individual B. As before, only transversion substitutions were considered. We estimated split times separately using the Mezmaiskaya genotypes (S3), by sampling a random sequence at each position, and by randomly sampling sequences that show evidence of deamination (see S11). Estimated split times are very similar for the genotypes and for the random read sampling. However, restricting the analysis to randomly sampled deaminated reads yields consistently higher split times in all comparisons, indicating that the selection of deaminated reads introduced a bias in the estimates. Nevertheless, all processing-variants for the Mezmaiskaya1 data show a closer relationship between Mezmaiskaya and Vindija33.19 than between Mezmaiskaya and Altai, and estimates of split times to Denisova and Mbuti fall within the range observed for the Vindija and Altai Neandertal (see Table S21).

Finally, we calculated the split times from two ancient Eurasians, Ust-Ishim and Loschbour, with evidence of human-Neanderthal admixture (Table S22). In both cases we observe more recent split times to Vindija33.19 than to Altai, congruent with a closer relationship of the Vindija Neandertal with the introgressing Neanderthal. Estimated split times for Ust'Ishim and Loschbour tend to be also more recent compared to Mbuti, consistent with Neandertal gene flow into these populations.

S8: Relationship of Modern Humans to Archaic Humans

Steffi Grote, Pontus Skoglund, Kay Prüfer* and Michael Dannemann

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

To investigate the relationship between modern humans and the archaic humans, we calculated D-statistics comparing the Denisovan individual, and the Altai and Vindija Neandertals to individuals from different present-day modern human populations. We found that both Neandertals share more derived alleles with all modern human populations as compared to Denisovans. The Vindija Neandertal shares more derived alleles with Out-of-Africa populations as compared to the Altai Neandertal. The higher allele sharing with the Vindija Neandertal is to a smaller extent also observed in comparisons with sub-Saharan Africans, possibly due to a small amount of Out-of-African ancestry in these populations. Neandertal ancestry estimates in present-day non-African populations outside Oceania range from 1.8% to 2.6%.

Datasets and D-Statistics Calculation

We used the genotype calls of the Denisovan genome and the Altai and Vindija Neandertal genomes as described in S3. Sites that fall into tandem repeat regions or that did not pass the GC-corrected coverage filter for all three archaic genomes were excluded (see also S3). Modern human genotypes were extracted from 271 individuals of the Simons Genome Diversity Panel (SGDP) (*22*), 2504 individuals of the 1000 Genomes Project Phase III release (*69*), and 35 Papuans from Vernot et al. (*70*). Low quality genotypes were removed from all three datasets (SGDP genotypes labeled with quality 'N', 1000 Genomes not labeled with 'Filter=PASS', and Papuan genotypes marked as 'Filter=LowQual'). The genotypes of the ancient modern humans Loschbour (8,000 ybp, Luxembourg) (*47*), and Ust'Ishim (45,000 ybp, Siberia) (*46*), described in S3, were added to the SGDP dataset.

The ancestral state for any given site was defined as the allele present and identical in the aligned genomes of chimpanzee (pantro4), and orangutan (ponabe2). We prepared two datasets, combining the archaic individuals with: (i) SGDP individuals, Loschbour and Ust'Ishim, and (ii) individuals from the 1000 Genomes combined with the 35 Papuans. Indels and non-bi-allelic sites (i.e. sites where more than one alternative to the ancestral state is observed) were excluded separately in each of these two combined datasets.

D-statistics (*3, 23, 71*) were calculated by comparing archaic individuals against collections of modern human genomes that were grouped by geographical origin. Genotype calls in the archaic humans were assigned values of 0, 0.5 and 1 for homozygous ancestral, heterozygous and homozygous derived states, respectively. Modern human populations were assigned the derived allele frequency. D-statistics were then calculated following Durand et al. (*72*). With the derived frequencies f_A, f_B and f_C, D was calculated as:

$$D = \frac{\sum f_A(1-f_B)f_C - (1-f_A)f_B f_C}{\sum f_A(1-f_B)f_C + (1-f_A)f_B f_C}$$

Similar to previous studies (*1-3, 71*), standard errors were calculated using a weighted block jackknife procedure (*73*) over all autosomes, with each divided into 20 equally sized blocks. The blocks were weighted by the denominator of the D-statistic.

Neandertal-Denisovan comparison

The D-statistics D(Neandertal, Denisovan, ModernPop, Chimpanzee + Orangutan) show that all modern human populations share significantly more derived alleles with each of the two Neandertals than with the Denisovan (with both human datasets: D > 6.6, Z > 12; Figure S27). The Vindija Neandertal shows slightly, but significantly, higher allele sharing with modern human populations compared to the Altai Neandertal (on average 0.5% more, Wilcoxon signed-rank test p < 1e-9,Figure S27; see next section for a direct comparison of Vindija and Altai). Most human populations give similar values for D in the range of 9.4% to 10.8% (Figure S27). The excess of derived allele sharing of the Neandertals with non-Africans compared to the Denisovan can be explained by Neandertal introgression into the ancestors of non-Africans (*2*).

Oceanians have lower D-values than other non-Africans (mean D = 8.0%), which is likely caused by an additional component of Denisovan ancestry (*1, 24, 74*) that lowers the excess of derived allele sharing with Neandertals compared to the Denisovan. Note that, although Oceanians have higher Denisovan than Neandertal ancestry, the Denisovan sample is thought to be much more divergent from the introgressing Denisovan population than the Neandertal samples are from the introgressing Neandertal population (*2*), resulting in an excess of matching alleles to Neandertals versus Denisovan.

The lowest values for D are observed in Africans (mean D = 7.0%). An excess of Neandertal allele-sharing with some specific African populations compared to the Denisovan could be due to low levels of Out-of-African ancestry in African populations (*75-77*). However, we observe a consistent signal of more allele sharing with Neandertals than with the Denisovan across all African populations (Figure S27). This signal has been previously observed in the initial analysis of the Altai Neandertal genome (D(Altai, Denisovan, African, Chimp) = 7%, Z > 11) (*2*) and interpreted as a sign of admixture between Denisovans and an unknown archaic hominin which diverged 0.9 - 1.4 million years ago from the lineage that led to humans, Neandertals and Denisovans, thereby increasing the genetic distance of Denisovans to modern humans (*2*). Another source could be gene flow from the ancestors of modern humans into the Neandertal as also previously suggested (*13*).

Neandertal-Neandertal Comparison

D(Vindija, Altai, ModernPop, Chimpanzee + Orangutan) shows that the Vindija Neandertal has an excess of derived allele sharing with modern humans compared to the Altai Neandertal, with an average D of 4.1% for all non-African populations (Figure S28). This excess of allele sharing suggests that the Vindija Neandertal is

more closely related to the introgressing Neandertal population than the Altai Neandertal. In non-African populations D ranges from 2.8% (North Ossetian, Caucasus) to 5.2% (Bougainville, Oceania) (Figure S28).

The levels of allele sharing with the Vindija as compared to the Altai Neandertal correlate with the estimated fraction of the genome called as Neandertal-introgressed (Pearson's $r = 0.55$, $p < 1e\text{-}9$ for the correlation with SGDP Neandertal ancestry estimates for non-Africans from (*78*) and Pearson's $r = 0.48$, $p < 1e\text{-}6$ for the correlation with Neandertal estimates for non-Africans by F4-ratios (see below)) (Figure S29). This correlation could be explained by a single introgression event from a Neandertal more closely related to the Vindija than to the Altai Neandertal, followed by drift or natural selection leading to population differences. Alternatively, the population differences may stem from multiple independent introgression events from Neandertal groups that were overall more closely related to the Vindija Neandertal than to the Altai Neandertal, or could be explained by some populations receiving admixture from a source that carried less Neandertal ancestry leading to a dilution of the signal of sharing with the Vindija Neandertal.

Among populations of African ancestry from the 1000 Genomes dataset the highest signals of allele sharing differences between the two Neandertals were observed in the two African American groups 'African Ancestry in Southwest US' (ASW, D = 1.9%, Z = 2.3) and 'African Caribbean in Barbados' (ACB, D = 1.5%, Z = 1.9), which is likely a result of non-African ancestry in these populations (Figure S28B) (*69*). Non-African ancestry is likely also the reason for comparatively high values of D for Sahrawi, Somali, Mozabite and Masai among the African populations from the SGDP (D = 2.1% - 3.1%, Z = 2.3 - 3.3; Figure S28C, Figure S33). Interestingly, D is larger than zero for all comparisons with African populations, although this difference is not significant in most comparisons (Figure S28). This consistent trend is unlikely to be caused by quality differences between the Altai and Vindija genomes since the lower-coverage Vindija-individual would be expected to be attracted to the outgroup, as illustrated by the statistic D(Altai, Vindija, Chimpanzee, Orangutan) which yields a D-value of 4% (Z = 6.8).

We speculate that the higher levels of derived allele sharing of Africans with the Vindija Neandertal could be due to a low fraction of introgressed Neandertal DNA in these populations, possibly due to back-migration of Eurasian groups to Africa (*75-77*). Alternative scenarios include gene flow from modern humans into the ancestors of the Vindija Neandertal, to a higher extent than the recently proposed gene flow into the Altai Neandertal (*13*), or gene flow from a more divergent hominin into the Altai Neandertal, like proposed for the Denisovan sample (*2*) but to a lower extent. However, we note that the signal is weak and further data will be needed to test whether Africans truly share more derived alleles with the Vindija Neandertal than with the Altai Neandertal.

Comparing Modern Human Populations to Neandertals

We used the Altai and Vindija Neandertal genomes to compare the degree of Neandertal-allele-sharing among modern human populations (Figure S30). The most pronounced signal of the D-statistic D(ModernPop1, ModernPop2, Neandertal, Chimpanzee + Orangutan) recapitulates the observation that non-African populations share significantly more derived alleles with Neandertals than African populations, with D-values between 3.6% and 8.0% ($Z > 14$). These values are in line with previous studies that used the low-coverage

Neandertal draft genome and the high-coverage Altai Neandertal genome and reported D-values between 4.2% and 7.3% for the comparison between African and non-African individuals (*2, 3*).

Our results are in agreement with the previous findings of higher Neandertal ancestry in East Asians compared to Western Eurasians and South Asians (Figure S30), which have been proposed to be the result of additional pulse(s) of Neandertal admixture into the ancestors of East Asians (*8, 25, 79, 80*). The SGDP dataset shows no significant differences in allele sharing with the Neandertals between East Asians and Central Asians / Siberians (Figure S30A,B) indicating a similarly increased Neandertal ancestry in these populations.

The highest allele sharing with Neandertals was observed in Oceanians (Figure S30), likely due to the additional gene flow from Denisovans into these populations (*1, 24, 70, 78*): Denisovans are a sister group of Neandertals, and the elevated D-values for Neandertal ancestry could be explained by the shared ancestry of Denisovans and Neandertals. See below for additional tests in support of this explanation.

The results obtained with the Vindija Neandertal and those obtained with the Altai Neandertal are matching closely, as indicated by strong positive correlation (pooled SGDP and 1000 Genomes: Pearson's $r > 0.999$ and $R^2 = 0.991$ for the model $D_{altai} = D_{vindija}$), which is higher than the correlation between results from the Altai Neandertal and the Denisovan (Pearson's $r = 0.83$, $R^2 = 0.6$).

Comparison of Modern Human Populations to the Denisovan

The amount of allele sharing of modern human populations with the Denisovan yields similar results to the D-statistics with the Neandertals, although with lower values for D. These results are not surprising given the close relationship between Neandertals and Denisovans, leading to similar allele sharing results with the Denisovan. However, this general pattern does not hold for Oceanians where sharing compared to other modern humans is larger for the Denisovan than for the Neandertals, with a mean absolute D-value of 5.7% for the Denisovan, compared to 3.6% for the Neandertals (Figure S30, Figure S31).

Interestingly, we observe no significant differences in derived allele sharing with the Denisovan between Central, East, and South Asian populations ($D < 0.4\%$, $Z < 1.7$, Figure S31), even though South Asians showed less derived allele sharing with Neandertals than other Asian populations ($D = 0.4\%$-0.9%, $Z = 1.5$-3.4, Figure S30). Assuming that most of the signal is driven by Neandertal ancestry, a higher amount of Denisovan ancestry in South Asians as compared to other Asian populations may be masked in this statistic by a lower amount in Neandertal ancestry (*2, 24, 74, 78*).

Neandertal ancestry estimates

We used the F_4-ratio statistic (*1, 23, 24*) to estimate the amount of Neandertal ancestry in SGDP subpopulations, the 8,000 year old European hunter-gatherer Loschbour (*47*), and Ust'Ishim, a 45,000 year old modern human from Siberia (*46*). The F_4-statistic is the numerator of the D-statistic, which describes the excess of derived allele sharing with a Neandertal *N* of population *A* compared to population *B*, using the derived allele frequencies f_N, f_A and f_B:

$$f_4(A, B, N, outgroup) = \sum f_A(1 - f_B)f_N - (1 - f_A)f_B f_N$$

As with D-statistics, the aligned chimpanzee (pantro4), and orangutan (ponabe2) genomes were used as outgroups to infer the ancestral state at a given site. To obtain ancestry estimates $\hat{\alpha}$ for a population *A*, the excess of derived allele sharing of *A* with the Altai Neandertal, compared to a population *B* which is assumed to carry no Neandertal ancestry, is normalized by the excess of allele sharing between the two Neandertal individuals Vindija and Altai compared to population *B*:

$$\hat{\alpha} = \frac{f_4(A, B, Altai, outgroup)}{f_4(Vindija, B, Altai, outgroup)}$$

Note that exchanging Altai and Vindija in this equation would include in the numerator genetic drift that occurred after the Vindija-Altai split and before the split between Vindija and the introgressing Neandertal population, whereas the denominator would only include drift up to the Vindija-Altai split. This would lead to an overestimate of $\hat{\alpha}$, which is avoided in the orientation used here for the calculation of ancestry estimates.
We computed ancestry estimates for all subpopulations of the SGDP, Loschbour, and Ust'Ishim, using Mbuti as population *B*, since Mbuti was inferred to be the SGDP population with the least Neandertal ancestry by D-statistics. Standard errors were computed using a weighted block jackknife (*73*), with the weight set to the sum of informative sites (sites where the numerator and/or denominator of the f_4 statistic is not zero).

Our estimates of Neandertal ancestry range from 1.8% to 3.2% in non-African populations (Figure S32A). Exceptionally high values are obtained for the three Oceanian populations Australian, Bougainville Islanders and Papuans (3.1% - 3.2%), and are likely due to the substantial amount of additional Denisovan ancestry in these populations (*1, 24, 70, 78*). Among the non-Oceanian populations, the highest estimate was obtained for Ust'Ishim (2.8%), followed by Altaian and Loschbour (both 2.6%). The ancestry estimates for the Loschbour and Ust'Ishim individuals remain high when restricting the analysis to transversions and match the previously observed trend towards higher Neandertal ancestry in individuals from the past (*81*). Similar to the D-statistic estimates (Figure S30), Central and East Asians show higher values (on average 2.5% and 2.4%) than South Asians and West Eurasians (on average 2.2% both; Mann-Whitney-U $p < 1e\text{-}10$).

Among the African populations Mozabite, Sahrawi, Somali and Masai have the highest ancestry estimates (1.7%, 1.7%, 0.8%, and 0.4% respectively). The remaining African populations are estimated to have less than 0.4% Neandertal ancestry. We find a strong correlation between the amount of European ancestry in Africans (estimated by D(French, Han, African, Chimpanzee + Orangutan)) and their estimated amount of Neandertal ancestry (Pearson's $r = 0.99$, $p < 1e\text{-}14$), providing further evidence for European admixture as the (major) source of Neandertal ancestry in Africans (Figure S33) (*75-77*).

In general, the F_4-ratio ancestry estimates are higher than those reported in the Altai Neandertal publication, which ranged between 1.6% and 2.3% (*2*). This deviation could be caused by a combination of a less admixed human reference population (in (*2*) the reference was a combination of Dinka, Mbuti and Yoruba individuals), a different outgroup (in (*2*) the outgroup was the Denisovan genome), and the higher quality of some of the archaic genomes compared. Our Neandertal ancestry estimates correlate strongly with the fraction of the genome that was assigned to originate from Neandertal introgression by Sankararaman et al. (*78*) (Pearson's $r = 0.84$, $p < 1e\text{-}29$; Figure S34).

To further study the effect of Denisovan ancestry on Neandertal ancestry estimates, we restricted our analysis to sites where the Denisovan genome is homozygous ancestral, removing around half of the informative sites. The Neandertal ancestry estimates are lower for Australians, Papuans and Bougainville Islanders (by -1.7% to -1.3%), but differ only slightly from previous ancestry estimates for the remaining populations (by -0.5% to 0%) (Figure S32B). These lower Neandertal ancestry estimates of Australians, Papuans and Bougainville Islanders (1.6%, 1.7% and 1.8%) are similar to those in other Oceanian populations (Maori, Hawaiian, Dusun, Igorot: 1.8% - 2.1%). This suggests that Denisovan ancestry in these three populations is the primary cause for the high Neandertal ancestry estimate. Our Neandertal ancestry estimates still correlate strongly with previous estimates when restricting to Denisovan ancestral sites (Pearson's $r = 0.67$, $p < 1e\text{-}14$; Figure S34).

S9a: Testing for Signals of Human Admixture into Neandertals

Kay Prüfer*

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

Kuhlwilm et al. (2016) (*13*) reported a signal of modern human admixture into Neandertals using the Altai Neandertal genome. Further analysis on chromosome 21 captured data suggested that the recipients of this admixture were eastern Neandertals, represented by the Altai Neandertal, but not western Neandertals, represented by Vindija and El Sidron Neandertal DNA libraries. Comparing the Altai Neandertal and the Vindija Neandertal genome we found no evidence for a closer relationship of the Altai Neandertal to Africans. The absence of such a signal is incompatible with admixture of a modern human related population admixing exclusively into eastern Neandertals, but is compatible with admixture into both Neandertals.

We also compared divergence among Archaics and Africans in windows of 100kb and 0.1cM. For the most deeply divergent windows to Africans, Denisovans show an increased divergence to Neandertals. In contrast, in regions that are most similar to Africans, Neandertals are more different to Denisovans. These signals are stable, observed with different measures of divergence, and in agreement with the observations in (*13*). In contrast, we do not observe consistent differences in heterozygosity in Neandertals or Denisovans at deep or shallow divergence to Africans. However, coalescent simulations carried out here indicate that the window-based analyses is susceptible to branch-shortening differences between archaic individuals and effective population size differences and that results have to be interpreted with caution.

Datasets

We used the Vindija, Altai and Denisovan genotypes described in S3. Sites in these archaic genomes were filtered to pass coverage-criteria and to be outside of tandem repeats (see S3). The 1000 Genomes Phase III data for Luhya and Yoruba (*50*) were used to call sites that are polymorphic in Africans; sites that are not listed as polymorphic were assumed to be fixed in Africans and identical to the human reference. In addition, we used the previously published high-coverage genomes of an Mbuti, a San, a Yoruban, a French, a Papuan and a Han individual (panel B from (*2*)). Whole genome alignments to the human reference genome (GRCH37, hg19) produced by the UCSC genome browser (*56*) for the chimpanzee, gorilla, orangutan and rhesus macaque genomes (*82-85*) and alignments produced in-house for the bonobo genome (*86*) were used individually or together to call the ancestral state in analyses requiring phased variants.

D-Statistics

To gain insight into the relationship between Archaics and modern humans, we calculate D-statistics (*3*, *23*) using the high-coverage genomes of the Altai, Vindija and Denisovan, three African, three non-African individuals and chimpanzee as an outgroup. One allele was sampled at random from the genotypes of each individual and standard errors were calculated using a weighted block-jackknife procedure (*73*) over windows of 5Mb along the reference genome. In a comparison of the form $D(A,B,X,O)$ positive values indicate an excess of sharing of derived alleles between *A* and *X*, while negative values indicate an excess of sharing between *B* and *X*.

Comparing the three African genomes, we find little difference between Africans in their allele sharing to the Archaics (Table S23). However, the comparison Mbuti – Yoruba shows a consistent signal of increased sharing (|Z|>2).

Under a scenario in which the Altai Neandertal received gene flow from a population closely related to present-day humans while the Vindija Neandertal did not, we would expect that Africans show an increase in allele sharing with the Altai, but not Vindija. However, we find no significant difference in sharing with Africans (Table S25). In contrast, both Neandertals share significantly more derived alleles with Africans compared with Denisova, compatible with two different, not mutually exclusive scenarios: 1) superarchaic admixture reducing the sharing of derived alleles between Denisovans and the other two groups or 2) modern human-related admixture into Neandertals increasing the sharing between these two groups.

Since the signal of modern-human related admixture into the Altai Neandertal but not Vindija from (*13*) was based on the comparison of chromosome 21, we also recalculated D with the re-genotyped chromosome 21 capture data (see S6) in all pairwise comparisons among Neandertals and found no significant differences in sharing with Africans (Table S26). Note that the Vindija 33.15 (capture) and Vindija 33.19 (the high-coverage shotgun data analysed here) represent most likely the same Neandertal individual (see S6).

Pairwise Differences and Heterozygosity in Windows

We calculated pairwise differences and heterozygosity in non-overlapping windows of 100kb size along the genome in an attempt to identify outlier regions that may be compatible with modern human admixture into Neandertals. Pairwise differences between the archaic humans were calculated by sampling a random allele at heterozygous sites for all individuals. Since the 1000 Genomes Luhya and Yoruba data shows evidence for some low-level Neandertal ancestry, we restrict our analyses to sites where one of the African alleles reaches at least 90% frequency and calculate archaic-African divergence against this majority allele. Windows overlapping inbred regions of at least 2.5cM in length ("HBD regions"; see S5) and windows containing less than 50kb after filtering were excluded. We retained 11755 windows after filtering. Rates of pairwise differences (divergence) and heterozygous sites (heterozygosity) were calculated by dividing counts by the number of sites that pass filters for each window.

Figure S36 shows the pairwise comparisons of the divergence of archaic humans to Africans. Interestingly, both the Altai and the Vindija Neandertal show some windows with a very low divergence to

Africans. In total we find 9, 21 and 2 *outlier* windows with less than 1/10kb differences per basepair to Africans in Vindija, Altai and Denisova respectively. We further tested whether these windows show high heterozygosity or deeper divergence to the other archaic humans. However, Neandertal windows with low divergence show a reduced divergence to Denisovan of ~5/10kb compared to the average of ~8/10kb when summing over all windows. Similarly, Vindija, Altai and Denisova show a reduced heterozygosity of ~1/10kb as compared to ~1.5/10kb over all windows. These signals are not statistically significant when testing the outlier windows against the genome wide distribution of values (Wilcoxon rank test; two-sided; $p>0.3$ in all comparisons).

We repeated this outlier analysis in windows of 0.1cM length using the Decode (*87*) and African American (*59*) recombination maps. Over all comparisons we observed at most 5 windows with less than 1/10kb differences for the African-Neandertals divergence, and no such windows were observed in the Denisovan comparisons. Heterozygosity for the outlier windows was lower than the average (significantly lower heterozygosity for Vindija with Decode map (Wilcoxon rank test, two sided, p-value=0.03); non-significant in all other comparisons (p-values >0.15)), and divergence to Denisova was either reduced or comparable to the average (significantly reduced Vindija-Denisova divergence with the African-American map (p-value = 0.02); others not significantly different (p-values>0.09)).

We next sought to test for differences in the distribution of divergence and heterozygosity over 100kb windows. In order to increase power, we assign windows to 10 bins based on the 10-quantiles over the African-archaic divergence. Then, for each archaic-specific binning, we calculated the divergence to other archaic humans and the heterozygosity. Confidence intervals for divergence and heterozygosity were calculated as the 95% central interval on 1000 bootstrap samples over windows in a bin. Additionally, heterozygosity was normalized by the average heterozygosity over all windows.

Figure S37 shows that the quantiles of divergence to Africans (the x-axis measure of bins) differ consistently between archaic humans. The Altai Neandertal shows a consistently shallower divergence to Africans as compared to the Vindija Neandertal. This difference can be explained by a shorter branch of the Altai Neandertal caused by the older age as compared to the Vindija (see S7). However, Denisova shows a consistent shift to deeper divergence compared to the two Neandertals. Branch-shortening estimates suggest that the Denisovan individual lived more recently than the Altai Neandertals and earlier than the Vindija Neandertal, and cannot explain this difference (S7).

Divergence between the archaic humans and heterozygosity in the archaic humans show a significant positive correlation with the divergence in Africans over windows. This signal is also visible in the binned data and is likely due to common factors driving divergence and heterozygosity at all time scales, such as mutation rate variation along the genome or varying degrees of negative or background selection.

When comparing divergence between Vindija and Denisova, and between Altai and Denisova (Fig. S37A,B) we observe the largest differences for the bins with the highest and lowest divergence to Africa. For windows in the bin with the highest divergence to Africans Denisova shows deeper divergence to Neandertals than Neandertals to Denisova, while the opposite signal is observed at low divergence to Africans. We note, however, that the bins used here are not independent since an increase in one bin must be compensated by

windows with lower values moving to another bin. We show below using coalescent simulations that this effect can be sufficiently strong so that observations are not readily interpretable as superarchaic or modern human admixture, or both. The comparison between Altai and Vindija Neandertals shows no significant differences for these bins (Fig. S37C).

We observe matching patterns when repeating the analysis with a window size of 0.1cM using two recombination maps. However, the signal of high Denisova divergence to Vindija for the highest divergence to Africans did not reach significance using the African American recombination map.

Heterozygosity shows marginal differences in the extreme bins (Fig. S37D). However, these differences are not observed in windows of 0.1cM with either recombination map.

Closest Divergence to Africans in Windows

A potential signal of admixture may be strengthened when calculating divergence using the closest matching alleles at heterozygous sites instead of a random allele, since phase is unknown in these genomes (*2, 13*). A prerequisite for this, however, is that admixture occurred sufficiently recent, or introduced alleles at a sufficiently high frequency, so that the introgressed alleles are often encountered in a heterozygous state. A potential issue with this type of analysis is the influence of drift and branch-shortening: An archaic individual that died further in the past will more likely retain some ancestral variants in a heterozygous state since it samples alleles before they may have drifted to fixation or are lost, but will also have experienced less recombination (see below for simulations exploring this idea further).

We repeat the analysis in windows from the previous section by choosing the archaic allele that matches the African state whenever the archaic is heterozygous. We also use this particular allele for calculating divergence to other archaic humans. Signals of divergence between Denisova and Neandertals do not change qualitatively (Fig. S38, top). However, we find that the Altai Neandertal shows significantly deeper divergence to Vindija than Vindija shows to Altai in the bin with the closest divergence to Africans (Fig. S38, bottom).

In the bin with the smallest African divergence, we also observe significantly higher heterozygosity in Altai than Vindija and Denisova for 100kb windows and significantly different values in all three archaic humans for 0.1cM windows (Fig S39). A less pronounced, and not always significant, mirror-signal of elevated heterozygosity in Denisova and Vindija is observed for high African divergence.

Relative Divergence to Africans in Windows

In an attempt to reduce the effect of branch-shortening, we used a relative measure of divergence to Africans for the archaic humans. At any position with an African derived allele frequency of at least 90% (compared to the ancestral state determined by five primate genomes that are required to agree), a random archaic allele was chosen and compared to the ancestral and the derived state. The ratio of matching the ancestral state over the total of ancestral and derived sites gives a measure of divergence to Africans relative to the divergence to chimpanzee, and was used to bin the data. However, since the relative measure of divergence has less

resolution, a large fraction of windows have no observable divergence to Africans. For this reason, we merge the two bins with the lowest divergence, resulting in quantiles: 20%, 30%, 40%, …, 90%. Due to the requirement of additional outgroups to call the ancestral state, only 6485 windows remain for analysis when requiring at least 50kb of sites to pass filters. We caution that the choice of a random allele, the merging of bins and the reduced number of windows remaining for analyses reduce our power to detect signals in this analysis.

As before, we start by comparing heterozygosity in archaic humans and divergence between archaic humans ranked by African divergence in 100kb windows (see Figure S40, and Figure S37 for comparison). We are able to confirm that Denisova shows significantly deeper divergence to Neandertals than Neandertals to Denisovans for the bin with the deepest African divergence. The opposite signal for bins with the shallowest African divergence is also present. While the analysis of 0.1cM shows identical trends, these differences are not significant in all comparisons.

Heterozygosity shows no significant differences between archaic humans in the highest and lowest bin of African divergence, and the analysis of 0.1cM shows no consistently different signal (one map showed a signal of higher Denisova heterozygosity for the lowest, the other for higher Denisova heterozygosity for the highest bin).

Exploring Influences on Window-based Measures using Coalescent Simulations

We used the coalescent simulator *scrm* (*88*) to test how admixtures and differences in branch-shortening are reflected in the three window-based measures discussed above. For this, we simulated three archaic lineages as analogues to Vindija, Altai and Denisova together with 408 modern human lineages, corresponding to the 1000 Genomes Luhya and Yoruba, together with a human lineage that falls basal to all archaic and modern humans (superarchaic) and chimpanzee as an outgroup lineage. Vindija and Altai split 135kya, Neandertals and Denisova 415kya, archaic humans and modern humans 550kya, superarchaic, archaic humans and modern humans 1Mya and the chimpanzee split is 10Mya. The Vindija, Altai and Denisova branches were shortened by 55ky, 115ky and 70ky, respectively. The Vindija, Altai, Denisovan and the superarchaic effective population sizes were set to 2,000, the common ancestor of Vindija, Altai and Denisova to 5,000, and the modern human lineage to 10,000. The mutation rate was 1.45×10^{-8} per basepair per generation, the generation time 29 years, the recombination rate 1.3×10^{-8} per basepair per generation.

For each run, we simulated the equivalent of ~1.2Gb of sequence in 100kb chunks. In order to reflect the mutation rate variation in 100kb regions over the genome in our simulation, we estimated a scaling factor per 100kb chunk from the observed data. For each window, we calculated the number of chimpanzee-macaque differences per basepair and normalized this quantity by the average rate of chimpanzee-macaque differences over all windows. These factors were then binned in increments of 0.02 (see Fig.S41), and each simulation scaled the mutation rate according to the binned factors for the number of windows in each bin. We note that this approach to model mutation rate variation makes the implicit assumption that the variation is largely preserved over longer evolutionary times.

Distinguishing Superarchaic from Modern Human Admixture

We first tested whether the window-based measures can confidently distinguish between scenarios of superarchaic admixture and human admixture. For this, we ran one simulation adding a 5% admixture event from modern humans into the common ancestor of Neandertals at 170kya, and a second simulation adding a 5% admixture from the superarchaic lineage into Denisova at 250kya (Fig S42 and Fig S43). All three measures yielded results that suggest that both superarchaic and modern human admixture can produce significantly different Neandertal-Denisova divergence in bins with both high African divergence and low African divergence.

Branch-shortening

We next sought to test the effect of branch shortening on our comparisons of African divergence to heterozygosity, and simulated a dataset that contains branch-shortening and compared the results of this dataset to a simulation that has the branch shortening of all archaic lineages set to the identical value of 70ky. The simulations did not include admixtures. Figure S44 shows that differences in heterozygosity are observed for the bins with the lowest African divergence for the two pairwise divergence measures when differences in branch-shortening are simulated, but no significant differences were observed when branch-shortening was identical.

In the simulation, Altai and Vindija differed the most in branch-length. Interestingly, this difference is also reflected in the estimated divergence between Altai and Vindija when stratifying by divergence to Africa, in both the left-side and right-side ends of the distribution (Fig S45). Again, no difference is observed when the simulated branch-shortening did not differ between the archaic humans. No significant differences were observed for the Vindija-Denisova and Altai-Denisova comparisons.

Effective population size

Last, we tested whether differences in effective population size can influence the window-based measures, and simulated with an effective population size of 10,000 for the Denisova lineage while the effective population size of Neandertals was kept at 2,000. In the simulation, all archaic humans have an identical branch-shortening of 70ky. Figure S46 shows that a difference in effective population size can lead to significant differences in bins of highest or lowest divergence. The measurement of pairwise divergence where the closest allele is chosen at archaic heterozygous sites shows the most pronounced effect from different population sizes.

S9b: Modern-Human Allele Frequency Stratified D-statistics with the Archaics

Kay Prüfer*

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

We compare the sharing of derived alleles between Neandertals and Denisova stratified by frequency of the derived allele in present-day African and non-African populations. Comparisons with all populations show a signal of high sharing with Neandertals for low-frequency alleles, consistent with some level of Neandertal ancestry in all present-day human populations, including African populations. African alleles of higher frequency, in particular fixed-derived alleles, also show an excess of sharing with Neandertals. By fitting coalescent simulations, we find that a scenario involving modern human admixture into Neandertals and admixture from a deeply divergent human lineage (superarchaic) into Denisovans can fit the observed patterns. In agreement with previous results, we also find that a scenario with only superarchaic gene flow fits the data better than a scenario with only modern human gene flow into Neandertals, but including both kinds of admixture provides a better fit still.

We also observe differences in the sharing of derived alleles between the Vindija and the Altai Neandertal when stratifying by African allele frequency. Low frequency African alleles are more often shared with the Vindija Neandertal than with the Altai Neandertal. This observation can be explained by a closer relationship of the Vindija individual to the introgressing Neandertal population together with a small fraction of Neandertal ancestry (<0.1%) in Africans.

Background and Outline of Analyses

Previous analyses have shown that the genomes of non-Africans carry a small proportion of Neandertal ancestry (*3*). This finding has been supported by several analyses and led to the construction of maps of Neandertal ancestry (*6, 8*). The comparison of the Altai Neandertal and Denisovan genomes also provided some evidence that Denisovans may have received admixture from a deeply divergent human lineage (*2*) and that Neandertals may have received admixture from a population most closely related to present-day humans (*13*). However, due to the availability of only few archaic genomes, the evidence for the latter two admixture events is weaker.

Here we use D-statistics stratified by allele frequency in modern human populations and simulations to study these admixture events in more detail. We begin by showing the patterns of derived allele-sharing of different human populations with the Denisovan and a Neandertal when stratifying by frequency of the derived allele in these populations. To understand how different admixtures would affect the D-statistics stratified by allele-frequency, we simulate data analogous to the Yoruban and CEPH European population with gene flows

of different types and magnitudes. These simulations consider each admixture event separately, while the true scenario may be a mixture of several different admixture events. However, simulating all possible combinations of admixture events with a variety of different parameters would be computationally demanding. We therefore devise a way to combine simulations of different admixtures to explore which combination best matches the data. We also compare the relative fit of a scenario just involving Neandertal admixture into modern humans and superarchaic admixture to a scenario with Neandertal admixture into modern humans and admixture from modern humans into Neandertals.

To investigate potential differences in admixture involving the two high-coverage Neandertals, we also compare patterns of derived allele sharing of human populations to the Altai and Vindija Neandertals using a similar approach.

Data

We use the 1000 Genomes Phase III dataset (*50*), together with the genotype calls of the Vindija and Altai Neandertals, and the Denisovan (genotypes for the archaic humans are described in S3). The 1000 Genomes dataset only reports sites that are segregating in at least one 1000 genomes individual; sites that were not segregating in the 1000 Genomes dataset were assumed to be identical to the human reference. An ancestral allele was called when several primate outgroups showed the identical state (otherwise the site was excluded). We used whole-genome alignments of chimpanzee, gorilla, orangutan, rhesus macaque to the human reference genome (GRCH37, hg19) produced by the UCSC genome browser (*56*) and alignments produced in-house for the bonobo genome (*86*). Analyses in this section that use the SGDP dataset(*22*) employ the same processing as described in S8, but require a minimum number of 40 alleles covered in individuals of each population. Sites in all datasets were filtered according to the set of minimal filters described in S3.

Allele-Frequency Stratified D-statistics between Neandertals and the Denisovan

For each modern human population *P* in the 1000 Genomes dataset, we calculate the D-statistics D(Vindija,Denisova,*P*,Outgroups), separately for each allele frequency in population *P*. Non-African human populations show an excess of low-frequency derived alleles that are shared with the Neandertal but not the Denisovan, consistent with admixture from Neandertals into these populations (Fig. S47). At higher frequency, the opposite signal is observed, which can be explained if ancestral alleles from the Neandertal admixture that segregate at low-frequency are re-introduced. Comparisons with the Altai Neandertal instead of Vindija yield similar results.

African populations also show an excess of sharing with Neandertals at low-frequency, albeit of much smaller magnitude (Fig S48). This signal is compatible with the previously described low-level back-migration of Neandertal-admixed humans to Africa (*77*). In addition to this signal, African populations exhibit an increase in derived allele sharing with Vindija with increasing frequency (Fig. S49). This increase is statistically significant for all five African populations ($p<0.036$, $0.14<r<0.32$) and significant in four out of

five comparisons when substituting Altai Neandertal for Vindija ($r=0.11$, $p=0.09$ for Gambian, all others $p<0.01$, $0.18<r<0.29$).

All African and non-African populations show an excess of derived allele-sharing with the Neandertals over Denisova at fixed sites. This excess of sharing at fixed sites with Africans has been previously used to infer admixture from a deeply divergent human lineage into Denisovans(*86*). However, since the 1000 Genomes Project only reports polymorphic sites, we have used the human reference to infer sites where all 1000 Genomes individuals carry the same derived allele. To test whether the use of the human reference introduces a bias, we have recalculated the D-statistic for the 1000 Genomes Yoruban population by using a high-coverage Yoruban individual as reference (panel B Yoruban individual from (*86*)), and found no significant difference in the counts of ABBA and BABA counts for fixed sites (Fisher's exact test $p>0.88$). We have also recalculated stratified D-statistics using the SGDP dataset (*22*), which includes information about coverage for all individuals, and find qualitatively matching results (Fig S50).

Coalescent Simulations

We use the coalescent simulation software *scrm* (*88*) (version 1.7.2) to explore the effect of different admixture scenarios D-statistics stratified by allele frequency. Each simulation includes a Vindija, a Altai and a Denisova lineage, a modern human population (either 198 lineages corresponding to the 1000 Genomes CEU or 216 lineages corresponding to the 1000 Genomes Yoruban population), a deeply divergent human lineage ("superarchaic") and an outgroup lineage. The following parameters are fixed for all simulations:

- Data is simulated in 100kb chunks
- Mutation rate: 1.45×10^{-8} per basepair per generation
- Recombination rate: 1.3×10^{-8} per basepair per generation
- Generation time: 29 years
- Branch shortening: Vindija, Altai and Denisova were sampled 55, 115 and 70 kya, respectively
- Piecewise constant demography for the archaic humans from S7, and a constant effective population size of 2000 for the superarchaic lineage
- Piecewise constant demography for the 1000 Genomes YRI and CEU population according to the estimates from SMC++ (*89*) (rescaled to match the mutation rate)
- Split times (see S7):
    - Vindija-Altai: 135kya
    - Neandertal-Denisova: 415kya
    - Archaic-ModernHuman: 550kya
- An outgroup lineage is simulated with a split time of 10Mya to all other lineages and an effective population size of 20,000 individuals

Following previous analyses (*2, 13*), we explore three admixture scenarios:

1. admixture from Neandertals into modern humans,

2. admixture from Modern humans into Neandertals, and
3. admixture from superarchaics into the Denisovan ancestors.

For the first two admixture scenarios, we varied the time of admixture and the percentage of the admixture. In order to avoid introducing additional parameters, we assume that admixture occured directly between the modern human and the Neandertal lineages and do not involve intermediate populations. In particular, we assume that Africans received gene flow from Neandertals directly, and not through an intermediate Eurasian population. This gene flow is assumed to originate from the Vindija Neandertal population, since Vindija appears to be more closely related to the introgressing Neandertal(s) than the Altai Neandertal is (S8, S11).

For the superarchaic admixture, we varied the split-time of the superarchaic from the ancestor of all other humans and the percentage of admixture. We assumed the gene flow from the superarchaic occurred 250kya. We note that because only a single Denisovan lineage is sampled, drift on the Denisovan lineage after the introgression is not relevant for the signals of stratified D-statistics.

Exploring Allele-Frequency Stratified D-statistics with Simulations

We ran simulations as described above for the three admixture scenarios for Yorubans and Europeans, varying the admixture parameters as shown in Table S27.

To quantify the fit of the simulated and observed allele-frequency stratified D-statistics, we first calculate for each frequency bin the fraction of ABBA site counts over ABBA and BABA site counts in the simulated data. In other words, with $E_{ABBA}(f)$ and $E_{BABA}(f)$ the counts for a given frequency f in a simulation, we calculate $F(f) = E_{ABBA}(f)/(E_{ABBA}(f) + E_{BABA}(f))$. Using the observed ABBA and BABA counts at frequency f in the real data ($O_{ABBA}(f)$ and $O_{BABA}(f)$, respectively) we then calculate the binomial probability for observing these counts given that the true fraction of ABBA is that seen in the simulation: $P(f) = \binom{O_{ABBA}(f)}{O_{ABBA}(f)+O_{BABA}(f)} * F(f)^{O_{ABBA}(f)} * F(f)^{O_{BABA}(f)}$. Analogous to calculating log-likelihood, we sum these probabilities as: $LL = \sum_{0<f\leq 1} \log(P(f))$.

To understand how the different admixture scenarios affect the stratified D-statistics, we focus on the Altai-Denisova D-statistics. Since Altai is likely equally close to the introgressing Neandertal and Vindija, we used Vindija as the source population for admixture in our simulations.

Yoruban Allele-frequency Stratified D in Simulations

Figure S51 shows simulations for Neandertal to Yoruban admixture. The best matching simulation as measured by LL was a 0.04% admixture into Yorubans 55kya. This simulation matches the excess of allele sharing at low-frequencies and produces a substantially better match ($LL = -1679$) than simulations without admixtures ($-1980 \leq LL \leq -1913$ in 10 independent simulations without admixture).

Figure S52 and S53 show the simulations for superarchaic admixture into Denisovan and modern human admixture into Neandertals. The simulations with superarchaic admixture show the best match for the simulation with 8% admixture of a superarchaic that split 1 million years ago from the human lineage ($LL =$

-1155). For the modern human admixture, two different admixture fractions yield matches that are nearly identical in score and substantially better than other parameter combinations for this scenario: 10% admixture 150kya and 15% 300kya ($LL = -1198$ and $LL = -1196$, respectively). We note that a high D at fixed sites, which is seen in the real data, is not reproduced with any of the simulations with only modern human admixture, while superarchaic admixture can generate this feature. This observation fits with previous results that supported the hypothesis of superarchaic admixture (see SI16a and SI16b in (*2*)). However, both types of admixture can produce positive slopes with increasing allele frequencies.

CEPH European Allele-stratified D in Simulations

We repeat our analysis with the allele stratified D(Altai,Denisova,CEU,Outgroup). Figure S54 shows that the best fitting scenario ($LL = -2276$) for the Neandertal to European admixture has a 2.25% admixture 55kya. Assuming 2% admixture 55kya and 2.75% admixture 60kya fit the data nearly as well as this best scenario ($LL = -2282$ and $LL = -2284$, respectively). These estimates of admixture time match previous estimates well (*46, 90*) and the percentage of admixture is consistent with our estimates based on F4-ratios (S8). As expected from the strong departure from a D statistic of 0 at most frequencies, simulations without admixture provide a substantially worse fit to the data ($-16722 \leq LL \leq -16386$ in 10 simulations). Similarly, neither superarchaic admixture nor modern human admixture into Neandertals provide a good match to the data (best scenarios 10%, 1Mya split, $LL = -12368$ and 15%, 150kya, $LL = -11345$, respectively, Fig. S55 and Fig. S56).

Matching a Combined Scenario of Admixtures

The simulations in the previous section considered each admixture event separately, i.e. we studied the effect of only Neandertal admixture into modern humans, of only modern human admixture into Neandertals, and of only admixture from a superarchaic into Denisovans. However, if several of these admixtures occurred together, the signal of allele-stratified D-statistics would be expected to show a combination of the signals observed in these separate simulations. We used an implementation of the Nelder-Mead optimization algorithm to identify the best fitting combination of signals according to the *LL* measure. The search algorithm requires expected allele-frequency stratified D-statistics that fall in between the specific values that were simulated, and we used bi-linear interpolation per frequency to generate these expected curves. The allele-stratified D-statistics are then summed to arrive at the expected curve for a combined scenario of admixtures. We note that by summing D-statistics, we assume that the signals produced by the different admixtures generate signals at independent sites that never overlap, e.g. an admixture from Neandertals to modern humans would never carry sites that were introduced from modern humans into Neandertals. For instance, for the largest parameters searched in Yorubans and averaging over frequencies, we observe 6% higher values when summing D-statistics compared to a simulation with the combined parameters. This assumption will thus lead to an underestimate of parameters when the admixture percentages are large.

To help the algorithm find as many plausible scenarios as possible, we run the search several times from different starting points, placing the starting points at values 16.7%, 50% and 83.3% of the length between the lowest and highest simulated values for each parameter. Of the 729 independent searches, all values are combined that show a small difference in *LL* to the best matching search *($\Delta LL \leq 3$)*.

To test the search algorithm, we simulated 50 scenarios for the Yoruban D-statistics and 50 scenarios for the CEPH European D-statistics. In each case, we randomly sampled parameters from the range of parameters given in Table S27. Parameters were then estimated and compared to the simulated parameters. The comparison of simulated to estimated parameters for Yorubans and Europeans (Figs. S57-58) shows that the method does not work well for the estimation of most admixture parameters. However, the percentage of modern human admixture into Europeans shows a good match to the simulated parameters.

We next applied the algorithm to the observed allele-frequency stratified D-statistics in Europeans and Yorubans. Table S28 shows the best point estimates and ranges of estimates when considering all searches with $\Delta LL \leq 3$ compared to the best match. Simulations with these estimated parameters fit the observed data well (Figure S59, top row), and the Yoruban simulations show the previously observed positive correlation with increasing allele frequency (r=0.25, p=0.0002 for the mean D over all 50 simulations; see also Fig.S59 bottom row).

We also estimated parameters when removing either superarchaic admixture into Denisovans or modern human admixture into Neandertals, and find that a scenario with superarchaic admixture fits the data better than a scenario with modern human admixture (Table S28). To test whether such a large difference in *LL* could be encountered by chance, we simulated for YRI and CEU 50 datasets with only superarchaic admixture and 50 datasets with only modern human admixture into Neandertal, randomly sampling the parameters from the ranges shown in Table S27. For each simulation, we estimated the fit of a model with only superarchaic admixture and the fit of a model with only modern human gene flow into Neandertals. We find no case, in which modern human admixture alone would yield a difference in *LL* as large as observed in the real data (Fig. S60), indicating that the better fit of the superarchaic gene flow model is not due to an erroneous fit of the superarchaic model when only modern human gene flow into Neandertals occurred.

Allele frequency stratified comparison to Vindija and Altai

We also compared modern human populations in their frequency-stratified sharing of derived alleles between the Altai and Vindja Neandertal (Fig S61). An excess of low frequency derived alleles with Vindija and not Altai is observed for all populations, consistent with Neandertal-admixture originating from a population of Neandertals that were more closely related with Vindija than Altai. Interestingly, the comparisons with African populations (Fig S62) show that an increased sharing with Vindija is also present for high frequency bins. A significantly positive correlation between frequency and derived allele sharing with Vindija is observed for four out of five comparisons with African populations when considering all frequencies (not significant: Luhya, r=0.11, p=0.12; remaining: r>0.15, p<0.03) and in all comparisons when considering frequencies of over 20% (r>0.26 and p<0.001 for all comparisons) (Fig S63).

Similar to the analyses described above, we used simulations to match the observed African allele-frequency stratified D-statistics comparing the Vindija and Altai Neandertal. We considered two scenarios. First, gene flow from the Vindija population into Yorubans with identical parameters as shown in Table S27, and second, gene flow from CEU (assumed to have split from Yorubans 100kya) into the Vindija population. For this latter admixture scenario, we simulated admixture percentages of 0, 0.25, 0.5, 0.75, 1 and 2% at 60, 80, 100 or 120kya. Figures S64 and S65 show the simulations in comparison to the observed data. The best fitting scenario for Vindija to African admixture (with $LL = -837$) is for a magnitude of 0.02% at 50kya, similar to the best match for the Altai-Denisovan comparison (0.04%, 55kya; see Fig.S51). For modern human to Vindija gene flow an admixture of 0.25% at 100kya or 120kya fit best ($LL = -886$).

We next applied our search algorithm to identify a scenario involving a mixture of both gene flow scenarios or only Neandertal to African gene flow. Table S29 shows that a similar age and percent admixture is estimated for the gene flow from Neandertals into modern humans. Modern human admixture into Neandertals yields only a small improvement in the fit and is estimated to be of small magnitude (0.11%) and to have occurred 120kya. Figure S66 shows that the simulations with the best paramters largely resemble the stratified D in the real data. While none of the simulations showed a positive correlation between stratified D and allele frequency that was as strong as the one observed (correlation with a minimum frequency of 20%: data: r=0.30, p=0.00002; simulations: $0.30 \leq r \leq 0.29$, p>0.00003), we find one outlier (r=0.29, p=0.00004) in the simulations with only Neandertal to modern human gene flow that matches the trajectory of the real data well. Based on these observations, we cannot exclude the possibility that the Vindija-Altai stratified D signal is explained solely by admixture from a population closely related to Vindija into modern humans.

S10: Relationship between Neandertals and Denisova

Kay Prüfer*

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

Previous results indicated that the Denisovan genome harbors ancestry related to the Altai Neandertal (*2*). In this paper we find similar patterns. There is an excess of heterozygosity in Denisova for windows in the Denisovan genome that show the least sequence divergence from the Neandertals, while Neandertals show no such signal for windows least diverged from Denisovans. Furthermore, we found that the Altai Neandertal shares significantly more derived alleles with the Denisovan than the Vindija Neandertal does. This result indicates gene flow into Denisovans from a Neandertal that is more closely related to Altai than Vindija.

Window based analysis

We explored differences in patterns of heterozygosity when comparing Vindija-Denisova and Altai-Denisova divergence in a window-based approach (see S9a for details). Briefly, the genome was divided up into non-overlapping windows of 100kb physical length or 0.1cM genetic length using two recombination maps (*59, 87*). For each window, random alleles from Denisova, Altai and Vindija were chosen at all positions and compared to compute the average nucleotide difference between Altai-Denisova and Vindija-Denisova; heterozygous sites in the archaic humans were also counted for each window. Regions of recent inbreeding in Altai, Vindija and Denisova longer than 2.5cM were excluded (see S5). Windows were required to contain at least 50,000 sites for which five outgroup primate genomes show identical state and which pass other filters (see S3 and S9a).

Over all windows that passed filters we find that the Denisovan heterozygosity correlates less well with Vindija-Denisova and Altai-Denisova divergence (Pearson's r=0.086 and 0.085) than Vindija heterozygosity correlates with Vindija-Denisova divergence (r=0.150) and Altai with Altai-Denisova divergence (r=0.139).

We further divided windows into 10 equally sized bins of increasing Neandertal-Denisova divergence and counted heterozygosity in windows of each bin. Figure S67 shows heterozygosity binned by Neandertal-Denisovan divergence for 100kb and 0.1cM window sizes. The bin with the lowest Denisova-Neandertal divergence consistently shows a higher heterozygosity in Denisova compared to the two Neandertals. The reverse of this signal is observed for the bin with the highest Denisova-Neandertal divergence. The signals of increased heterozygosity for the most closely related bins are consistent with the previously suggested admixture from Neandertals into Denisovans that occurred sufficiently recently in the history of the Denisovan individual to maintain Neandertal alleles as polymorphisms (*2*). In agreement with this hypothesis, we observe

a strong difference in heterozygosity for the windows with the smallest divergence between Neandertals and Denisovan when we repeat the analysis in 100 bins (Figure S68).

In section S9a, we discussed several factors that could influence the window-based analysis of divergence. To test whether the observed signals could be created in the absence of admixture, we ran coalescent simulations as described in S9a and S9b. Briefly, the simulations include stepwise effective population sizes from PSMC estimates, split times and branch shortening, and model mutation rate variation in windows of 100kb. Out of 100 simulations with these parameters we find only one instance in which the simulated Denisovan shows significantly higher heterozygosity than Vindija and Altai for the least diverged bin (Fig. S69). The signal for this simulation, however, is less strong than observed in the real data. In the majority of simulations (53%), Vindija shows a higher heterozygosity than Denisova for the lowest divergence bin, while Altai shows lower heterozygosity (see Fig. S69 for an example).

D-statistics

The plots in Figure S68 show a minimally higher Denisovan heterozygosity for windows that have shallow divergence to Altai compared to windows with shallow divergence to Vindija. Although this signal is not statistically significant, it opens up the possibility that the introgressing Neandertal is more closely related to the Altai Neandertal, which was found in the Denisova cave, than the Vindija Neandertal from a cave in Croatia.

To test this hypothesis, we used the Vindija, Altai and Denisovan genotypes to calculated D-statistics as described in S9a and found that the Altai Neandertal shares more derived alleles with Denisova compared to Vindija (Table S30). A possible explanation for this signal may be a lower quality of the Vindija genome compared to the Altai genome. Such a quality difference could result in increased sharing between Altai and Denisova since the lower quality Vindija genome is expected to share the state of the outgroup more often due to errors at sites where all three archaic humans are identical and different from the outgroup.

We tested whether the higher error in Vindija is a possible explanation for the greater sharing between Altai and Denisova by repeating our analysis with increasingly diverged outgroups (Table S30). If the signal is solely caused by errors on the Vindija lineage that coincide with differences of the archaic humans to the outgroup, then the sharing of alleles should increase linearly with the length of the outgroup lineage. The orangutan lineage is for instance between 2-3x longer than the lineage leading to chimpanzee or bonobo and should increase the signal by this factor if the signal is solely due to random error. In contrast, we find that D only increases moderately with increasing lineage length, suggesting that the excess of sharing is not due to errors alone.

Assuming that admixture originated from the Altai Neandertal population, we explored what percentage of admixture would be required to result in a D(Altai, Vindija, Denisova, outgroup) of ca. 6%. For this, we simulated 100 times, each, scenarios with gene flow from the Altai population to the Denisova population at 115kya with percentages 0.5%, 2.5% and 5%, and tabulated the resulting D-statistics (Fig. S70). D-statistics as high as 6% were observed only in simulations with at least 5% gene flow. Earlier times of

admixture of 130kya yielded lower D-statistics than 115kya for all simulated admixture rates, while a later admixture date (71kya) did not change the result qualitatively. These results suggest that an admixture percentage of more than 2.5% would be required to produce a D statistics of 6%.

We also used these simulations to study the effect of admixture on the window-based divergence measures (Fig S71), and find that an admixture of 2.5% produces signals that resemble those observed in our data, while a 5% admixture does not. The window-based measure would thus suggest that an admixture of less than 2.5% would be required to qualitatively match the signals in the real data. We speculate that unaccounted forms of error, misspecifications in the simulated model and/or unaccounted admixtures are the cause for the discrepancy between expected D-statistics and the window-based signal.

S11: Relationship of Mezmaiskaya to other Neandertals

Mateja Hajdinjak, Kay Prüfer*

* To whom correspondence should be addressed (pruefer@eva.mpg.de)

We used the low-coverage Mezmaiskaya 1 Neandertal genome to study the relationship of the Mezmaiskaya 1 individual to the two Neandertals sequenced to high-coverage and found that Mezmaiskaya 1 is more closely related to the Vindija than to the Altai Neandertal. Results in S8 showed that the Vindija Neandertal is more closely related to the introgressing Neandertal(s) than the Altai Neandertal. In agreement with the longer shared ancestry between Vindija and Mezmaiskaya and consistent with the previous results, we found evidence that the Mezmaiskaya 1 Neandertal is also more closely related to the introgressing Neandertal(s) than the Altai Neandertal.

Data

To infer the relationships between Neandertals, as well as their proximity to the introgressing Neandertal(s), we used the genotypes of the Altai and Vindija Neandertals, the high-coverage Denisovan individual (*1*), and a number of present-day human populations from the Simons Genome Diversity project (*22*), as well as the genomes of several ape outgroups as described in S9a. For the Mezmaiskaya 1 Neandertal we used the genotypes from S3 (called Mez1_snpAD from here on) or a base from a randomly chosen read overlapping a mapable position (criteria: MQ≥25, BQ ≥30, Map35_100%) (Mez1_random).

The Mezmaiskaya 1 data was estimated to contain 2-3% of modern human contamination (S3). Previous analyses used the subset of sequences that show substitutions consistent with ancient DNA damage to enrich for endogenous sequences and reduce the influence of modern human contamination (*91*). Following this approach, we selected Mezmaiskaya 1 sequences that show C to T changes to the human reference at the first and last two bases of sequences for the UDG-treated libraries and within the first 3 basepairs from the 3' and/or the 5' end for the non-UDG treated libraries (see S1 and S2 for details on library preparation and sequencing). Again, at each position, we chose a base at random from overlapping reads and a new dataset was constructed by intersecting these positions with the above data (Mez1_random_deam).

Compared to the high-coverage genomes, the Mezmaiskaya 1 Neandertal only sums to a total of 2x genomic coverage, including 1.4x coverage from the non-UDG treated libraries that show a high frequency of C to T exchanges due to ancient DNA damage (see S2). In order to avoid false signals due to these C to T changes, we restricted all calculations to transversions by excluding any site at which C and T or G and A were observed among all individuals.

Mezmaiskaya 1 is closer to the Vindija than the Altai Neandertal

Comparing the high-coverage Vindija and Altai Neandertal genomes to the Mezmaiskaya 1 individual, we find that the Mezmaiskaya 1 Neandertal is significantly closer to the Vindija Neandertal (Table S31). This signal is stable to the choice of outgroup and present for all three types of processing of the Mezmaiskaya 1 data.

Comparison to modern human genomes

Comparing the Mezmaiskaya 1 individual, a high coverage Neandertal and two present-day African populations from the SGDP we observe different results depending on whether Mezmaiskaya 1 was restricted to deaminated sequences or not. The two unfiltered datasets, Mez1_snpAD and Mez1_random, showed higher allele-sharing with Yoruba compared to San and Mbuti ($2.8<|D|<4.1$, $2.8<|Z|<4.7$, Table S32). In contrast, the two Neandertals and the two African genomes more likely form independent clades when using randomly sampled bases from deaminated Mezmaiskaya 1 sequences ($0.6\% \leq |D| \leq 2.37\%$, $|Z| < 2.15$; Table S32). Out-of-African populations, the likely source of contamination in the Mezmaiskaya 1 sample, are more closely related to Yorubans than Mbuti or San. Contamination is thus expected to shift the derived-allele sharing towards increased sharing with Yorubans, as observed for the unfiltered Mez1_snpAD and Mez1_random datasets.

The Mezmaiskaya 1 individual shows a consistently closer relationship to present-day out-of-African populations compared to the Altai Neandertal (Table S33). The signal is smallest (D%) in the dataset of Mezmaiskaya 1 putatively deaminated sequences (Mez1_random_deam), consistent with a reduced modern human contamination of out-of-African descent contributing to the signal. The Vindija Neandertal shows no significant difference in its relationship to out-of-African populations compared to the deamination restricted Mezmaiskaya 1 dataset (Mez1_random_deam), but a consistent signal of increased sharing between out-of-Africans and Mezmaiskaya 1 Neandertal for the unrestricted datasets (Mez1_snpAD and Mez1_random) (Table S34). These results indicate that Mezmaiskaya1 is more closely related to the Neandertal contributing to present-day out of Africans than the Altai Neandertal, but approximately equally close to the introgressing Neandertal compared to Vindija.

Mezmaiskaya 1 is likely closer to the introgressing Neandertal(s) than Altai

We follow the method from Supplementary Information 14, ref.(*2*) in order to estimate the proportion of contamination in the Mezmaiskaya 1 data that would be required to generate an excess of sharing with out-of-Africans when compared to the Altai Neandertal. Assuming that the source of present-day human contamination is likely of Eurasian origin we can calculate the f_4 ratios $\frac{f_4(Altai,Mez1\_random,Han,Mbuti)}{f_4(Altai,French,Han,Mbuti)}$ and $\frac{f_4(Altai,Mez1\_random,French,Mbuti)}{f_4(Altai,Han,French,Mbuti)}$. These comparisons yield point estimates of between 3.32% and 3.33% contamination for Mez1_snpAD, and 3.92% and 4.05% contamination for Mez1_random (Table S35). These

point estimates lie outside of the confidence intervals for the maximum likelihood estimate of contamination for Mezmaiskaya 1 (1.7-2.6%, see S3) but within the wider confidence intervals for the estimate based on the high-frequency Neandertal variants (1.7-5.8%). In contrast, substituting the Altai by the Vindija Neandertal yields estimates in the range of 1.98% and 2.82% (Table S35), which are in better agreement with the estimates for modern human contamination. Putatively deaminated Mezmaiskaya 1 sequences yield estimates of contamination that are not significantly different from zero when comparing to the Vindija Neandertal.

S12: Identifying Introgressed Sequence using the S* Method

Benjamin Vernot*

* To whom correspondence should be addressed (benjamin_vernot@eva.mpg.de)

We previously developed a method for identifying introgressed sequence in a population of modern human individuals, given two archaic genomes for comparison. Here, we apply this model to three pairs of currently available high coverage archaic genomes (Vindija Neandertal, Altai Neandertal, Denisovan).

We observe that a) the sequenced Vindija individual was more closely related to the introgressing Neanderthal individuals, for all tested modern human populations, b) likely all introgressed Neandertal sequence was derived from one or more populations more closely related to the sequenced Vindija individual than Altai, c) due to the closer relationship between Vindija and the introgressing Neandertals, we are able to identify 10-20% more introgressed Neandertal sequence per individual, and 3% more Denisovan sequence in Melanesians, when using Vindija rather than Altai.

Methods Overview

We first repeat the methods of Vernot et al, 2016 (*70*), which identified introgressed fragments by comparison with the Altai Neandertal and Denisovan genomes. We performed this calculation on East Asians, Europeans and South Asians from the 1000 Genomes Project Phase III (*50*), and Melanesians from Vernot et al, 2016 (*70*). In addition to the filters used originally, we also applied the general filters described in Section S3.

The method entails the following steps – the only conceptual difference from Vernot et al 2016 (*70*) is the addition of the Vindija Neandertal genome:

1. Calculate S* in 50kb windows (10kb step), in phased modern human individuals, without comparison to an archaic genome.
2. Select an "S* callset" of haplotypes with a significantly high S* value, compared to simulations (99^{th} percentile).
3. For each putative introgressed haplotype from this S* callset, calculate an empirical p-value of its match ("match p-value") to the Altai, Vindija and Denisovan sequences, by comparison with similar haplotypes in Yorubans.
4. Using a likelihood model on the joint distribution of match p-values for a pair of archaics, assign haplotypes to either Archaic 1, Archaic 2, ambiguous (i.e., we can't distinguish between Archaic 1 and Archaic 2), or null (i.e., not enough evidence to be classified as introgressed). In the previous

analysis only the pair Altai Neandertal-Denisovan was considered. Here we use Vindija-Denisovan and Vindija-Altai in addition.

As part of step 4, various demographic models are simulated, and archaic match p-values are calculated for true introgressed and non-introgressed haplotypes in these simulations. Each simulated joint p-value distribution for a pair of archaics is then compared to the observed distribution, and using a maximum likelihood method, we co-estimate the best fit demographic model, and the proportion of S* haplotypes introgressed from Archaic 1 or Archaic 2, or non-introgressed. The models considered, and the implications of the estimated archaic proportions, are discussed below.

Demographic Models

Demographic parameters are generally as in Vernot et al, 2016: Neandertal-Denisovan split time (N-D), Altai-Vindija Neandertal split time (NA-NV), split time between the sequenced Neandertal and the introgressing Neandertal popultion (N-Intr N), and the same for Denisovan (D-Intr D) and both Neandertals (NA-Intr NA; NV-Intr NV). For comparisons with Denisova, we used the simulations from Vernot et al, 2016, which constrained N-D to 400kya. For comparisons between Altai and Vindija, we generated new simulations using the same general parameters, but varied NA-NV from 75kya to 250kya, and varied NA-Intr NA and NV-Intr NV from 50kya to 225kya, constrained such that the introgression date is always less than the split times. For each demographic model, we consider proportions from each archaic between 0 to .65, and select the pair of proportions that maximizes the likelihood of the model. For comparisons between Neandertal and Denisovan, D-Intr D was infered from Melanesians (MEL), and then fixed for the other populations. The best-fit demographic models are reported in Table S36, and the top likelihood for each model is shown in Figures S72-73 It should be noted that these simulations are not an accurate representation of the true demographic history – most notably in that they do not simulate branch shortening. For example, an estimate of the divergence of the introgressing Neandertal and the sequenced Neandertal of 50kya or 100kya is somewhat difficult to interpret. As shown in Vernot et al (2016) (*70*), however, the identification of introgressed haplotypes is robust to variation in the simulated demographic model.

Proportion of S* Callset Derived from each Archaic

We find that the maximum likelihood archaic proportions are broadly consistent when comparing Altai vs Denisova and Vindija vs Denisova, with slightly higher estimates in the Vindija comparison: for Eurasian populations, between 44.2% and 49.6% of the S* callset is estimated to come from Neandertals when using Altai, and between 46.3% and 52.1% when using Vindija (Table S37, Figures S74a and S74b). Interestingly, in the comparison of Altai and Vindija, the estimated proportion from Altai is 0.0%-0.1% in all populations, consistent with all of the introgressed sequence from Neandertals deriving from one or more archaic populations more closely related to Vindija than to Altai. The corresponding estimated proportions from Vindija, and thus the estimate of the total amount of Neandertal sequence in the S* callset, are consistently lower than estimates when considering either Neandertal against Denisovan, suggesting that the contrast with

Denisovan adds power to detect Neandertal sequence even in populations with no appreciable Denisovan introgression.

Confidence intervals and non-independence of introgressed haplotypes

The likelihoods used to estimate the proportions for each archaic suffer from non-independence in two ways. First, nearby haplotypes are non-independent due to shared coalescence; this non-independence is corrected via jackknifing (*73*). Second, an introgressed haplotype is often present in more than one individual, and thus is considered multiple times in the population likelihood calculation. This is mostly an issue when estimating confidence intervals on the archaic proportions. Thus, we estimate archaic proportions for each individual separately, and then consider the average per population. To jackknife we split the genome into 10 regions of equal size, and re-estimate the archaic proportions after removing each region. From these resamplings, the standard errors can be computed as in ((*73*), eqns 1&2), and confidence intervals are reported as +/- 2*se (Table S37).

Vindija Neandertal allows for the identification of more introgressed sequence

The closer relationship between the sequenced Vindija individual and the introgressing population of Neandertals allows for the identification of additional introgressed sequence. In addition, some haplotypes that previously were classified as "ambiguous introgressed" (i.e., the null model of "non-introgressed" could be rejected, but they could not be classified as Neandertal or Denisovan), can now be classified as Neandertal or Denisovan, leading to an increase in identified Denisovan sequence in Melanesians (Figure S75). Overall, this allows for the identification of ~11% more Neandertal sequence in Eurasian populations, ~20% more Neandertal sequence in Melanesians, and 3% more Denisovan sequence in Melanesians.

Deserts of Neandertal ancestry

We next calculated Neandertal ancestry in 10mb windows across the genome, requiring at least 80% of the window to be "callable" in our introgression analysis. The amount of Neandertal sequence identified per window is highly correlated between analyses using the Altai or Vindija high coverage genomes (R^2=.99; Figure S76). The slope of this correlation is significantly greater than one (p<2.2e-16, linear regression), consistent with the observation that we identify more Neandertal sequence when using the Vindija genome. Windows overlapping previously identified archaic deserts (*70*) are also outliers in the overall distribution (Figure S76).

S13: Detecting Structural Variation using BayesTyper

Laurits Skov* and Mikkel Heide Schierup*

* To whom correspondence should be addressed (lskov@cs.au.dk, mheide@birc.au.dk)

13.1 Abstract

With the aim to improve the detection of structural variants, including long indels, in the archaic genomes (Altai Neandertal, Vindija Neandertal and Denisova), we use a probabilistic framework to determine the genotypes of 90 million known variants. We first show that genotype calls have very high concordance with the genotypes called for four modern humans in the 1000 genomes project, and that we call, as expected, very few heterozygous variants in Chimpanzee and Orangutan which are expected to rarely share polymorphisms with humans and archaic genomes. Further, our method calls very few heterozygous variants in the inbred tracts of the Altai individual suggesting a low false positive rate. We show that our method can improve the robustness of indels called with GATK on ancient DNA and can identify around 8500 polymorphic indels in the Vindija, 8000 in the Altai Neandertal and 7500 in the Denisova genome above the size of 8 bp, which is the upper bound for indels found using GATK. We identified 4537 indel variants with a putative functional effect, of which 425 were new. We also identify 259 transversions, which were annotated as "probably damaging" with a Polyphen score above 0.9. Of these 75 are new.

13.2 Introduction

Methods for calling structural variants using read mapping typically do not discover variants longer than the read length. Since the read length in ancient genomes is short, the power to detect structural variants is therefore even more limited and mapping based variation callers such as GATK (*48*) only calls very short indels (<8 bp) with an unknown false positive rate. Here we search for evidence of structural variants known from other sources by interrogating the high coverage of short reads for evidence of the ancestral variant, the derived variant, or both. We employ a K-mer based approach as implemented in the BayesTyper software (https://github.com/bioinformatics-centre/BayesTyper). The approach tabulates the K-mer footprint of a set of known human variants and seeks support for this profile as well as the reference profile in the primary sequence reads from each individual to be genotyped. This is done in a fully probabilistic framework in BayesTyper. We apply this approach to a variant call set of SNPs and structural variants previously determined in humans, Neandertals and Denisovan. To evaluate the sensitivity and specificity and to calibrate filtering, we first genotype variants in a set of modern humans, one Chimpanzee and one Orangutan. We then use these filters to genotype variants in three archaic human genomes: the Vindija and Altai Neandertals and the Denisova.

13.3 Genotyping variants with BayesTyper – a K-mer approach

BayesTyper is a probabilistic genotyper that mainly consists of three steps: constructing variants graphs, constructing K-mer tables and genotyping based on K-mer footprint of each path through the variant graph as shown in Figure S77. The steps are described in more detail below.

13.3.1 Variant graph construction

We merge all variant call sets into a variant graph as shown in Figure S77a. We use variants from Vindija, Altai, Denisova and human populations. The list of call sets is shown in TableS38. The combined variant graph consists of around 89 million variants.

13.3.2 K-mer construction

K-mers are constructed from the reads of each individual of interest as seen in Figure S77b. Here we chose a K-mer size of 31 because that is below the minimum read length of 35 bp used as cut-off for the ancient genomes. We constructed K-mer profiles for the Altai Neandertal, Denisova and Vindija Neandertal, along with 4 modern humans and 2 great apes. The complete list of individuals are shown in Table S39.

13.3.3 Genotyping

For each variant, which corresponds to a path through the variant graph BayesTyper counts the number of K-mers that support this path as is shown in Figure S77c.

To balance the number of false positive (FP) variants while keeping high sensitivity we investigated the consequence of varying the number of supporting K-mers for calling a given allele. We used an overall K-mer size of 31 bp to stay below the read size cut-off for ancient genomes of 35 bp, and looked at the number of variants called as a function of the number of K-mers (NOK) that a new variant induces are actually observed in the reads (the NOK variable ranges between 1 and 31). We expect that the false positive (FP) rate decreases with increasing value of NOK because of accumulating evidence for the K-mer, but also that the false negative (FN) rate will increase because some K-mers are not found either because of the region being low coverage, associated with sequencing errors or very low complexity.

In order to find a good balance between FP and FN we exploited three predictions based on other genomes as our quality metrics.

First, we expect humans to share very few polymorphism with non-human great apes by descent since variation segregating in humans is generally less than 2 million years old and the split with Chimpanzee is more than 6 million years ago (*67*). Thus, typing human variant call sets in Chimpanzee and Orangutan by BayesTyper should produce very few heterozygous calls.

Second, individuals from the 1000 genomes project have already been genotyped with other methods for the complete set of human SNVs. While this genotyping may not be perfect, we expect that the concordance between our genotyping in the same individuals and their 1000 genomes genotypes should increase with increasing NOK value.

Third, we expect to see no or only few heterozygous sites in the tracts of homozygosity by descent (see S5) and we use the highly inbred Altai Neandertal for this test.

Figure S78 upper panels show that the decrease in heterozygosity in one Orangutan (40X coverage) and one Chimpanzee (40X coverage) of both genotyped deletions and insertions as a function of NOK on chromosome 17. The lower panels of Figure S78 show that the number of variants genotyped decreases with higher values of NOK. For NOK=25 around half of the variants in Chimpanzee and around a third of the variants in Orangutan are still genotyped but the heterozygosity is below 0.5%. We genotype fewer variants in Orangutan because we are using the human reference genome to "fill" the edges between the nodes. Humans and Orangutans diverged further back in time than humans and Chimpanzee and because of this more variants that are not in humans have accumulated in the Orangutan genome and other variants within 31 bp of the focal variants will decrease the NOK.

The upper panel in Figure S79 shows the concordance rate between the 1000 genomes genotypes of indels and the genotypes called with BayesTyper for four individuals of different ethnicities as a function of the NOK value and the lower panel the number of variants genotyped as a function of NOK on chromosome 17. Concordance increases with NOK to a certain point. For NOK=25, concordance is above 99.5% for deletions and above 99% for insertions, while around 70% of the deletions and 40% of insertions are still genotyped. The concordance rate drops for insertions when the NOK value reaches 30, due to the low number of variants remaining.

Figure S80 shows that the decrease in heterozygosity in the Altai Neandertal of both genotyped deletions and insertions as a function of NOK on chromosome 14 from 50 Mb to 100 Mb. We expect other tracks of homozygosity by descent (HBD) to behave similar to this region, and therefore we only use this one for our filtering.

Based on these three tests we chose a filter with NOK=25 for the analysis of ancient human and archaic human genomes, using the same K-mer size of 31. This filter minimizes the error associated with the genotypes to ~0.5% and maximizes the number of polymorphic sites retained.

13.4 Structural variants in archaic genomes

13.4.1 Number of variants called and transition issues

We firstly checked how many of the variants from the human, Altai, Vindija and Denisova call sets we could genotype using our approach. The SNPs from the archaic humans were called using snpAD (see S3) and the indels form the archaic humans were called using GATK (*48*). We show a summary of how many variants could be genotyped in Figure S81. For Altai, "known" variants are those that originate from the Altai call set and were genotyped in the Altai individual. "New" variants are variants coming from the Vindija, Denisova or 1000 genomes call set – they are "new" in the sense that they could not be found using GATK or snpAD in the original Altai call set.

With regard to indels, we only found support for around 25% of the indels called using GATK, suggesting a high false positive rate when using GATK for calling indels in ancient DNA. With regard to SNPs (transitions

and transversions), we are able to genotype around 93% of the transversions and 97% of the transitions, showing that this genotyping approach agrees with the snpAD.
We were also able to genotype variants that were not found using GATK (for indels) and snpAD (for SNPs). We find around 25,000 indels that originate primarily from the 1000 genomes variants. We also find around 300,000 transversions that originate primarily from the 1000 genomes variants. For the new transversions we find that 98% (Altai), 94% (Denisova) and 98% (Vindija) are in regions that were removed using the minimal set of filters mentioned in S3.

For transitions, we see twice as many novel variants in Vindija Neandertal compared to Altai Neandertal, likely because the majority of DNA libraries from Vindija are not UDG treated (see S1). Because the Vindija reads are mostly not UDG treated (it will look like we have more transitions than we actually do) and we are trying to genotype all the variants from the 1000 genome project (there are more than 50 million transitions in this data set), we expect to call many false positives for transitions. Of the genotyped variants, the heterozygous/homozygous ratio for transitions is also different from the ratio for transversions and indels (p-value=2.2E-16) as can be seen in Figure S82. This indicates that there are many false positives variants when looking at transitions for Vindija and Denisova, but to a lesser extent Altai. We conclude that BayesTyper does not perform well when calling transitions in ancient DNA and we thus exclude transitions for the rest of the analysis in this section.

13.4.2 Number of heterozygous sites across the archaic genomes

We counted the number of heterozygous indels and transversions in bins of 100 kb across the entire genome for each archaic individual (Figures S83-88). We observe long tracks of homozygosity by descent (HBD) both for indels and transversions, even though we only minimized the number of heterozygous indels on chromosome 14 in our genotype filtering (see section 12.3.3 Genotyping). This implies that our indel call set should have a low false positive rate. The original indel call set using GATK show a high heterozygosity in the autozygous regions suggesting a high rate of false positives. Our results also suggest that BayesTyper has a low false positive rate when calling transversions in ancient DNA.

13.4.3 Length distribution of newly detected indels

The longest indel called in the Denisova, Altai and Vindija indel call sets was 8 bp for Altai, 7 bp for Vindija and 7 bp for Denisova. We were able to genotype variants from the 1000 genomes project that were longer than 8 bp providing an improved catalogue of indels. We show the size distributions of the new variants in Figure S89

13.5 Annotation of structural variants

A large proportion of the genotyped indels and transversions were shared with the 1000 Genomes project (Table S40), allowing us to use the 1000 Genomes functional annotation. For each individual, we show the number of indels from each functional class in Figure S90.

We identified 4537 variants (out of a total of 193356 indels) with a putative functional consequence (i.e. frameshift_mutation, inframe mutation, coding sequence, regulatory region and transcription factor binding site). Of these 425 were new. We find a frameshift deletion in the OR51B4 gene (an olfactory receptor) that is present in Vindija and Altai but not in Denisova. This mutation is only present in South East Asia at an allele frequency of less than 1%. We also find a two indels in close proximity of each other in the ABHD16B gene that is only found in Denisova and South East Asia both at a frequency of 0.3%.

We also identify 259 transversions that are annotated as "probably damaging" with a Polyphen score above 0.9. Of these 75 are new.

13.6 Differences in sharing of human variation along the genome.

We counted the number of variants that were shared between humans (present in 1000 genomes) and the archaic individuals for bins of 100 kb across the genome. Human populations generally share more indels with Neandertals than with Denisova.

Figure S91 show the number of shared indels and Figure S92 show the number of shared transversions per window of 100 kb across the entire genome.

In the HLA region, we generally observe more indels that are shared with any of the Neandertals than with Denisova. This is shown in Figure S93. For the HLA genes around 33 Mb it seems that modern humans only share indels with the Vindija Neandertal.

S14: Inference of Copy Number Variation in the Vindija Neanderthal Genome

PingHsun Hsieh*, Brad Nelson, and Evan E. Eichler*

* To whom correspondence should be addressed: PingHsun Hsieh (hsiehph@uw.edu) and Evan E. Eichler (eee@gs.washington.edu)

Recent studies have revealed great diversity in copy number variants (CNVs) among human populations, hominins, and non-human primates, and suggested their roles in evolution(*1, 2, 92-96*). Of particular interest are lineage-specific CNVs. Indeed, there has been evidence that some *H. sapiens*-specific CNVs, such as *BOLA2*, might be selectively advantageous in our species(*97*). In this study, we aim to identify and refine lineage-specific CNVs in hominins using whole genome data, including the newly sequenced Neanderthal genome (~30X) from the Vindija cave in Croatia.

Read depth-based CNV inference

To identify CNVs across the entire genome, we used the published read depth-based digital comparative genomic hybridization (dCGH) method(*92, 93*). In short, sequencing reads were decomposed into 36 base-pair (bp) sequence fragments, which were in turn mapped to the repeat masked human reference genome (Build 37/hg19) using mrsFAST(*98*). We allowed up to two mismatches per 36-mer in order to increase our mapping sensitivity and accurately infer the aggregate paralogous copy number within repetitive regions. Because Illumina sequencing coverage often correlates with GC content due to biases in library constructions and sequencing, we corrected for underlying GC content based on a regression procedure applied to known copy number invariant (i.e. copy number state of 2) regions(*92, 93*). Finally, we estimated copy number across the genome in overlapping windows of 500 unmasked-bases with a sliding size of 100 unmasked-bases.

We identified CNVs in a discovery panel of 20 genomes, including the Vindija Neanderthal, Altai Neanderthal and Denisovan genomes, as well as 17 diverse, high coverage genomes from Simons Genome Diversity Project(*22, 93*)(SGDP) (**Figure S94**). These 17 SGDP genomes were specifically selected because they show the lowest variance in genome sequencing coverage(*93*). Genotypes and CNV boundaries were determined based on a scale-space filtering algorithm as described previously(*92*).

Initial call set and quality control

We initially inferred 4,693 CNVs from our discovery panel of 20 genomes. We assessed the performance of our CNV inference against negative-control loci that have a known normal diploid copy number (CN) state of 2 in the genome. These control loci represent 4,836 genomic blocks with length >100 kb that are free of known structural variants in the database of genomic variants (DGV) and primate segmental duplications based on previous analyses(*92*). We estimated 92.6% accuracy for the Vindija genome; yet the accuracy decreasing significantly with increasing GC composition (Figure S95). The lower accuracy of CN prediction for

sequences with higher GC content is consistent with a bias towards lower sequencing coverage with higher GC content in ancient genomes (Supplementary S2). This control experiment suggests a relatively higher false discovery rate consistent with the observation of an excess of Vindija CNV calls compared to the other genomes in the discovery panel (**Figure S94A**). Many of these calls are short (<10 kb) and show high variance in read-depth. Because read-depth variance decreases with increasing call length, we found the distributions of CNV counts are compatible across the 20 genomes after applying a minimum length cutoff of 10 kb to the initial call set (**Figure S94B**). This resulted in a conservative set of 904 CNV calls from the 20 genomes in the discovery panel (mean±s.d.: 328.9±14.95 CNV calls). The median size of these CNV calls is 17,410 bases, but skewed towards small events (**Figure S96**).

Hominin-specific CNV loci

To identify hominin specific CNVs, we genotyped the 904 CNV loci identified from the discovery panel for 316 publicly available genomes, including the three archaic genomes, 224 SGDP genomes (AMH), 3 ancient *H. sapiens* (ANC: LBK(*47*), Loschbour(*47*), and Ust-Ishim(*46*)), and 86 non-human primates (bonobo, chimpanzee, gorilla, and orangutan)(*92*). We defined a CNV as "hominin-specific" if it is found in the hominins, but not in any non-human primate genome included in this study, and identified 197 hominin-specific CNVs from our initial call set (**Figure S97**). The Vindija genome shows evidence for copy number variation at 40 of the 197 loci (**Table S41**). We found that the two Neanderthal genomes share more CNV calls than either of them shares with the Denisovan individual (**Figure S97**). This suggests that the two Neanderthals are more closely related to one another than either of them is to the Denisovan individual, in agreement with their putative phylogenetic relationship based on nucleotide differences (see earlier parts of materials and methods). Interestingly, we found that among these 197 hominin-specific CNVs, there is a significant reduction in deletion calls at exonic sequences compared to duplications (**Table S42**), possibly the result of purifying selection against deletions at genic loci(*93, 94*). We did not, however, find any significant difference in exonic/non-exonic CNV calls between modern human and archaic samples. While the greater number of AMH-specific CNV calls (n=130) is simply a consequence of the relatively larger sample size in our analysis, 27 of these AMH-specific calls have reached at least 5% in frequency in our samples (Table S52).

Duplications

Of the 78 hominin-specific duplications (**Figure S97, Table S52**), 13 and 33 loci were only found along the Neanderthal-Denisovan branch and AMH branch, respectively. We found no duplication call shared only among the three archaic genomes. Interestingly, the duplication locus chr1:234911364-234956952 shows a large expansion in copy number in our hominin samples (>96% of all samples have more than 4 copies), including the two Neanderthals (Vindija CN=9, Altai CN=11) and Denisovan (CN=5) (**Figure S98**). This intergenic locus overlaps with a known segmental duplication, whose known paralog (chr1:17088430-17125658; >99% identity) encompasses the first exon of *MST1L*, which encodes putative macrophage stimulating 1-like protein. Among the Vindija-specific duplication calls, the locus chr20:25343369-25413592 overlaps with two genes: *ABHD12* (Abhydrolase domain-containing protein 12; exon 1/13), and *GINS1* (DNA

replication complex GINS protein subunit 1; exon 1-4/7) (**Figure S99**). Mutations in this region are reported to be associated with the neurodegenerative disorder PHARC(*99*). Among the AMH-specific duplication calls, the locus chr7:143439085-143572547 shows a large expansion in modern humans. The gene *TCAF1* encompassed in this locus encodes a protein binding to *TRPM8* channel, a cold sensor that is highly expressed in prostate and other non-temperature-sensing organs in humans, and has been associated with prostate cancer and Dentin sensitivity(*100*).

In addition, we re-examined the 17 hominin-specific duplications previously reported(*2*) using the same genotyping panel of 316 genomes used in this study. This analysis allowed us to reclassify three loci previously identified as Altai Neanderthal-specific (n=2) and Denisovan-specific (n=1) duplications because they are shared with at least one of the human/non-human primate genomes in the genotyping panel (**Table S43**). We also found that all three archaic genomes have the putative ancestral copy number state of 2 at the locus chr16:30200098-30206185, which encompasses the gene *BOLA2* (bolA family member 2). This is consistent with the recent evidence that the rapid expansion of the *BOLA2* segment in copy number only occurred in the modern human lineage(*97*).

Deletions

Within the 119 hominin-specific deletions identified from our genotyping panel (**Figure S97A**), we identified 27 archaic- and 82 AMH- specific deletions (**Table S52**). Similar to the duplications, we observed no common deletion specific to all the three archaic genomes. Of the four deletion calls shared between the two Neanderthals, the locus chr13:100776502-100796519 encompasses the 3rd intron of *PCCA* (Propanoyl-CoA) and shows evidence of (H3K27Ac) regulatory activity in the umbilical vein endothelial cell type (**Figure S100**). Mutations in this gene are known to be associated with propionic acidemia, a metabolic disorder, which results in neurologic damage and mental retardation(*101*). Among the AMH-specific deletion calls, 34 loci overlap with genic sequences (**Table S52**). While much of these gene-encompassing deletions are segregating at relatively low frequencies among modern human samples (<5%), we identified four common deletions that cause the loss of the entire *OR52N5* (olfactory receptor, family 52, subfamily N), and partial loss of *NME7* (NME family member 7), *ZBTB20* (zinc finger and BTB domain containing 20), and *RIMS1* (regulating synaptic membrane exocytosis 1). Interestingly, we noticed that the deletion (both homozygous and hemizygous) locus chr3:114658237-114672775 is particularly common among Africans (20/36) and Oceanians (14/21) compared to other populations (**Figure S101**).

We previously reported the complete loss of the genes *GSTT1*, *MRGPRG*, and *C11orf36* and the partial loss of the genes *GSTTP2* and *SPINK14* in both the Altai Neanderthal and Denisovan genomes, while they show copy number polymorphism among modern humans(*2*). Here we found that the Vindija genome matches the genotypes of the Altai Neanderthal at these loci (**Table S44**). We also confirmed that, as previously reported(*2*), the homozygous deletion of the *GSTT1* locus is common among Asian populations (frequency(CN=0) = 0.378, 0.429, 0.217, and 0.318 for East-Asians, Oceanians, South-Asians, and Siberians,

respectively, in the SGDP sample); though it is also common among Africans (frequency(CN=0) = 0.306) as well as Europeans (frequency(CN=0) = 0.237).

S15: Phenotypic Associations of Introgressed Neandertal Alleles

Stéphane Peyrégne, Manjusha Chintalapati, Kay Prüfer*
* To whom correspondence should be addressed (pruefer@eva.mpg.de)

We identified sets of Vindija, Altai and Denisovan alleles that are shared with modern humans outside of Africa as a result of introgression. To determine whether these alleles influence modern human phenotypes, we intersected them with alleles reported in the NHGRI-EBI Genome-Wide Association Studies (GWAS) Catalog. We identified 85 haplotypes showing a significant association with at least one phenotype. The majority of all phenotype-associated alleles were shared with both Neandertals. However, we identified 11 additional associations with Vindija-specific alleles and 5 associations with Altai-specific alleles, highlighting that more information about the functional impacts of introgression is obtained by using the Vindija Neandertal which is more similar to the introgressing Neandertal population than the Altai is.

Identification of Vindija and Altai alleles segregating in modern humans

We used the 1000 Genomes Project Phase 3 dataset (*50*) to identify variants segregating in modern humans and shared with the high-coverage Vindija and/or the Altai Neandertals (*2*). We first extracted allele counts at single nucleotide polymorphism (SNP) sites using vcftools (*102*), removing both indels and sites with more than two alleles. We then focused on sites where African populations are fixed while non-African populations carry an allele that is shared with at least one of the four Neandertal chromosomes. We excluded African populations with non-negligible European admixture (Americans of African Ancestry in SW USA, ASW, and African Caribbeans in Barbados, ACB). We further restricted our analysis to Neandertal sequences with confident mapping to the human reference genome (hg19) ($MQ \geq 25$, Map35_100%) and only used genotype calls (Supplementary S3) in regions that are not tandem repeats and that passed the GC-corrected coverage filter (see S3). This resulted in 304,350 sites enriched for non-African variants introduced through Neanderthal gene flow.

Intersection with variants reported in genome-wide association studies

To determine whether these variants have previously been associated with any phenotype we downloaded the latest version of the NHGRI-EBI GWAS Catalog (on 13/11/2016) (*103*). This catalog reports 32,041 SNP-trait associations with p-values < 1.0×10^{-5}. We used liftover (*104*) to convert coordinates to hg19 and intersected the GWAS loci with our set of variants. Since introgressed archaic alleles are likely to occur on extended haplotypes leading to multiple associations being inferred we removed sites that were in linkage disequilibrium ($r^2>0.8$), keeping the variant with the lowest GWAS p-value. A total of 65 haplotypes associated with 69 phenotypes were shared by both Neandertals (Table S45). We also identified associations specific to

Vindijia or Altai Neandertal alleles. Interestingly, we find more Vindija-specific associations (11, see Table S45) than Altai-specific associations (5, see Table S46). Table S48 lists the frequency of introgressed Vindija and Altai-specific alleles in non-African populations in the 1000 Genomes Phase 3 data.

Haplotype patterns and validation of the introgressed candidates

Previous studies analyzed haplotypes to test for a signal of extended sharing with Neandertals to support introgression (*105*). To further support that phenotype-associated variants entered modern human populations through introgression, we used the introgression maps from S12. Three Vindija-specific and two Altai-specific alleles show no evidence for introgression in those maps (Table S48). We tested the length of the apparent introgressed haplotypes around the two most significant associations; a Vindija-specific allele associated with variations in LDL levels (*106*), and an Altai-specific allele that influences Type 2 diabetes risk and which was previously shown to be of archaic origin (*6, 10*). Restricting our analysis to individuals carrying Neandertal alleles in the 1000 Genomes dataset, we looked for other Neandertal-like alleles in a window of 200kb up- and downstream from the phenotype-associated variants (Figure S102). All consecutive sites with a Neandertal-like allele shared by at least 80% of the individuals were considered to belong to the haplotype. The lengths of the haplotypes are 210,687bp and 58,832bp for the Vindija-like haplotype influencing LDL cholesterol levels and the Altai-like haplotype influencing Type 2 diabetes risk, respectively. Both are significantly longer than expected under incomplete lineage sorting ($p = 0.0145$ and 7.163e-13, respectively) as calculated by (*105*) using the parameters from (*11*) (with the slow mutation rate of 0.5×10^{-9} $bp^{-1}.year^{-1}$), S7 (branch shortening) and local recombination rates of 0.0735 cM/Mb and 1.4345 cM/Mb respectively (*59*).

Comparison to allele sharing with the Denisovan

We repeated the analysis with the aim to detect Denisovan alleles shared with modern humans outside of Africa. We required the Denisovan variants to be absent in Africans but present in any non-African individual in the 1000G, and then overlapped the identified alleles with the GWAS catalog. We note that populations of the 1000G dataset are expected to carry little or no Denisovan ancestry, as there are no Oceanian populations in this dataset. Of the 60,370 Denisovan-specific alleles that we identify, only three were associated with phenotypes in the GWAS catalog (Table S49) compared to 161 among 223,334 Neandertal-specific alleles (15,633 Altai-specific, 49,789 Vindija-specific and 157,797 shared between Altai and Vindija). This is significantly less overlap with functional variants identified by GWAS than for Neandertals (Fisher's exact test, odds ratio=14.507, p-value= 2.114e-13). This low fraction could be explained by many of the Denisovan-shared variants representing long standing neutral variation instead of introgressed haplotypes.

To determine whether the alleles identified are consistent with having an introgressed origin we analyzed the haplotype lengths of all three phenotype-associated Denisovan-shared variants (Figure S103). We found that two of the three variants fall on significantly longer haplotypes than expected under a model of incomplete lineage sorting (Table S50) suggesting that they are likely introgressed. Denisovan-introgression is expected to be rare in non-Oceanians but we find these alleles at up to 6% frequency in Europeans (Table

S51), who have been inferred to lack Denisovan ancestry. It may therefore be that these haplotypes represent variation present in the introgressing Neandertal population that was not captured by the two sequenced Neandertals.

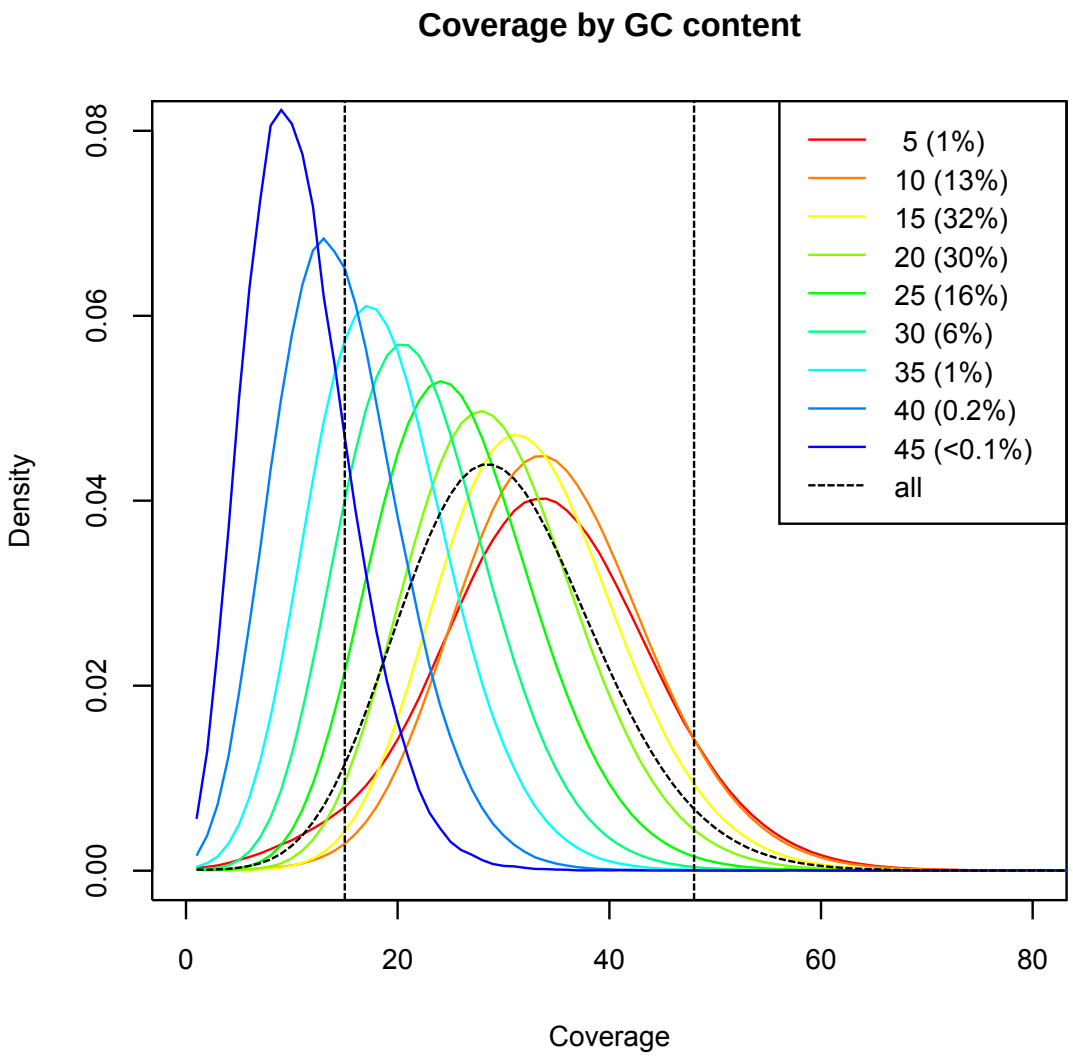

Fig. S1.
Coverage estimates per GC bin for Vindija 33.19. Dashed line shows the coverage distribution summing over all GC bins and the vertical dashed lines mark the edges of the 95% central part of this distribution. Solid lines show the coverage distribution in bins of 5 bases (e.g. 5 stands for GC content between 5 and 9 bases among the 51 bases in a window). The percentage of the mapable genome in each GC bin is given in brackets in the legend.

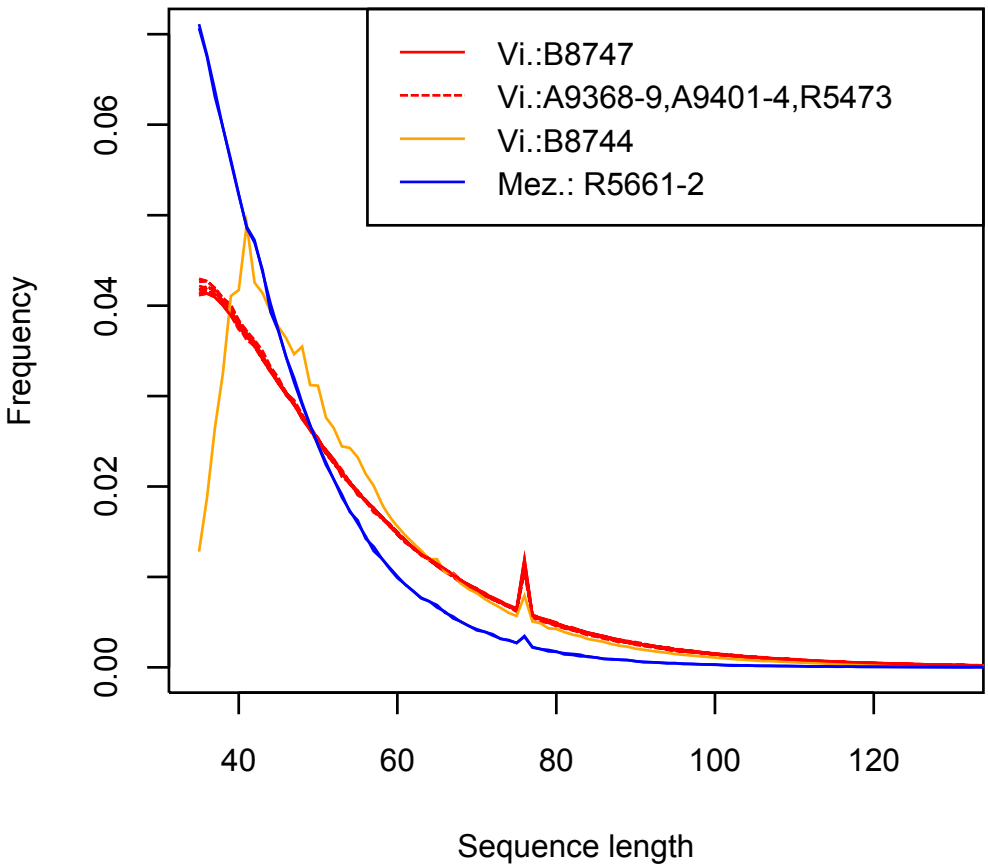

Fig. S2.

Sequence length distributions for Vindija 33.19 (Vi.) and Mezmaiskaya 1 (Mez.) libraries.

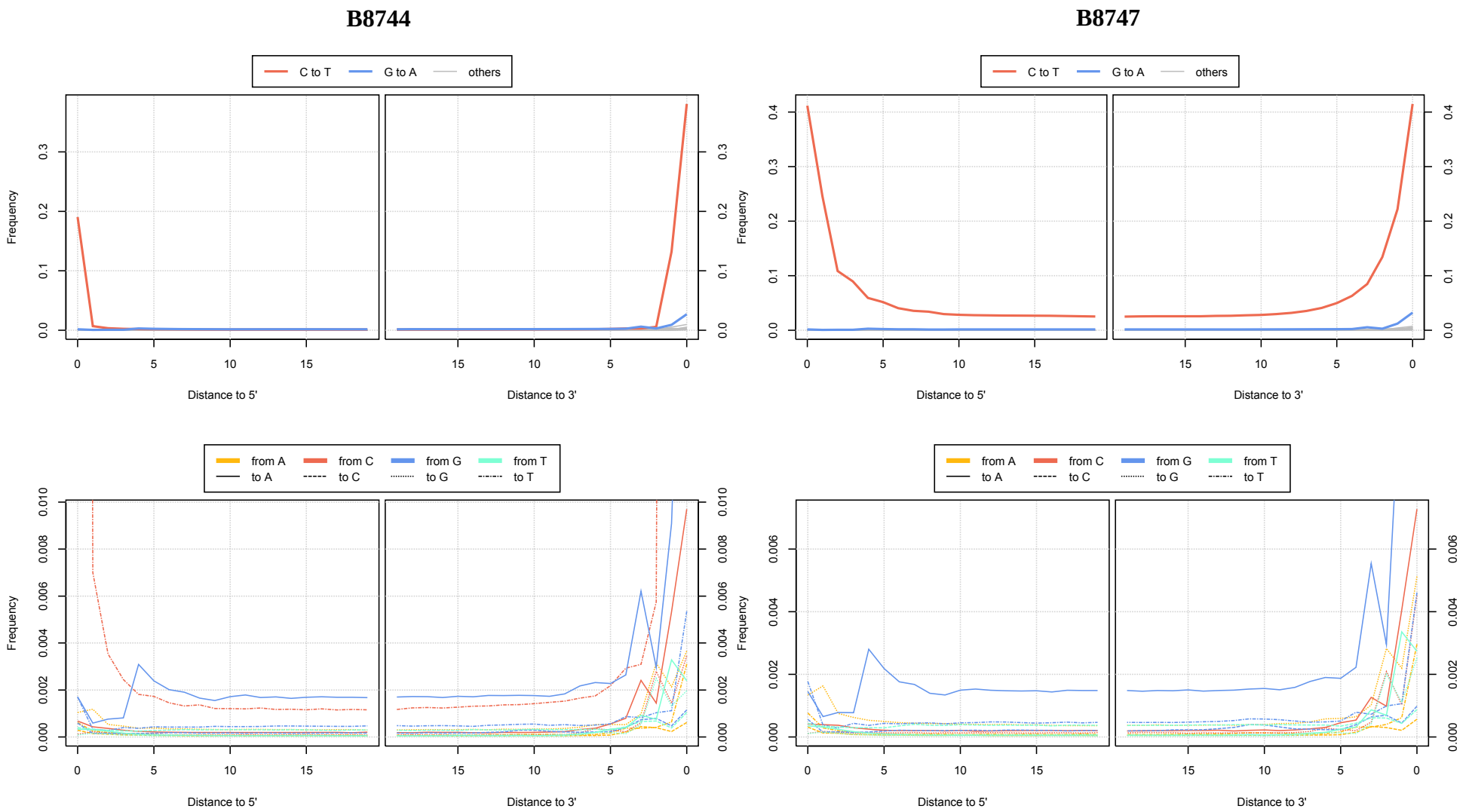

Fig. S3.

Patterns of substitutions along sequences for two Vindija 33.19 libraries: UDG-treated library B8744 (left) and untreated library B8747 (right). Top row shows the strongest substitution rate, C to T, in comparison to the reverse complement substitution, G to A. Other substitutions are shown on a finer Y-scale in the bottom row. Other untreated libraries from Vindija 33.19 look similar to B8747.

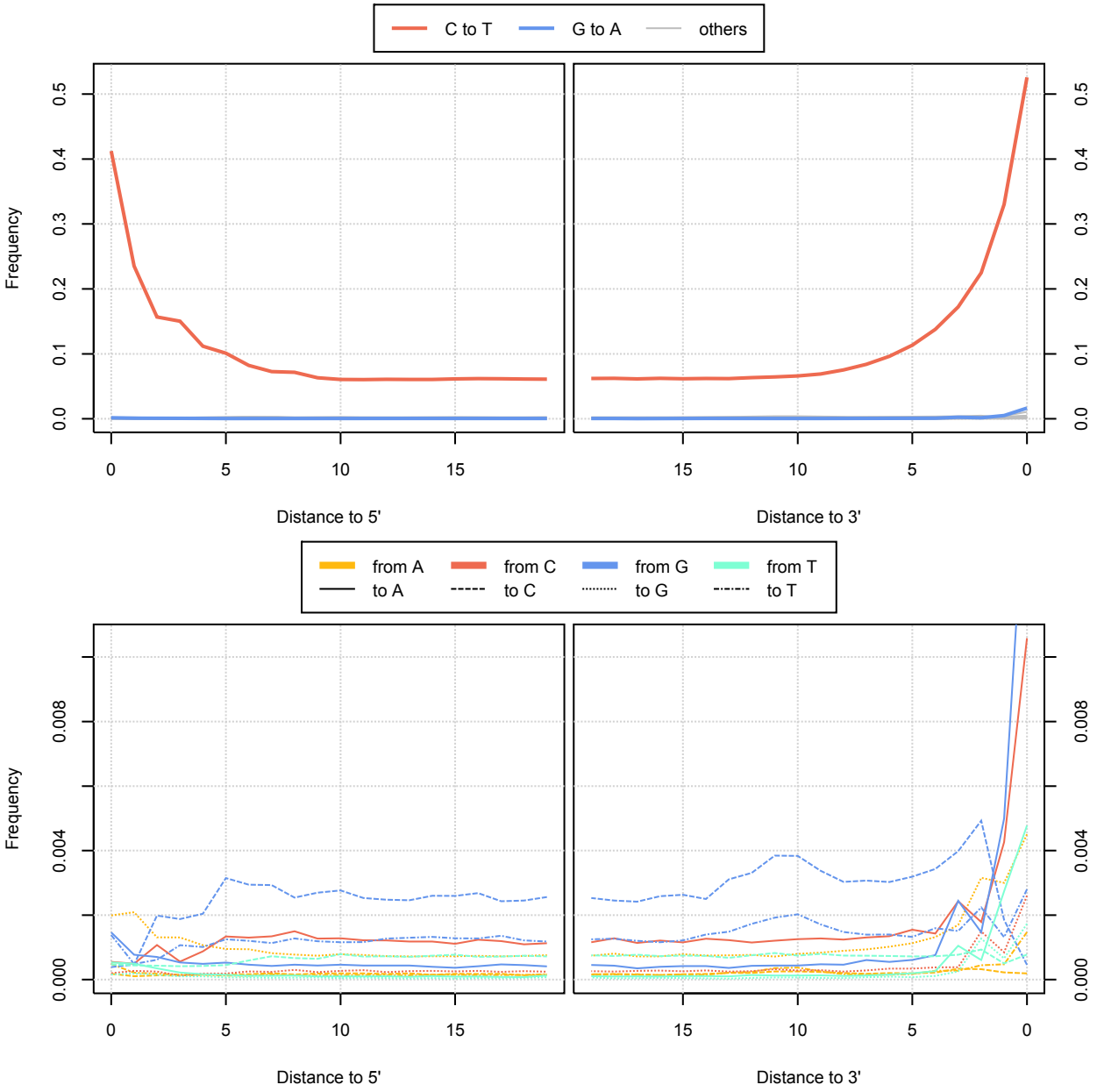

Fig. S4.
Patterns of substitutions along sequences for Mezmaiskaya 1 library R5661. R5662 shows similar patterns.

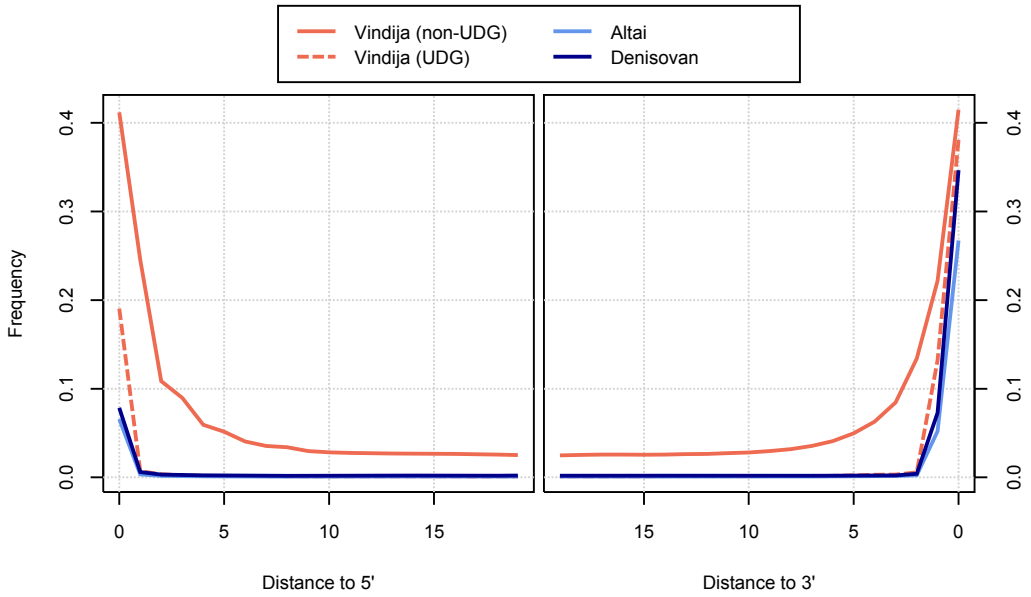

Fig. S5.
C to T changes along sequences from Vindija 33.19, Altai and Denisova. Untreated data ("non-UDG") constitutes with 76% the majority of data from Vindija 33.19.

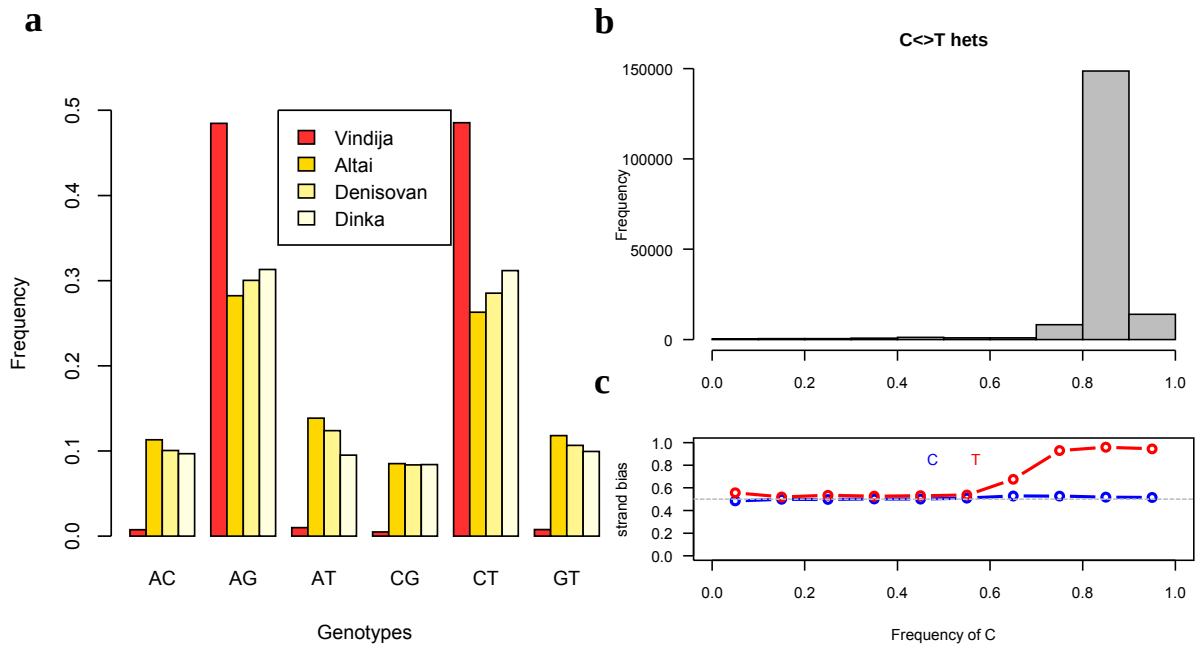

Fig. S6.

a) Fraction of heterozygotes called as each genotype using GATK in Vindija 33.19, Altai Neandertal, Denisovan and a present-day Dinka. **b)** Allelic frequencies at C/T heterozygous sites. **c)** Fraction of forward orientation among sequences showing either C (red) or T (blue) by frequency of C at C/T heterozygous positions. Values around 0.5 indicates that there is an equal proportion of sequences on the forward and reverse strand, while values above and below 0.5 indicate a bias for sequences on the forward and reverse strand, respectively.

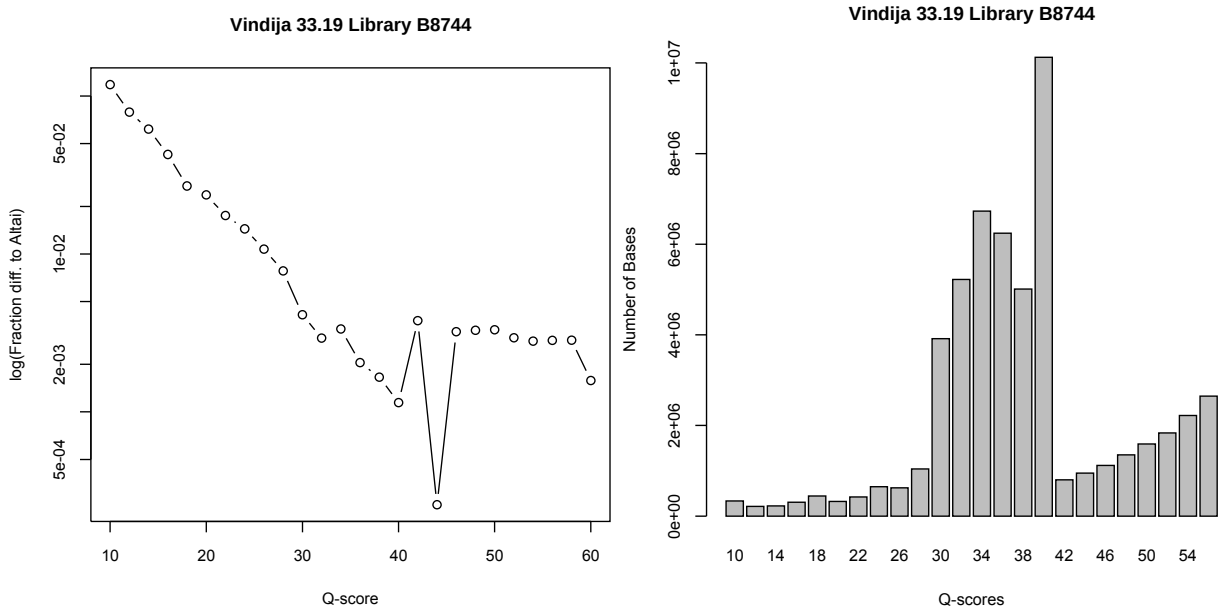

Fig. S7.

Left: Quality scores compared to differences to the Altai Neandertal genome (log-scale). Right: Number of bases in each quality score bin except Q60, which contains ~94% of all bases compared. The library used (B8744) is enzyme treated to remove ancient DNA associated misincorporations.

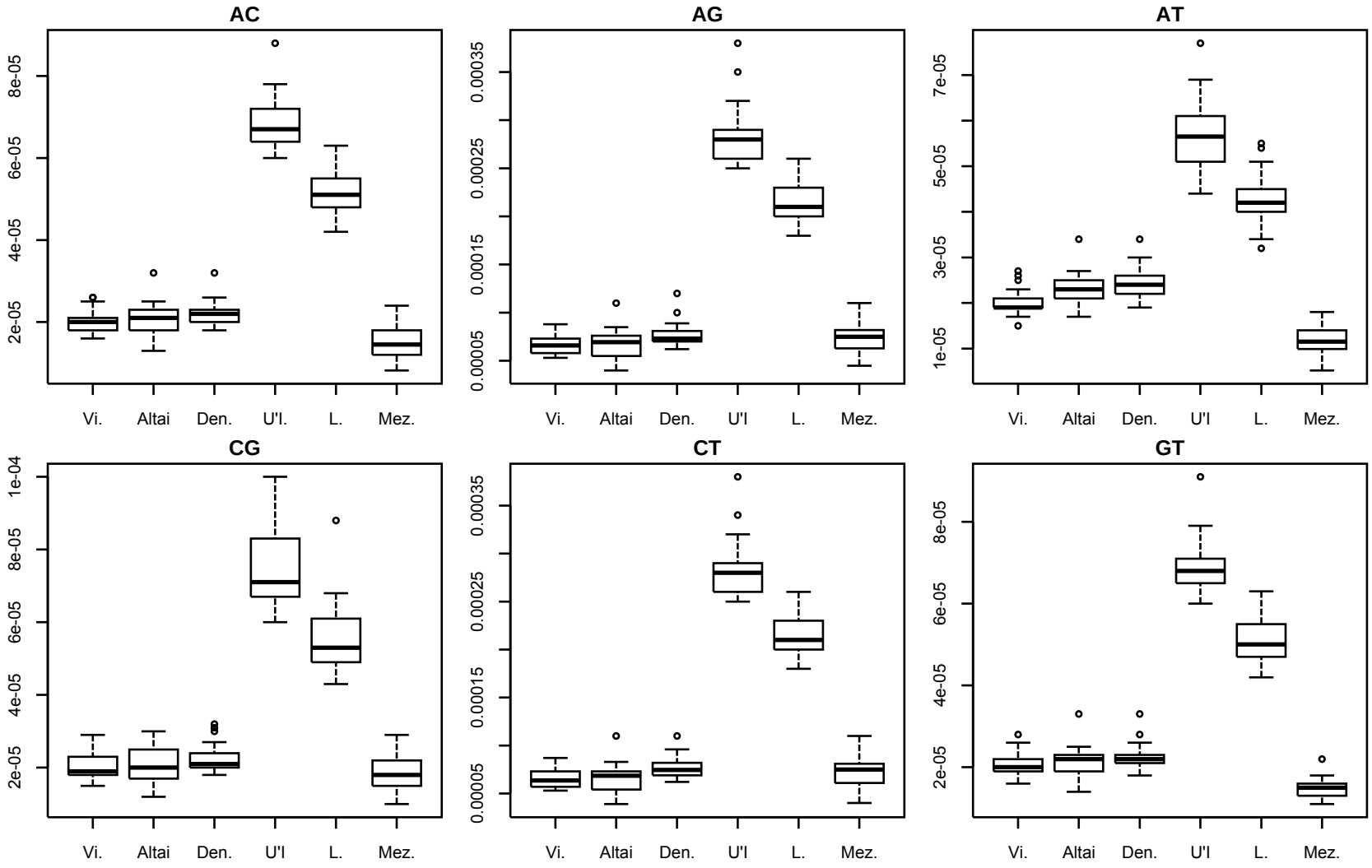

Fig. S8.
Estimated prior probabilities for each genotype over all autosomes for Vindija 33.19 (Vi.), Altai, Mezmaiskaya (Mez.) Neandertals, Ust'Ishim (U'I) and Loschbour (L.), and Denisova (Den.).

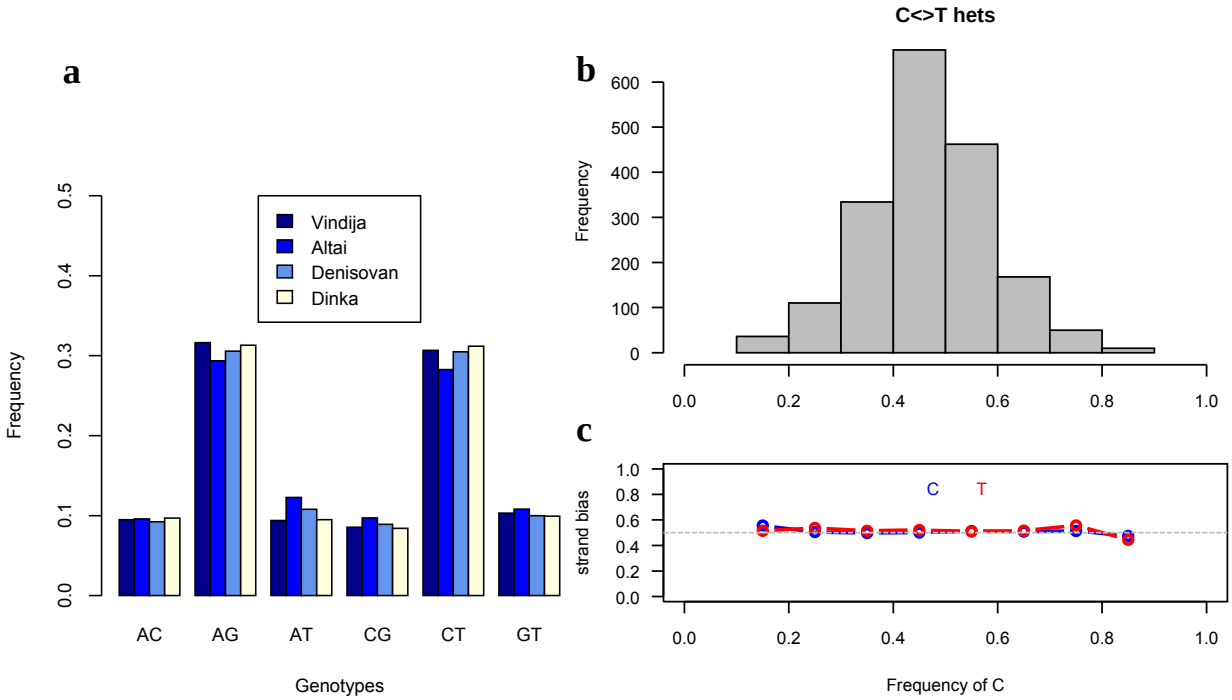

Fig. S9.

a) Fraction of heterozygotes called as each genotype using *snpAD* for Vindija 33.19, Altai Neandertal, Denisovan and GATK for a present-day Dinka. **b)** Allelic frequencies at C/T heterozygous sites. **c)** Fraction of forward orientation among sequences showing either C or T by frequency of C at C/T heterozygous positions.

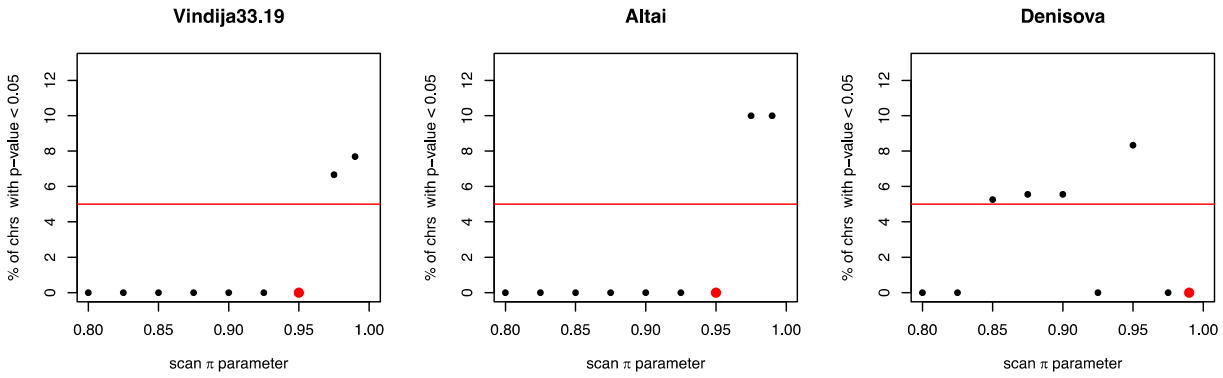

Fig. S10.
Percentage of chromosomes (y-axis) with significant clustering of HBD tracts (p-value < 0.05) longer than 2.5cM for each value of the parameter π (x-axis). The largest π where the percentage of chromosomes is below 5% (red line) and that was chosen for the analyses is shown with a red dot.

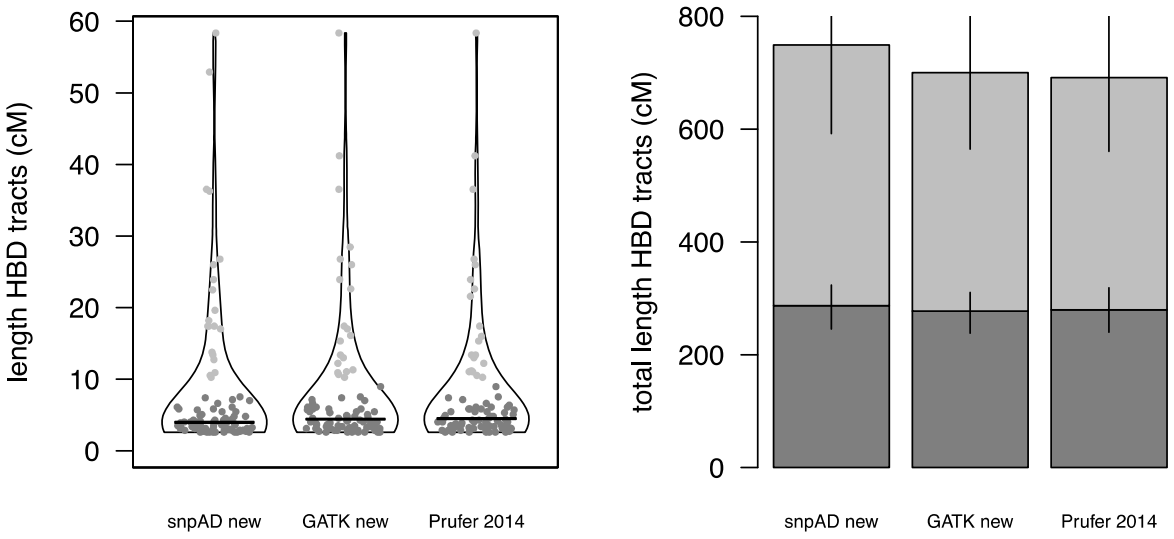

Fig. S11.

HBD chunks comparison among different genotype calls and processing of the Altai Neandertal. "snpAD new" refers to the genotypes from S3, "GATK new" to genotypes with GATK on the decoy aligned sequences, as described in S2. Left: Distribution of putative HBD tracts. Each dot represents a single HBD tract. The black line shows the median of the distribution. Right: Total length of the HBD tracts divided in those >10cM (light gray) and those ≤10cM (dark gray). Vertical lines show the 95% C.I. calculated from 100 bootstrap samples of the tracts.

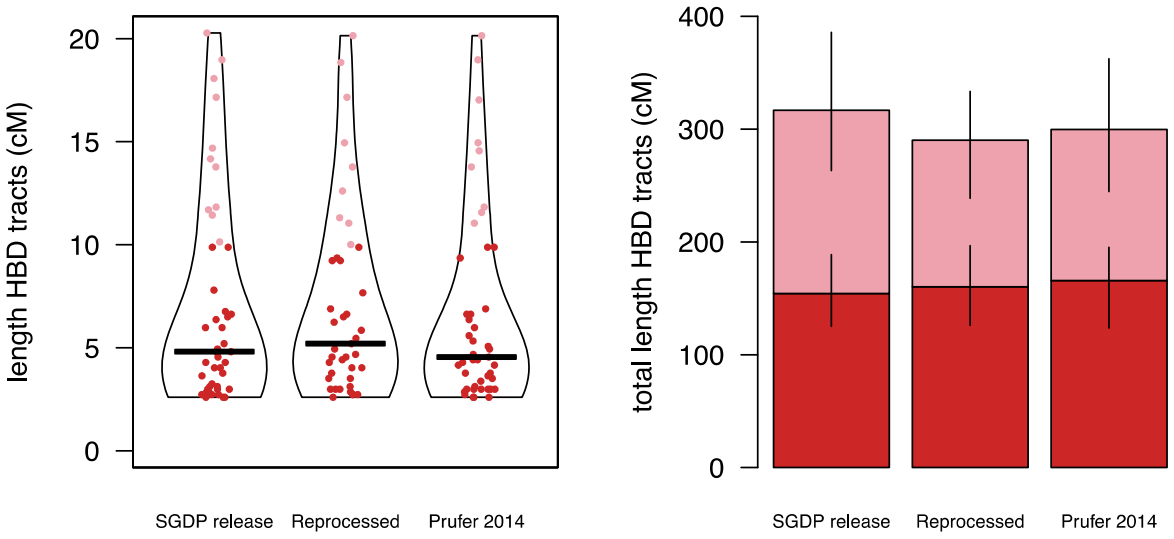

Fig. S12.
HBD chunks comparison among different genotype calls of the same Karitiana individual. Two different filters on genotype quality were applied: for 'SGDP release' GQ ≥ 1, and for the 'Reprocessed' and 'Prufer 2014' GQ ≥ 40. On the left plot is the distribution of putative HBD chunks, where each dot represents a single HBD chunk and the black line the median of the distribution. On the right plot the total length of the HBD chunks divided in those >10cM (red) and those ≤10cM (pink), with the vertical lines representing the 95% C.I. calculated from 100 bootstraps of the tracts.

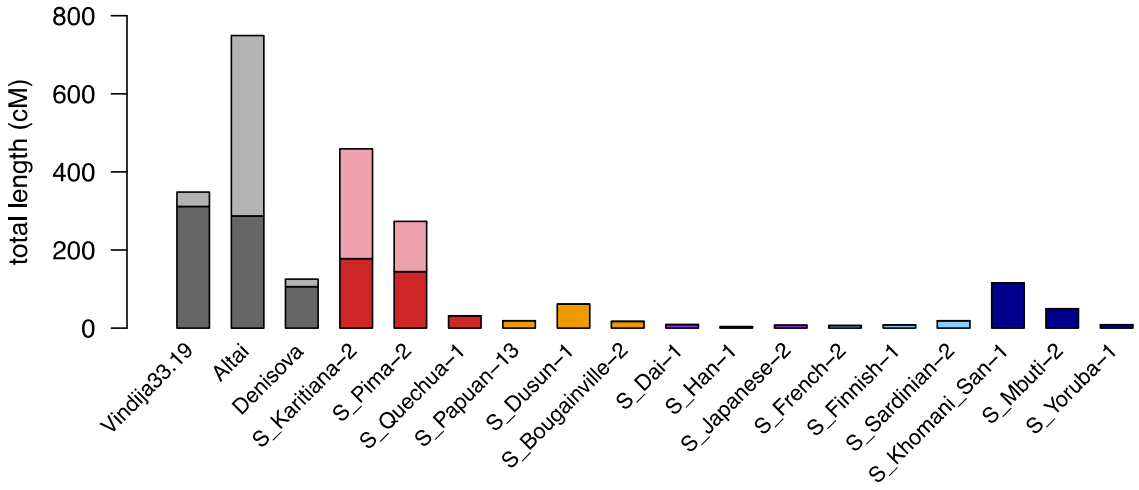

Fig. S13.
Total length of HBD tracts using average recombination rate of 1.3cM/Mb. The darker colors refer to the sum of chunks of length between 2.5 and 10cM in length, and the lighter colors refer to the sum of chunks higher than 10 cM. The colors were chosen to represent the continental locations as labeled in Figure S15.

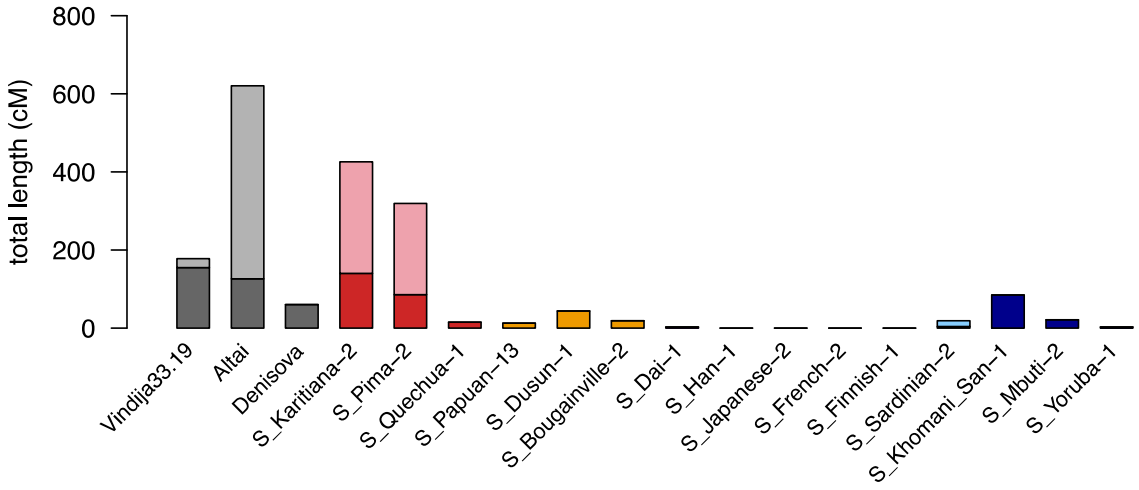

Fig. S14.
Total length of HBD tracts as in Figure S13 but using the African American recombination map (*59*).

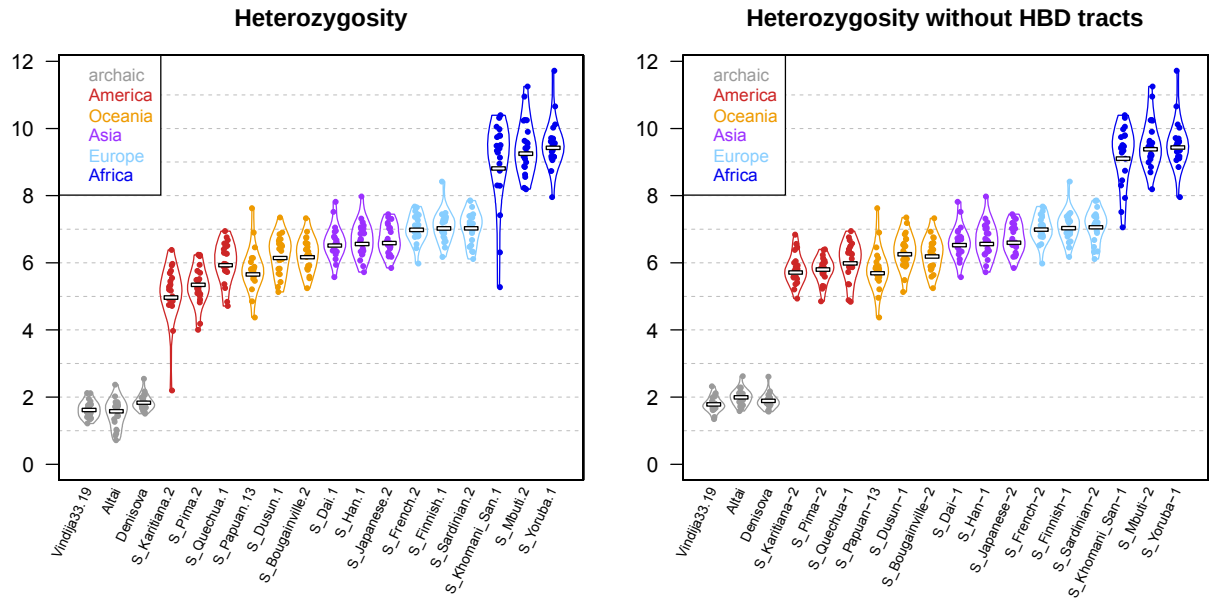

Fig. S15.

Heterozygosity distributions. Each violin plot represents the heterozygosity distribution across the 22 autosomes (which are also shown as in dots), and the white rectangles are the overall genome-wide heterozygosity as reported in Table S14.

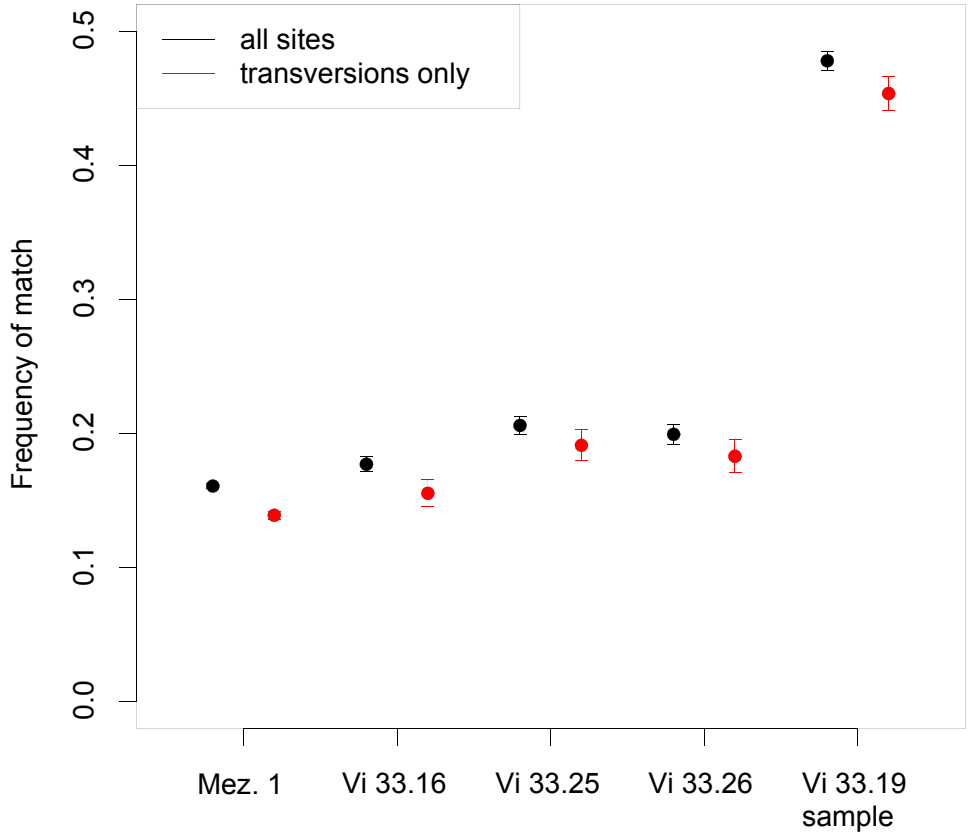

Fig. S16.
Matching of Vindija 33.19 heterozygous sites in low-coverage Neandertal data. Error bars give the 95% binomial confidence interval.

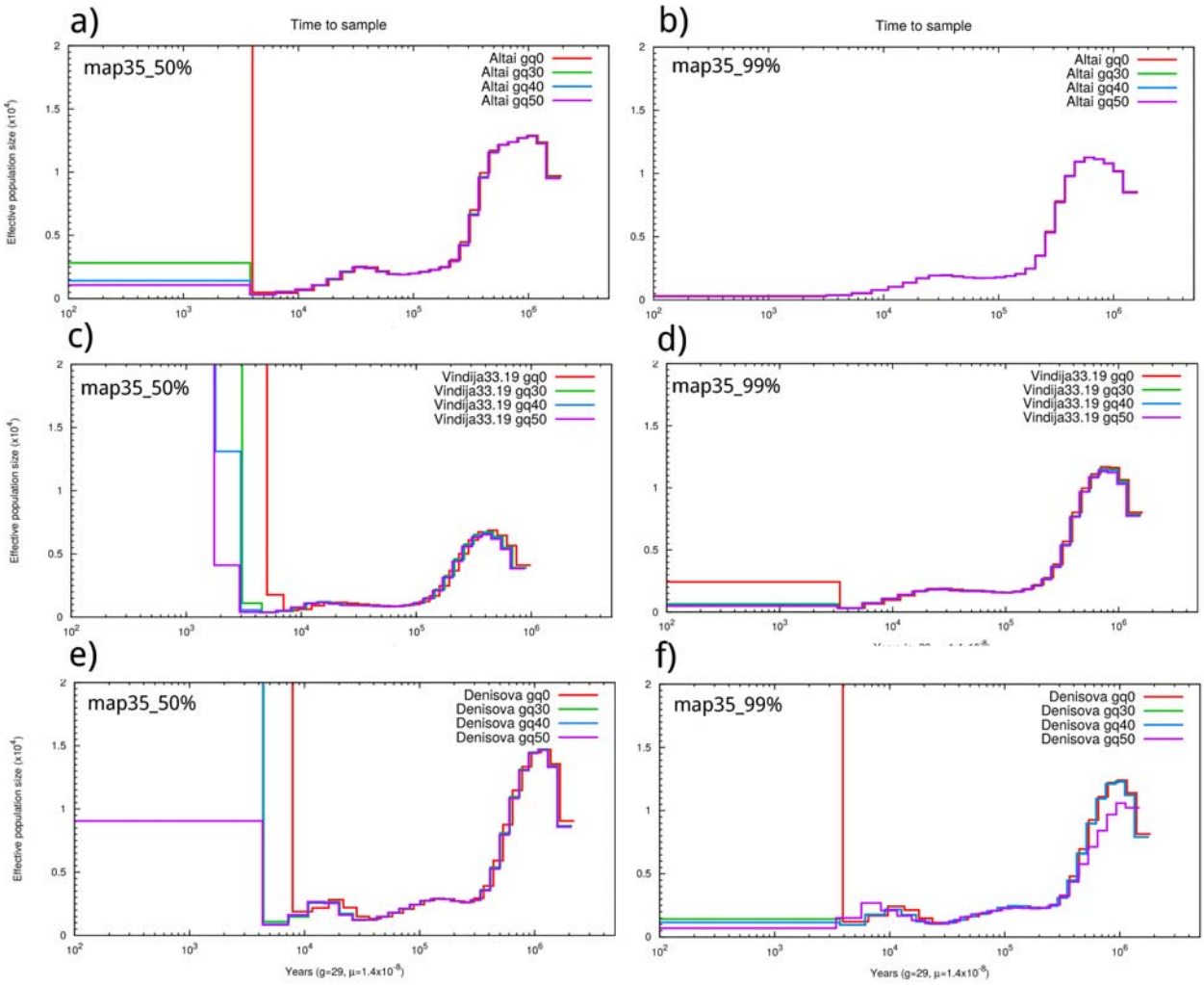

Fig. 17.
PSMC curves for Altai (a-b), Vindija33.19 (c-d) and Denisova (e-f) filtered with different genotype quality thresholds and with different mappability filters, map35_50% (a,c,e) and map35_100% (b,d,f).

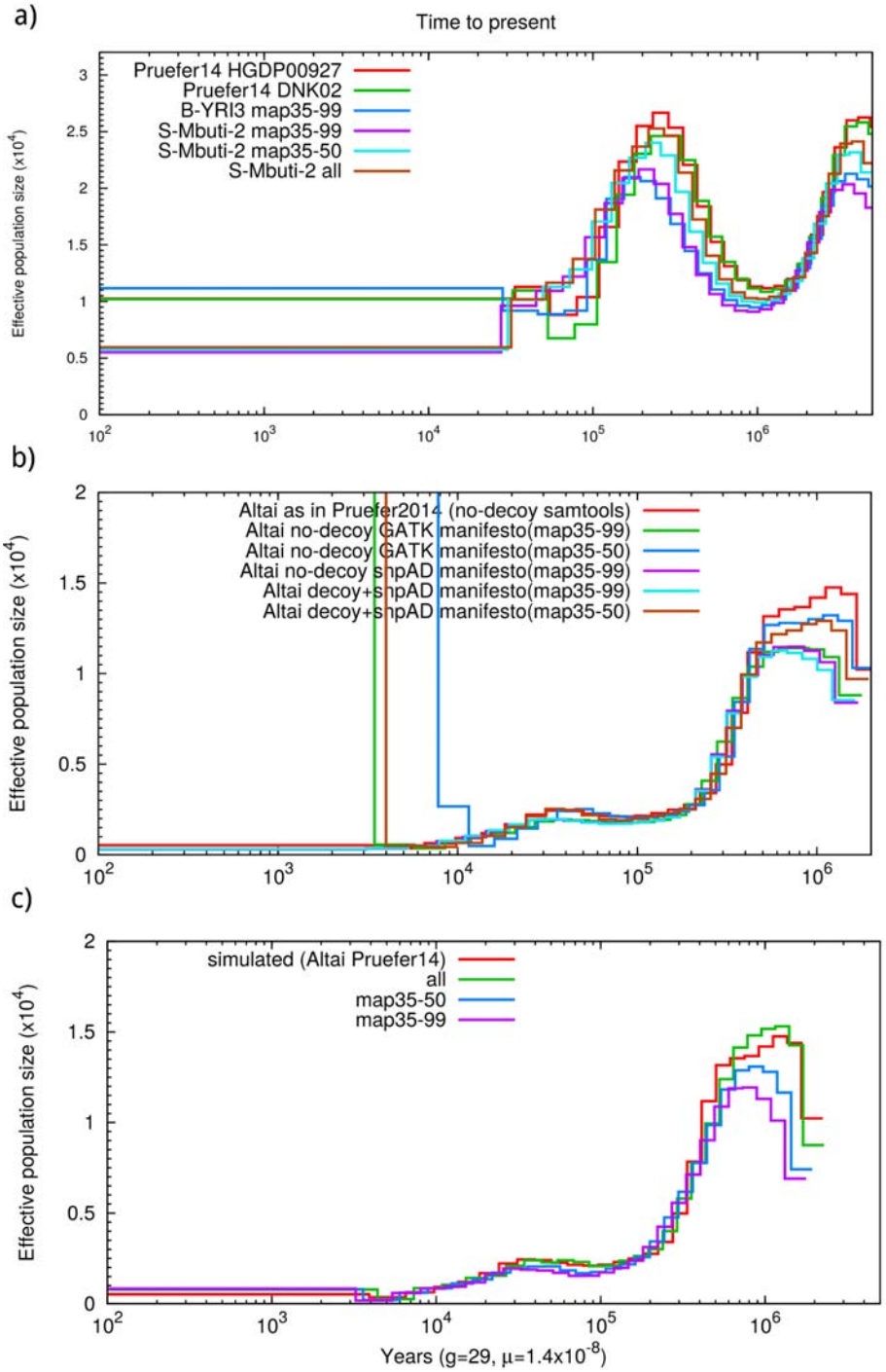

Fig. S18.
Effects of more stringent filtering on PSMC estimates for a) modern, b) ancient and c) simulated genomes.

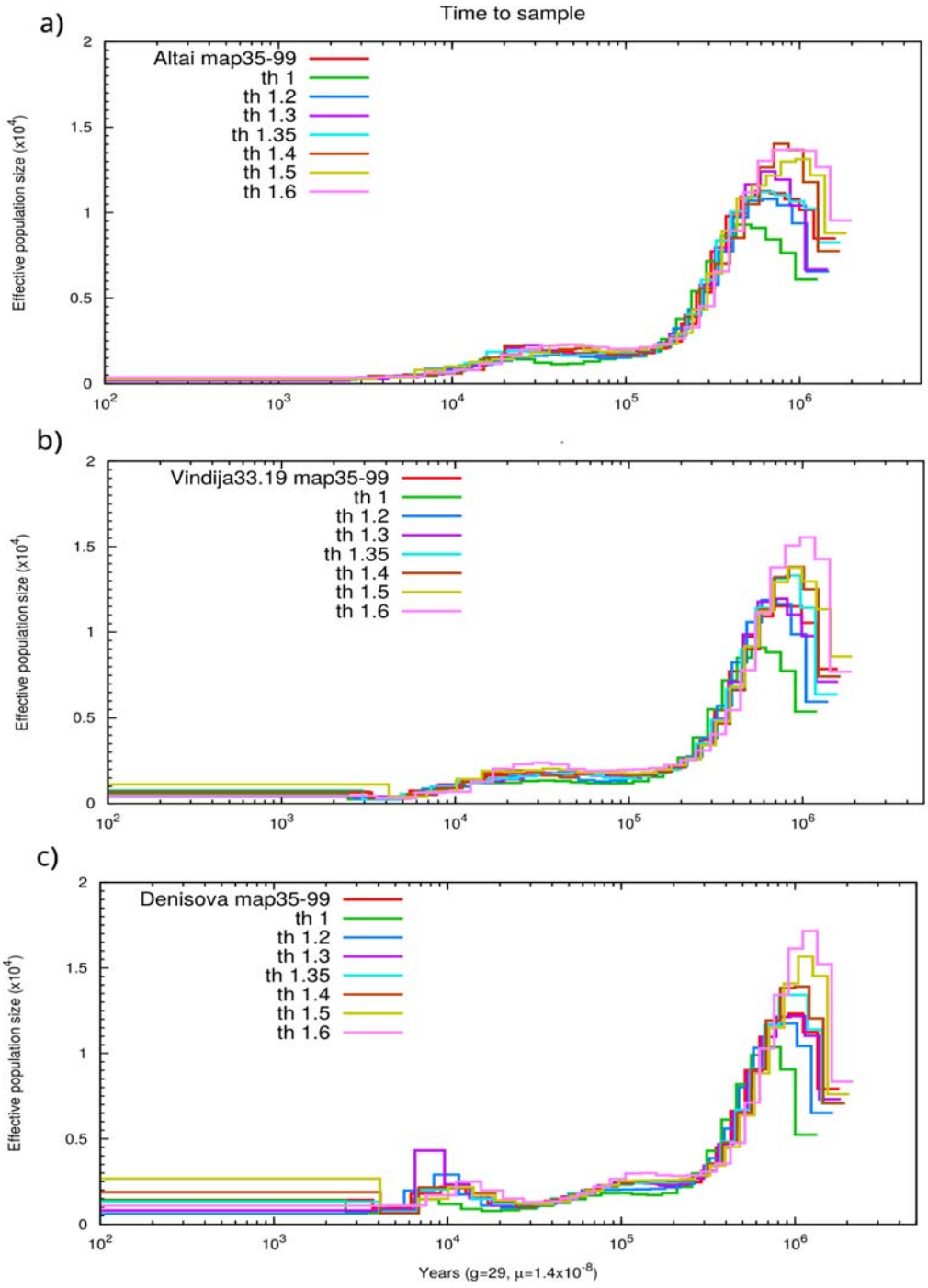

Fig. S19.

A sample of rescaling curves obtained to fit the corrected demography for Altai (a), Vindija (b) and Denisova (c). The estimated curves from the filtered data with mapability filter map35_100 (red) are rescaled by multiplying theta by a factor 'th' as shown in the legend for the different curves.

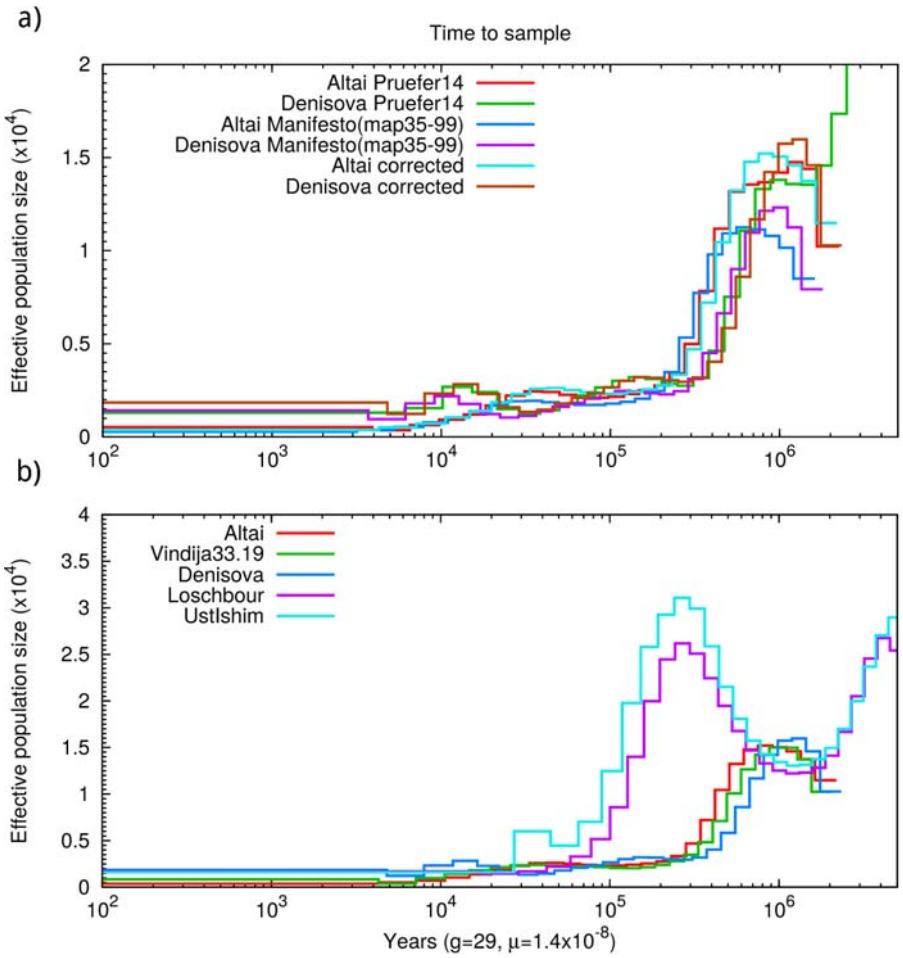

Fig. S20.

PSMC curves after correction. (a) Curves for Altai and Denisova as shown in Prüfer 2014, after applying filters including map35_100 and after correction. (b) Demographic histories of ancient hominin samples after correction.

```
ms 2 1 -t 645.6901 -r 95.580962351238 30000000 -en 0.0997 1 1.3763 -en 0.1659 1 1.9359 -en 0.2462 1 2.8299 -en 0.3434 1 3.7586 -en 0.4613 1 5.0165 -en 0.6041 1 6.2030 -en 0.7773 1 6.9742 -en 0.9871 1 7.0692 -en 1.2415 1 6.7771 -en 1.5497 1 6.4688 -en 1.9233 1 6.2562 -en 2.3760 1 6.2782 -en 2.9247 1 6.5188 -en 3.5898 1 6.8857 -en 4.3958 1 7.4881 -en 5.3727 1 8.9953 -en 6.5566 1 12.6641 -en 7.9916 1 19.4238 -en 9.7306 1 28.1659 -en 11.8384 1 35.6656 -en 14.3929 1 39.7922 -en 17.4889 1 40.9909 -en 21.2412 1 40.5491 -en 25.7889 1 39.3030 -en 31.3005 1 37.0012 -en 37.9805 1 30.9487 -en 55.8887 1 101.6739

ms 2 1 -t 1480.264 -r 228.766112986448 30000000  -en 0.0436 1 0.4496 -en 0.0724 1 1.1386 -en 0.1073 1 1.6976 -en 0.1494 1 2.1414 -en 0.2004 1 2.5273 -en 0.2620 1 2.7677 -en 0.3365 1 2.8236 -en 0.4266 1 2.7121 -en 0.5355 1 2.6260 -en 0.6671 1 2.6174 -en 0.8264 1 2.5544 -en 1.0189 1 2.4373 -en 1.2516 1 2.4051 -en 1.5331 1 2.5286 -en 1.8733 1 2.8193 -en 2.2848 1 3.2856 -en 2.7822 1 4.1041 -en 3.3837 1 5.6754 -en 4.1109 1 8.3609 -en 4.9902 1 11.8388 -en 6.0533 1 14.8836 -en 7.3388 1 16.7310 -en 8.8930 1 17.6350 -en 10.7722 1 17.6137 -en 13.0443 1 16.1529 -en 15.7915 1 12.0577 -en 23.1293 1 44.8768

ms 2 1 -t 3215.941 -r 497.848889692575 30000000 -en 0.0224 1 0.6683 -en 0.0373 1 1.2633 -en 0.0552 1 1.5349 -en 0.0769 1 1.2425 -en 0.1031 1 0.8649 -en 0.1349 1 0.7299 -en 0.1733 1 0.7947 -en 0.2197 1 0.9664 -en 0.2758 1 1.1379 -en 0.3437 1 1.2766 -en 0.4258 1 1.4448 -en 0.5252 1 1.6290 -en 0.6452 1 1.7332 -en 0.7905 1 1.7182 -en 0.9661 1 1.6411 -en 1.1786 1 1.6021 -en 1.4354 1 1.7259 -en 1.7461 1 2.1842 -en 2.1218 1 3.1602 -en 2.5762 1 4.6545 -en 3.1258 1 6.3211 -en 3.7903 1 7.6928 -en 4.5940 1 8.5163 -en 5.5660 1 8.6435 -en 6.7415 1 7.8947 -en 8.1631 1 5.5635 -en 11.9614 1 25.2126
```

Fig. S21.
Command lines to generate corrected demographies for the Altai Neandertal, Vindija Neandertal and Denisovan, respectively.

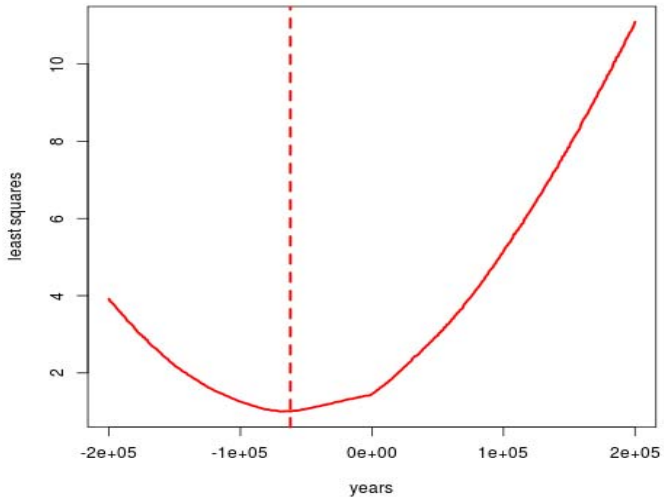

Fig. S22.
Alignment of the PSMC curves of Vindija33.19 and Altai using a least squares fit. 10000 estimated Ne values are sampled logarithmically in time analogously to the time subdivision of PSMC intervals according to the formula ti=i log(1+10Tmax) − 1), with i from 1 to 10000. Least square values are normalized by the minimum value of the fit.

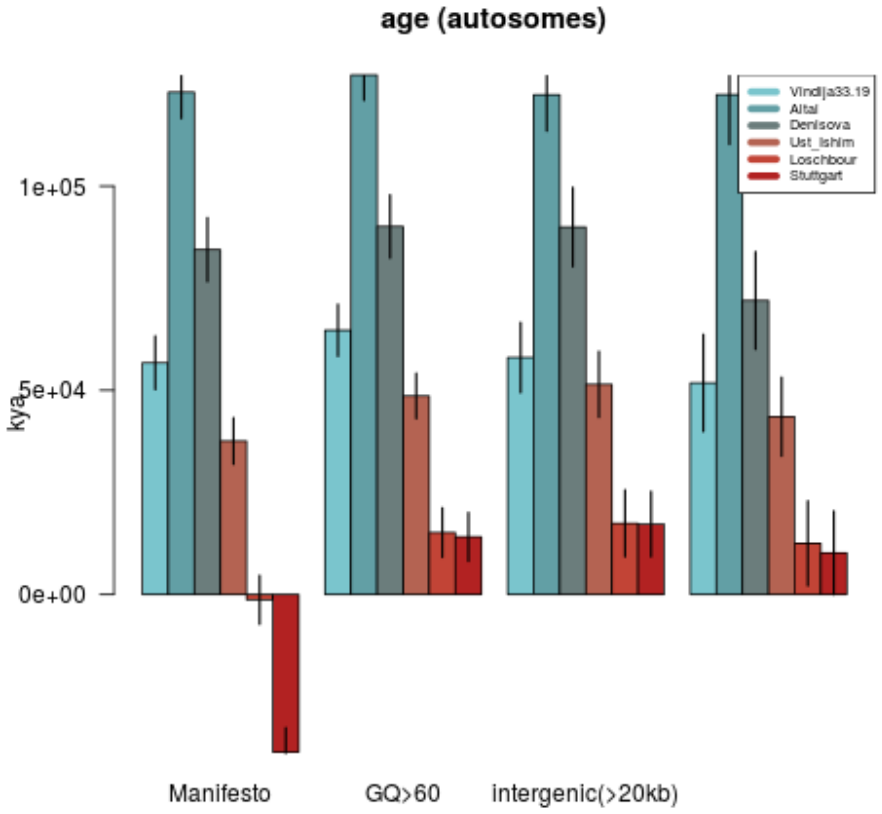

Fig. S23.
Autosomal branch shortening estimates for Vindija33.19, Altai, Denisova, Ust'Ishim, Loschbour and Stuttgart versus a present day African genome.

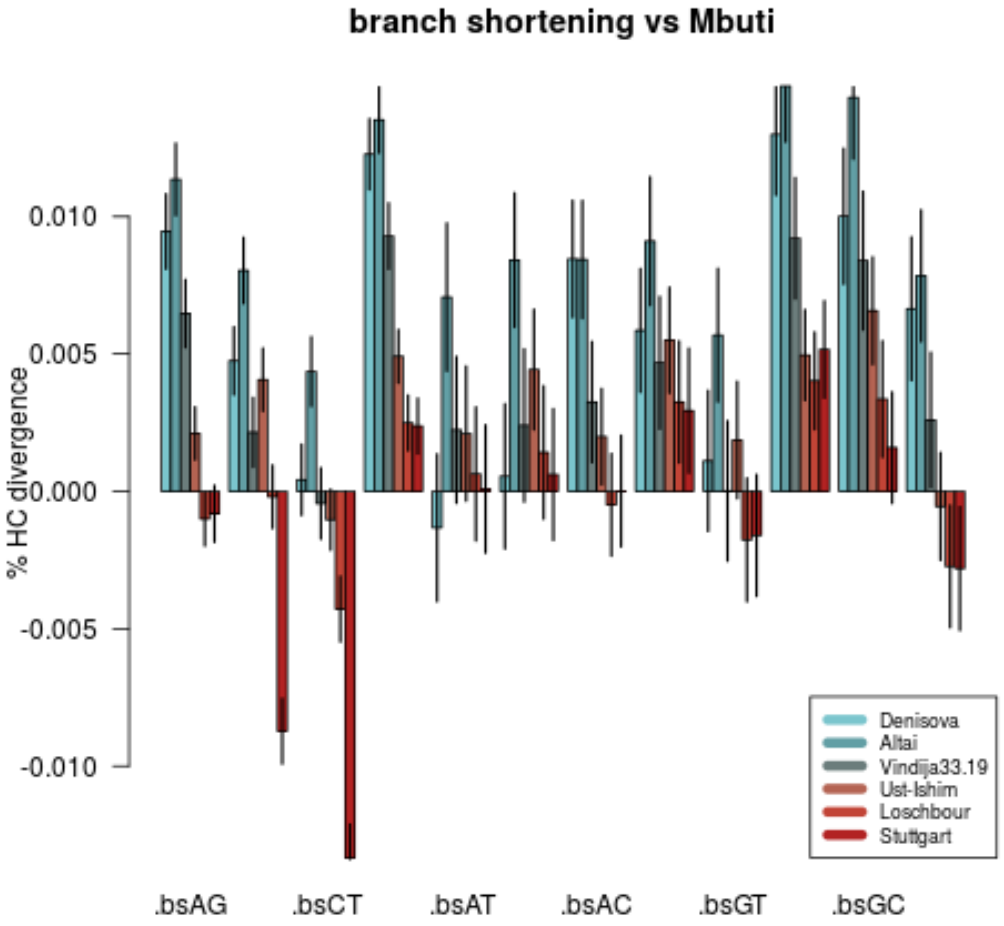

Fig. S24.
Branch shortening across substitution types in terms of human-chimpanzee divergence fraction.

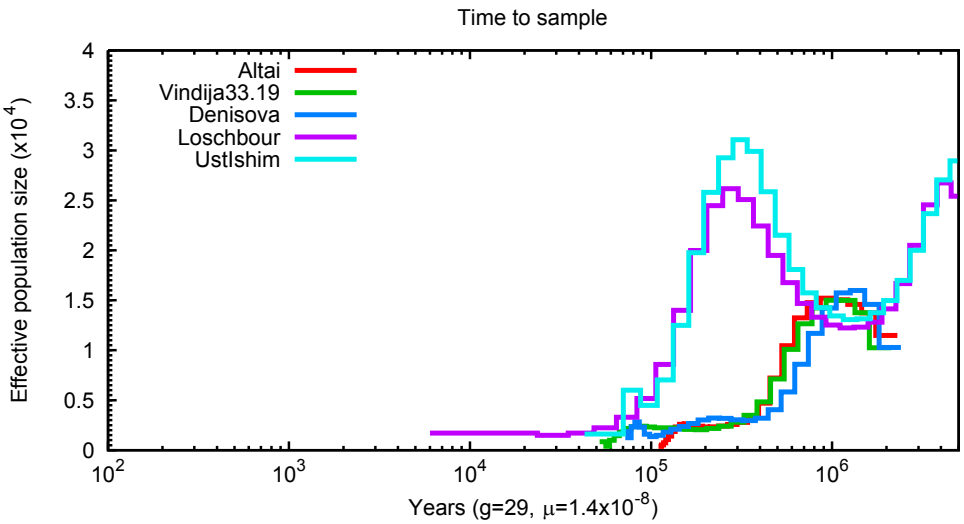

Fig. S25.
PSMC curves for the ancient genomes in FigS7.4 in terms of time before present rescaled by branch shortening.

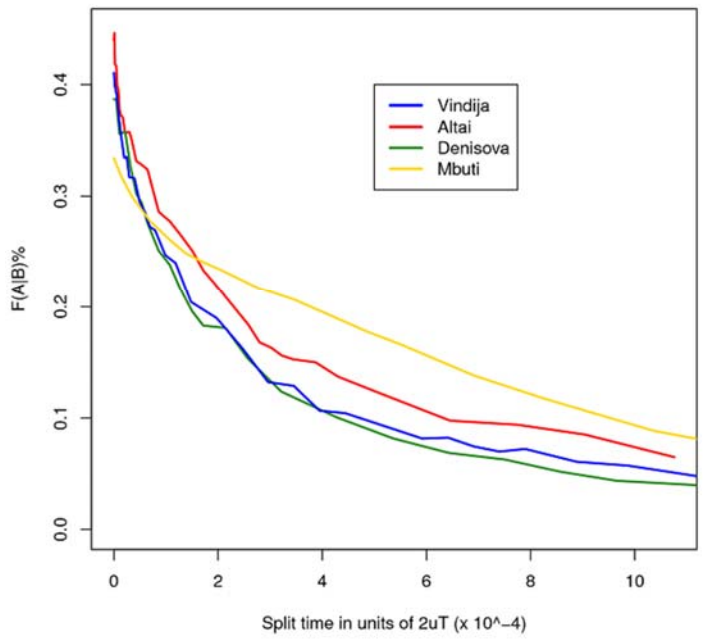

Fig. S26.
Calibration curves for the F(A|B) split times estimates. The observed F(A|B) values are fitted against this calibration curve in order to obtain a split time in terms of human-chimpanzee divergence. This is later corrected by branch shortening and multiplied by the split time between human and chimpanzees.

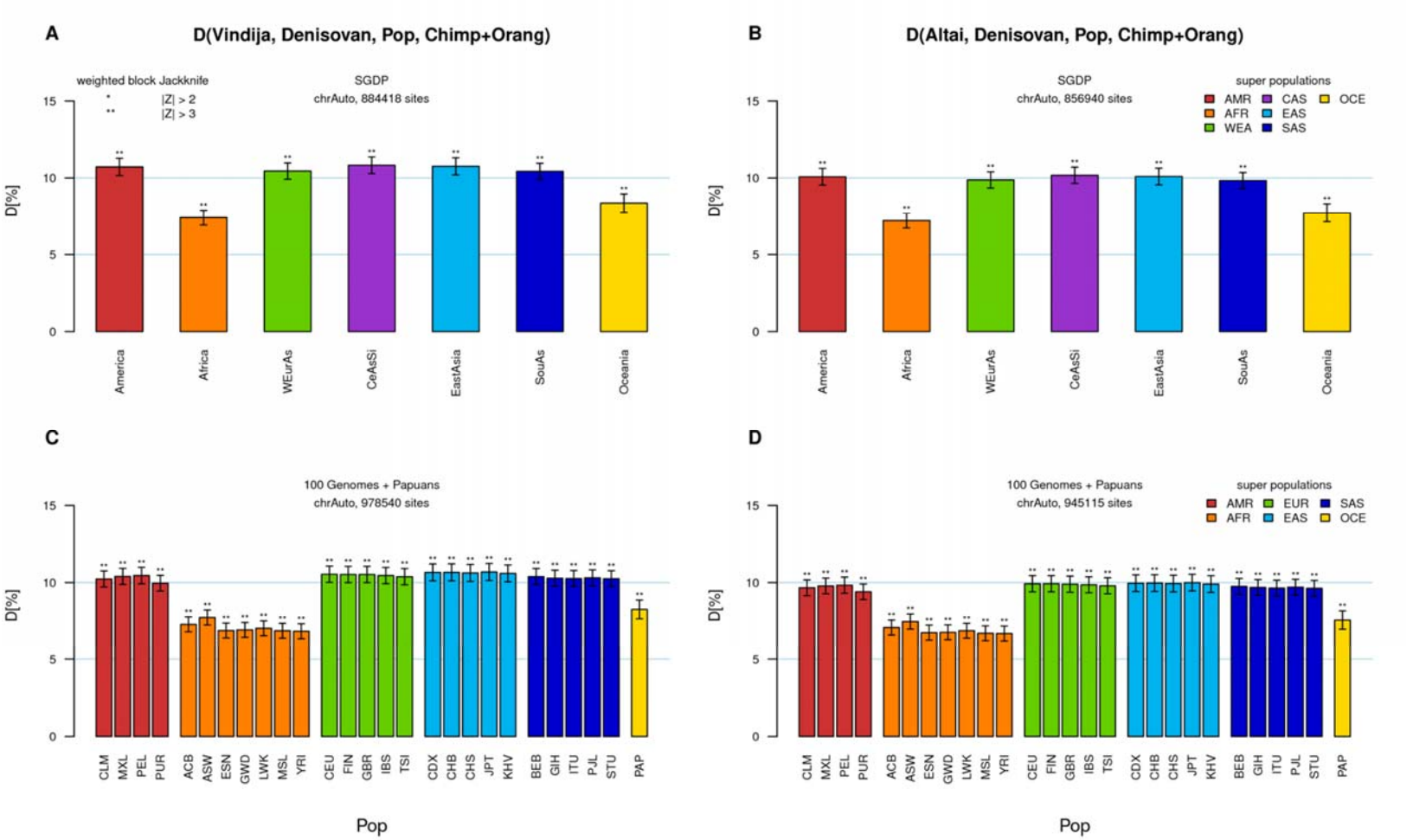

Fig. S27.
D-statistic comparing derived allele sharing with modern human populations between the two Neandertals (A,C: Vindija, B,D: Altai) and the Denisovan. Values above zero indicate a larger amount of derived alleles shared with the Neandertal compared to the Denisovan. Data on modern humans: SGDP superpopulations (A,B), 1000 Genomes Project and Papuan samples (C,D). Error bars indicate 1 standard error. Abbreviations: AMR=Americans, AFR=Africans, WEA=West Eurasians, EUR=Europeans, CAS=Central Asians and Siberians, EAS=East Asians, SAS=South Asians, OCE=Oceanians, PAP=Papuans. For abbreviations of the 1000 Genomes populations see http://www.1000genomes.org/category/population/ .

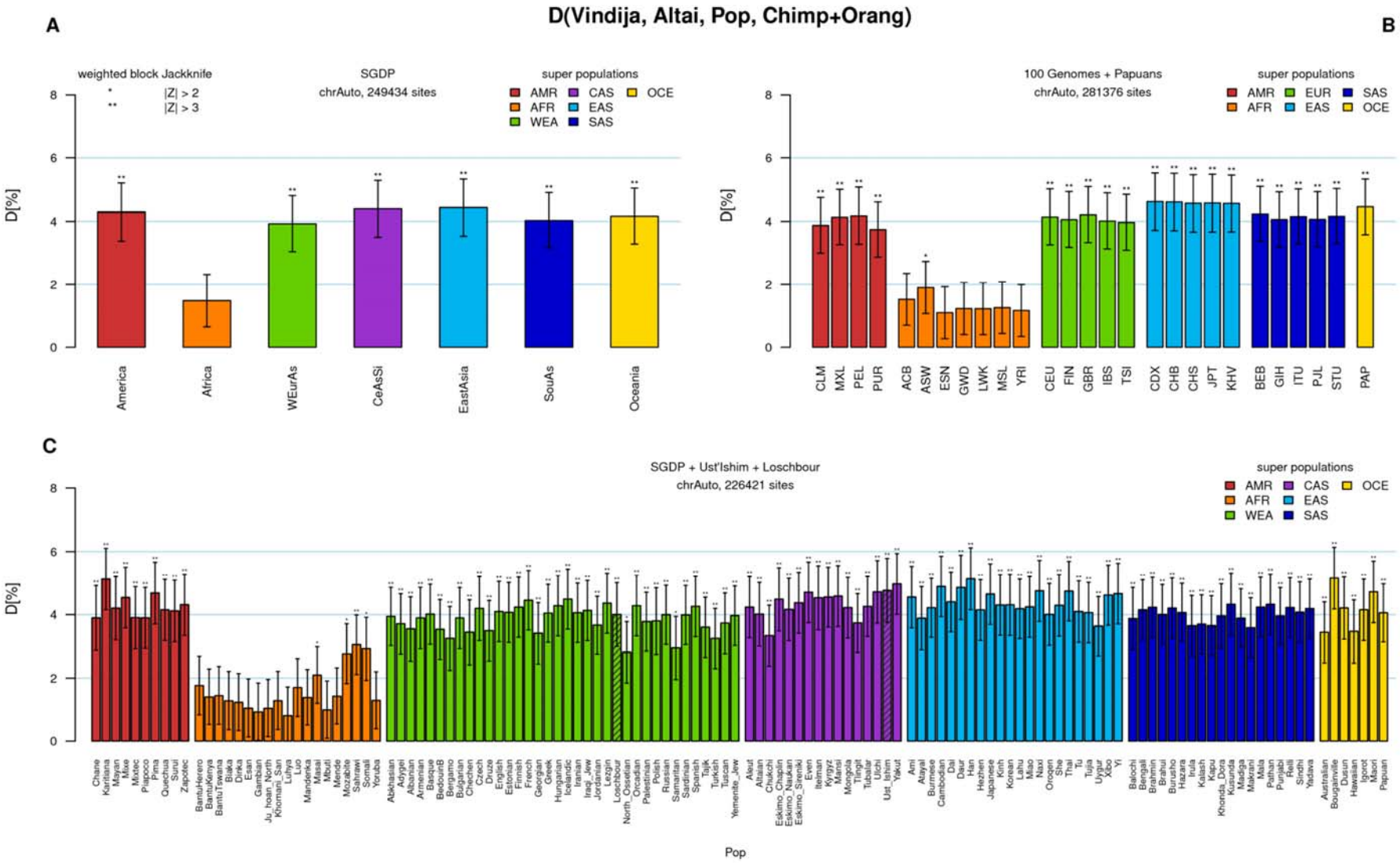

Fig. S28. D-statistic comparing derived allele sharing with modern human populations between the Altai Neandertal and the Vindija Neandertal. Values above zero indicate a larger amount of derived alleles shared with the Vindija Neandertal compared to the Altai Neandertal. (C) The ancient modern humans Ust'Ishim and Loschbour are shown as striped bars and are included into the CAS and WEA superpopulations, respectively. Error bars indicate 1 standard error.

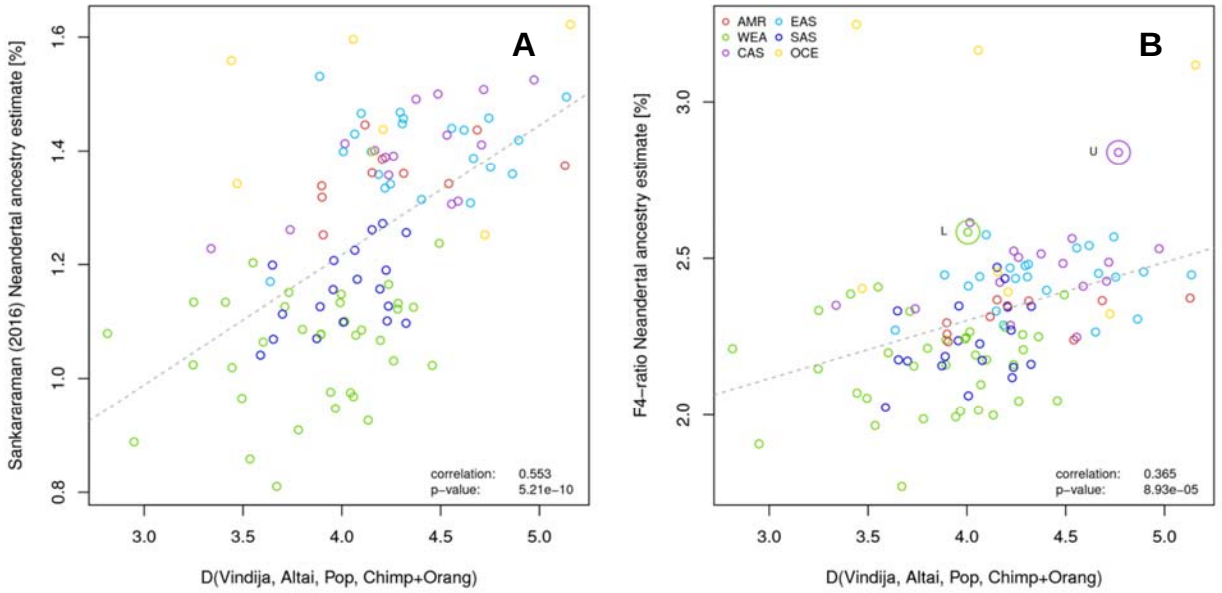

Fig. S29.

Correlation (Pearson's r) of the excess of derived allele sharing of modern human populations with the Vindija as compared to the Altai Neandertal (Fig. S28.), with the Neandertal ancestry estimate for the SGDP populations obtained from (*76*) (A) and from F4-ratios (B, see below). L: Loschbour, U: Ust'Ishim.

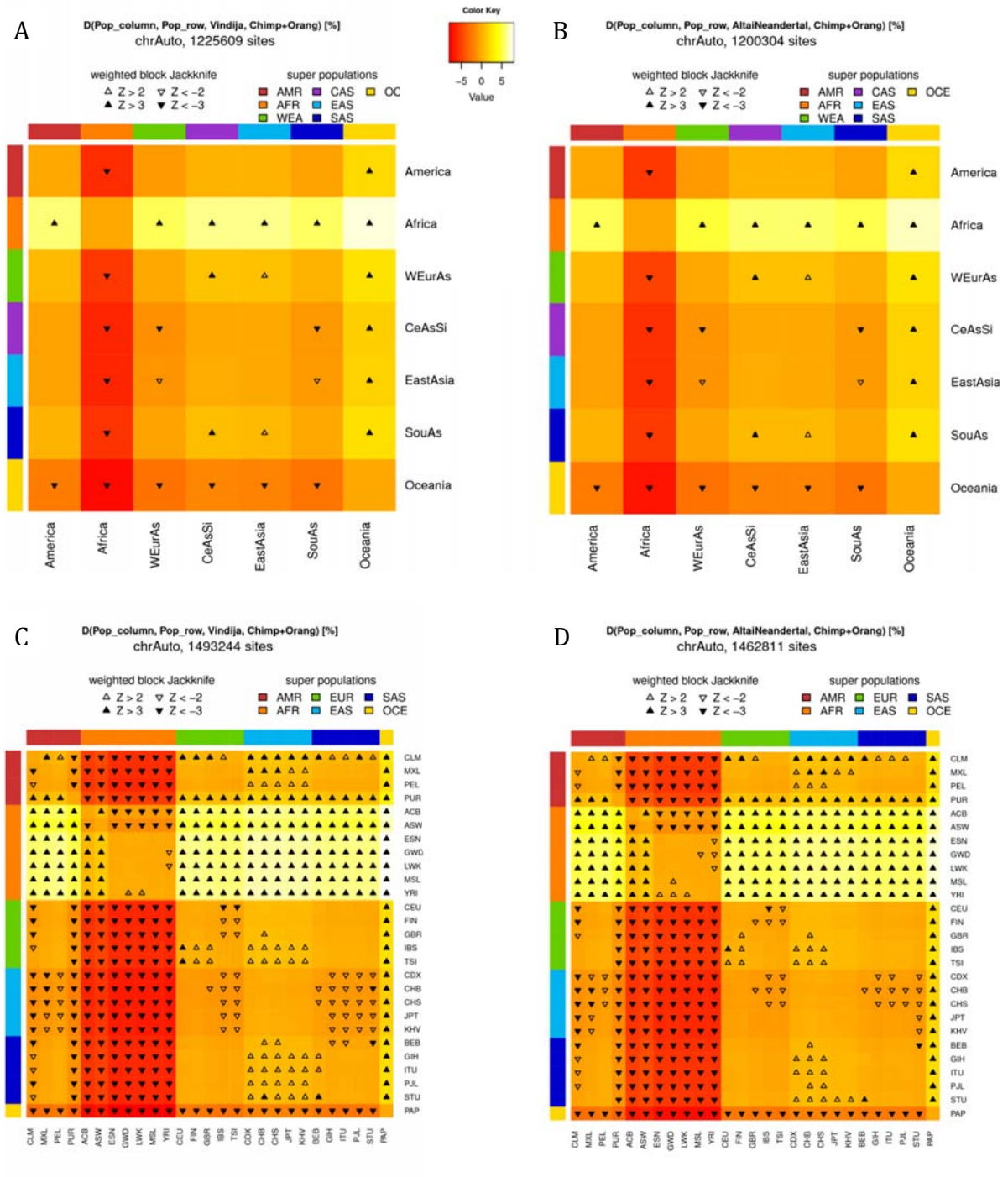

Fig. S30.

D-statistics comparing derived allele sharing with the Vindija Neandertal (A,C) and the Altai Neandertal (B,D) between pairs of modern human populations. Results are based on the autosomes. Values below zero indicate more derived allele sharing with the Neandertal for the population in the row compared to the population in the column. The sign of the Z-scores depends on the sign of the D-statistic which is indicated by arrows for significant results. Data on modern humans: SGDP superpopulations (A,B), 1000 Genomes Project and Papuan samples (C,D).

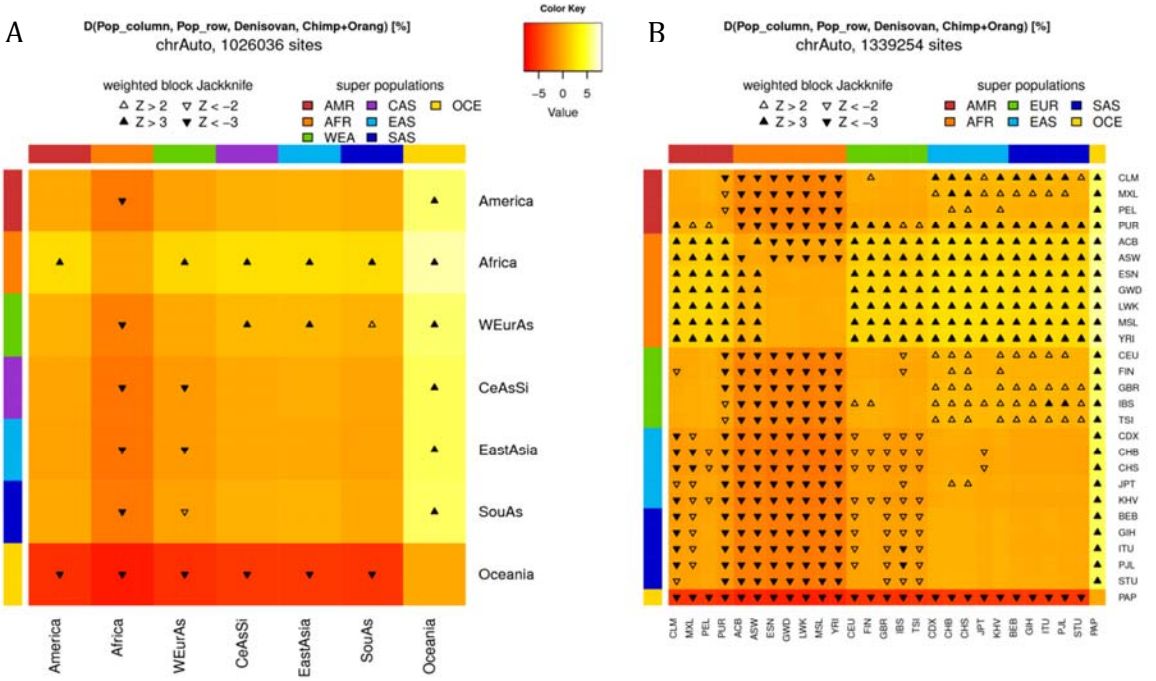

Fig. S31.

D-statistics comparing derived allele sharing with the Denisovan between pairs of modern human populations. Results are based on the autosomes. Values below zero indicate more derived allele sharing with the Denisovan for the population in the row compared to the population in the column. The sign of the Z-scores depends on the sign of the D-statistic which is indicated by arrows for significant results. Data on modern humans: SGDP superpopulations (A), 1000 Genomes Project and Papuan samples (B).

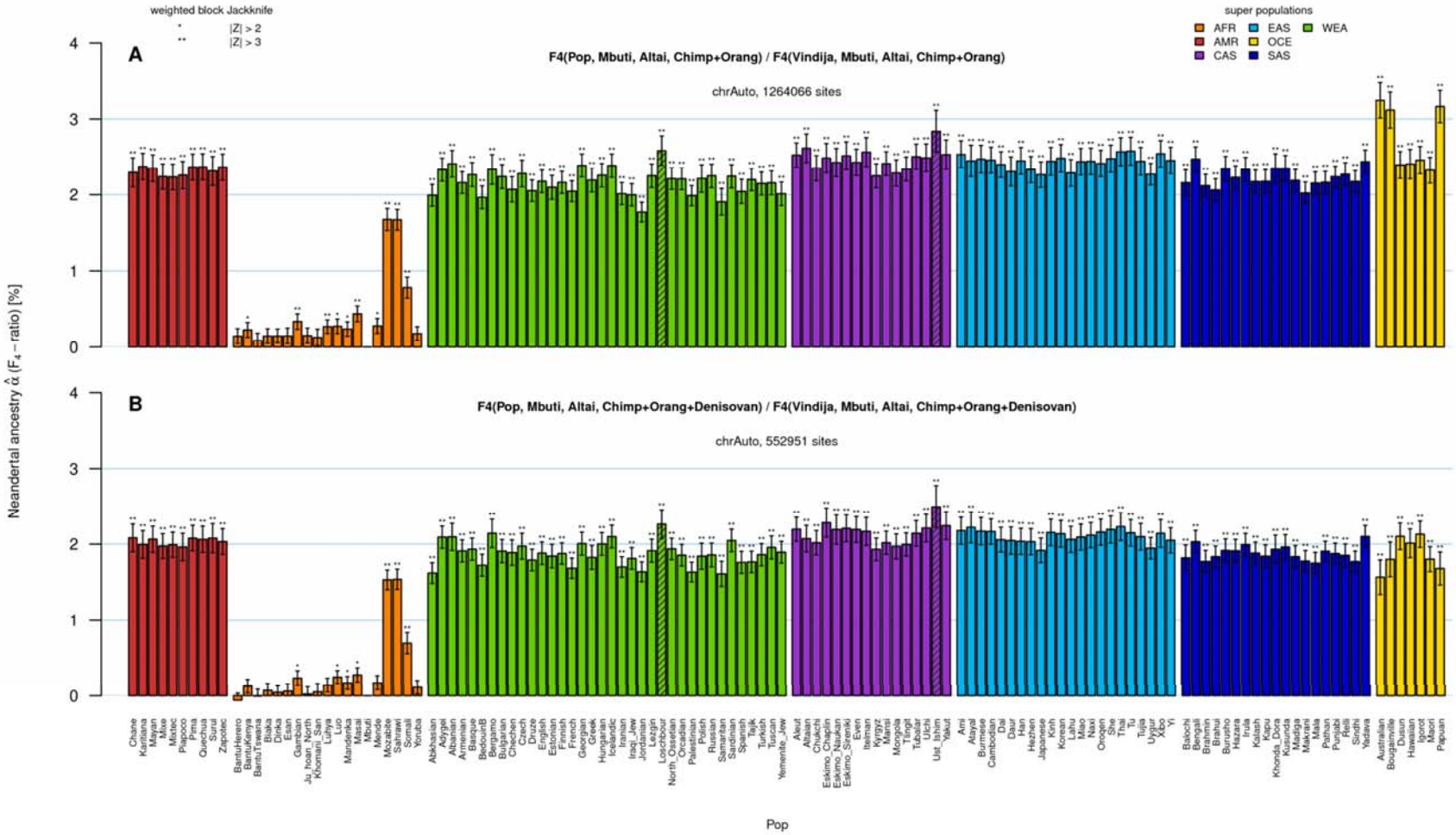

Fig. S32.
Neandertal ancestry for the SGDP subpopulations, together with Loschbour and Ust'Ishim, estimated by the excess of allele sharing with the Altai Neandertal compared to Mbuti and compared to Vindija. In (B) the analysis was restricted to sites where the Denisovan is homozygous ancestral. Error bars indicate 1 standard error.

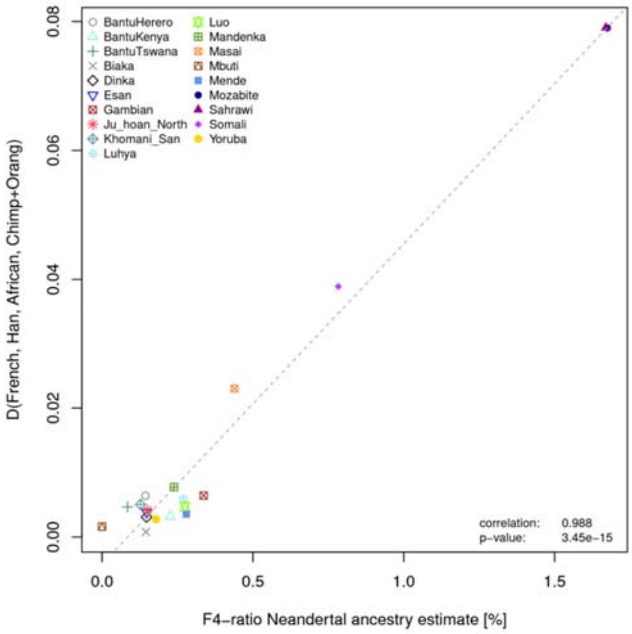

Fig. S33.
Correlation (Pearson's r) between the Neandertal ancestry in Africans estimated with F_4-ratios (Figure S32, top) and the European ancestry inferred by D(French, Han, African, Chimpanzee + Orangutan).

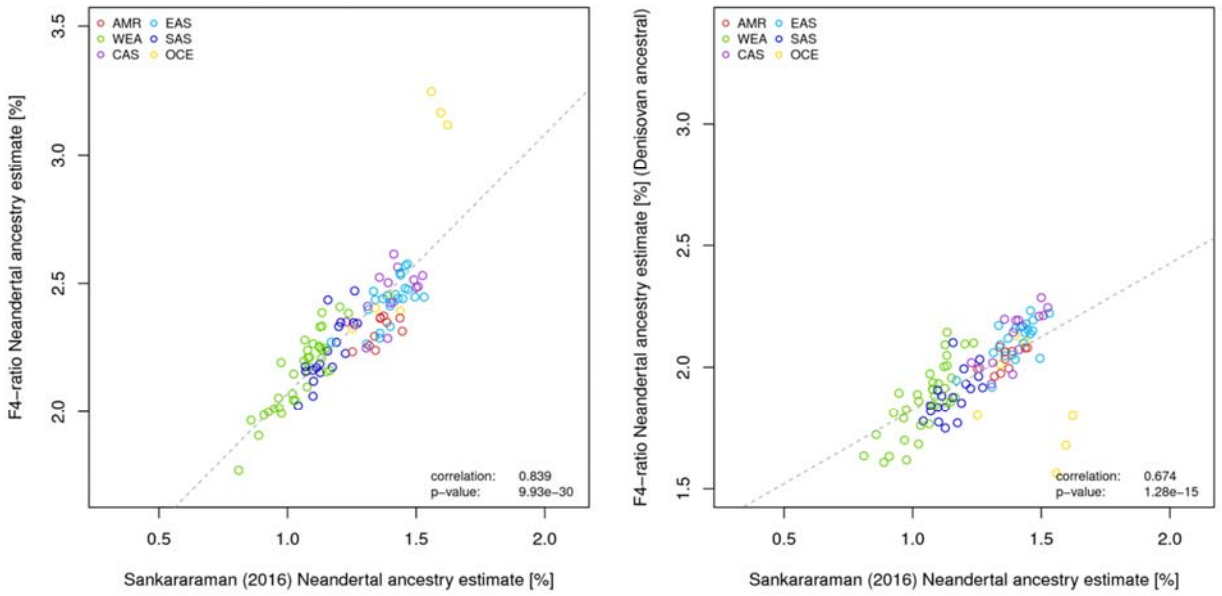

Fig. S34.
Correlation (Pearson's r) of the Neandertal ancestry estimate obtained with F_4-ratios with the Neandertal ancestry estimate for the SGDP populations obtained from Sankararaman et al. (2016) (*78*). A: F_4-ratios from Fig. S32A . B: F_4-ratios restricted to sites where the Denisovan is homozygous ancestral (Fig. S32B).

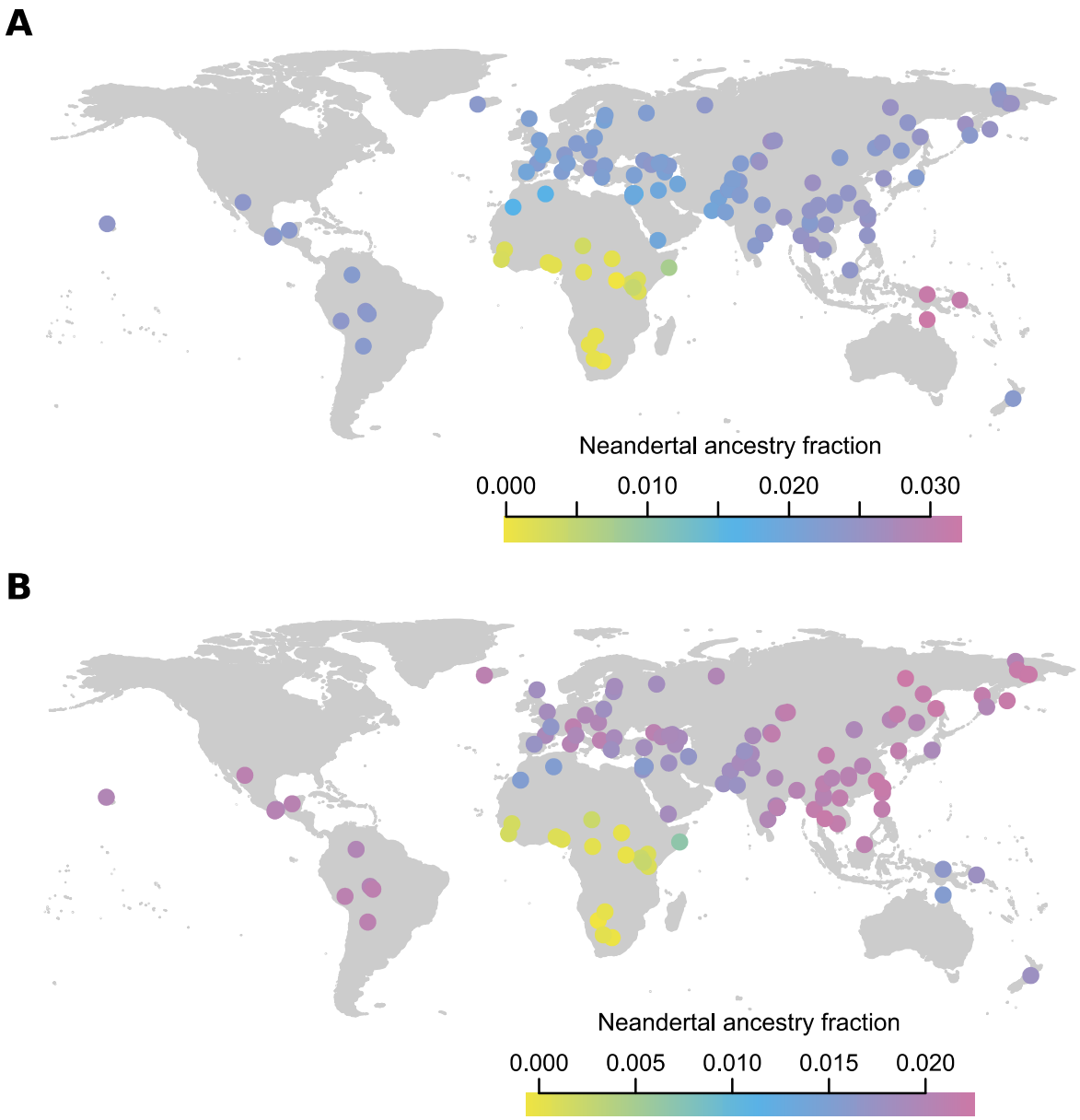

Fig. S35.
Neandertal ancestry estimates for SGDP populations estimated by F4-ratios from Fig. S8.6A (**A**) and S8.6B (**B**).

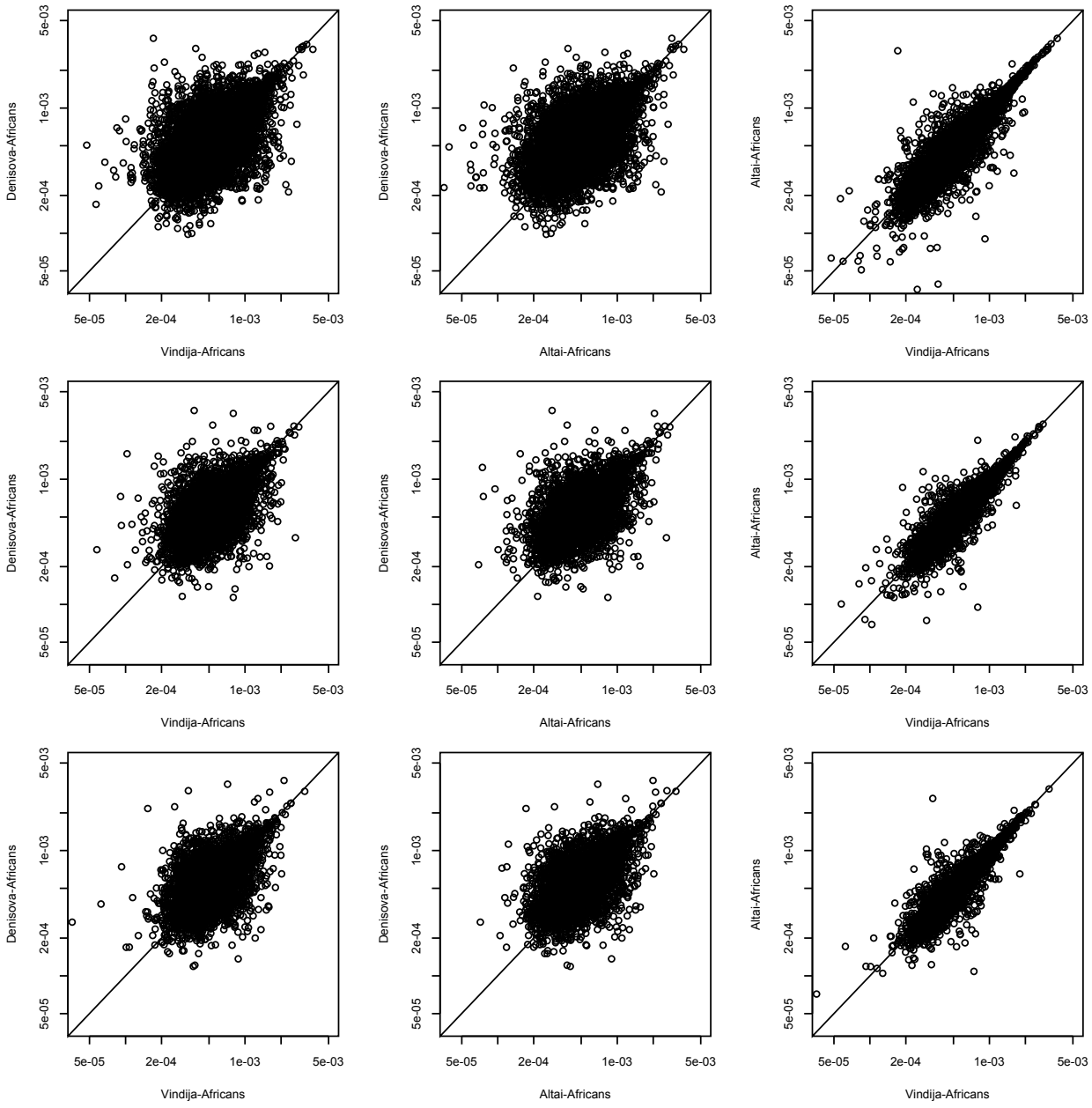

Fig. S36.
Pairwise comparison of African divergence between Vindija, Altai and Denisova. X and Y axes are in log-scale. Top row: 100kb window size; middle row: 0.1cM with the African American recombination map; bottom row: 0.1cM with the Decode recombination map.

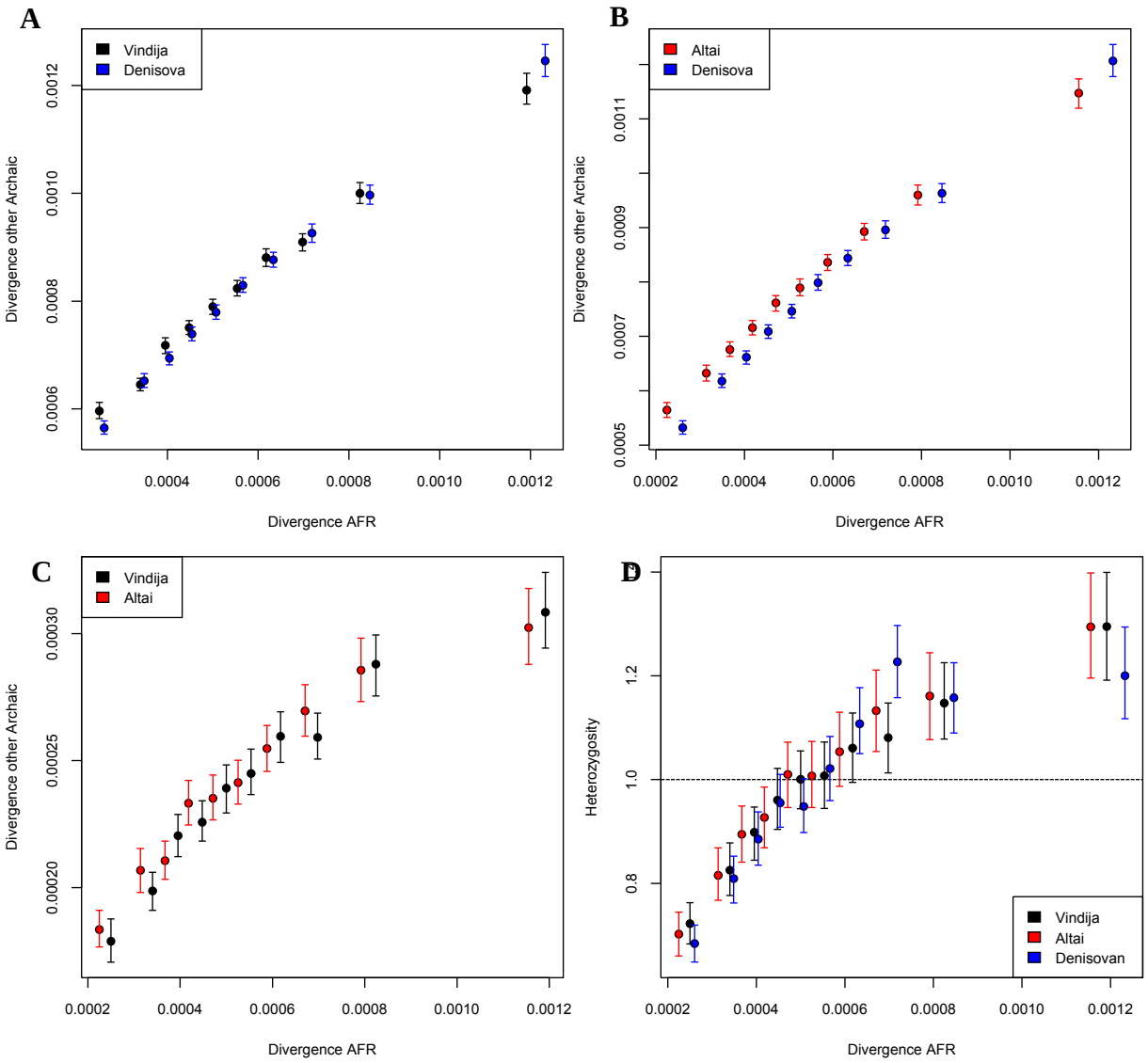

Fig. S37.

Pairwise divergence and heterozygosity over 100kb windows binned by African divergence. (Note that the y-scale is not matched between plots and that the axis does not start at 0.) **(A)** Vindija and Denisova ranked by divergence to Africans (x-axis) compared to Vindija-Denisovan divergence (y-axis). **(B)** Altai compared to Denisova (as in (A)). **(C)** Vindija compared to Altai (as in (A)). **(D)** Relative heterozygosity in Vindija, Altai and Denisova (y-axis) ranked by divergence to Africans. Dashed line represents the expected heterozygosity over all windows.

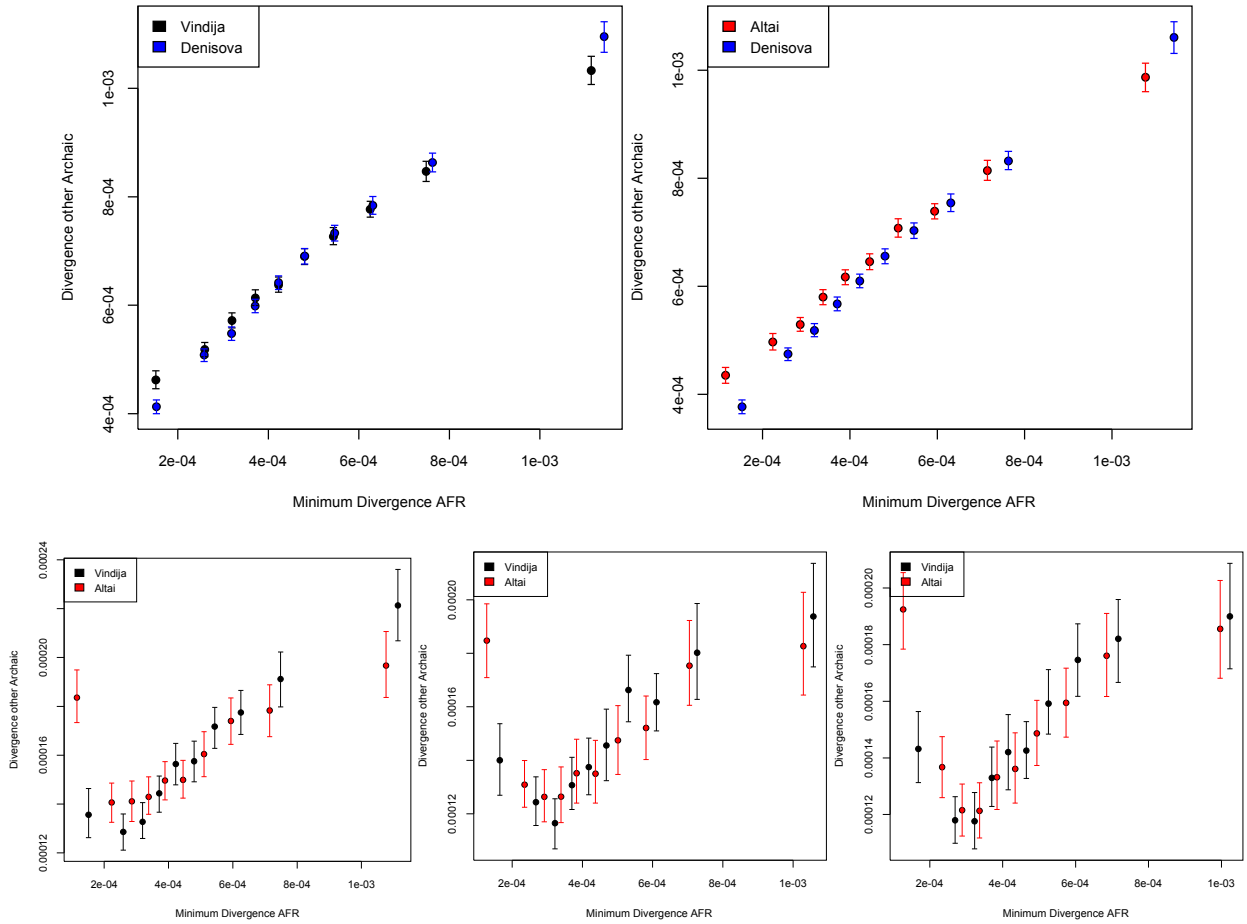

Fig. S38.
Pairwise divergence binned by the minimum divergence to Africans. Top: Neandertal-Denisova comparisons in 100kb windows; Bottom: Vindija-Altai comparison in 100kb window (left), 0.1cM windows AA recombination map (middle), 0.1cM windows Decode recombination map (right).

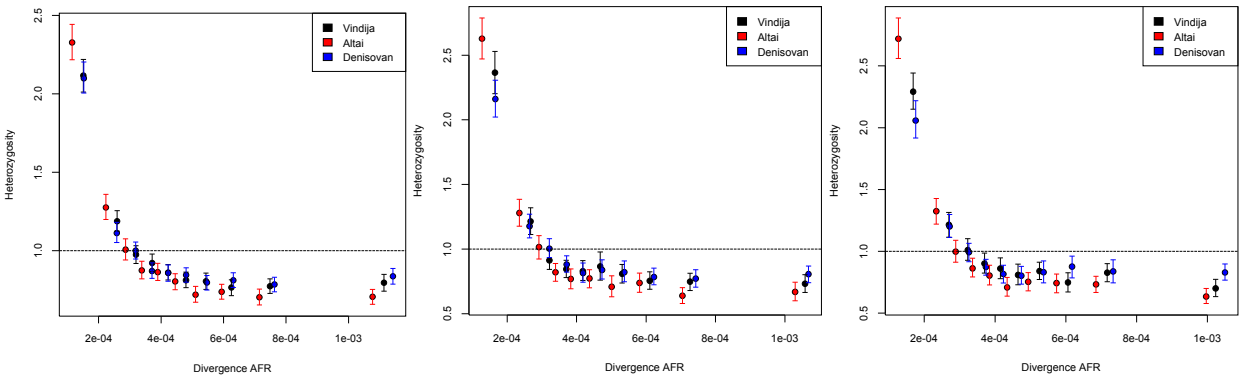

Fig. S39.
Heterozygosity binned by the minimum divergence to Africans. Left: 100kb window; middle: 0.1cM AA recombination map; right: 0.1cM Decode recombination map.

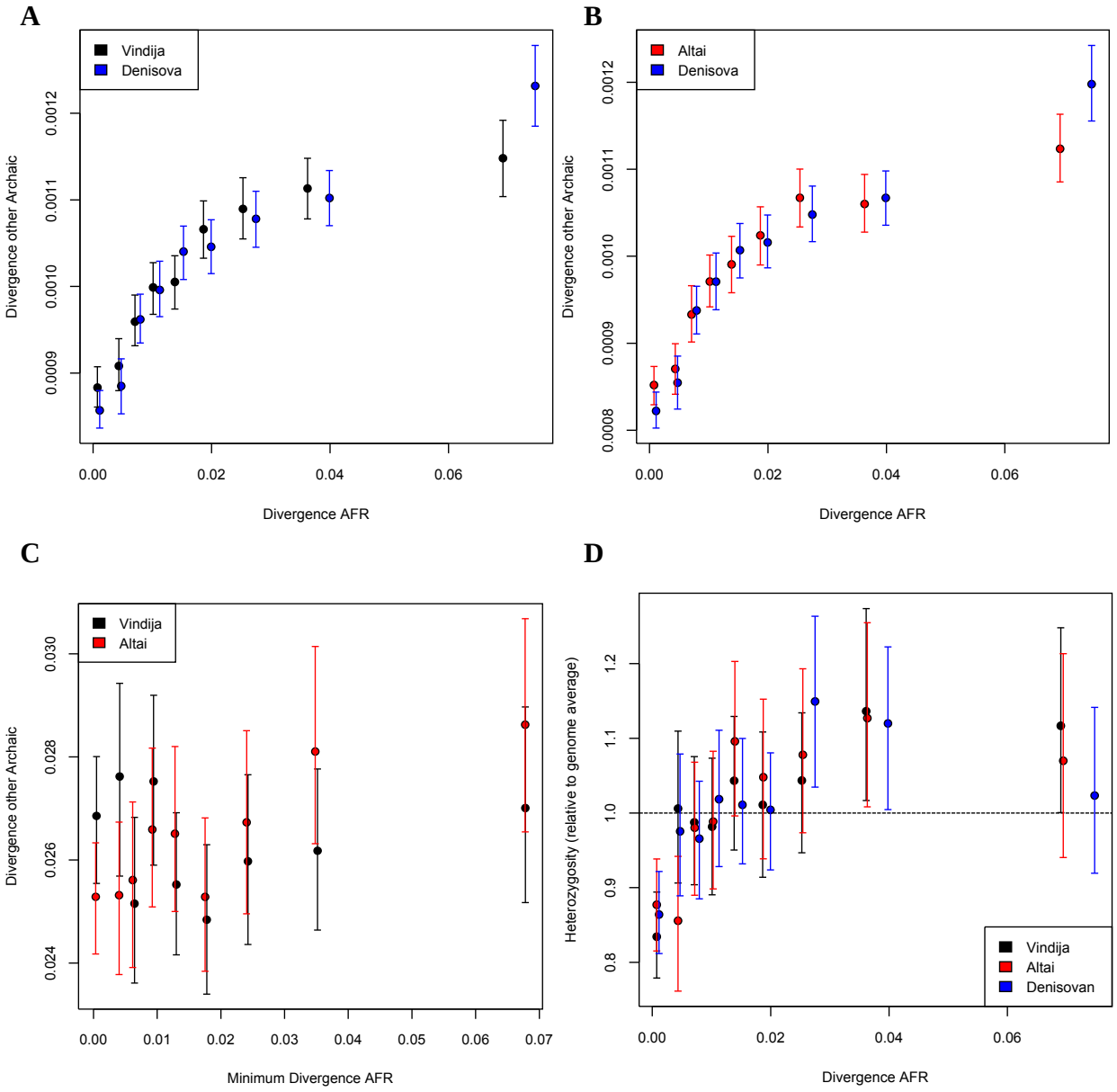

Fig. S40.

Pairwise divergence and heterozygosity over 100kb windows binned by relative African divergence. Y-axes **(A)** Vindija and Denisova ranked by divergence to Africans (x-axis) compared to Vindija-Denisovan divergence (y-axis). **(B)** Altai compared to Denisova (as in (A)). **(C)** Vindija compared to Altai (as in (A)). **(D)** Relative heterozygosity in Vindija, Altai and Denisova (y-axis) ranked by divergence to Africans. Dashed line represents the expected heterozygosity over all windows.

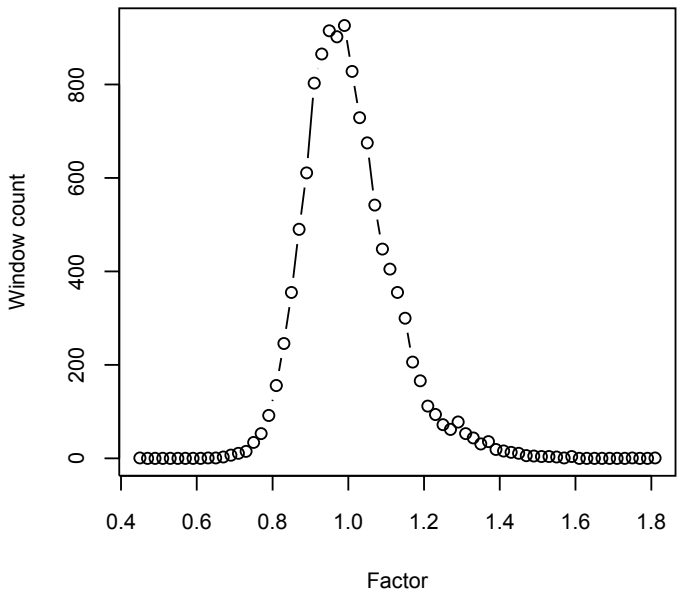

Fig. S41.

Histogram of mutation rate scaling factors.

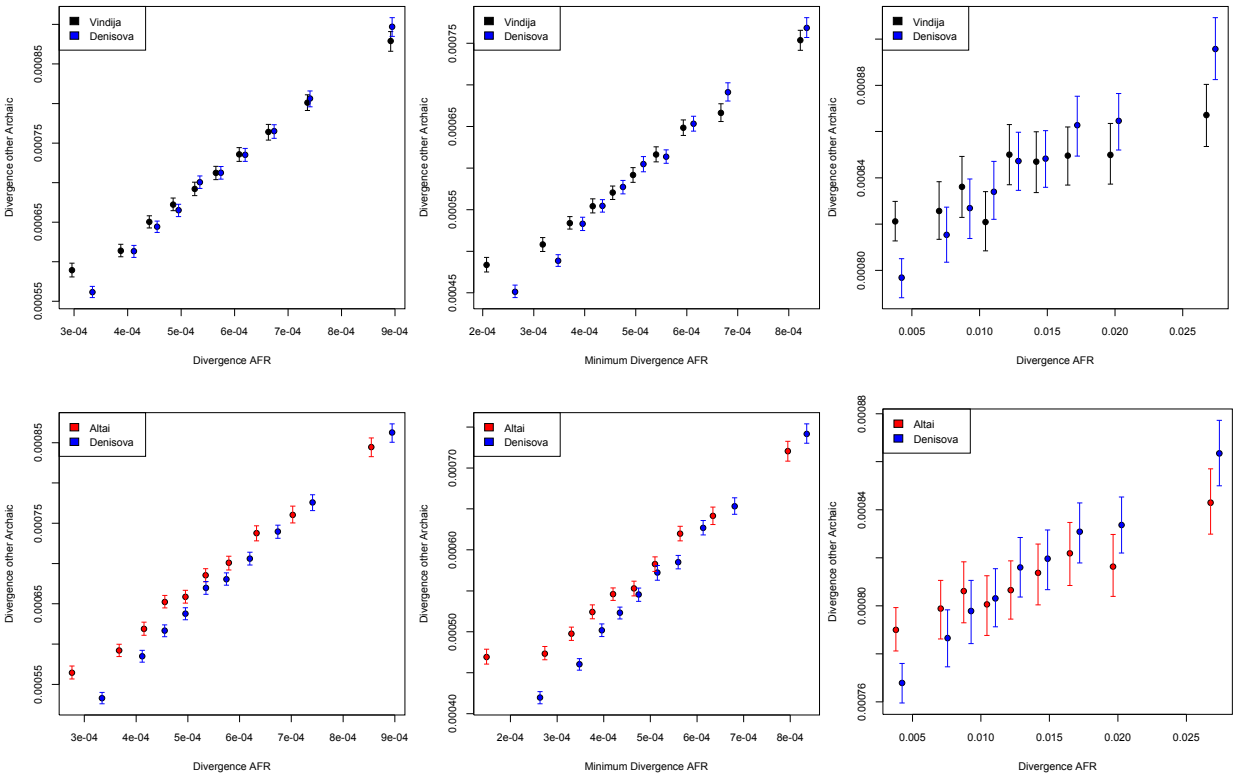

Fig. S42.
Simulations with 5% modern human admixture into Neandertals and measuring pairwise differences (left), pairwise differences choosing the closest allele to Africans (middle) and relative divergence to Africa (right).

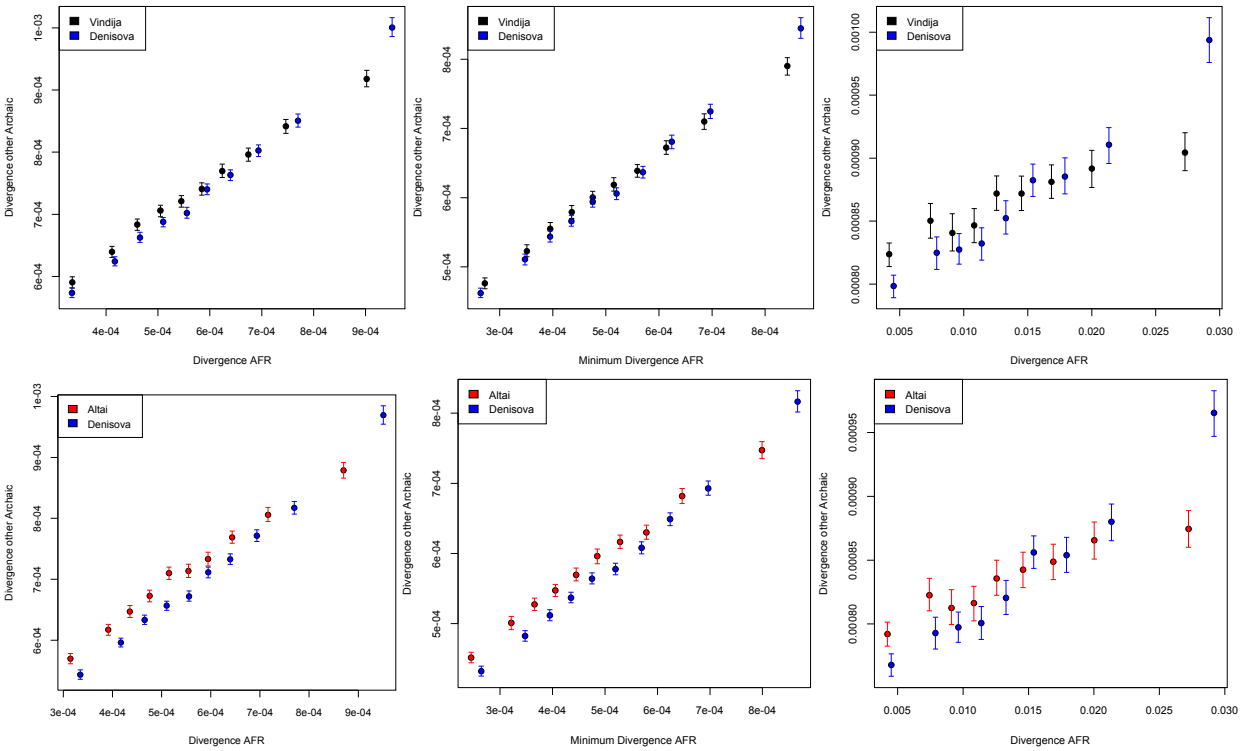

Fig. S43.
Simulations with 5% superarchaic admixture into Denisovans and measuring pairwise differences (left), pairwise differences choosing the closest allele to Africans (middle) and relative divergence to Africa (right).

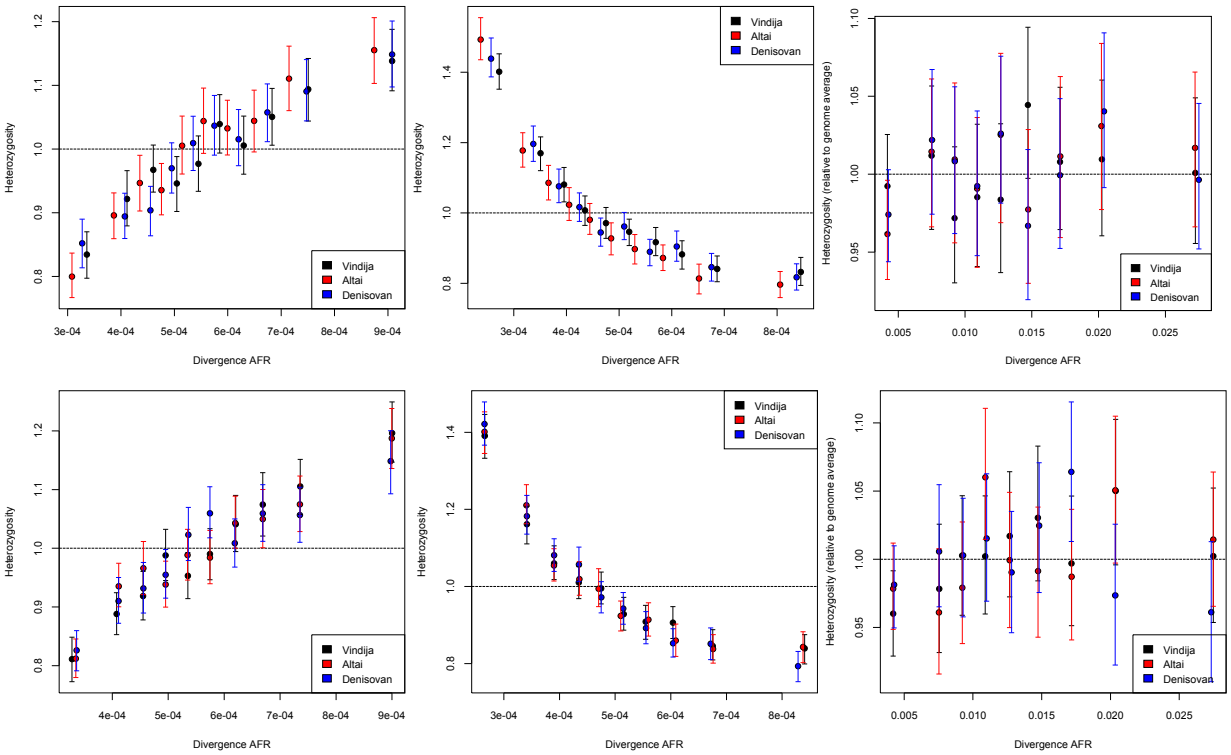

Fig. S44.
Simulations with branch-shortening differences between archaic humans (top row) and simulations with identical branch-shortening for all archaic humans (bottom row) for the measure of pairwise divergence (left), pairwise divergence choosing the closest allele to Africans (middle) or a relative measure of divergence (right).

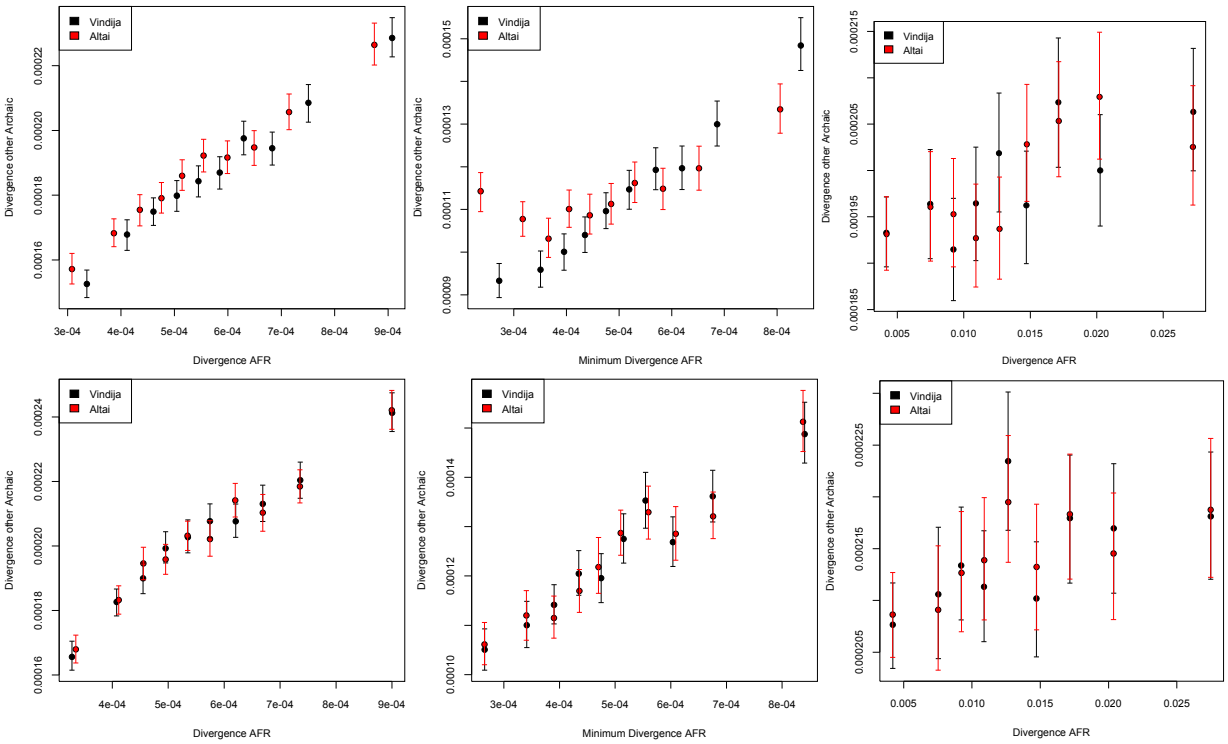

Fig. S45.
Simulations with and without branch-shortening differences between archaic humans. Figure order as in Fig S44.

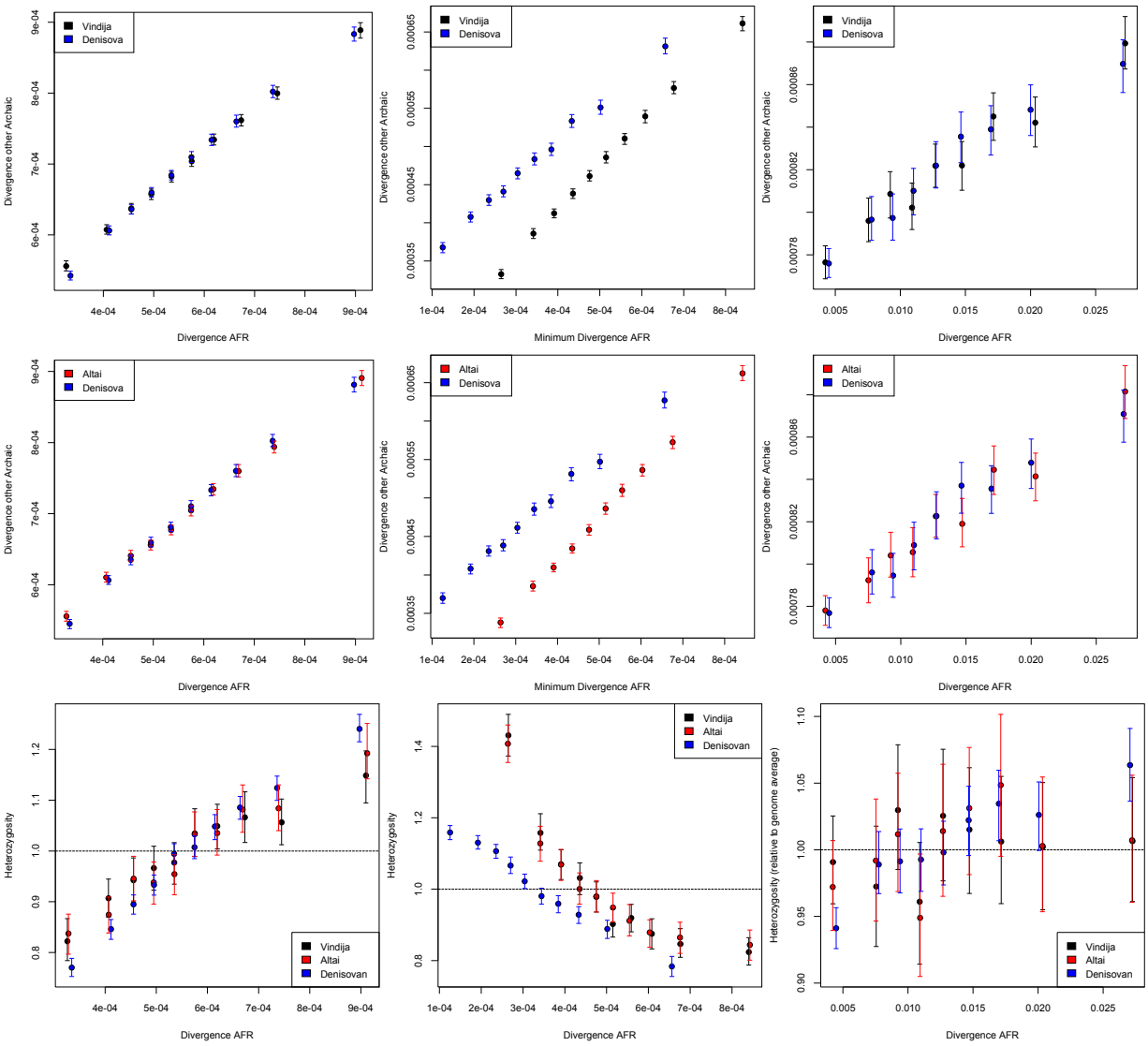

Fig. S46.

Simulations with effective population size difference between Neandertals and Denisovan. Left: pairwise divergence; middle: pairwise divergence using the closest allele at archaic heterozygous sites; right: relative divergence.

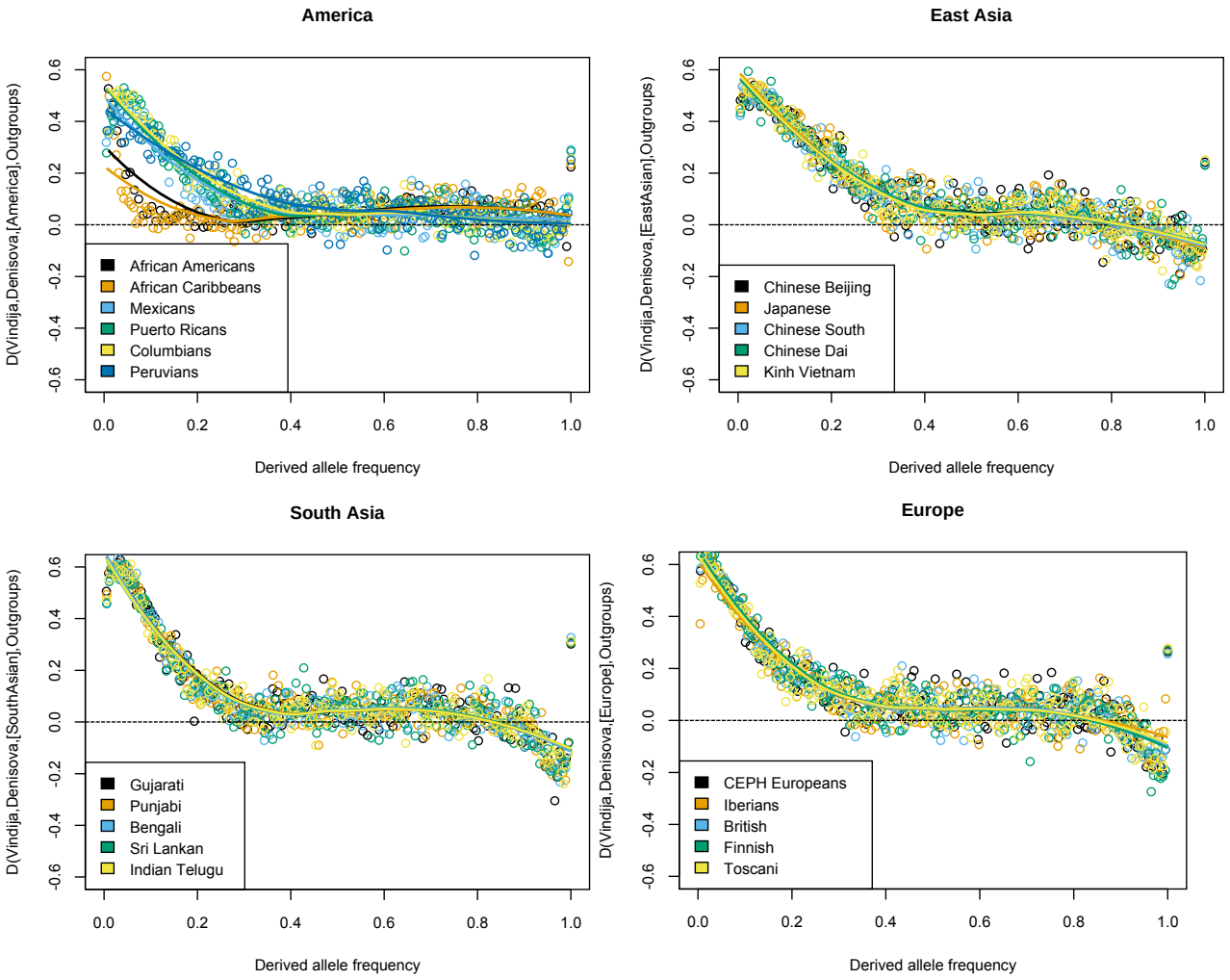

Fig. S47.

Allele frequency stratified D-statistics in 1000 Genomes non-African populations. Points indicate the D-statistics values for each frequency bin; lines show a smoothing (loess-curve) over these points.

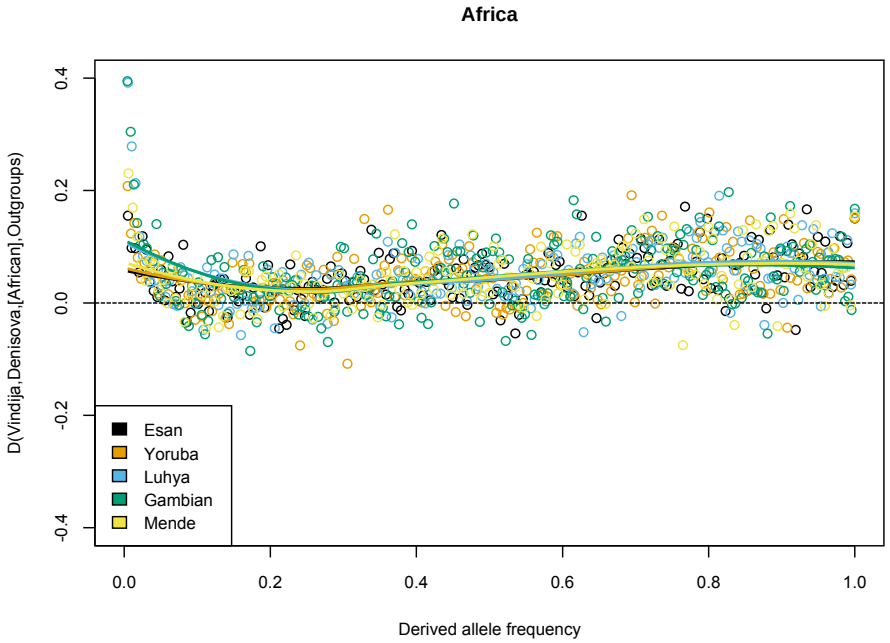

Fig. S48.

Allele frequency stratified D-statistics in 1000 Genomes African populations. Points indicate the D-statistics values for each frequency bin; lines show a smoothing (loess-curve) over these points.

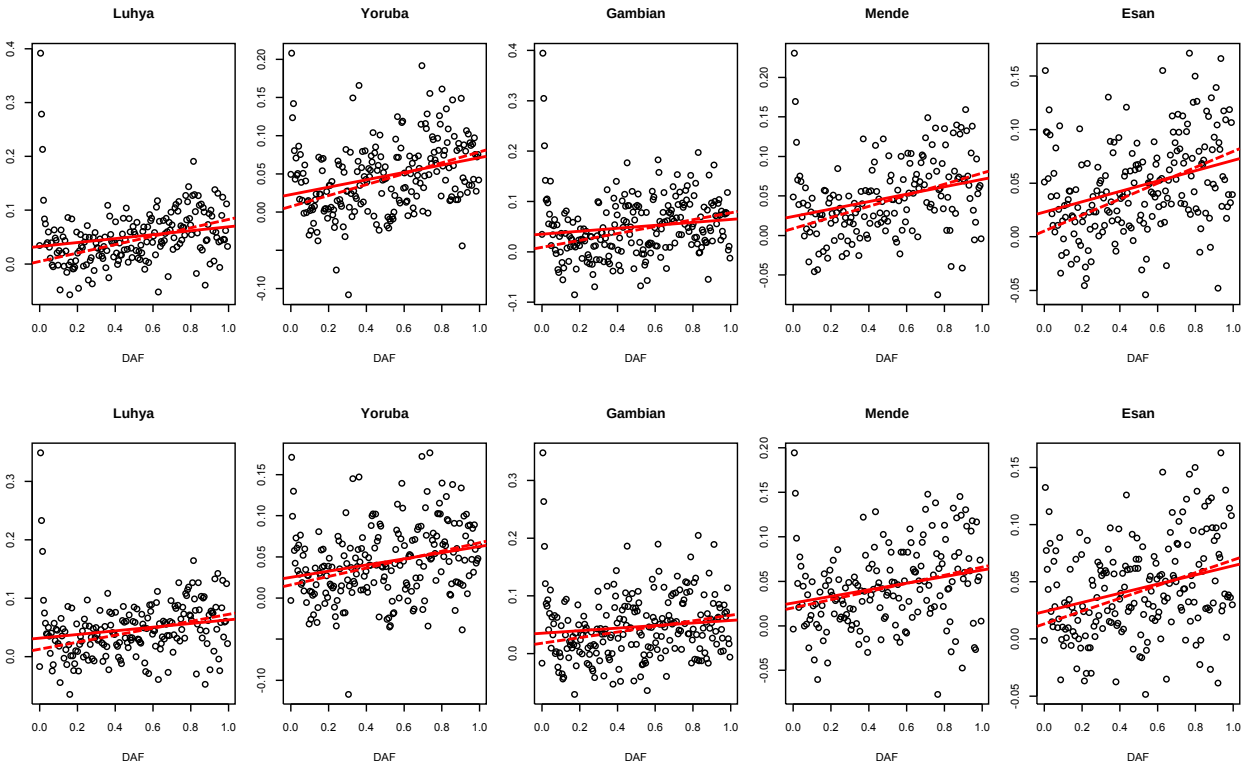

Fig. S49.

Linear regression for stratified D(Vindija,Denisova,[African population],Outgroup) and allele frequency (top row) and D(Altai, Denisova, [African population],Outgroup) and allele frequency (bottom row). Solid lines show the line fit to all frequencies; dashed lines show the fit for frequencies > 0.2.

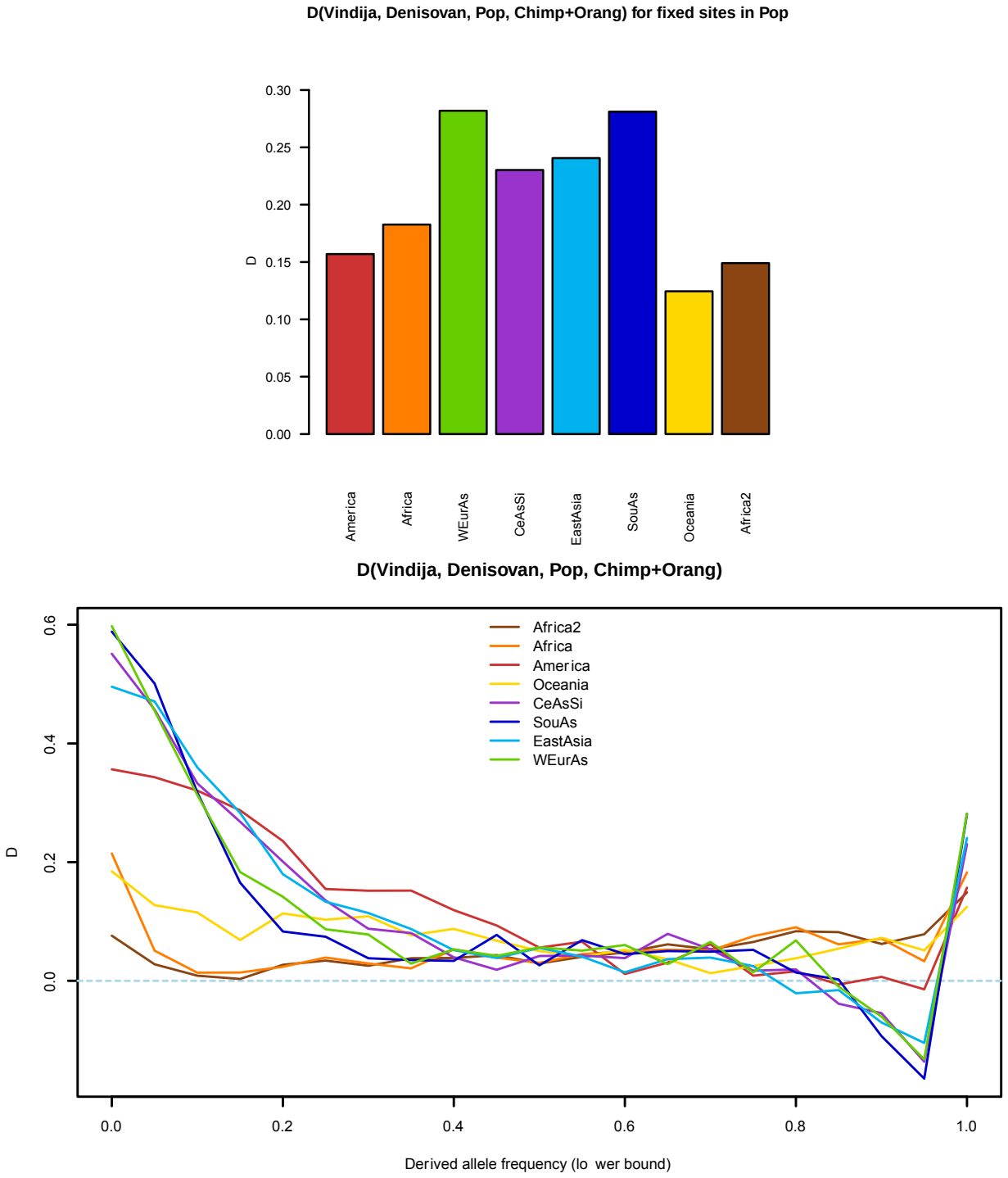

Fig. S50.
D(Vindija,Denisova,[SGDP population],Outgroup) for fixed sites (top) and by allele-frequency (bottom). The population "Africa2" refers to individuals from all African populations except for Sahrawi, Mozabite, Somali and Masai, which showed evidence for substantial non-African ancestry (See S8). Lines in the lower plot are averages over 5% frequency bins.

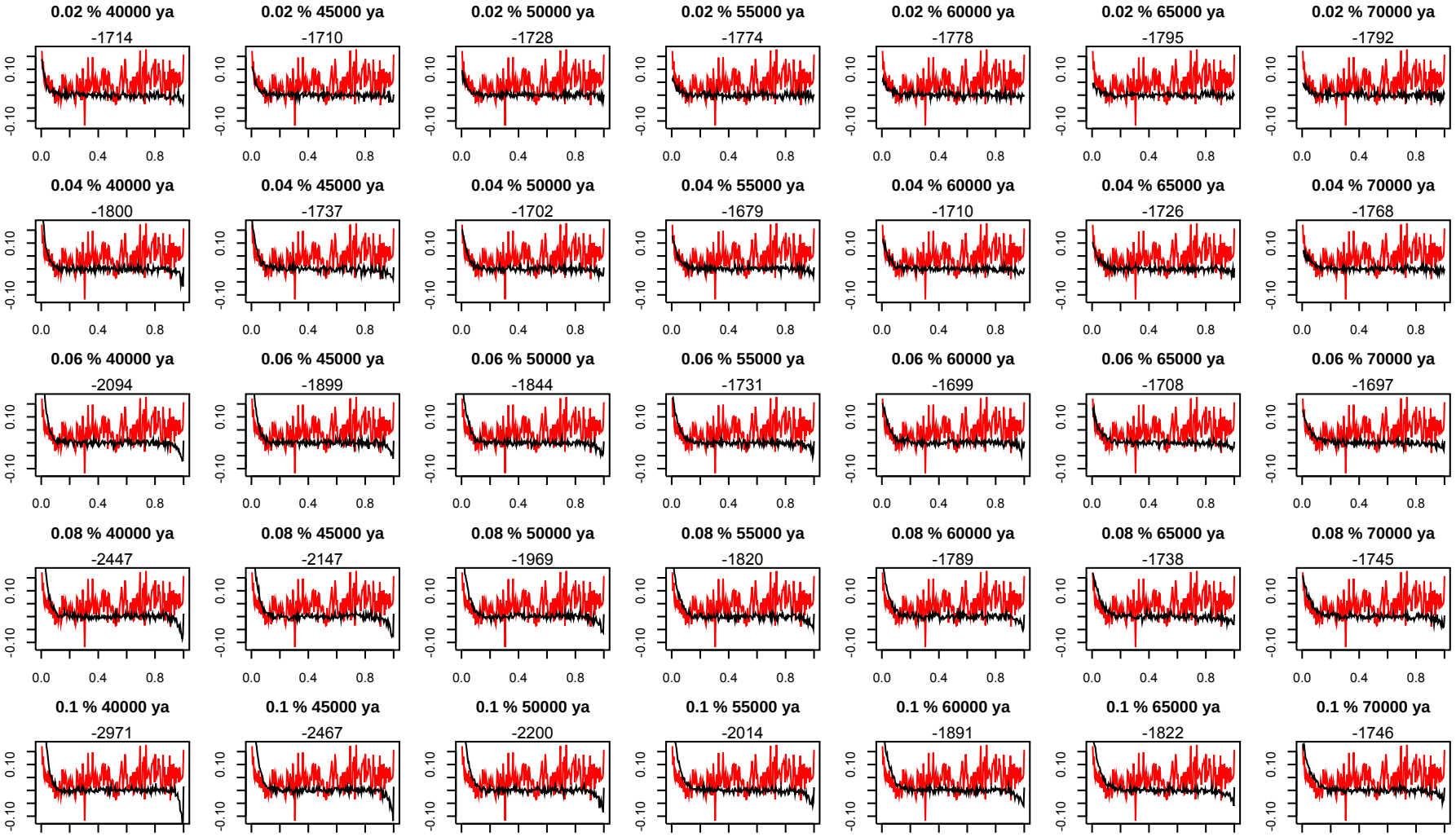

Fig. S51.

Simulated (black line) and observed (red line) allele-frequency stratified D(Altai, Denisova, Yorubans, Outgroup) for a scenario of admixture from Neandertals into Yorubans at different times and magnitudes (noted in the title of each subplot). The measure of fit (LL) is given above each plot.

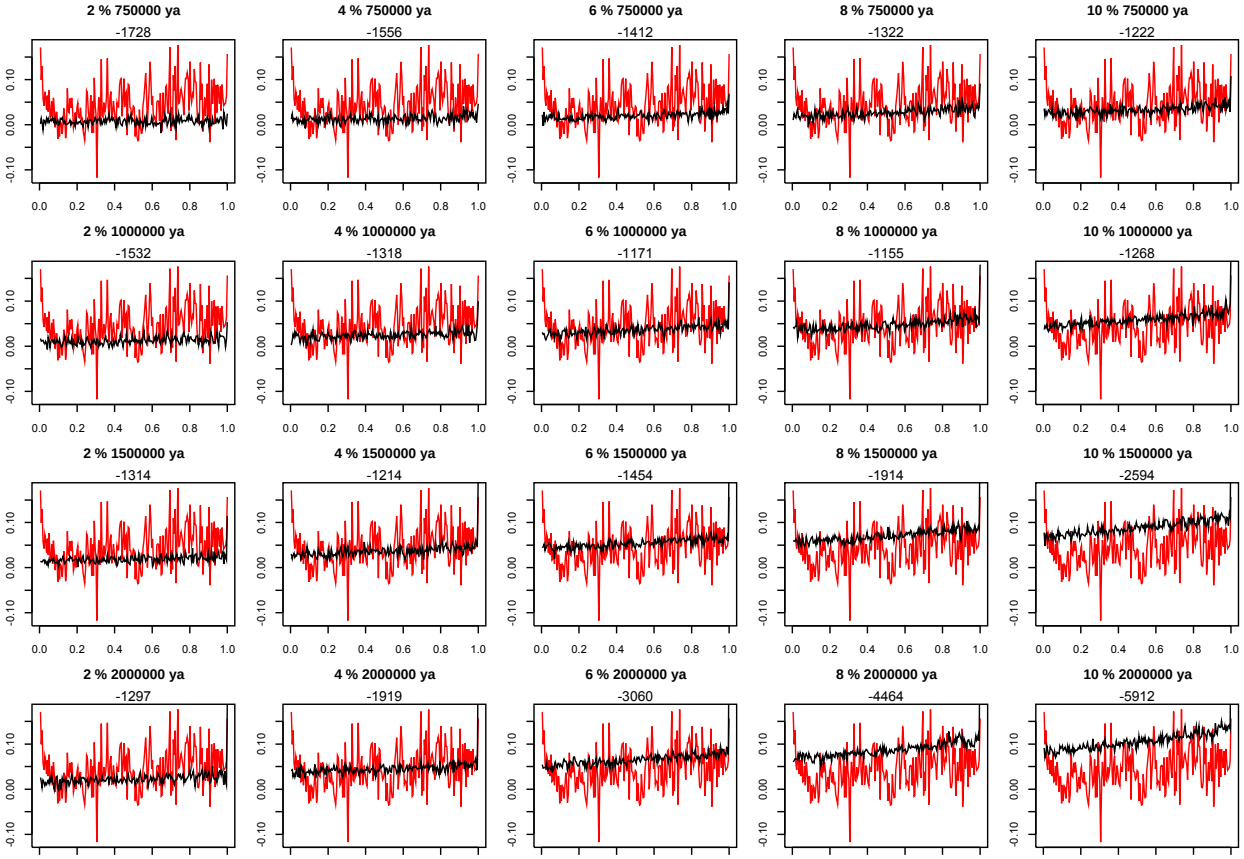

Fig. S52.
Simulations for superarchaics admixture into Denisovans. Labelling as in Fig S51.

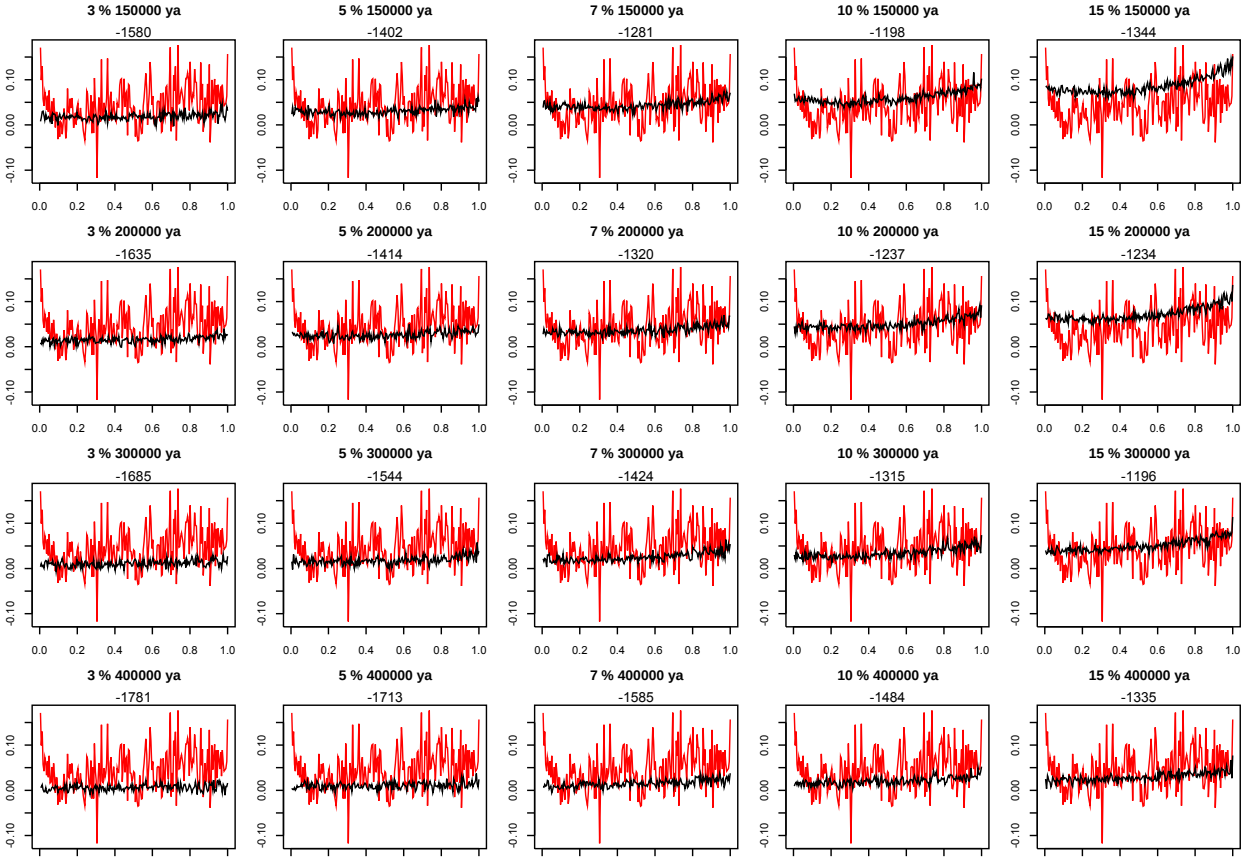

Fig. S53.
Simulations for modern human admixture into Neandertals. Labelling as in Fig S51.

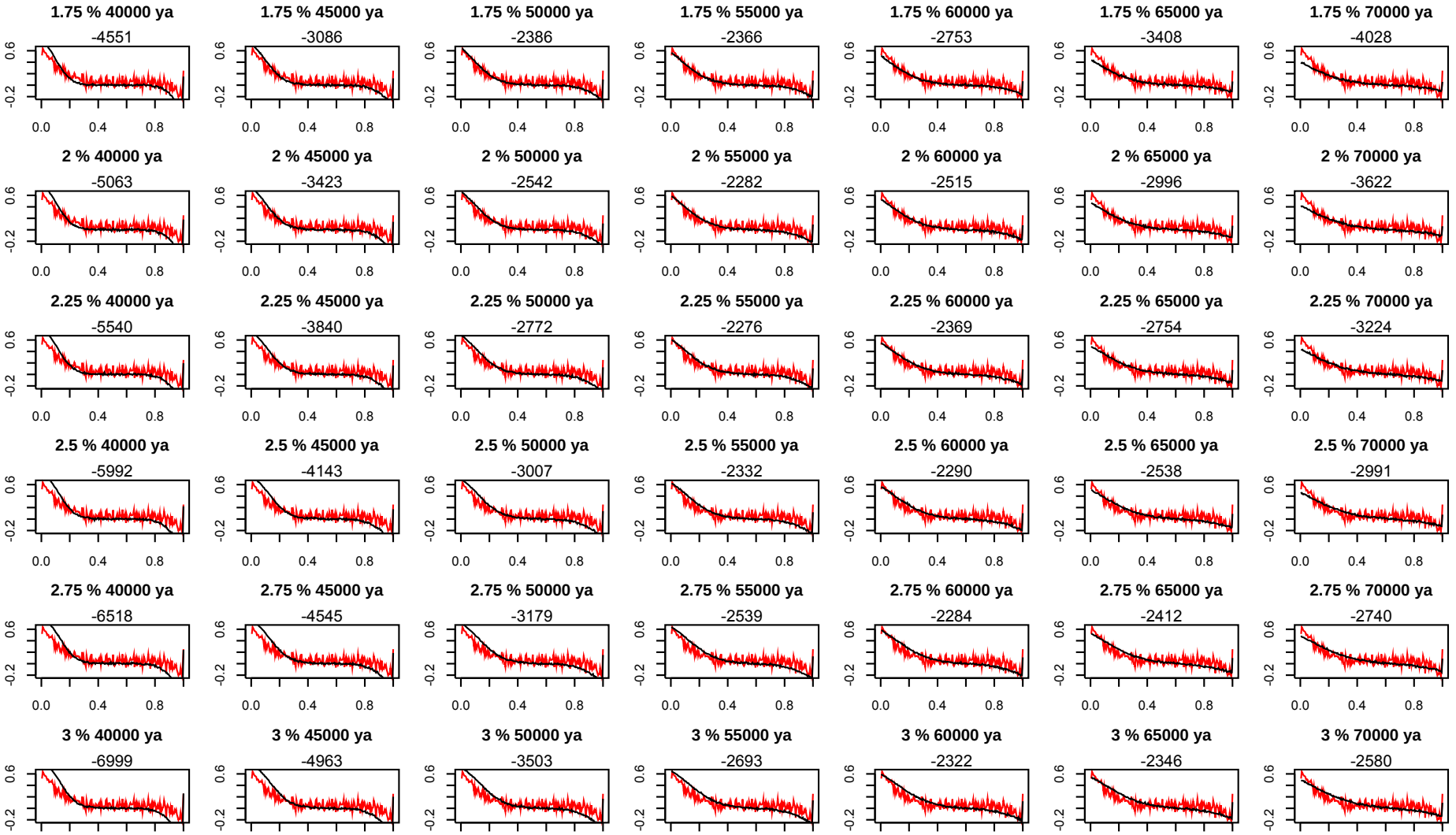

Fig. S54.

Simulated (black line) and observed (red line) allele-frequency stratified D(Altai, Denisova, CEPH European, Outgroup) for a scenario of admixture from Neandertals into Yorubans at different times and magnitudes (noted in the title of each subplot). The measure of fit (LL) is given above each plot.

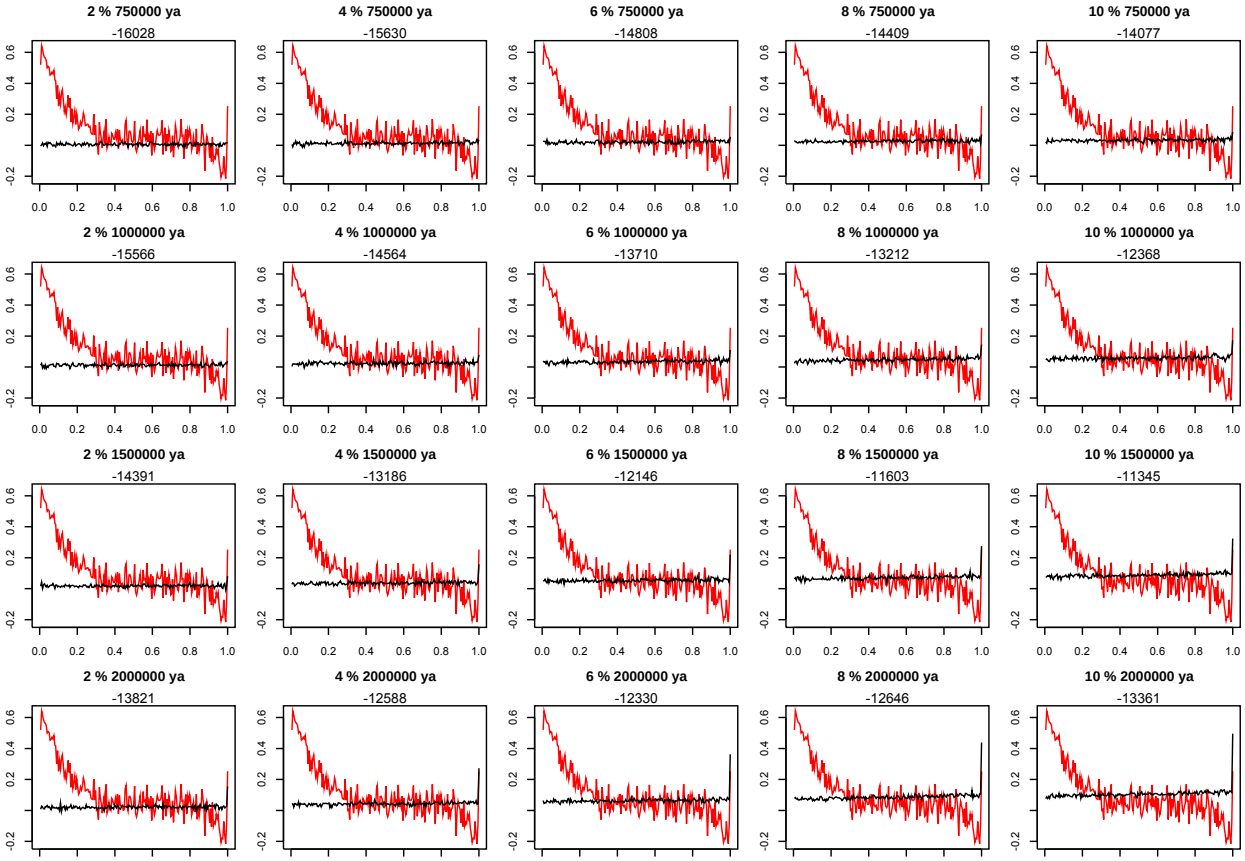

Fig. S55.
Simulations for superarchaic admixture into Denisovans. Labelling as in Fig S54.

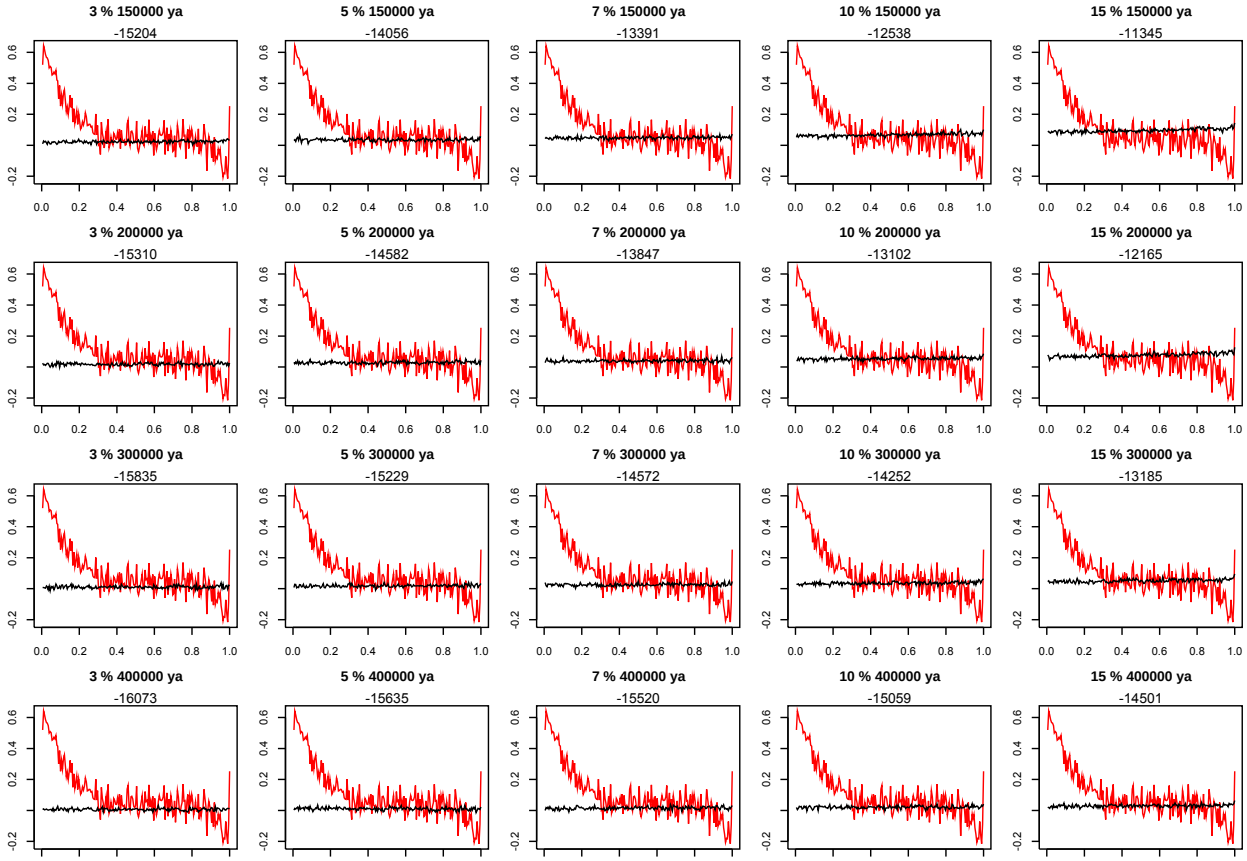

Fig. S56.
Simulations for modern human admixture into Neandertals. Labelling as in Fig S54.

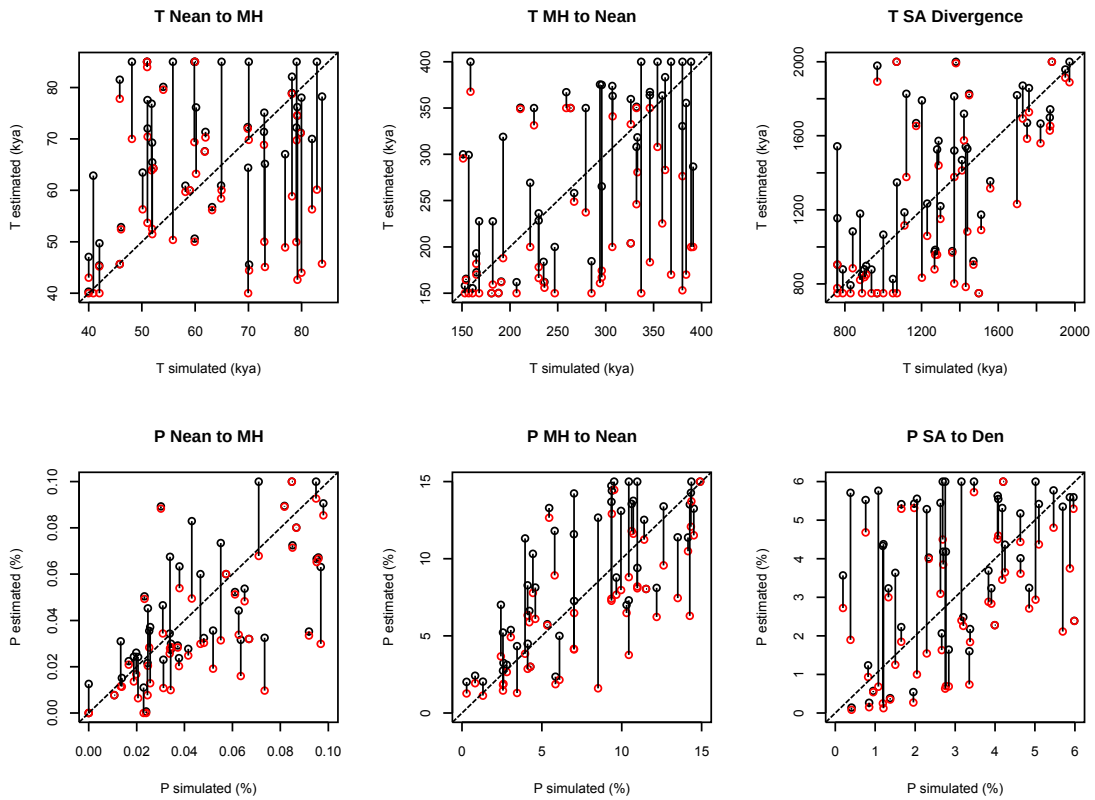

Fig. S57.

Simulated and estimated parameters for simulations with 50 randomly chosen combinations of admixtures for Yorubans. Red and black points give lowest and highest estimate for a given simulation when considering the range of estimated parameters with $LL \leq 3$.

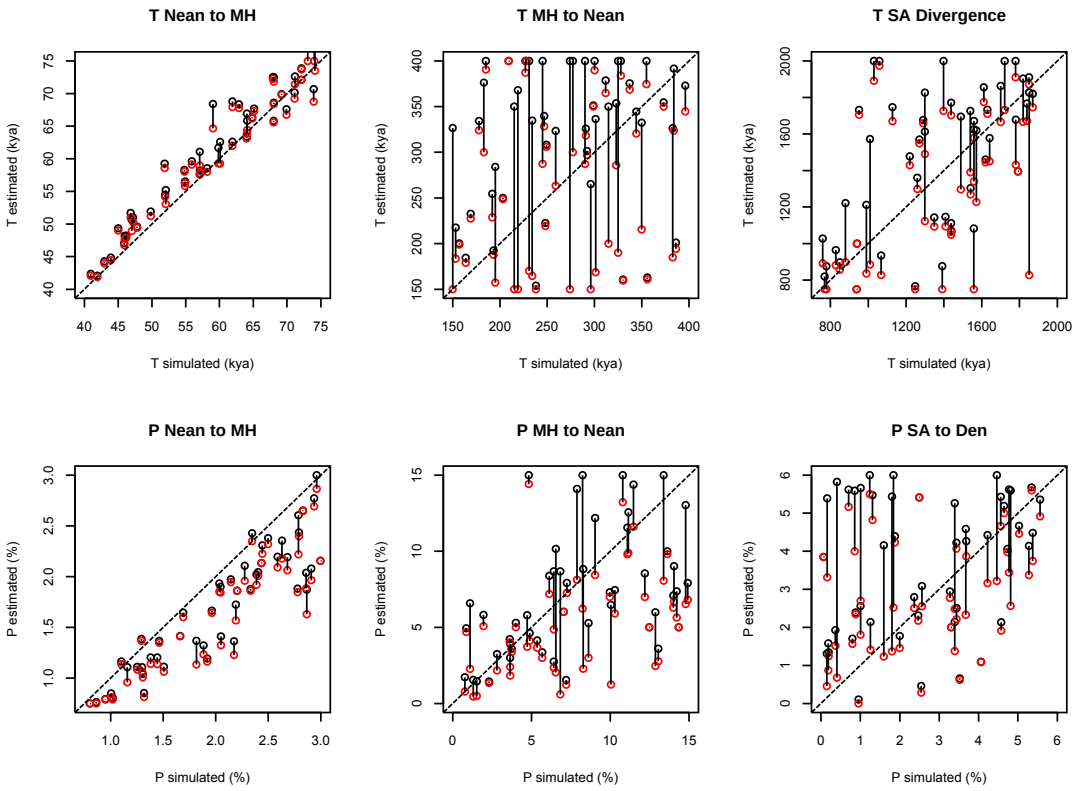

Fig. S58.
Simulated and estimated parameters for simulations with 50 randomly chosen combinations of admixtures for CEPH Europeans. Labelling as in S9b.10.

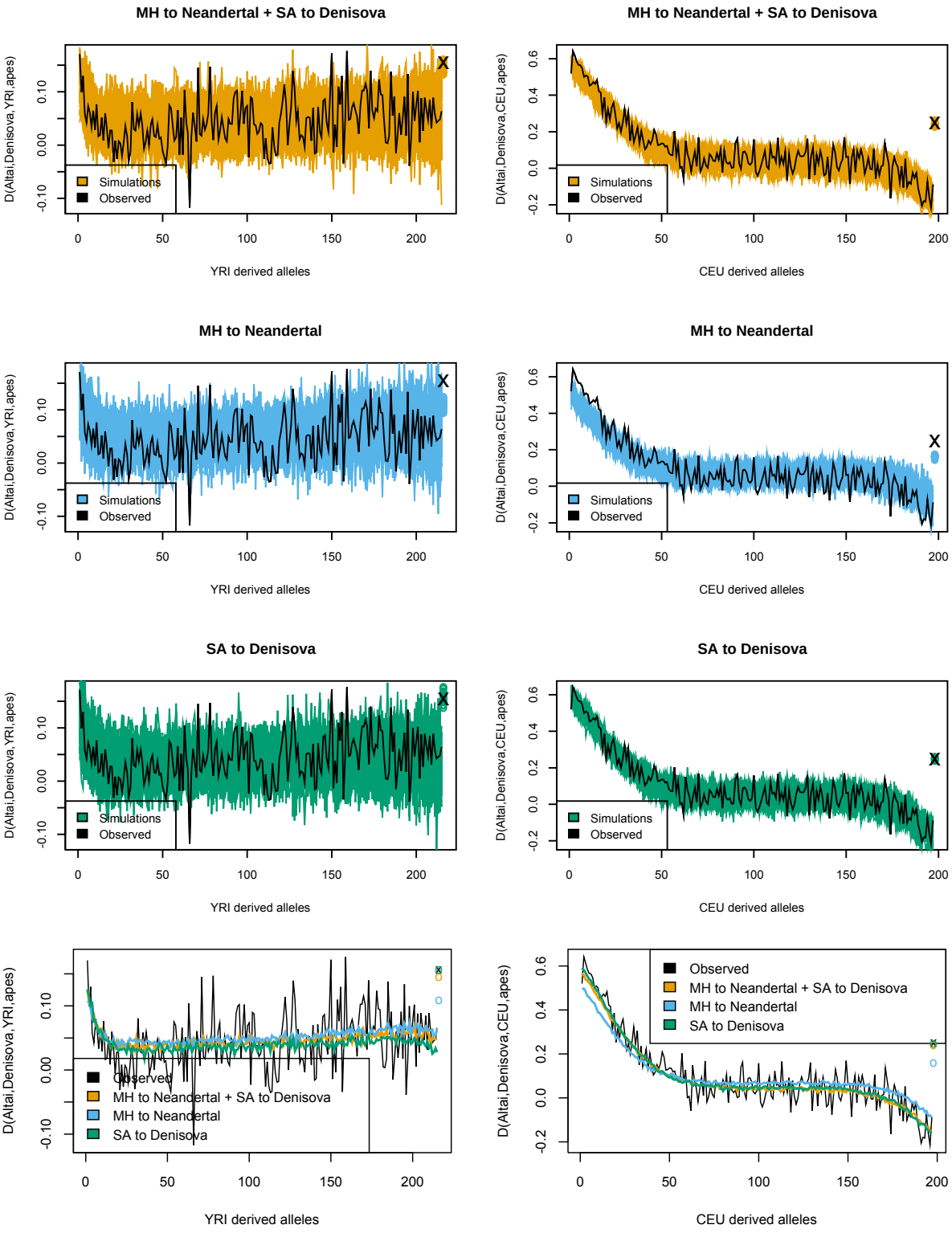

Fig. S59.
Three top rows: Simulations with the best estimated admixture scenarios for YRI (left) and CEU (right) as shown in Table S28. Colored bands show lines for 50 individual simulations of 1.2Gb of simulated sequence, each. **Bottom row:** Average over the individual simulations in the top 3 rows.

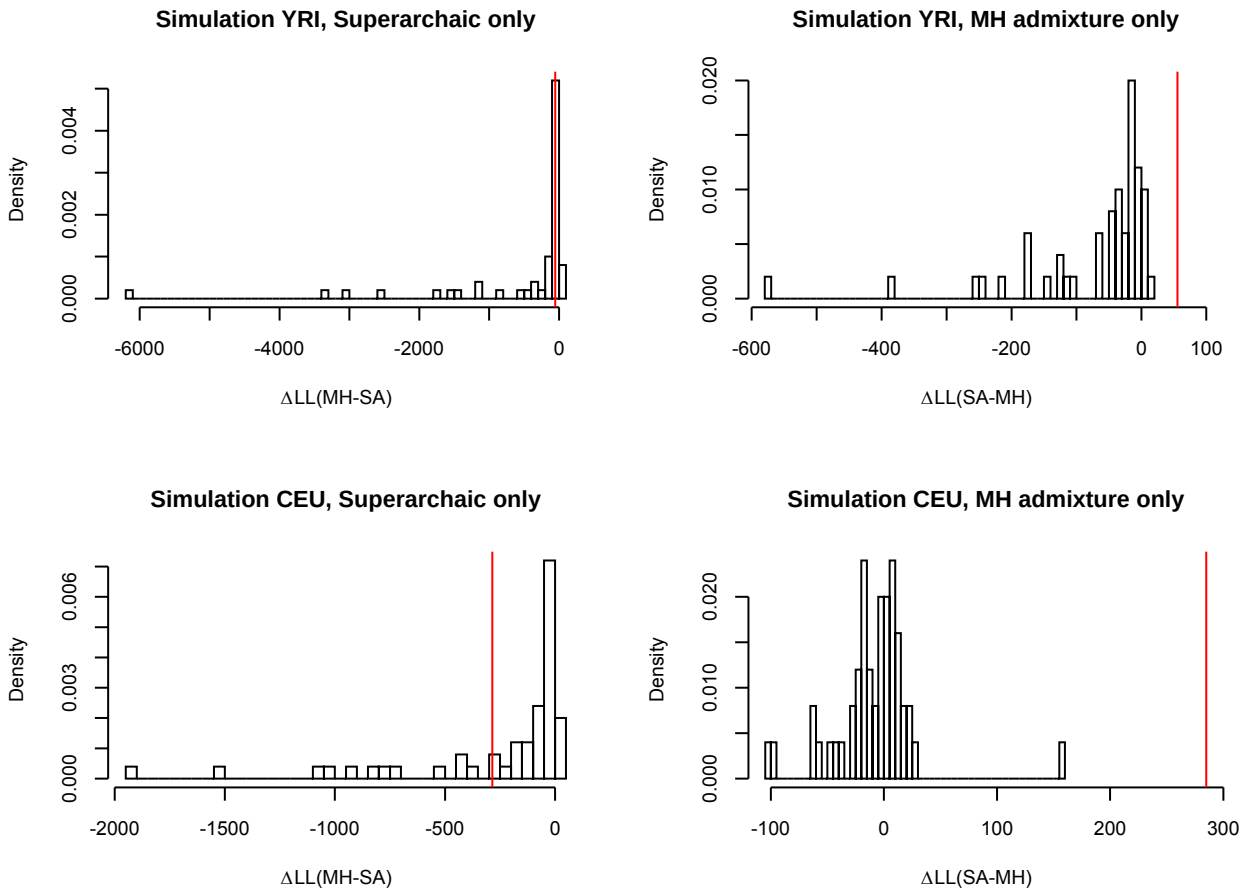

Fig. S60.
Simulations with only superarchaic admixture into Denisova or only modern human admixture into Neandertals for YRI and CEU. X-scales are oriented so that a negative value indicates a better fit of the model that matches the simulation. Red line indicates the difference in *LL* observed in the real data (see Table S28).

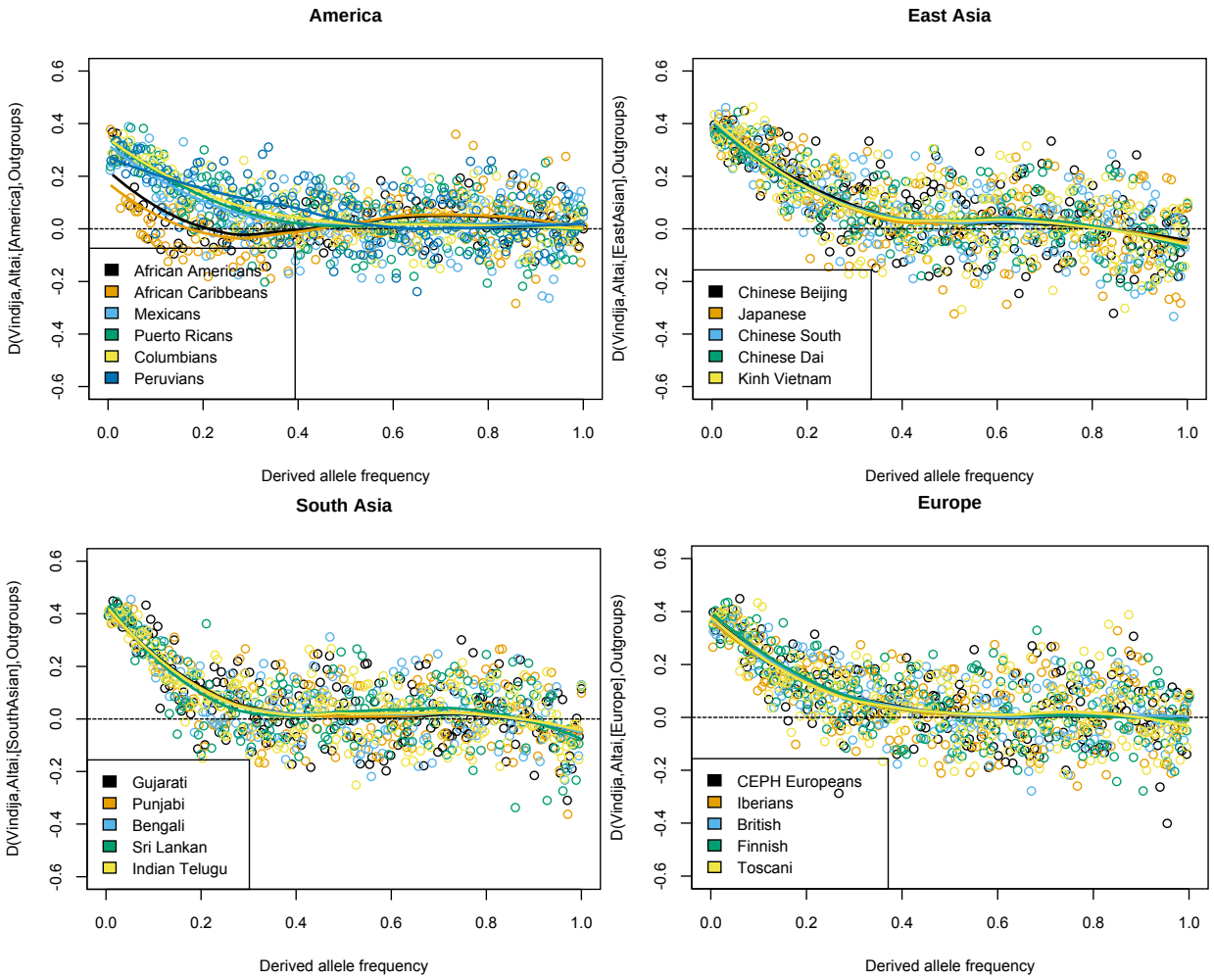

Fig. S61.
Allele frequency stratified D(Vindija,Altai,[population],Outgroup) in different non-African 1000 Genomes populations.

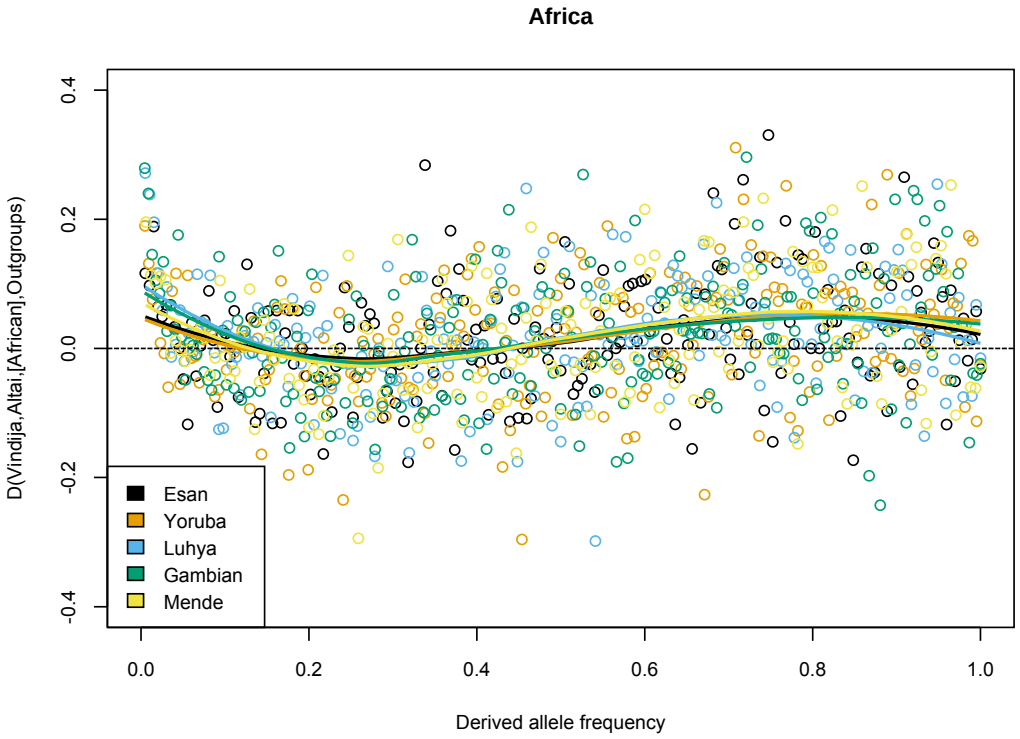

Fig. S62.
Allele frequency stratified D(Vindija,Altai,[population],Outgroup) in African populations.

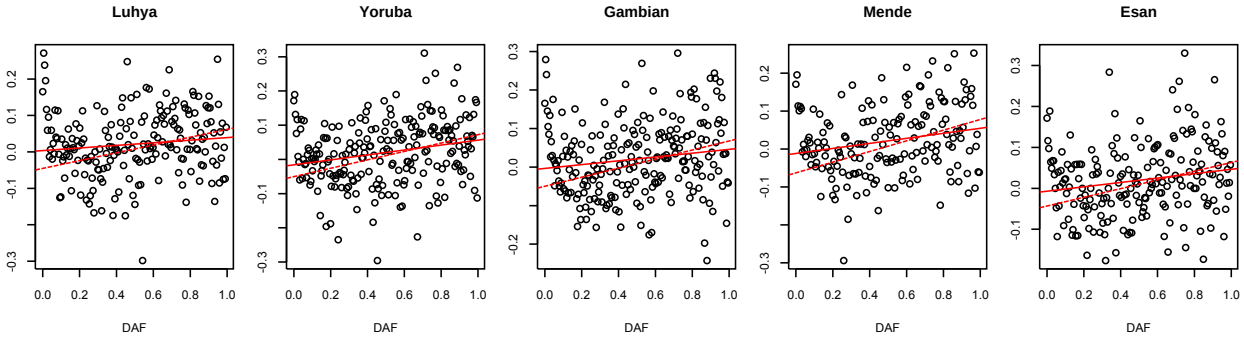

Fig. S63.
Linear regression for stratified D(Vindija,Altai,[African population],Outgroup)

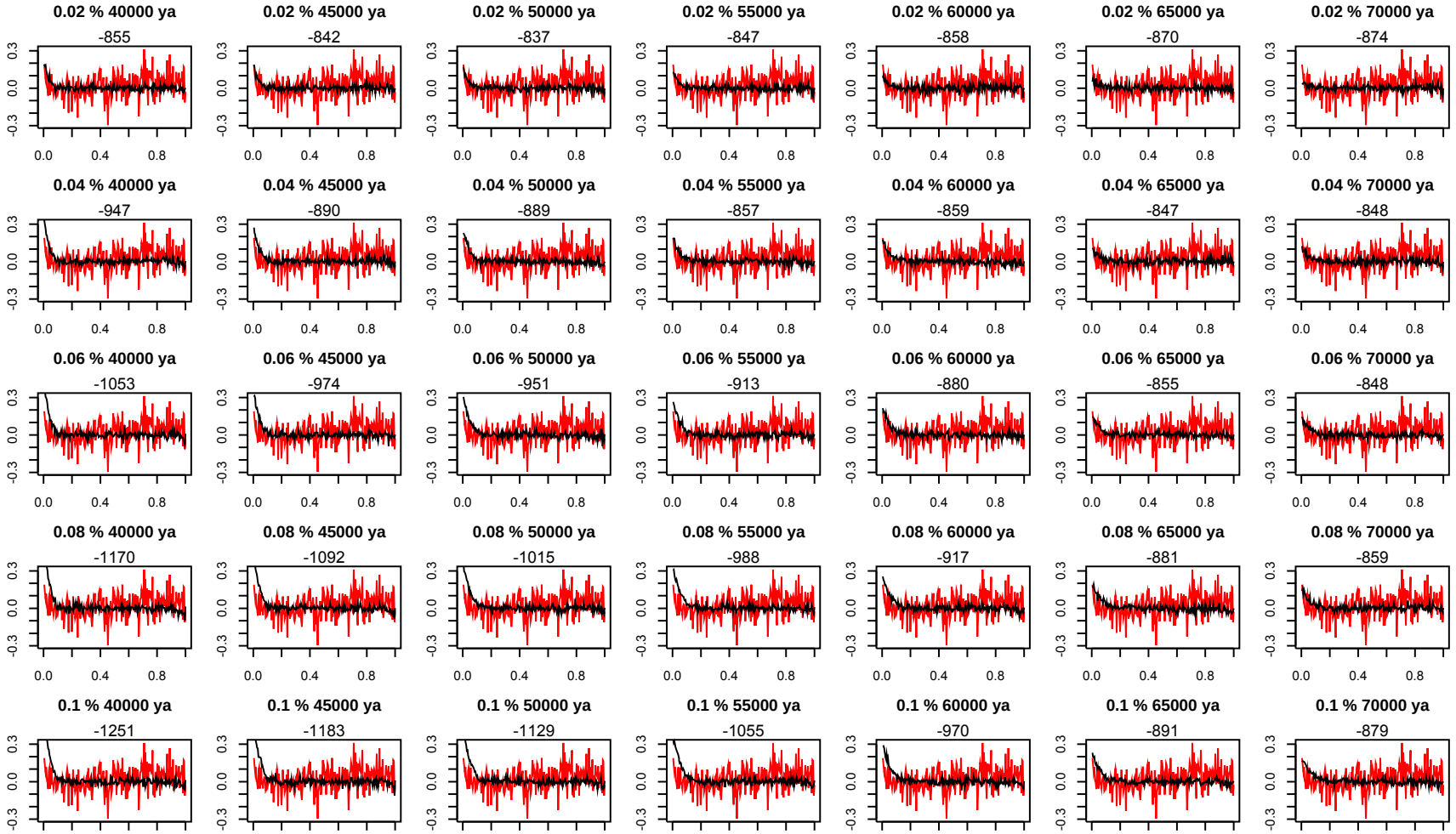

Fig. S64.
Simulated (black line) and observed (red line) allele-frequency stratified D(Altai, Denisova, YRI, Outgroup) for a scenario of admixture from the Vindija Neandertal population into modern humans at different times and magnitudes (noted in the title of each subplot). The measure of fit (LL) is given above each plot.

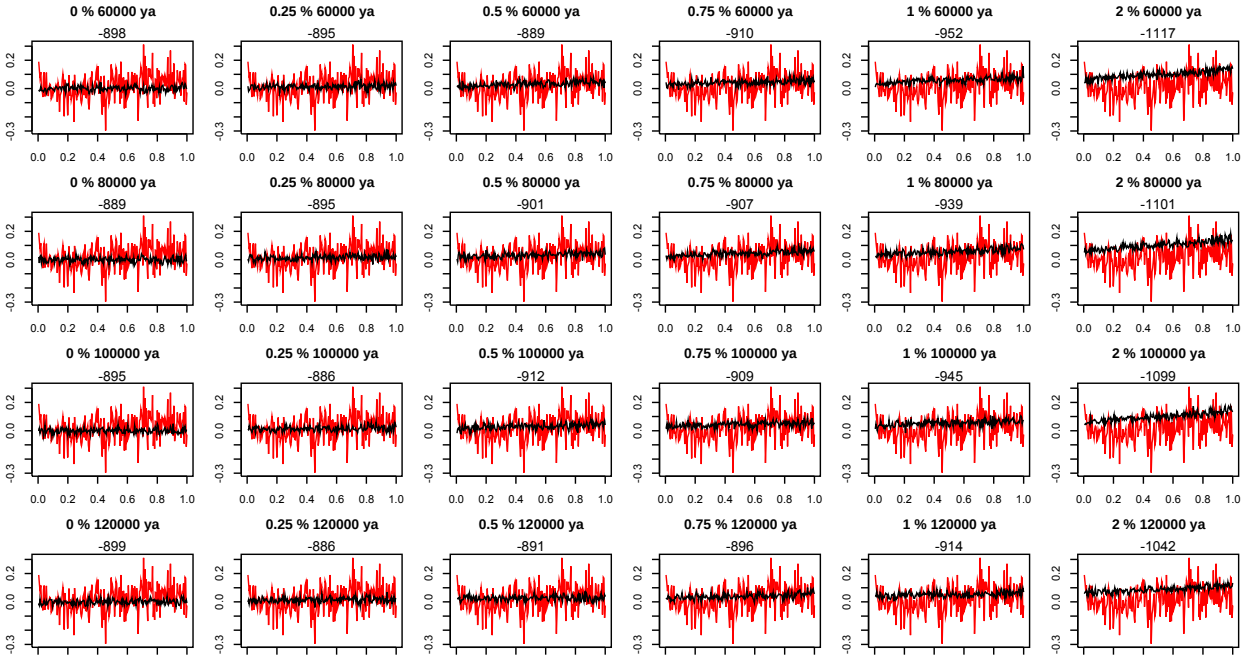

Fig. S65.
Simulated admixture from modern humans specifically into the Vindija Neandertal population. See also Fig. S9b.16.

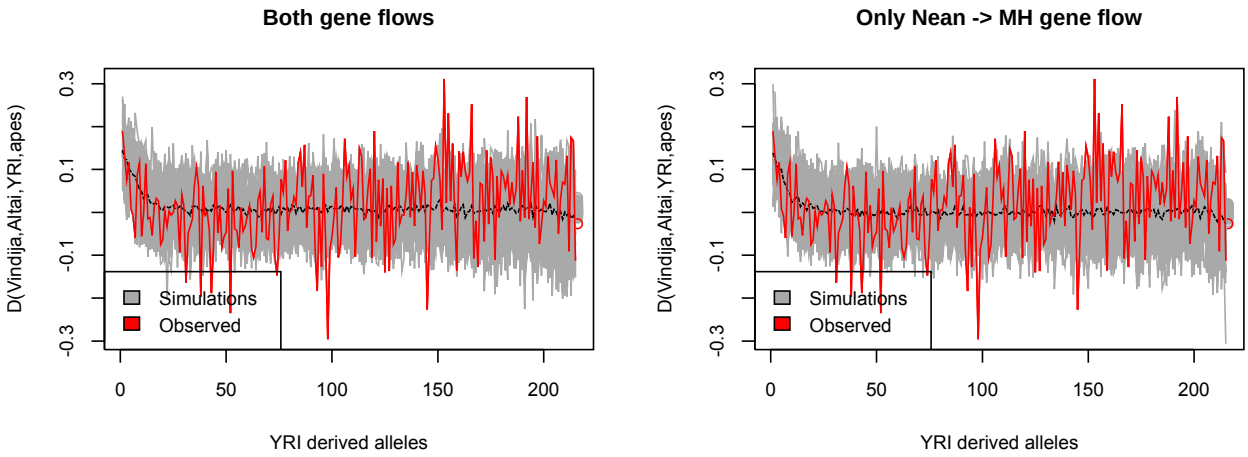

Fig. S66.
Simulations with the best fitting parameters (Table S29).

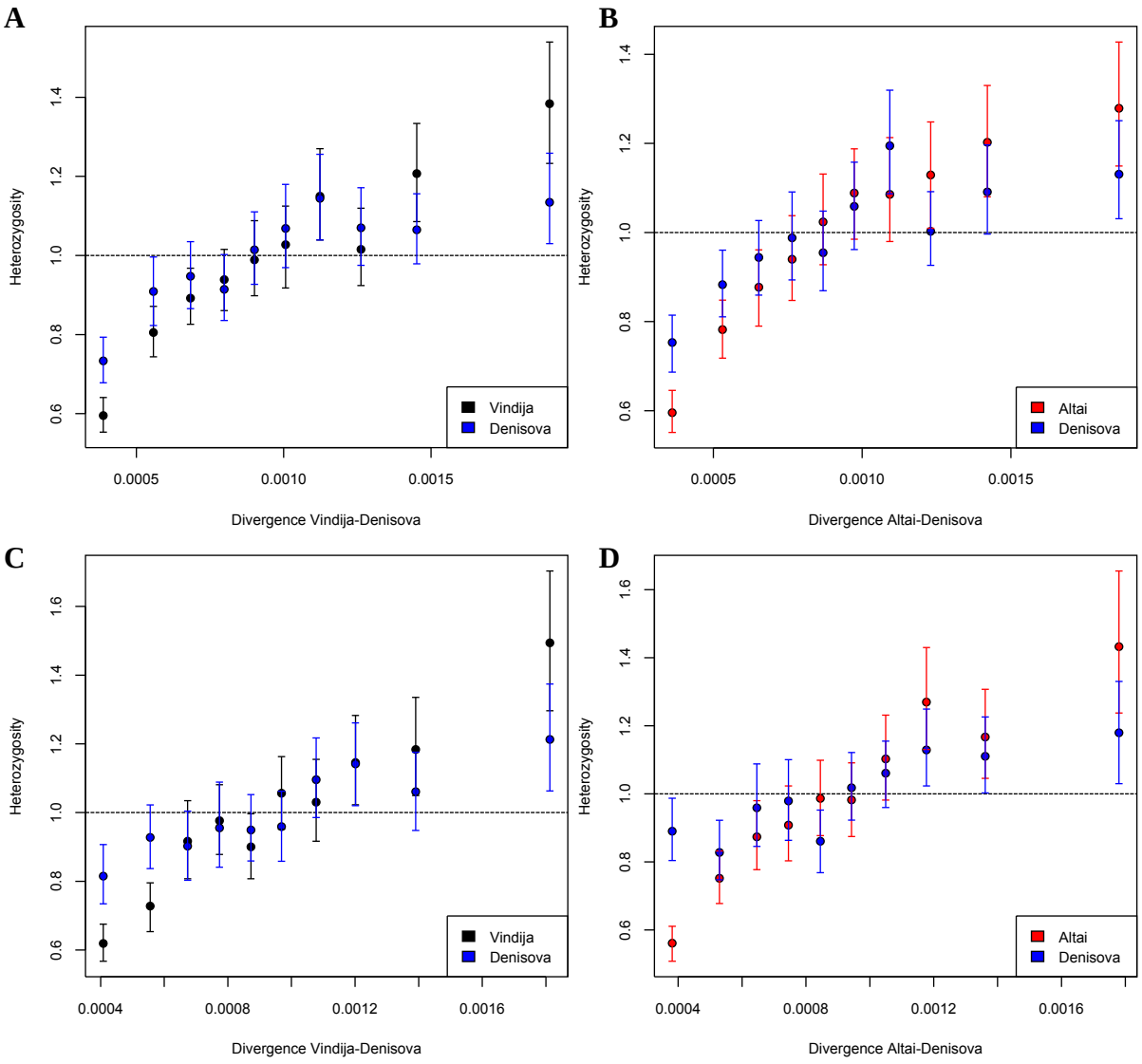

Fig. S67.

Heterozygosity in bins of divergence. **(A)** and **(B)**: 100kb bins; **(C)** and **(D)**: 0.1cM bins using the Decode recombination map. African American recombination map shows patterns intermediate between (A) and (C), and (B) and (D). Error bars represent the 95% central confidence intervals over 1000 bootstrap samples in each bin. Heterozygosity was normalized by the average heterozygosity over all windows and the dashed line represents the average heterozygosity in all individuals.

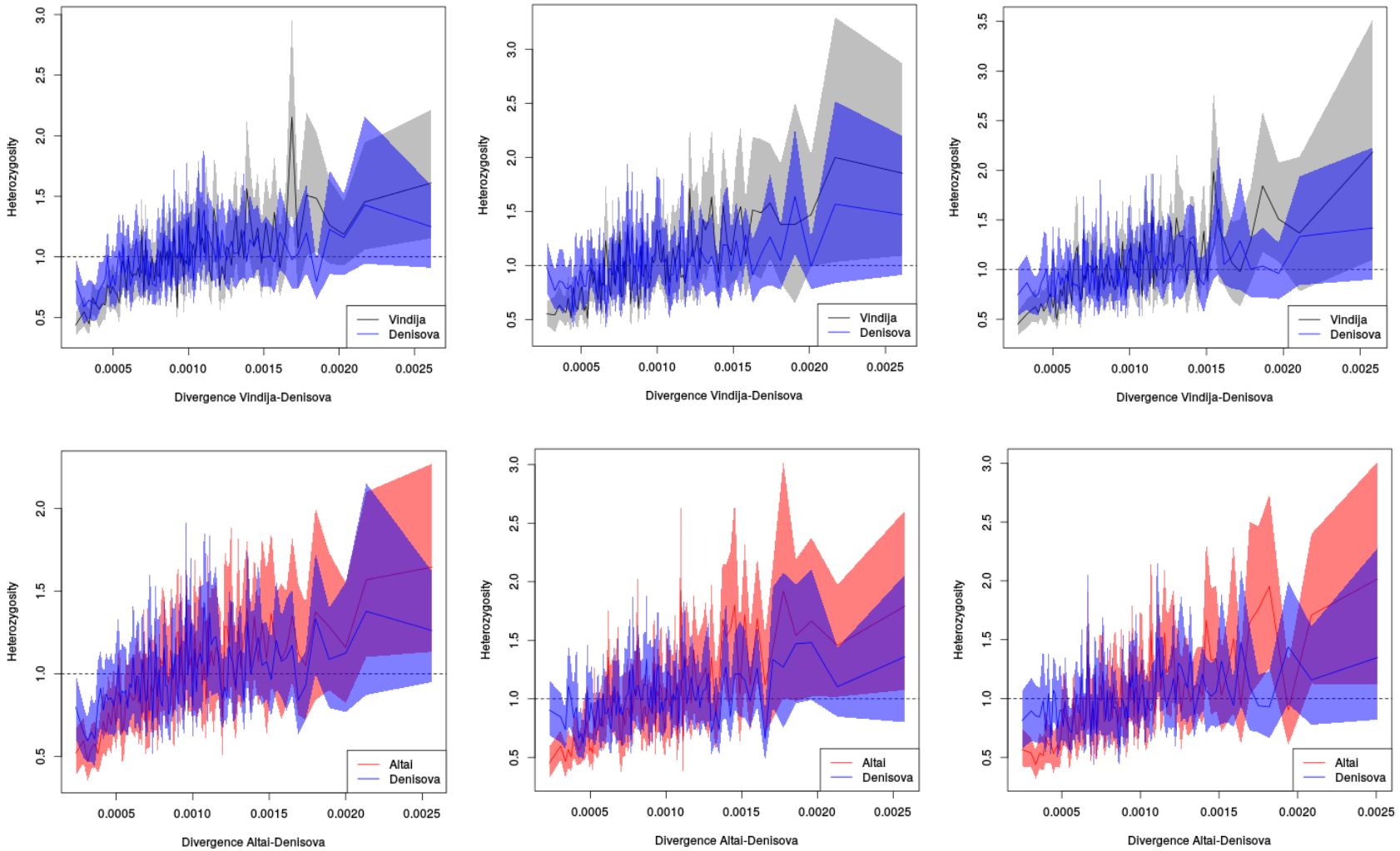

Fig. S68.
Heterozygosity in 100 bins of divergence. Shaded areas show the 95% central confidence intervals based on 1000 bootstrap samples. Left column shows plots for 100kb windows, middle column for 0.1cM windows according to the African American recombination map and the right column 0.1cM windows for the Decode map.

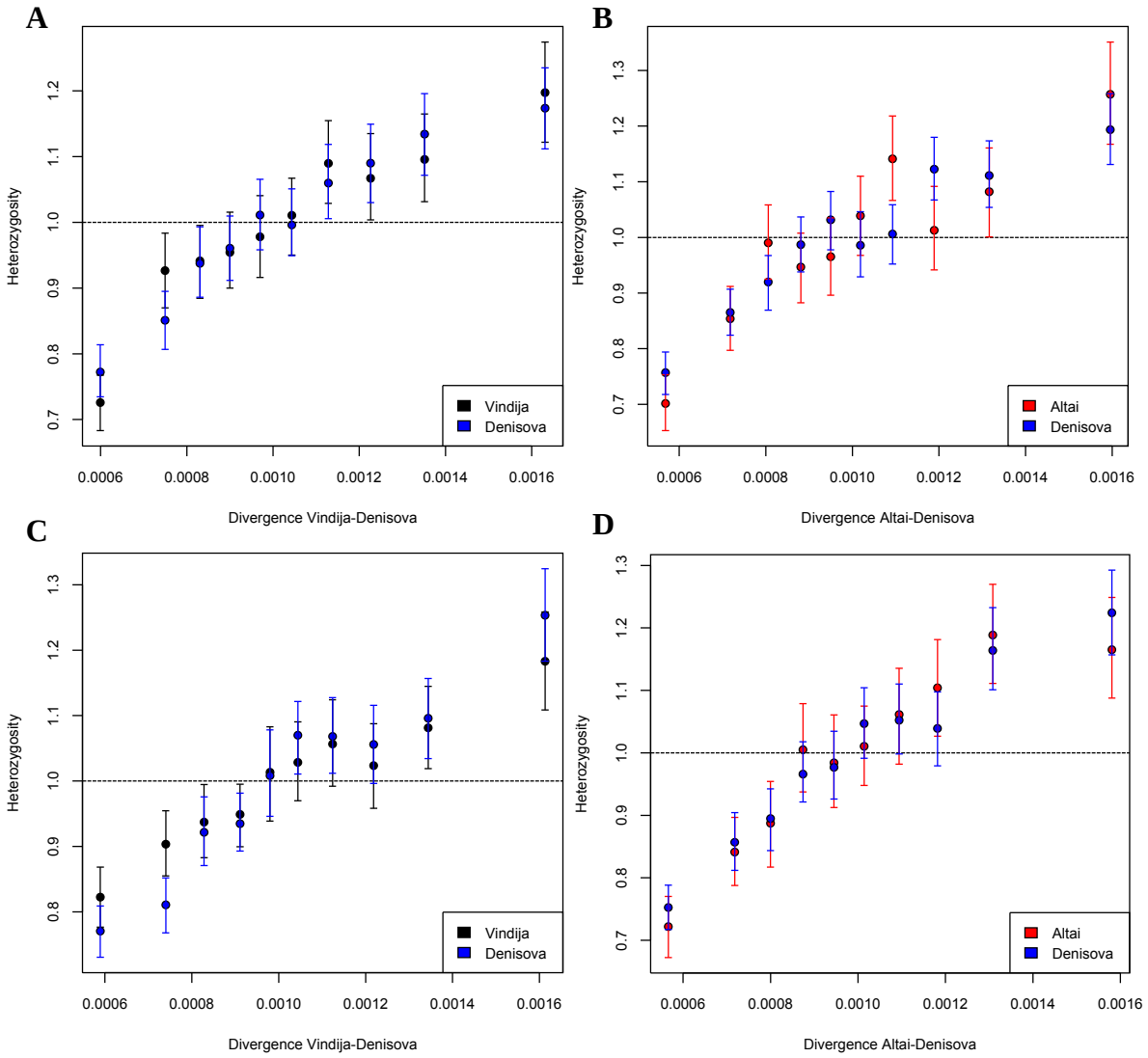

Fig. S69.
Heterozygosity in bins of divergence for two simulated datasets. **(A)** and **(B)**: Outlier simulation that shows significantly higher heterozygosity for the lowest bin; **(C)** and **(D)**: Randomly chosen simulation.

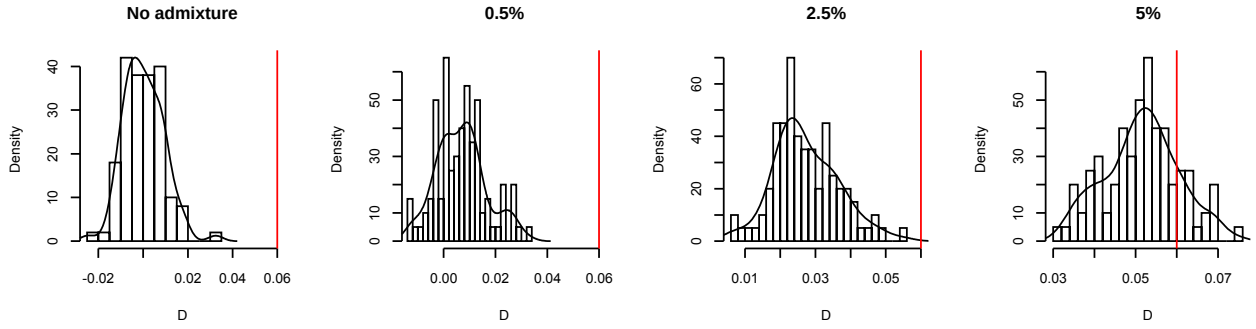

Fig. S70.
Histograms of D(Altai,Vindija,Denisova,Outgroup) for simulations with different admixture percentages from the Altai population into the Denisovan population.

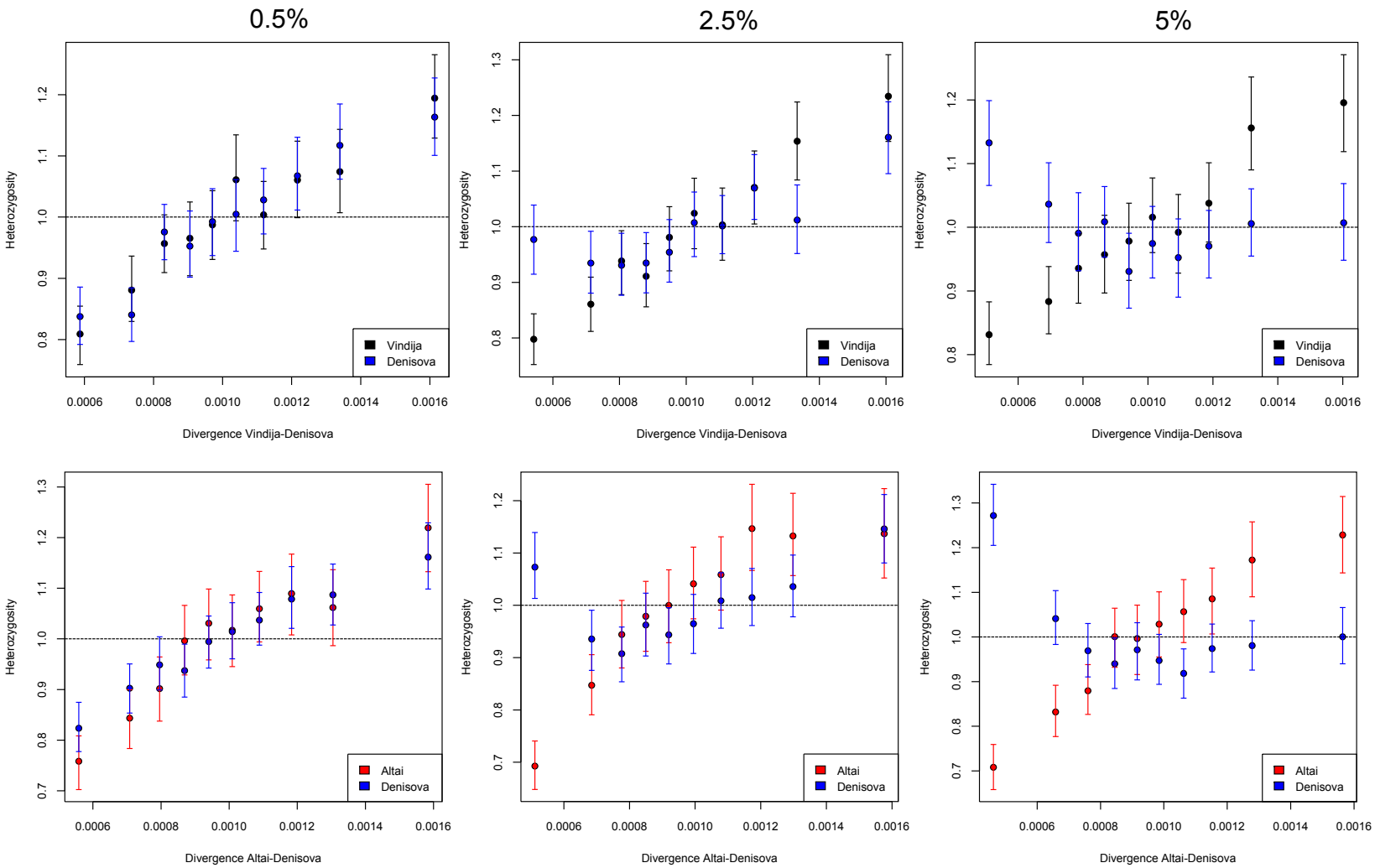

Fig. S71.
Simulations with 0.5%, 2.5% and 5% admixture from Altai to Denisova at 115kya. D(Altai,Vindija,Denisova,Outgroup) was 0.4% (left), 2.3% (middle) and 6.3% (right).

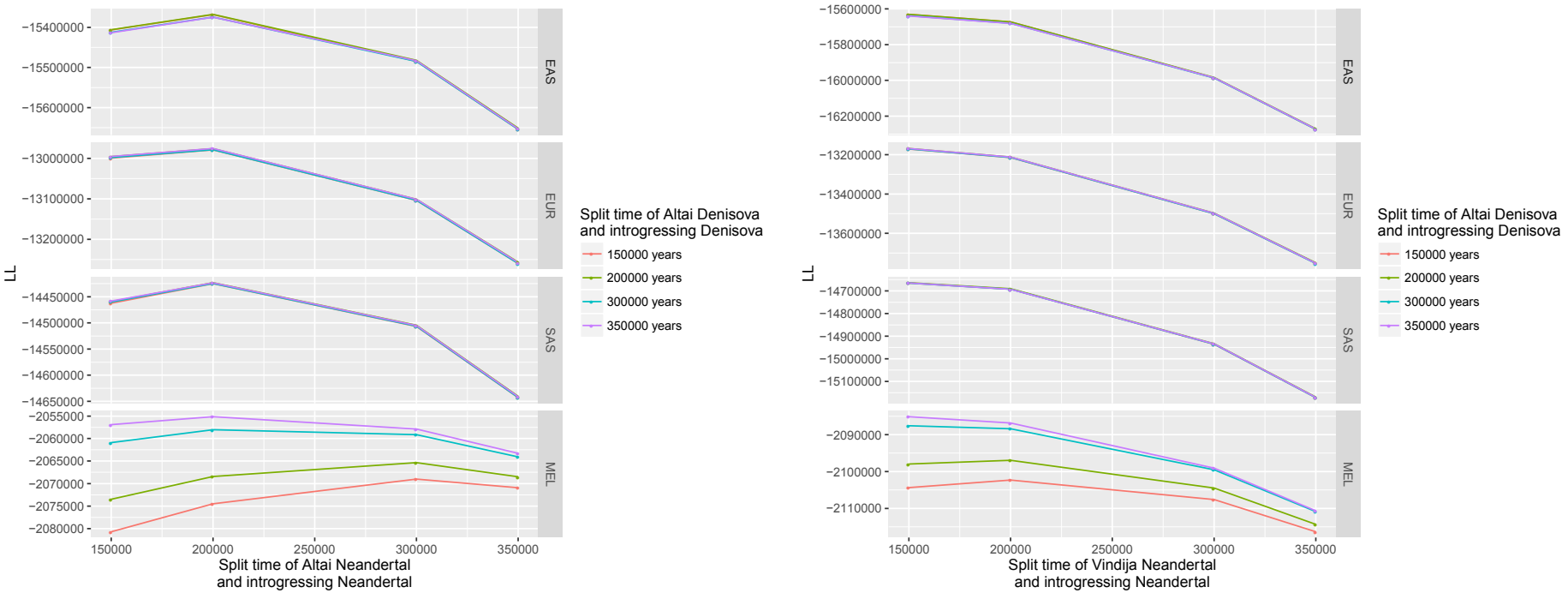

Fig. S72.

Demographic model likelihoods; left: Altai vs Denisova; right: Vindija vs Denisova

Fig. S73.

Demographic model likelihoods; Vindija vs Altai. Unlike comparisons with Denisova, the split time of the archaics was varied; each column is a different value of this parameter.

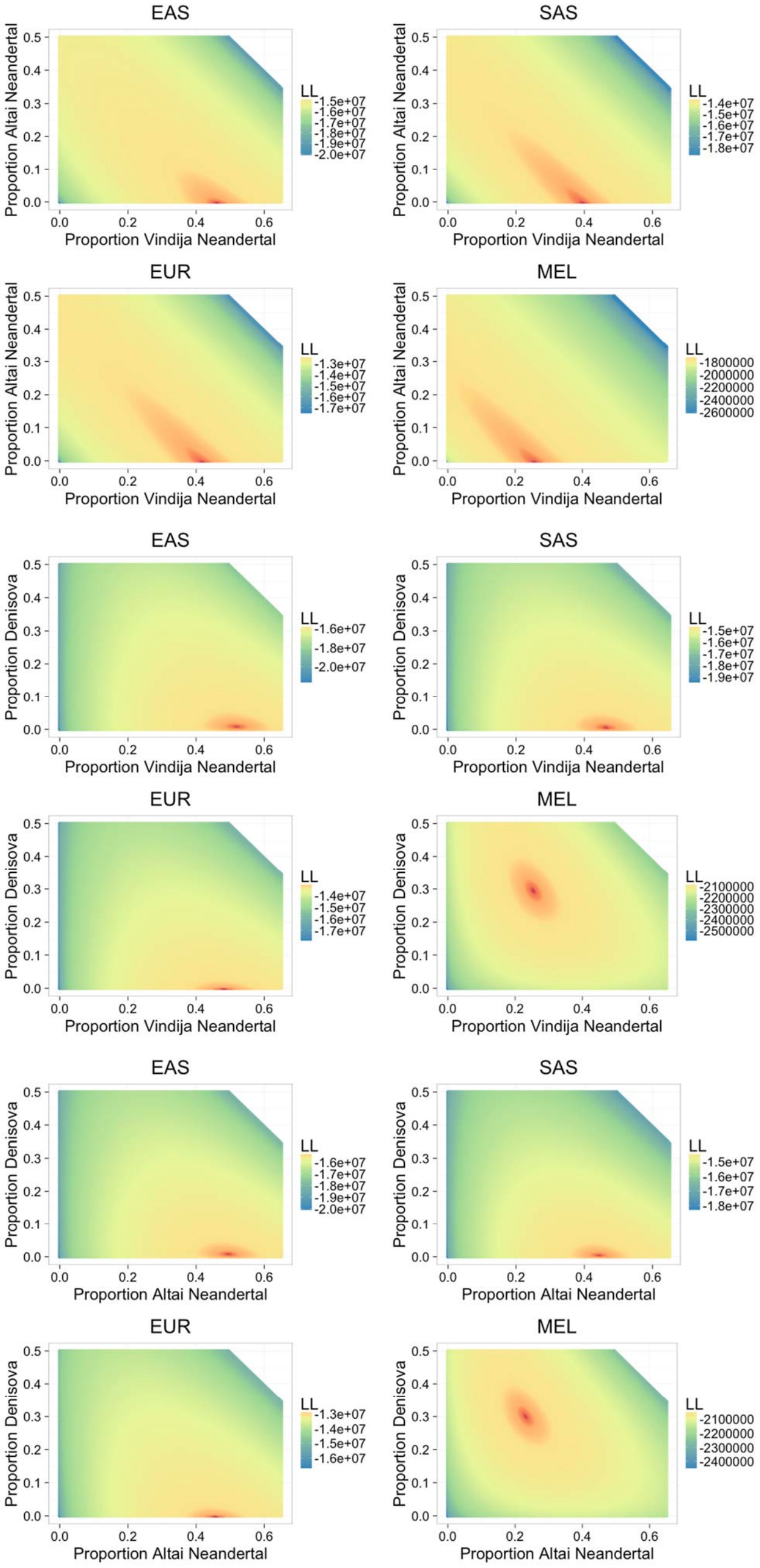

Fig. S74.
Likelihoods for the proportion of S* sequence from each archaic, for all archaic pairs and all populations.

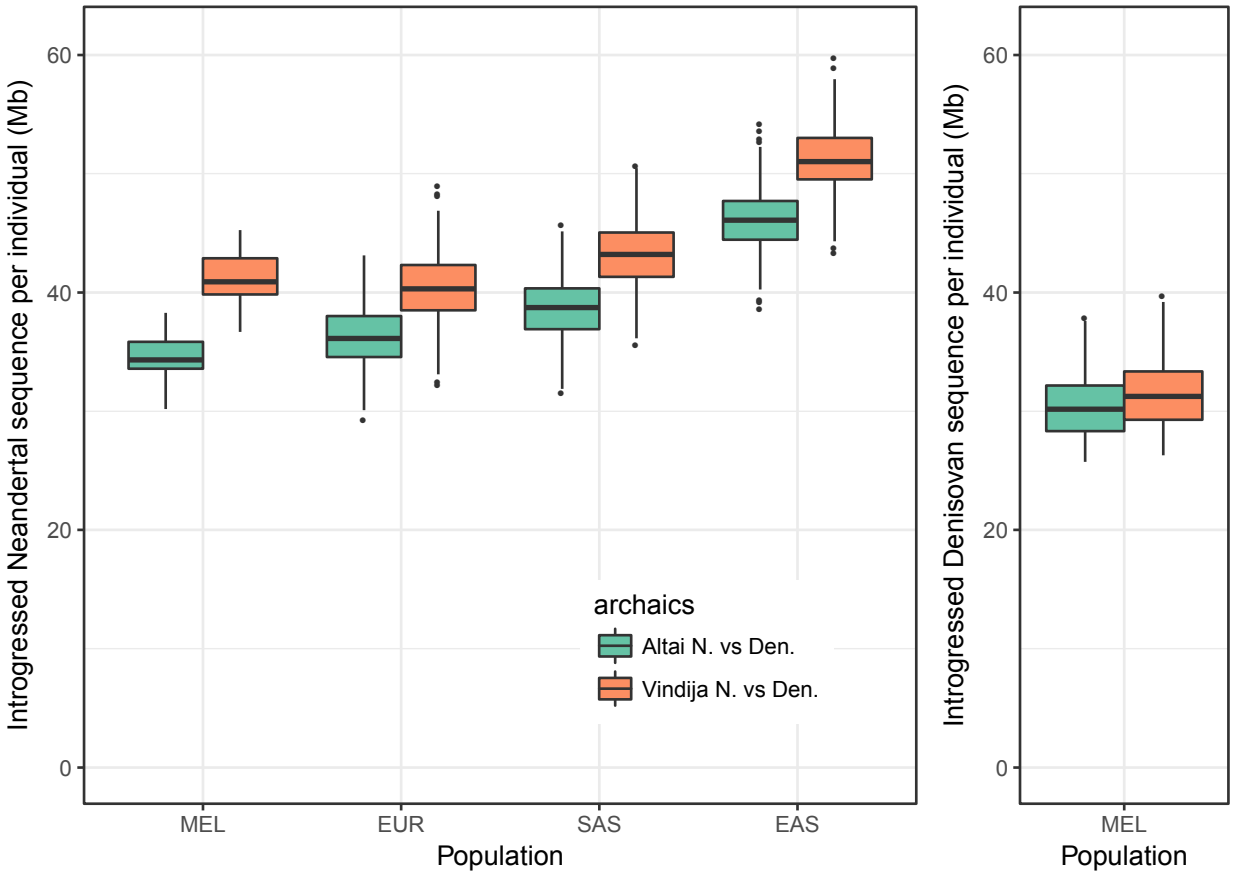

Fig. S75.
Additional introgressed sequence is identified by using Vindija instead of Altai.

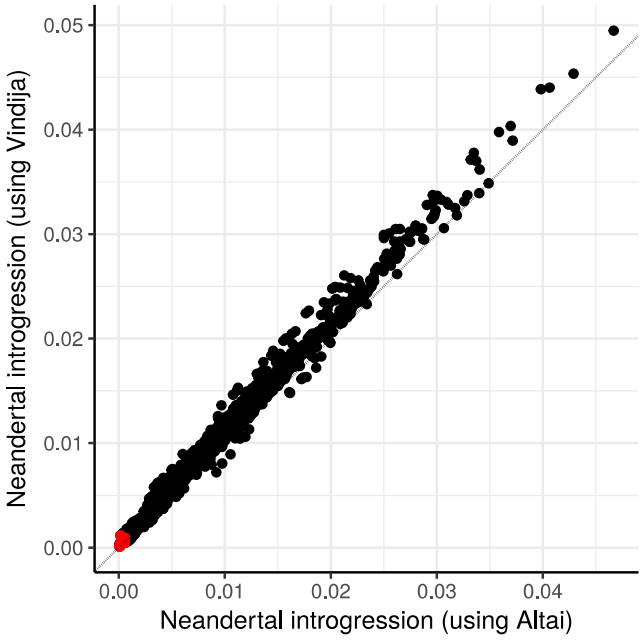

Fig. S76.
Neandertal ancestry in 10Mb windows. Red points show windows previously identified as archaic deserts. Dotted line shows slope=1.

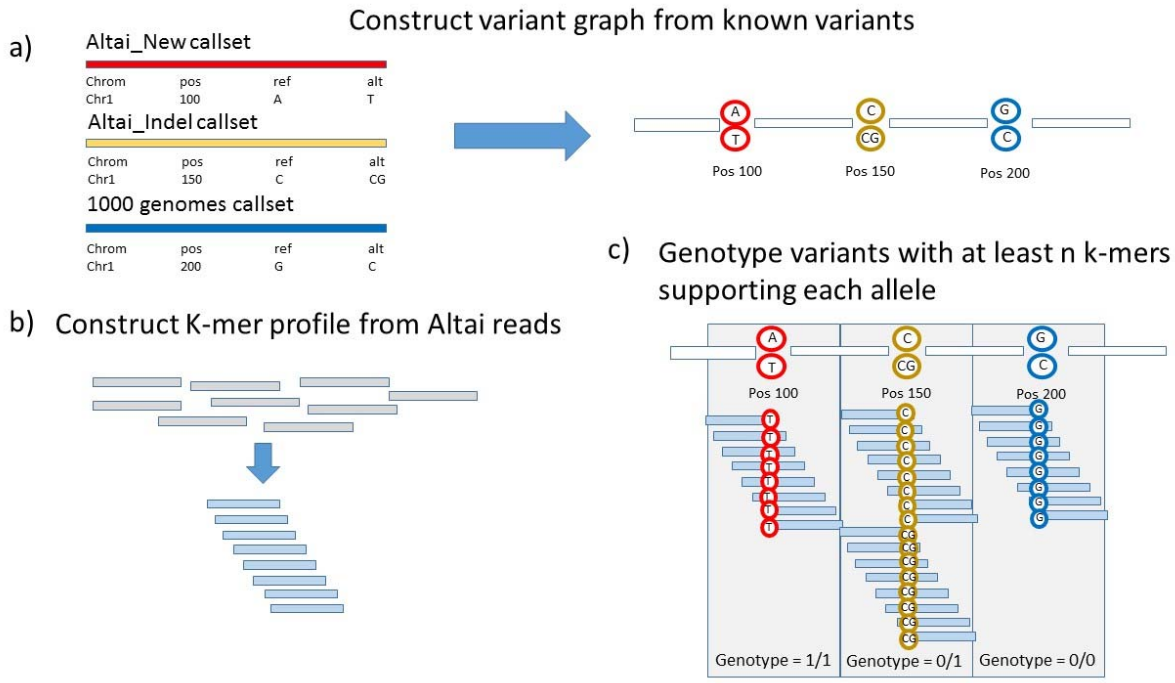

Fig. S77.

A small example of the BayesTyper genotyping process using three variants and the Altai individual. a) We merge the call sets into a variant graph using the human reference genome hg19. Nodes.in the graph represent the different variants while edges represent the sequence in-between variants. We keep track of the call set origin of each variant. b) We construct K-mer profiles for all individuals considered, in this case only the Altai Neandertal. c) We genotype the three variants. For each variant, we count the number of K-mers (NOK) that support the path through the graph that contains this variant. For the variant at position 100 all K-mers contain the alternative allele T so the genotype is 1/1. For the variant at position 150 we observe eight K-mers supporting the reference allele and eight K-mers supporting the alternative allele, therefore the genotype is 0/1 at this position. If we had observed 8 K-mers supporting the alternative allele and 5 K-mers supporting the reference allele, and our K-mer cutoff was 6 the variant would have been genotyped "1/." and therefore excluded. For the variant at position 200 we observe that all K-mers contain the G, meaning that the genotype is 0/0.

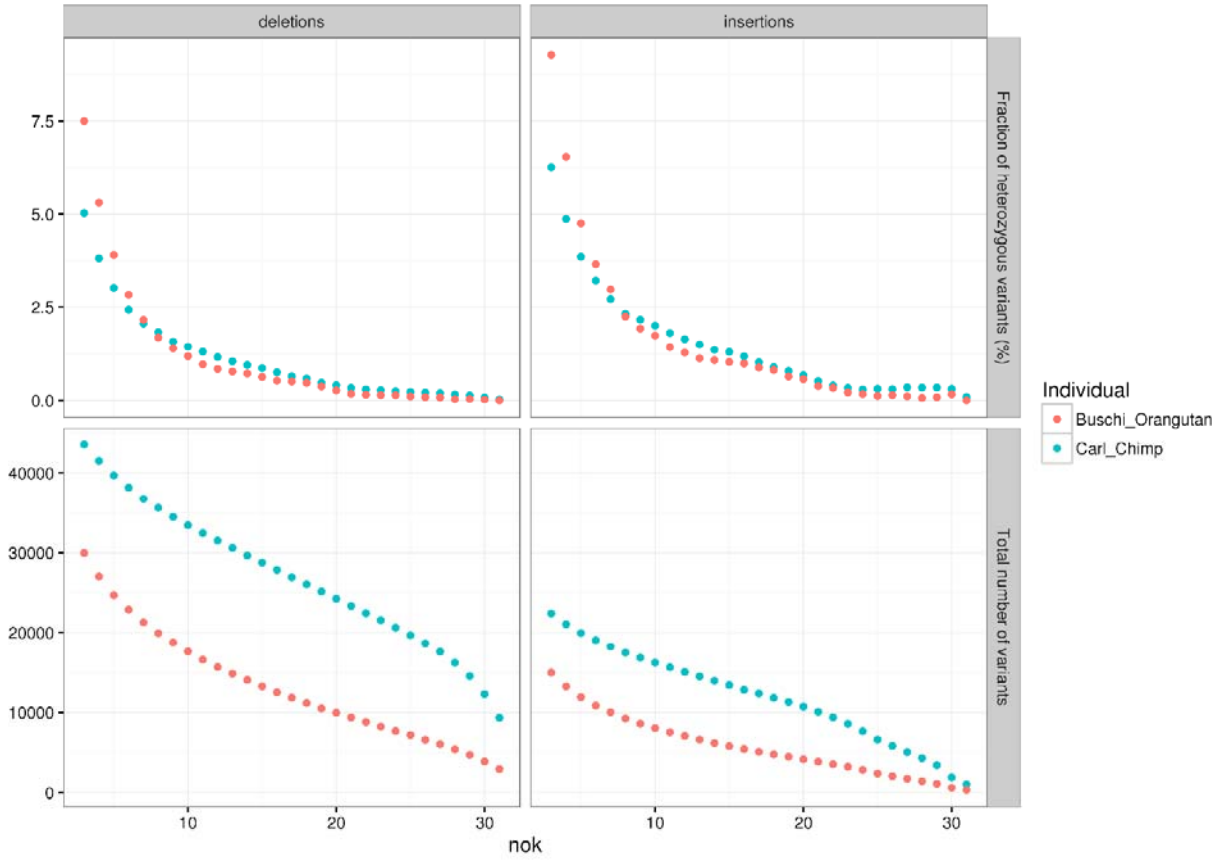

Fig. S78.
The percentage of heterozygous sites and number of genotyped human structural variants in Chimpanzee and Orangutan as a function of the number of supporting K-mers (NOK).

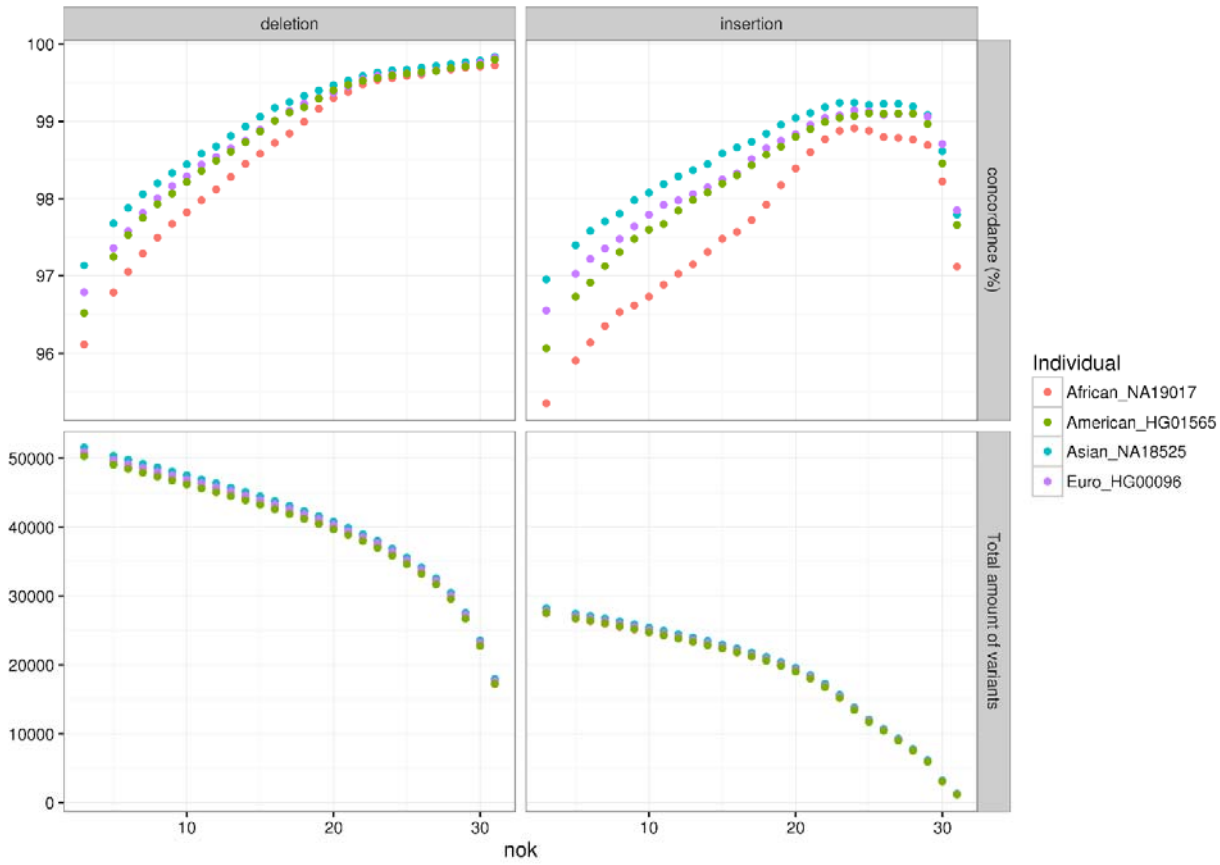

Fig. S79.
Concordance with 1000 genomes project genotypes are shown in the top panels for insertions and deletions. The lower panels shows the total amount of variants that could be genotyped.

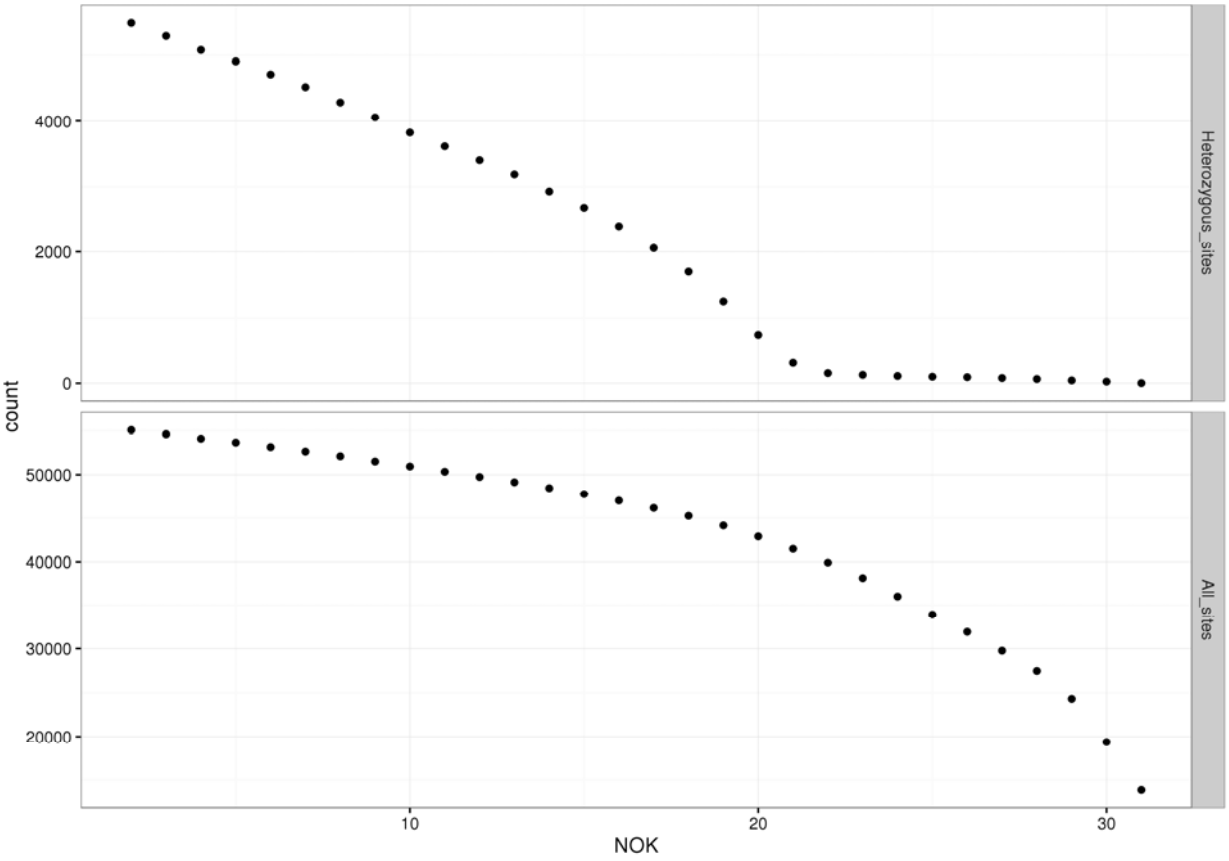

Fig. S80.
Number of heterozygous sites and total number of variants found in a tract of homozygosity in the Altai Neandertal on chromosome 14 from 50 Mb to 100 Mb. The number of heterozygous variants genotyped in the long track of homozygosity drastically decreases with more than 20 supporting K-mers.

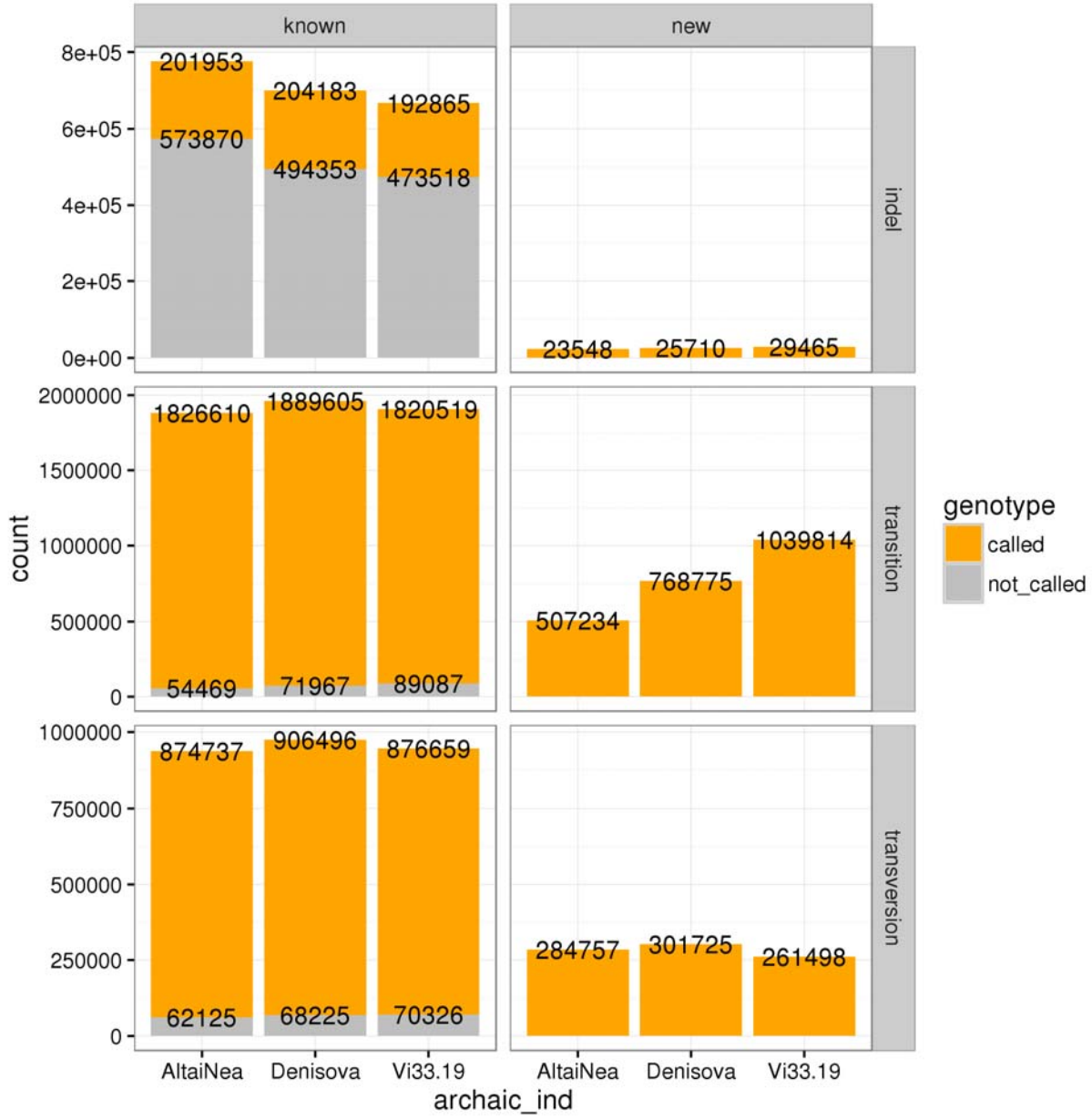

Fig. S81.

The number of known and new variants. Each column is an archaic individual. The known variants are variants from the Altai, Denisova and Vindija call sets. The known indels would be the indels, called with GATK in each individual. For example, for all the indels in the Altai indel dataset we found enough K-mer support to genotype 201953 variants – but could not find K-mer support for 573870 variants. For Altai we also find enough K-mer support for genotyping 23548 "new" indels.

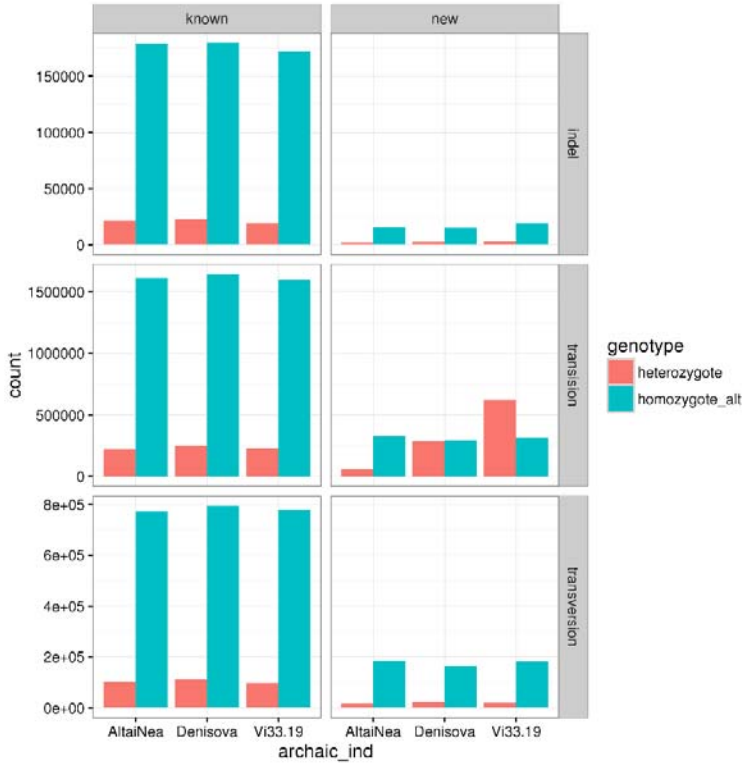

Fig. S82.
The number of known and new variants are shown and each column is an archaic individual. The number of heterozygous and homozygous variants for Altai, Denisova and Vindija.

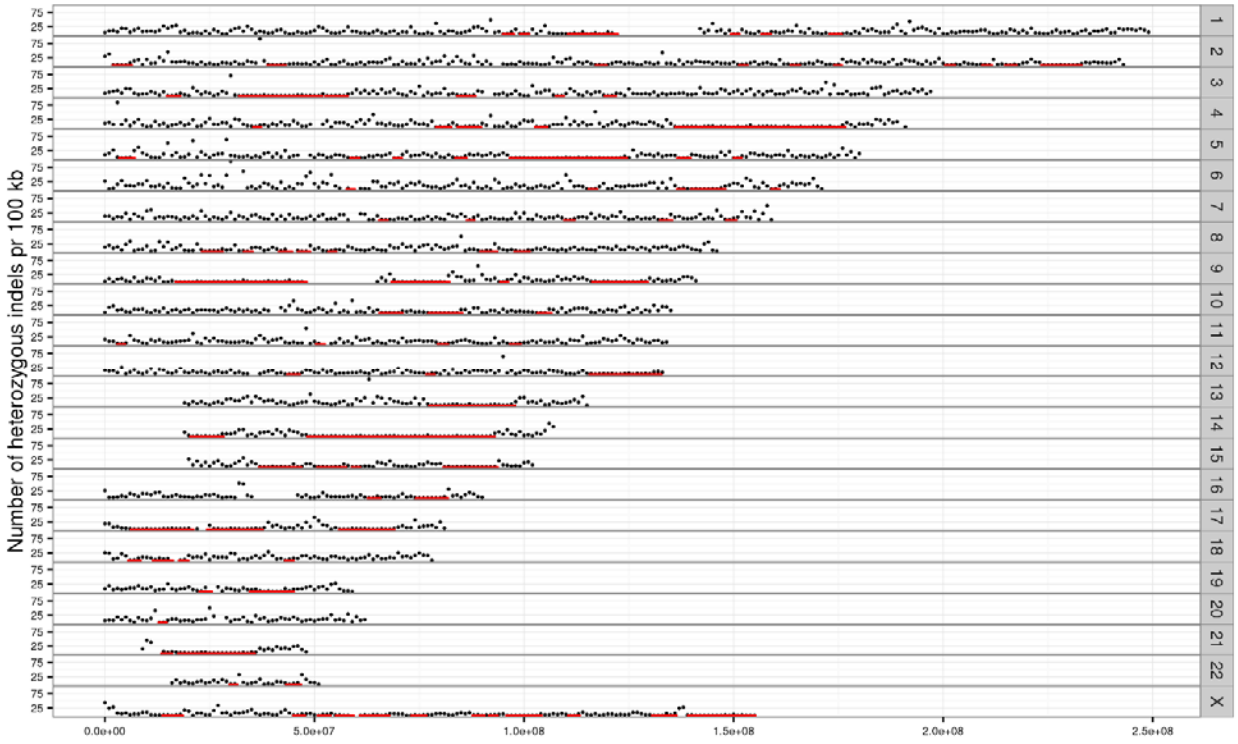

Fig. S83.
Number of heterozygous indels along each chromosome for Altai Neandertal. Y-axis is truncated at 100 variants. Red lines are tracks of inferred homozygosity.

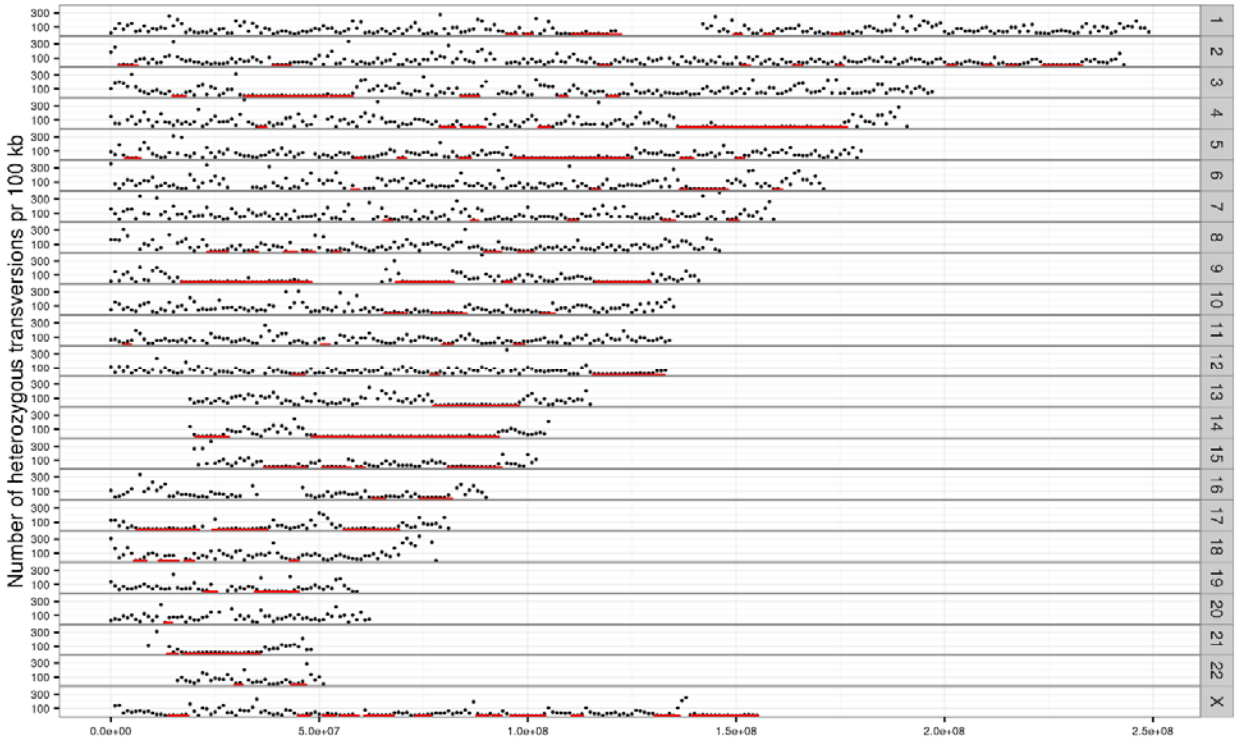

Fig. S84.
Number of heterozygous transversions along each chromosome for Altai Neandertal. Y-axis is truncated at 400 variants. Red lines are tracks of inferred homozygosity.

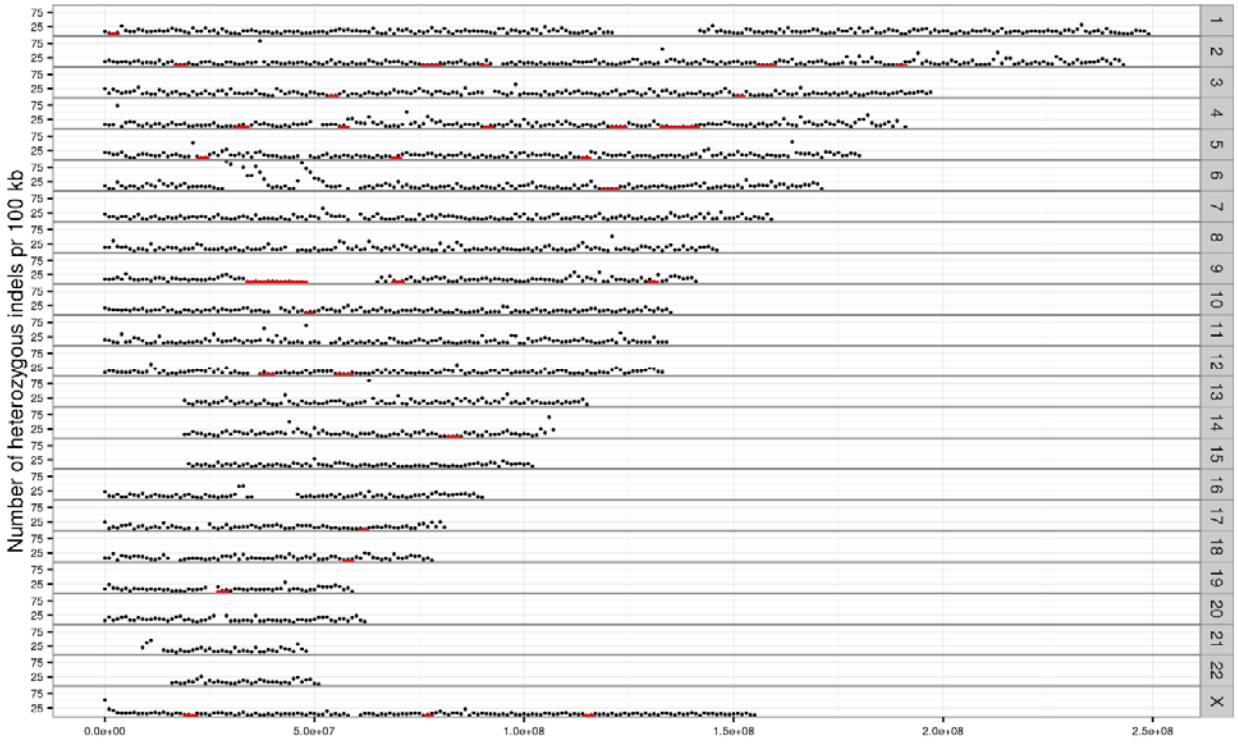

Fig. S85.
Number of heterozygous indels along each chromosome for Denisova. Y-axis is truncated at 100 variants. Red lines are tracks of inferred homozygosity.

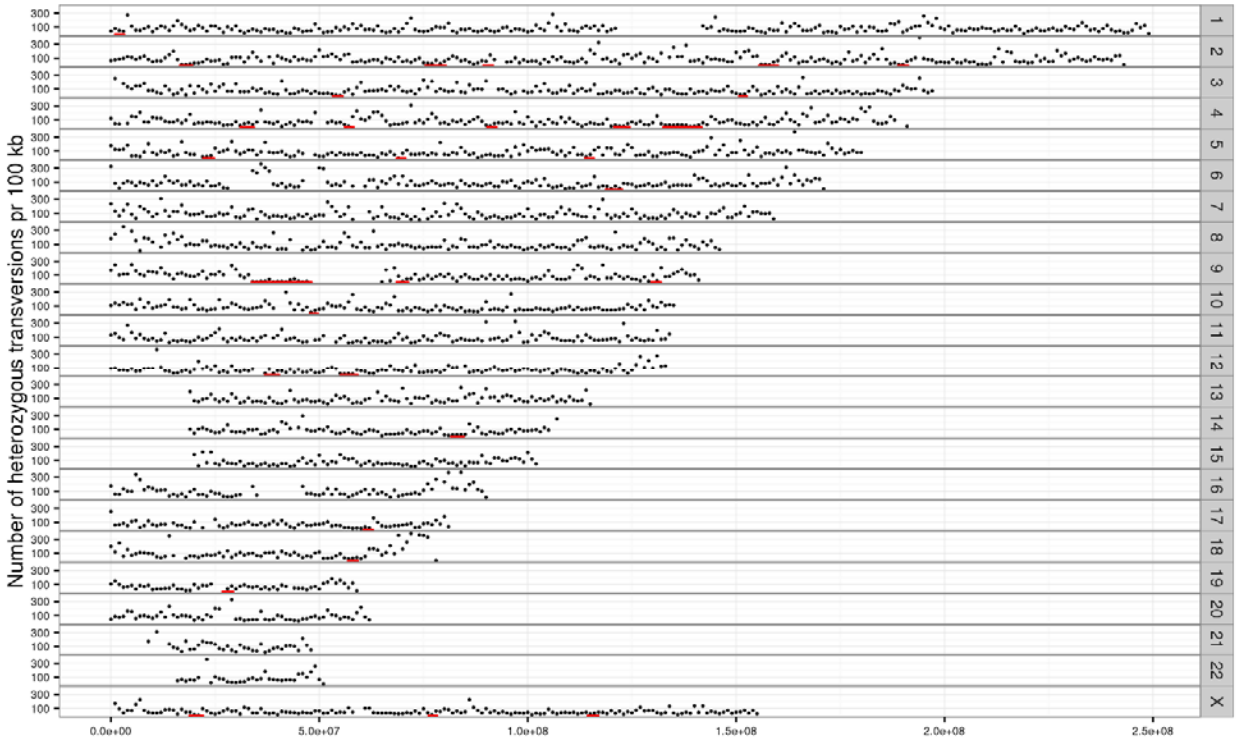

Fig. S86.
Number of heterozygous transversions along each chromosome for Denisova. Y-axis is truncated at 400 variants. Red lines are tracks of inferred homozygosity.

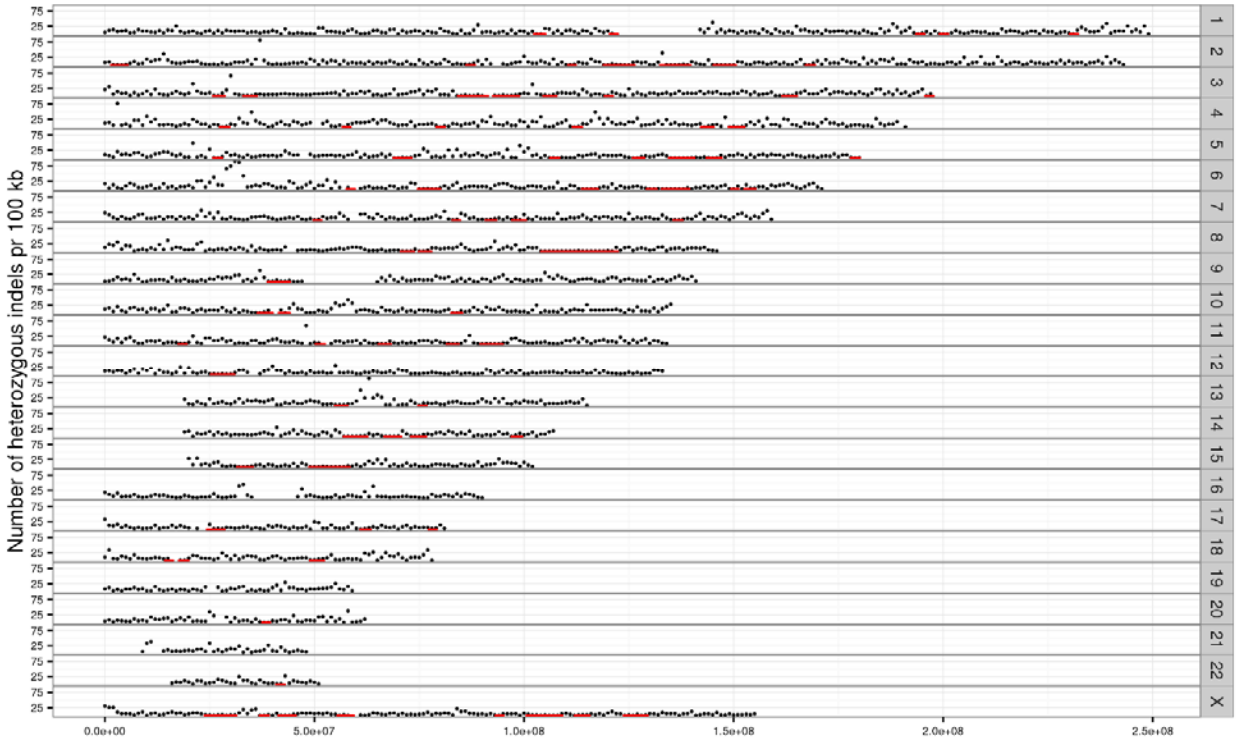

Fig. S87.
Number of heterozygous indels along each chromosome for Vindija Neandertal. Y-axis is truncated at 100 variants. Red lines are tracks of inferred homozygosity.

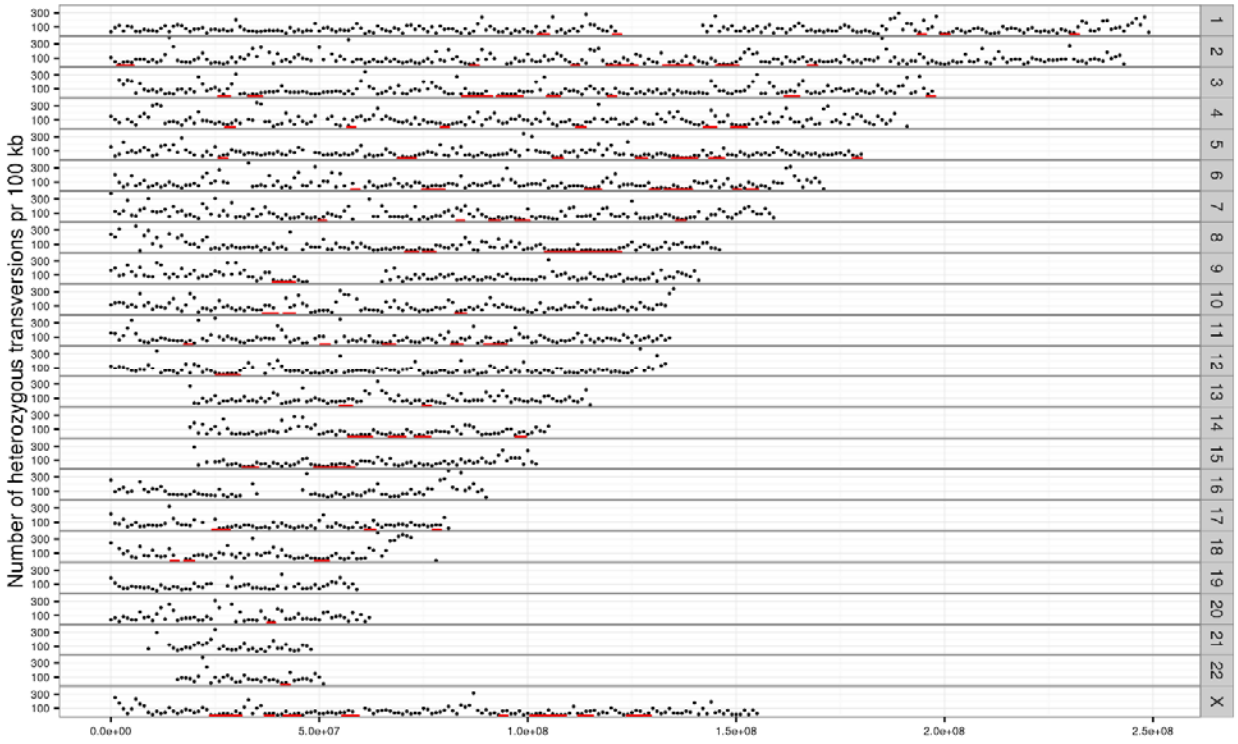

Fig. S88.
Number of heterozygous transversions along each chromosome for Vindija Neandertal. Y-axis is truncated at 400 variants. Red lines are tracks of inferred homozygosity.

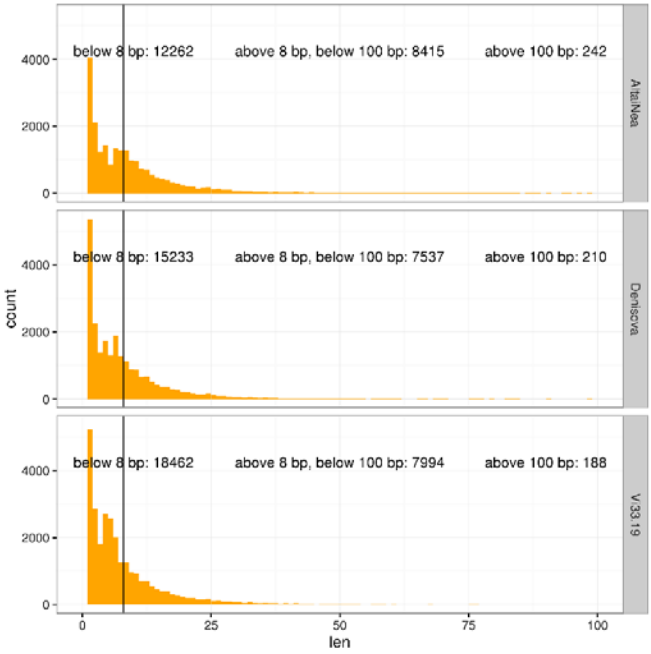

Fig. S89.
The size distribution of new indels for Altai, Denisova and Vindija. We show both heterozygous indels and homozygous indels for the alternative allele. The black vertical line indicates the length of the longest variant (8 bp) called in any of the original Altai, Denisova or Vindija indel call sets. The x-axis is truncated at 100 bp. The number of variants below 8 bp, above 8 bp and below 100 bp and variants above 100 bp are shown for each individual.

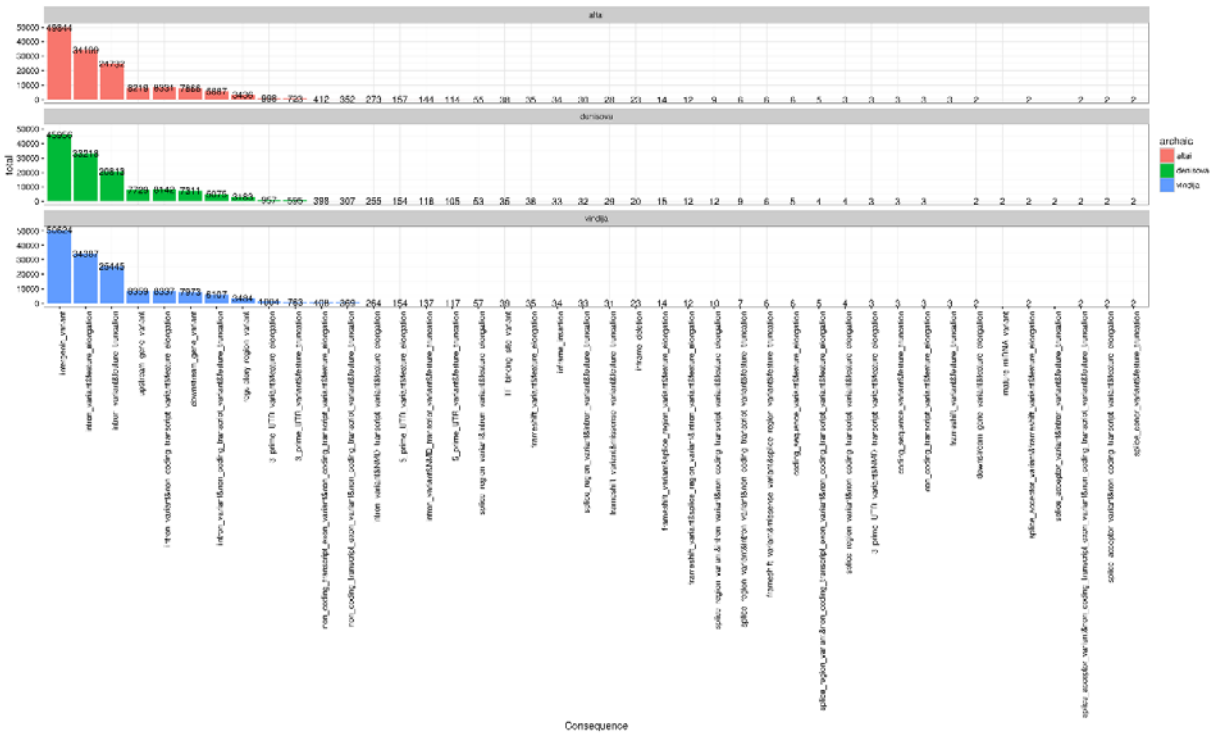

Fig. S90.
Functional annotation of the new and known variants. The functional annotation is taken from the 1000 genomes project.

Fig. S91.
The number of shared indels with the archaic individuals in 100 kb windows.

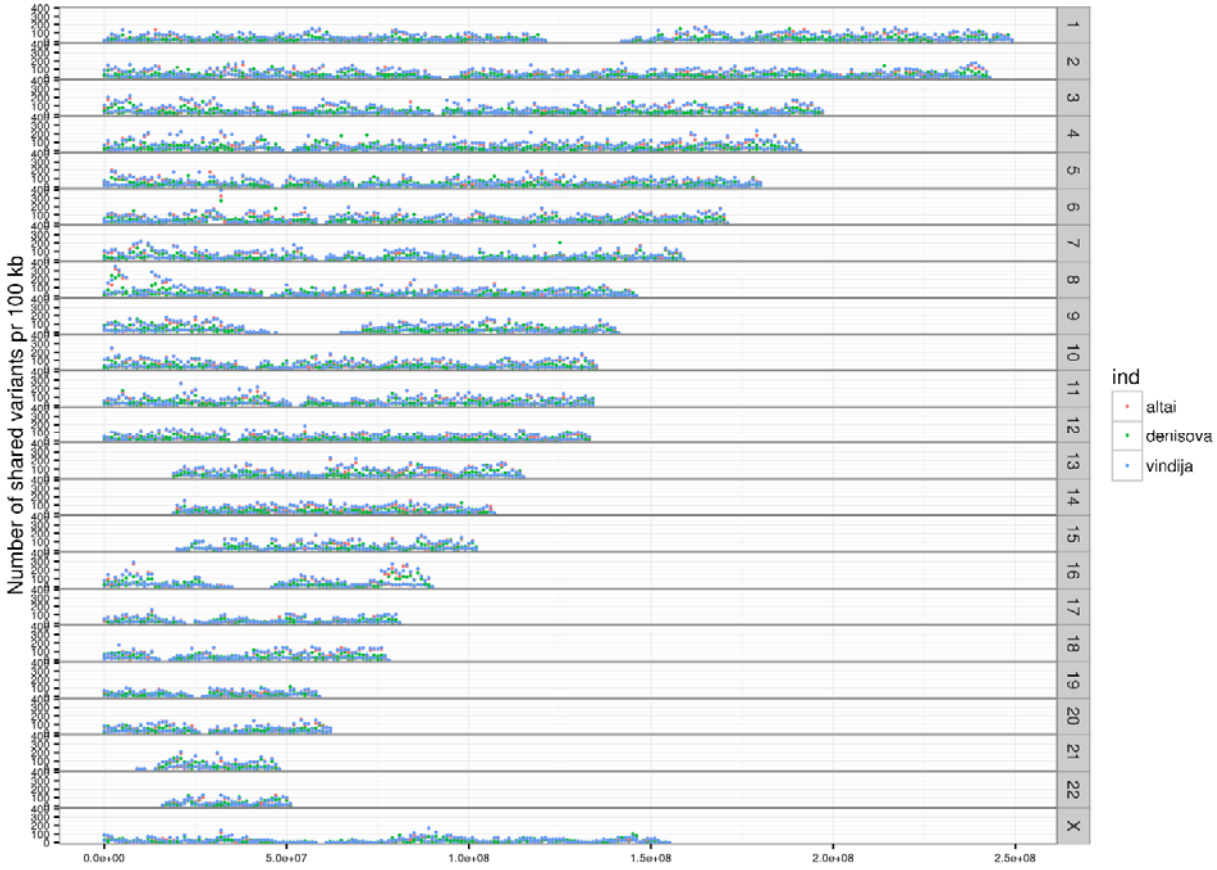

Fig. S92.
The number of shared transversions with the archaic individuals in 100 kb windows.

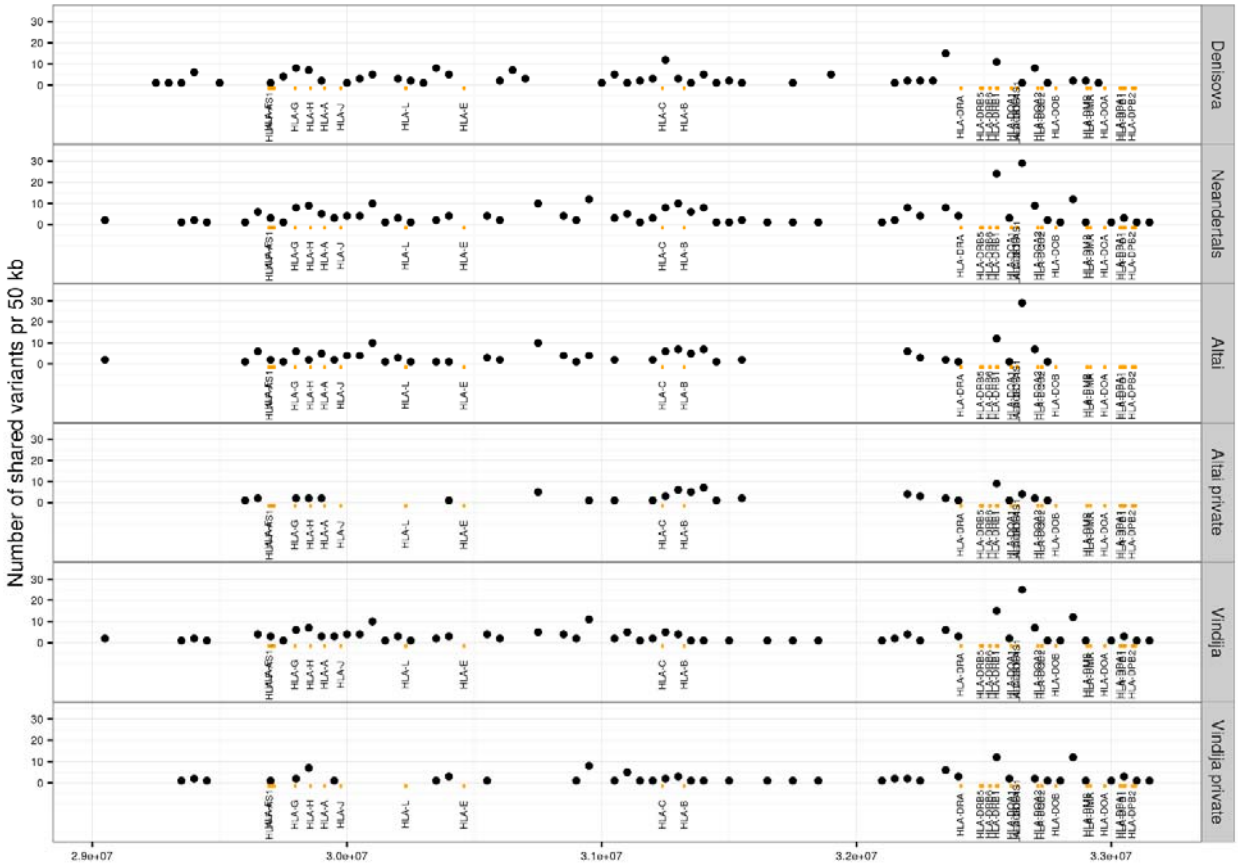

Fig. S93.
Number of shared indels between Denisova, both Neandertals, Vindija Neandertal and Altai Neandertal in the HLA region. HLA protein coding genes are shown in orange. For Denisova we counted variants private to Denisova. For Altai and Vindija we counted the variants that were present in the given individual and absent in Denisova (had to be homozygous for the reference allele). The numbers for Altai private and Vindija private are indels private to that individual.

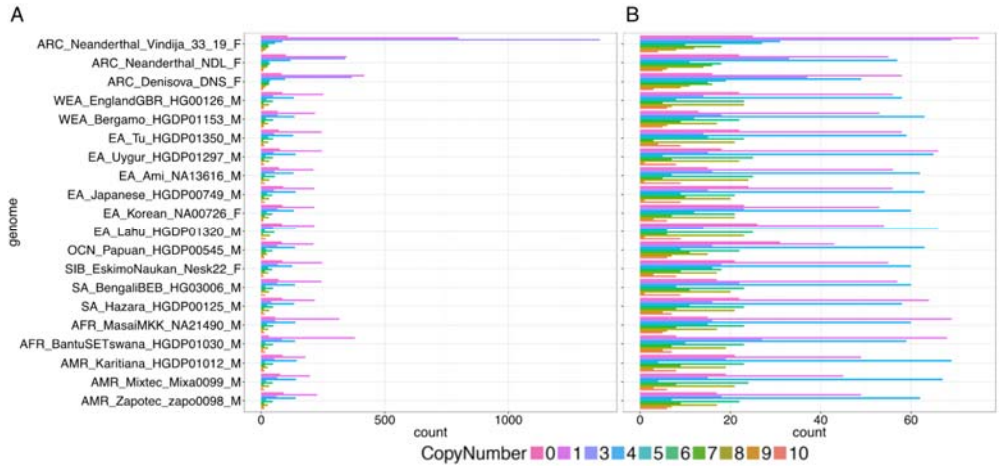

Fig. S94.
CNV genotype distribution for 20 genomes in the discovery panel. Before (A) and after (B) application of a 10-kbp length threshold. For simplicity, the most likely ancestral state CN=2 calls are excluded from the plots. We observe a significant excess of CNV calls within the Vindija genome before the application of a minimal length cutoff.

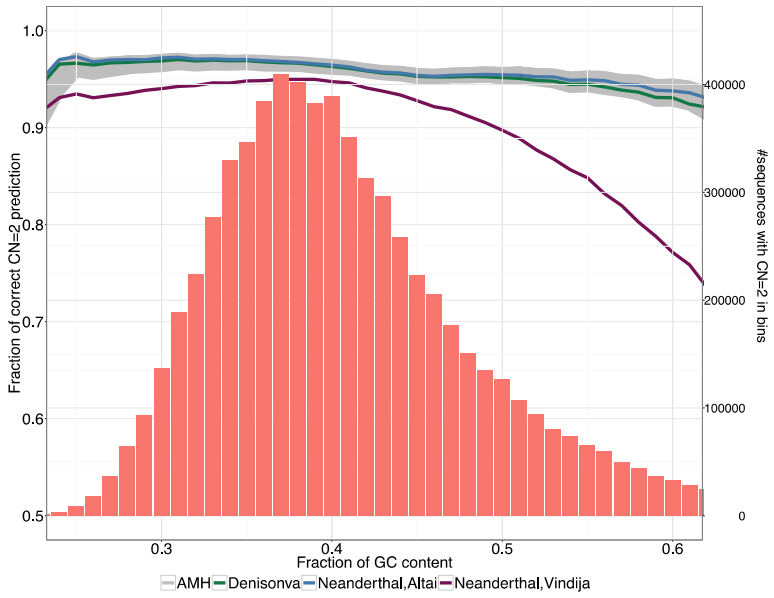

Fig. S95.
Accuracy of the dCGH inference for the 20-genomes discovery panel using known copy number invariant sequences (CN=2) across the spectrum of GC content. Sequences were grouped into intervals of 1% GC-content bins. Lines represent the fraction of correct prediction for the CN=2 control loci (the left y-axis) in the 20 genomes. Bar-chart indicates the numbers of known CN=2 sequences in the GC-content bins (the right y-axis).

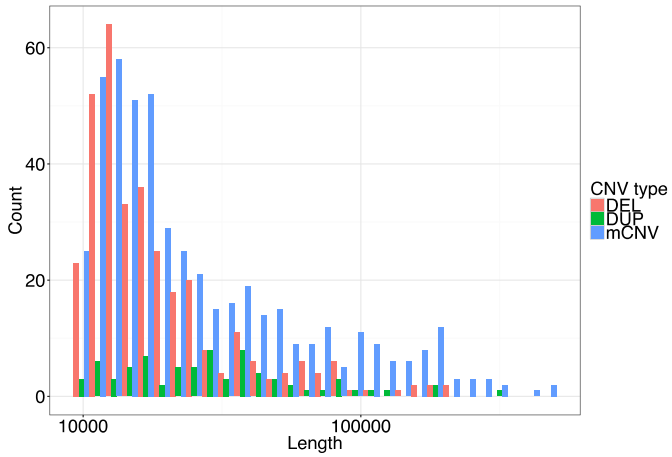

Fig. S96.
Log-scaled size distribution of the 904 CNV calls. DEL: biallelic deletion (CN=0, 1, or 2), DUP: biallelic duplication (CN=2, 3, or 4), mCNV: multi-allelic CNV.

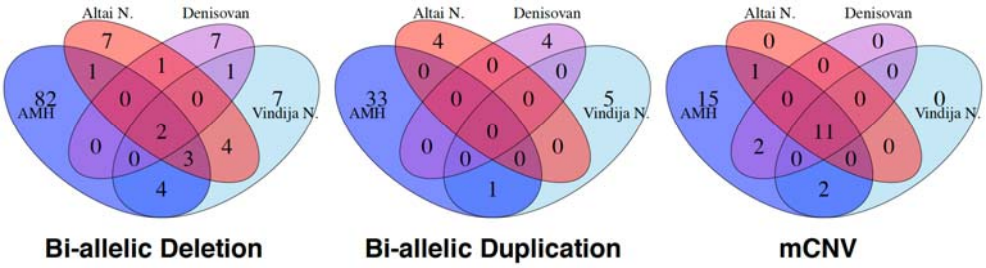

Fig. S97.

Venn diagrams of 197 putative hominin-specific CNVs. Numbers in areas refer to the amounts of non-reference (CN≠2) calls shared among and/or specific to lineages. Biallelic deletions and duplications are loci called with CN=0, 1, or 2, and CN=2, 3, or 4, respectively, across samples. Multi-allelic CNVs (mCNVs) refer to other more complicating variants in copy number. AMH: anatomically modern human; Altai N.: Altai Neanderthal; Vindija N.: Vindija Neanderthal.

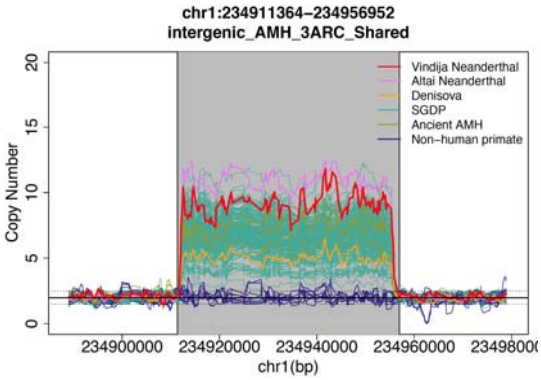

Fig. S98.

Expansion in copy number at chr1:234911364-234956952 (shaded area) in the hominin samples of this analysis. Note that the two Neanderthal individuals have almost the highest copy number (Vindija CN = 9 and Altai CN =11) among the samples tested in the current study.

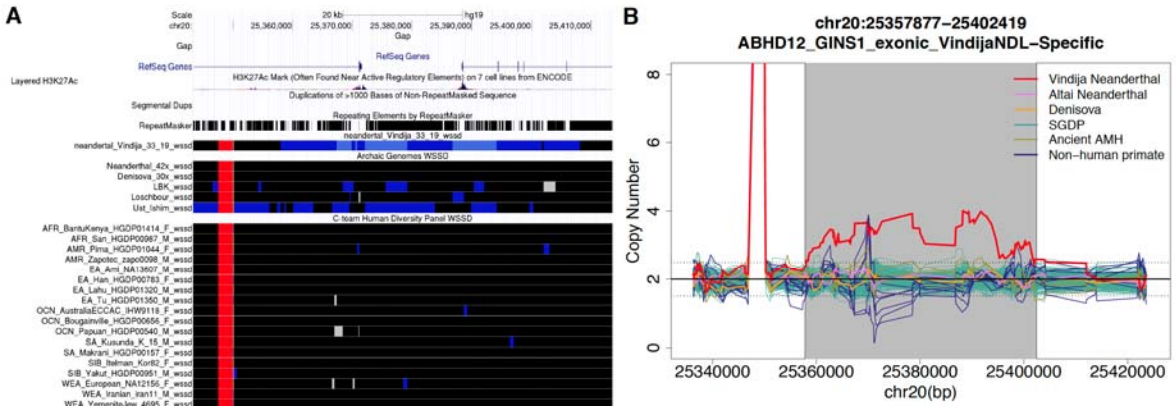

Fig. S99.
Vindija Neanderthal-specific duplication at chr20:25343369-25413592. (A) Genome browser screenshot with tracks of whole genome shotgun sequence detection (wssd) for a subset of genomes from the genotyping panel. The black, dark blue, and light blue colors in the tracks correspond to copy number status of 2, 3, and 4, respectively. (B) Trajectories of copy number variation across chr20:25343369-25413592 (shaded area) for samples in the genotyping panel.

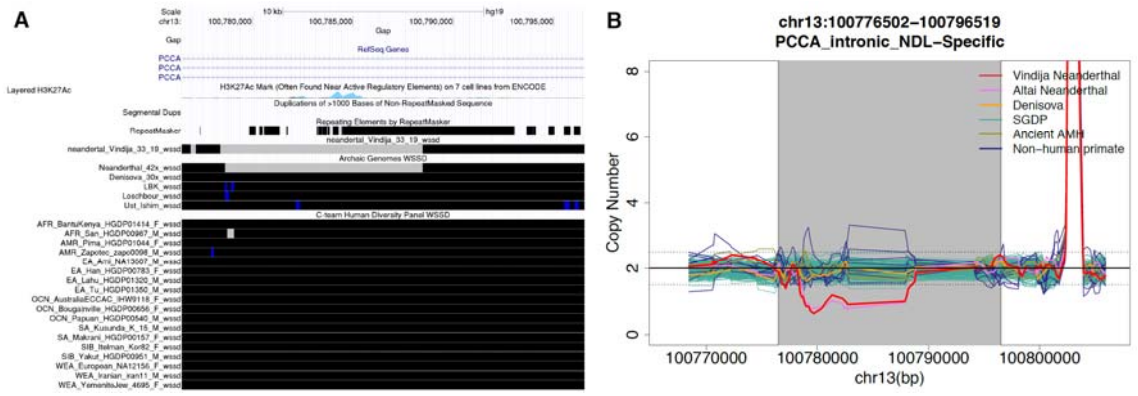

Fig. S100.

Neanderthal-specific deletion at chr13:100776502-100796519. (A) Genome browser screenshot with tracks of whole genome shotgun sequence detection (wssd) for a subset of genomes from the genotyping panel. The grey and black colors in the tracks correspond to copy number status of 1 and 2, respectively. (B) Trajectories of copy number variation across chr13:100776502-100796519 (shaded area) for samples in the genotyping panel.

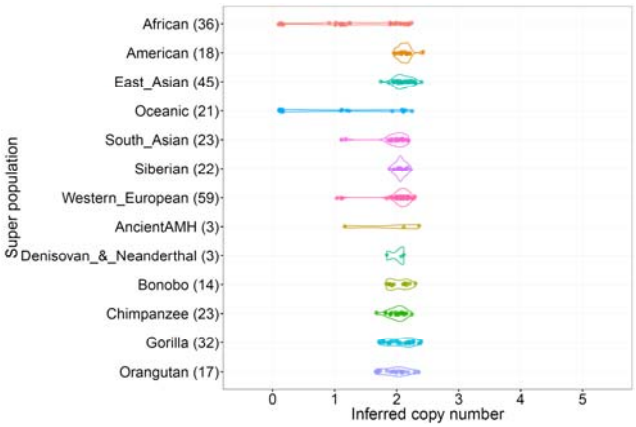

Fig. S101.
Common deletion of chr3:114658237-114672775 in African and Oceanic populations. This locus encompasses the gene *ZBTB20*, a protein coding gene with zinc finger and BTB domains. Both homozygous/hemizygous deletions are segregating at relatively high frequencies in the Africans and Oceanians in our sample.

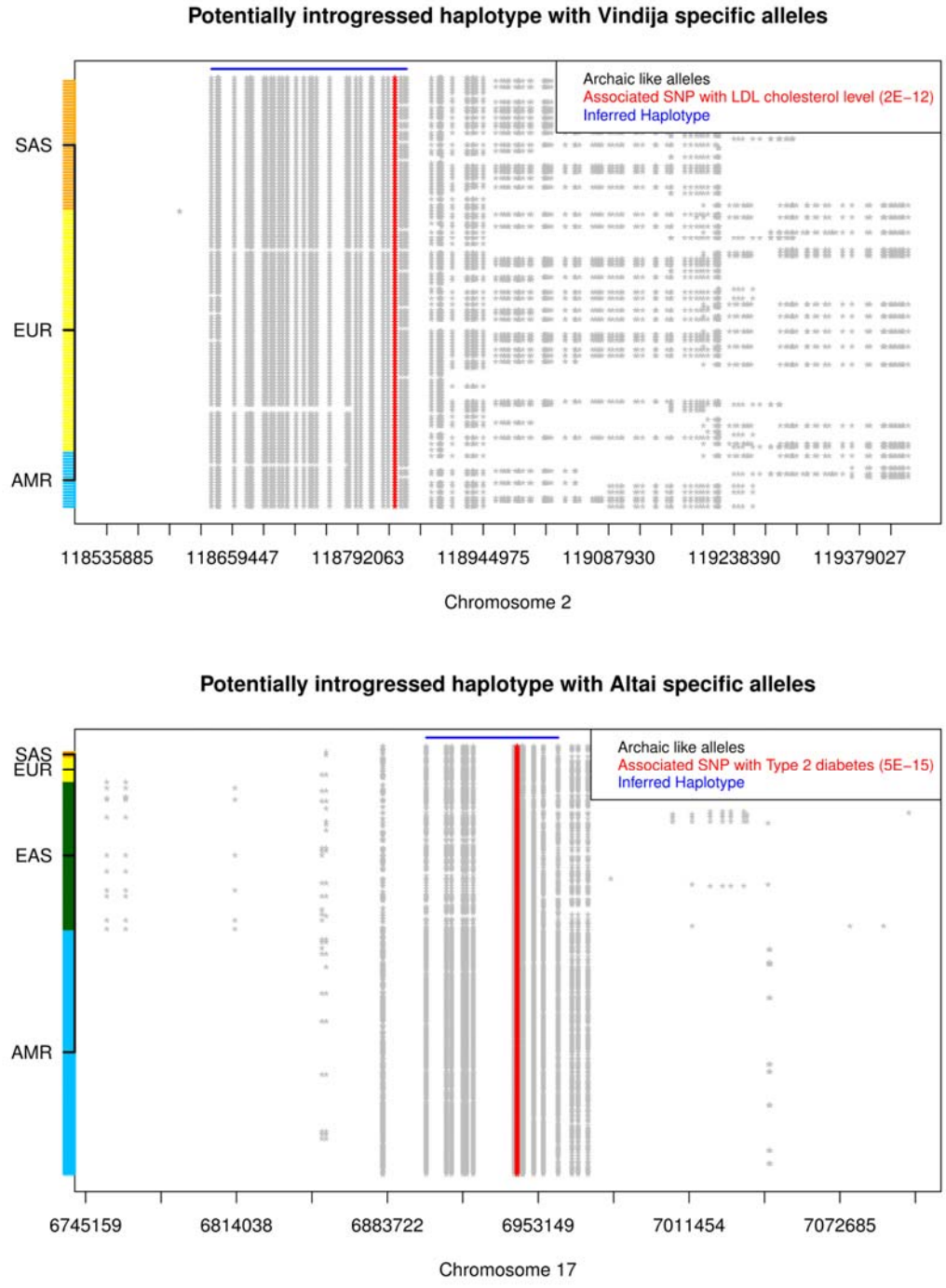

Fig. S102.

Haplotype patterns around a Vindija-specific shared allele associated with variation in LDL cholesterol levels (top) and an Altai-specific shared allele associated with Type 2 diabetes (bottom). Each line corresponds to an individual chromosome (EAS= East Asians; EUR=Europeans; AMR=Americans; SAS=South Asians). Grey stars represent archaic-like alleles (absent in Africa but shared between non-Africans and the Archaics). Both haplotypes are significantly longer than expected under incomplete lineage sorting (p = 0.0145 and 7.163e-13 for Vindija and Altai-specific haplotypes, respectively).

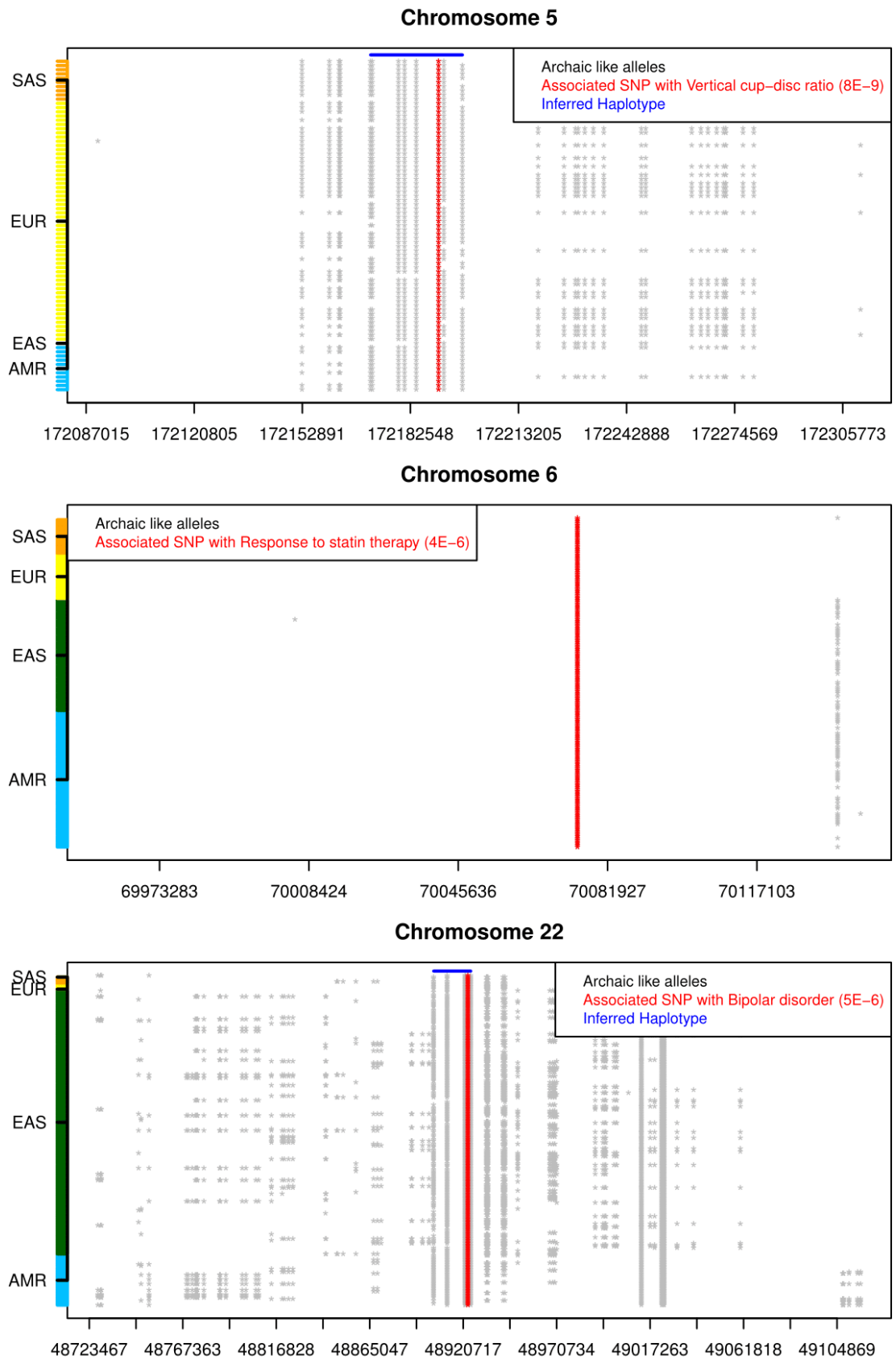

Fig. S103.
Haplotype patterns around three Denisovan-specific shared allele associated with human phenotypic variation. Focusing on the carriers of the GWAS variant, all consecutive sites with an archaic-like allele shared by at least 80% of the individuals were considered to belong to the haplotype. As Figure S102, each line corresponds to an individual chromosome (EAS= East Asians; EUR=Europeans; AMR=Americans; SAS=South Asians). Except for the variant on chromosome 6, that does not appear to be part of a haplotype, both variants on chromosome 5 and 22 belong to haplotypes significantly longer than expected under incomplete lineage sorting (p-value= 0.0325 and 2.786e-14 respectively).

Table S1.

Mitochondrial DNA coverage and contamination estimates determined for 19 bone fragments from Vindija Cave previously identified as Neandertal. Vindija 33.19 is shown in bold.

Sample	Mitochondrial genome coverage (unique sequences)	Present-day human contamination (%)
Vi 11.34	0.5	0.0
Vi 33.10	1.4	5.4
Vi 33.11	22.3	5.8
Vi 33.15	33.2	8.0
Vi 33.17	56.1	6.1
Vi 33.19	**52.1**	**1.7**
Vi 33.20	42.6	7.0
Vi 33.20	1.2	1.9
Vi 33.22	3.9	1.7
Vi 33.22	1.7	0.0
Vi 33.23	5.3	1.2
Vi 33.26	57.5	0.2
Vi 33.26	52.7	1.1
Vi 33.28	0.8	0.0
Vi 33.29	17.8	1.1
Vi 33.30	34.3	0.3
Vi 33.30	15.7	0.1
Vi 33.31	192.4	0.1
Vi 33.31	29.8	0.1
Vi 33.36	0.5	0.0
Vi 33.38	40.7	2.9
Vi 33.38	29.2	1.8
Vi 33.39	15.6	0.5
Vi 33.39	13.7	0.0
SP2688	4.5	0.5
SP2712	10.2	1.4

Table S2.
Summary of extracts and single-stranded libraries prepared from Vindija 33.19. The extract that was further used to prepare libraries for the high coverage genome is shown in bold.

Extract ID	mg bone used	Sequencing library ID	UDG treated	Fraction of sequences that align to hg19 (%)	Nuclear genome coverage (in whole library)
E1912	15	A5521	no	1.3	0.07
E1913	11	A5522	no	3.8	0.37
E1914	10	A5523	no	1.5	0.06
E1915	13	A5524	no	4.8	0.86
E2412	49	A2363	yes	2.2	1.56
E2413	60	A2364	yes	1.2	0.90
E2414	**41**	**A2365**	**yes**	**9.7**	**19.42**
E2415	34	A2366	yes	2.4	5.00
E2416	35	A2369	no	2.8	3.19
E2417	34	A2370	no	0.5	0.46
E2471	44	A2400	yes	1.8	1.55
E2472	35	A2401	yes	1.7	2.18

Table S3.

Samples, extracts and single stranded libraries prepared for the generation of a high coverage Vindija 33.19 Neandertal genome and a 1.4x coverage Mezmaiskaya 1 Neandertal genome.

Sample	Extract ID	Sequencing library ID	Extract used (in µL)	UDG treated	Gel fractionation
Vindija 33.19	E2414	A9368	2.5	no	no
Vindija 33.19	E2414	A9369	2.5	no	no
Vindija 33.19	E2414	A9401	2.5	no	no
Vindija 33.19	E2414	A9402	2.5	no	no
Vindija 33.19	E2414	A9403	2.5	no	no
Vindija 33.19	E2414	A9404	2.5	no	no
Vindija 33.19	E2414	B8744	15	yes	yes
Vindija 33.19	E2414	B8747	10	no	no
Vindija 33.19	E2414	R5473	10	no	no
Mezmaiskaya 1	E733	R5661	5	no	no
Mezmaiskaya 1	E733	R5662	5	no	no

Table S4.
Runs and lanes sequenced from Vindija 33.19 libraries. Runs marked in grey were base-called with Bustard.

Run-ID	Lane(s)	Library(-ies)
SN7001204_0448_BC7T9KACXX	1-4	A9368
SN7001204_0465_BHCHTCBCXX	1-2	A9368
SN7001204_0466_AH7CN5BCXX	1-2	A9368
SN7001204_0467_BH7CL5BCXX	1-2	A9368
SN7001204_0468_AHCCTVBCXX	1-2	A9368
SN7001204_0471_BC8CJUACXX	1-8	A9368,B8744,B8747,R5473
SN7001204_0475_BHG7TKBCXX	1-2	A9368,B8744,B8747,R5473
SN7001204_0448_BC7T9KACXX	5-8	A9369
SN7001204_0454_AC7RCNACXX	1-8	A9369
D00594_0037_AC820PANXX	1-2	A9401
SN7001204_0472_AC8FJAACXX	2	A9401
D00594_0037_AC820PANXX	3-4	A9402
SN7001204_0472_AC8FJAACXX	3-4	A9402
D00594_0037_AC820PANXX	5-6	A9403
SN7001204_0472_AC8FJAACXX	5-6	A9403
D00594_0037_AC820PANXX	7	A9404
SN7001204_0472_AC8FJAACXX	7-8	A9404
D00594_0038_AC824MANXX	1-8	A9401,A9402,A9403,A9404
D00594_0039_BC8277ANXX	1-8	A9401,A9402,A9403,A9404
SN7001204_0473_AHCFFGBCXX	1-2	A9401,A9402,A9403,A9404
SN7001204_0474_BH7CNLBCXX	1	A9401,A9402,A9403,A9404
SN7001204_0476_BH7HLHBCXX	1-2	A9401,A9402,A9403,A9404
SN7001204_0477_AHG7JCBCXX	1-2	A9401,A9402,A9403,A9404
SN7001204_0493_BHJJWNBCXX	2	A9401,A9402,A9403,A9404
SN7001204_0304_BHA41VADXX	1	B8744
SN7001204_0322_AHB2KMADXX	1	B8744
SN7001204_0325_AHB0GEADXX	2	B8744
SN7001204_0424_BC68ABACXX	1-8	B8744
SN7001204_0426_BC7CJVACXX	1-8	B8744
SN7001204_0434_BC5YKCACXX	1-8	B8744
SN7001204_0436_BC7R2LACXX	1-8	B8744
SN7001204_0458_AH77FGBCXX	1-2	B8744
SN7001204_0460_AH77H7BCXX	1-2	B8744
SN7001204_0461_BH7CMTBCXX	1-2	B8744

SN7001204_0464_AH7C3TBCXX	1-2	B8744
SN7001204_0402_AH3K3YBCXX	1-2	B8747
SN7001204_0405_AH5C2YBCXX	1	B8747
SN7001204_0413_AH57WMBCXX	1	B8747
SN7001204_0414_BH5C5LBCXX	1-2	B8747
SN7001204_0417_AH5HGHBCXX	1-2	B8747
SN7001204_0423_AC68KKACXX	1-8	B8747
SN7001204_0425_AC7M8KACXX	1-8	B8747
SN7001204_0433_AC6M7MACXX	1-8	B8747
SN7001204_0435_AC7RM7ACXX	1-8	B8747
SN7001204_0447_AC7VCDACXX	1-8	R5473
SN7001204_0455_BC8C89ACXX	1-8	R5473
SN7001204_0456_AC8C55ACXX	1-8	R5473
SN7001204_0457_BC8C53ACXX	3-8	R5473

Table S5.
Runs and lanes sequenced from Mezmaiskaya 1 libraries.

Run-ID	Lane(s)	Library
151218_SN7001204_0470_AC8D12ACXX	1-4	R5661
151218_SN7001204_0470_AC8D12ACXX	5-8	R5662

Table S6.
Coverage estimates per sequencing library.

Sample	Library	Coverage Autosomes	Coverage on Chromosome X
Vindija 33.19	A9368	1.85	1.87
	A9369	1.95	1.97
	A9401	1.42	1.41
	A9402	1.56	1.56
	A9403	1.44	1.42
	A9404	1.36	1.36
	B8744	7.08	7.23
	B8747	6.88	7.02
	R5473	6.47	6.54
	Sum:	30.01	30.38
Mezmaiskaya 1	R5661	0.77	0.81
	R5662	0.67	0.69
	Sum:	1.44	1.50

Table S7.
Groups of libraries/read groups considered separately for the calculation of error profiles. Reference gives the published VCF file used for comparison.

Sample	Reference	Library(-ies)/Read Group(s)	Library Prep.	UDG-treatment
Vindija 33.19	Altai Neandertal	B8744	single	yes
		A9368-9, A9401-4, B8747, R5453	single	no
Mezmaiskaya 1	Altai Neandertal	R5661-2	single	no
		L4533, L4677-8, L4740-1	double	yes
Altai Neandertal	Altai Neandertal	L9105	double	yes
		L9198-9, L9302-3	single	yes
Denisova	Denisova	SL3003-4	double	yes
		B1107-10, B1128, B1130, B1133	single	yes
Ust'Ishim	Ust'Ishim	B5347, B3899-B3907	single	yes
Loschbour	Loschbour	Loschbour	single	yes
		Loschbour1-4	double	yes

Table S8.

Estimated prior probabilities (×10e-5) for heterozygous genotypes for Vindija 33.19 (Vi.), Altai Neandertal (Al.), Denisovan (D.), Ust'Ishim (U.I.), Loschbour (L.) and Mezmaiskaya 1 (M.). Note: Ust'Ishim and Loschbour are males and calls on their haploid chromosomes X and Y are only listed for completeness.

	Genotypes																																	
	AC						AG						AT						CG						CT						GT					
	Vi.	Al.	D.	U.I.	L.	M.	Vi.	Al.	D.	U.I.	L.	M.	Vi.	Al.	D.	U.I.	L.	M.	Vi.	Al.	D.	U.I.	L.	M.	Vi.	Al.	D.	U.I.	L.	M.	Vi.	Al.	D.	U.I.	L.	M.
chr1	1.8	2.0	2.0	6.1	4.2	1.5	6.4	7.0	7.3	25.7	17.6	7.5	1.9	2.2	2.3	5.1	3.4	1.1	1.9	2.0	2.0	6.3	4.3	2.0	6.2	6.9	7.4	25.6	17.6	7.3	1.9	2.2	2.1	6.2	4.2	1.7
chr2	1.9	1.9	2.0	6.7	4.6	1.6	6.1	6.0	6.7	27.0	19.2	8.0	2.0	2.3	2.3	5.7	4.1	1.3	1.8	1.8	2.0	7.1	4.7	1.8	6.0	6.1	6.6	26.9	19.3	7.9	1.9	1.9	2.0	6.8	4.7	1.6
chr3	1.8	2.1	1.9	6.7	4.8	1.6	5.8	6.5	6.4	26.2	19.0	7.7	1.9	2.4	2.3	5.8	4.2	1.2	1.7	2.0	1.8	6.6	4.8	2.0	5.7	6.6	6.6	25.8	18.9	7.9	1.9	2.2	2.0	6.7	4.8	1.6
chr4	2.5	2.1	2.2	6.4	5.5	2.0	8.2	6.1	6.9	25.1	21.4	10.4	2.7	2.6	2.6	6.0	5.1	1.7	2.1	1.7	1.9	6.0	5.0	2.0	8.2	6.1	6.8	25.3	21.2	10.5	2.6	2.1	2.1	6.5	5.6	1.8
chr5	1.7	1.9	2.1	6.5	5.1	1.5	5.8	6.0	7.1	25.5	20.2	7.7	1.9	2.2	2.5	5.7	4.4	1.5	1.6	1.6	2.0	6.4	5.0	1.6	5.6	5.8	6.9	25.4	20.4	8.0	1.8	1.9	2.2	6.4	5.1	1.5
chr6	2.6	3.2	3.2	7.2	6.1	2.2	8.8	10.8	11.6	29.3	25.0	10.9	2.5	3.4	3.4	6.4	5.5	1.8	2.4	3.0	3.1	7.0	6.0	2.6	8.7	11.1	11.4	29.2	25.0	11.0	2.6	3.3	3.3	7.3	6.2	2.2
chr7	2.0	2.4	2.2	6.8	5.1	1.4	6.6	7.7	7.2	27.7	20.9	8.4	2.0	2.7	2.6	5.9	4.4	1.4	1.9	2.3	2.1	7.3	5.4	1.8	6.4	7.8	7.5	27.9	20.6	8.1	2.0	2.4	2.2	6.9	5.1	1.5
chr8	1.8	2.4	2.6	7.5	5.4	1.8	5.7	7.8	8.1	29.1	21.6	8.2	1.9	2.7	2.8	6.4	4.6	1.3	2.0	2.5	2.5	8.8	6.2	2.2	5.6	7.5	8.3	29.1	21.6	7.7	1.9	2.4	2.6	7.6	5.5	1.6
chr9	2.0	1.5	2.1	7.2	5.7	1.4	6.7	4.5	7.3	28.3	22.0	6.1	2.0	1.8	2.4	5.6	4.5	1.0	2.2	1.6	2.2	8.3	6.5	1.8	6.3	4.4	7.4	28.6	22.2	7.3	2.1	1.6	2.2	7.1	5.5	1.2
chr10	2.3	2.2	2.2	7.8	5.1	1.6	7.7	7.1	7.3	32.0	21.0	7.7	2.3	2.3	2.4	6.5	4.1	1.3	2.3	2.2	2.1	7.8	5.1	1.5	7.6	7.1	7.6	32.0	21.1	8.7	2.4	2.3	2.1	7.9	5.0	1.7
chr11	1.9	2.3	2.3	6.7	5.0	1.3	6.1	7.1	8.1	27.1	20.8	6.1	1.9	2.5	2.6	5.5	4.2	1.0	1.8	2.1	2.3	7.1	5.2	2.0	6.0	7.2	8.0	27.2	20.6	6.0	2.0	2.2	2.3	6.8	5.0	1.5
chr12	1.6	1.7	1.9	6.4	4.8	1.3	5.3	5.5	6.2	26.4	20.9	7.1	1.7	2.1	2.2	5.4	4.1	1.1	1.5	1.6	1.8	6.5	5.0	1.5	5.3	5.4	6.2	25.7	20.3	6.8	1.7	1.8	1.8	6.3	4.8	1.4
chr13	2.0	2.1	2.2	7.3	4.9	1.8	6.8	7.1	7.6	29.5	19.4	9.3	2.3	2.5	2.7	6.9	4.5	1.5	1.8	2.0	2.0	6.7	4.4	2.2	6.5	7.2	7.4	28.9	18.6	9.1	2.1	2.2	2.3	7.3	4.7	1.7
chr14	1.7	1.3	2.1	6.1	4.6	1.3	5.4	4.0	7.2	25.6	19.6	6.3	1.9	1.8	2.4	5.2	4.0	1.0	1.7	1.2	2.0	6.8	4.9	1.6	5.6	3.9	7.1	25.6	19.7	6.2	1.7	1.4	2.1	6.1	4.7	1.3
chr15	1.7	1.6	1.8	6.9	5.3	1.0	5.6	4.7	6.8	27.5	22.5	6.4	1.8	1.8	2.1	5.4	4.4	1.0	1.7	1.6	2.0	7.9	6.4	1.4	5.3	4.8	6.6	27.7	22.6	6.2	1.6	1.6	2.0	6.7	5.4	1.1
chr16	2.1	2.3	2.4	6.7	6.3	1.0	7.0	7.1	8.9	27.1	25.6	5.8	1.9	2.2	2.1	5.0	4.9	1.0	2.9	2.9	3.2	10.3	8.8	2.2	7.0	6.8	8.3	26.2	24.7	5.9	2.2	2.2	2.3	6.8	6.3	1.2
chr17	1.6	1.7	2.0	6.0	4.9	0.8	6.2	5.3	7.5	27.6	23.0	4.9	1.5	1.7	1.9	4.4	3.6	0.7	1.9	1.9	2.3	7.1	5.7	1.0	5.8	5.4	7.5	26.7	22.3	5.2	1.7	1.8	2.1	6.0	4.9	1.3
chr18	2.0	2.2	2.2	6.9	4.6	1.2	6.6	7.6	7.0	28.8	20.0	6.9	2.1	2.6	2.6	6.1	4.1	1.5	1.8	2.1	1.9	6.7	4.5	1.2	6.6	7.3	6.9	28.1	19.6	7.9	2.0	2.4	2.1	6.9	4.9	1.3
chr19	2.3	2.1	2.2	6.5	4.7	1.8	7.7	6.9	8.1	30.7	21.9	6.6	2.1	2.1	2.0	4.8	3.5	1.1	2.8	2.5	2.7	8.7	6.1	2.9	7.4	6.9	7.8	30.3	21.9	5.9	2.4	2.2	2.3	6.7	4.7	1.2
chr20	2.0	2.5	2.2	6.4	5.5	1.0	6.7	8.5	8.5	28.9	24.3	4.5	1.9	2.4	2.2	5.0	4.0	0.9	2.1	2.6	2.4	7.3	5.9	1.0	6.5	8.3	8.2	27.9	23.7	4.0	1.9	2.5	2.3	6.6	5.6	1.1
chr21	2.6	1.8	2.6	8.8	6.0	2.4	8.7	5.5	8.4	38.2	26.0	8.6	2.6	2.4	3.0	7.7	5.4	1.3	2.3	1.8	2.4	9.1	6.1	1.5	8.4	5.2	8.3	38.4	25.6	8.2	2.8	2.0	2.8	9.1	6.0	1.5
chr22	2.1	2.0	2.6	7.0	5.2	1.2	7.3	7.8	10.0	34.9	24.7	7.5	1.8	2.0	2.1	4.5	3.2	0.5	2.4	2.6	3.0	9.2	6.8	2.2	7.3	7.5	9.6	33.8	24.2	6.1	2.0	2.3	2.4	6.9	4.7	1.6
chrX	1.2	1.3	1.3	0.3	0.3	0.9	3.4	3.7	4.1	1.5	1.2	4.8	1.3	1.6	1.8	0.4	0.3	1.0	1.0	1.1	1.3	0.4	0.4	1.2	3.6	3.8	4.3	1.4	1.2	5.6	1.2	1.3	1.4	0.3	0.3	1.1
chrY				1.6	1.3					3.4	3.2					1.7	1.3					1.3	1.1					3.4	3.0					1.2	0.8	

Table S9.

SNPs called within the longest inbred (chr21:17081807-35881807) and outside of this inbred region on chromosome 21 in Altai Neandertal. *GATK* calls used decoy aligned Altai Neandertal data to match the input alignments to those used in the *snpAD* calls. Fisher exact test was used to calculate p-values for the four blue and four green values.

Filters	Region	GATK	snpAD	Fisher Exact Test
Mapability	Inbred	1092	701	p-value = 3.2e-09
	Non-inbred	3962	3489	
Mapability, coverage, simple repeats, indels	Inbred	206	101	p-value = 3.5e-08
	Non-inbred	2431	2339	

Table S10.
Mitochondrial contamination estimates per sequencing library. Lower and upper CI give the binomial confidence interval. Last row shows estimates for Mezmaiskaya 1 including previous sequencing data.

Sample	Library	Sequences matching modern human state	Sequences matching Neandertal state	%Contam.	%lower CI	%upper CI
Vindija 33.19	A9368	58	2327	2.43%	1.85%	3.13%
	A9369	40	2450	1.61%	1.15%	2.18%
	A9401	19	1727	1.09%	0.66%	1.69%
	A9402	34	1917	1.74%	1.21%	2.43%
	A9403	29	1813	1.57%	1.06%	2.25%
	A9404	21	1750	1.19%	0.74%	1.81%
	B8744	139	8024	1.70%	1.43%	2.01%
	B8747	112	7578	1.46%	1.20%	1.75%
	R5473	112	7544	1.46%	1.21%	1.76%
	All lib. combined	564	35130	1.58%	1.45%	1.71%
Mezmaiskaya 1	R5661	43	2116	1.99%	1.45%	2.67%
	R5662	44	2099	2.05%	1.50%	2.75%
	R5661+R5662	87	4215	2.02%	1.62%	2.49%
	R5661+R5662+old data	152	10024	1.49%	1.27%	1.75%

Table S11.
Male contamination estimates per sequencing library.

Sample	Library	Coverage Autosomes	Coverage chr. Y	Y chr. Contamination Est.
Vindija 33.19	A9368	1.84	0.006	0.70%
	A9369	1.95	0.007	0.76%
	A9401	1.42	0.005	0.73%
	A9402	1.56	0.006	0.78%
	A9403	1.44	0.005	0.71%
	A9404	1.36	0.005	0.70%
	B8744	7.07	0.027	0.77%
	B8747	6.87	0.024	0.71%
	R5473	6.46	0.024	0.74%
	All lib. combined	29.97	0.111	0.74%
Mezmaiskaya 1	R5661	0.77	0.008	2.18%
	R5662	0.66	0.008	2.33%
	R5661+R5662	1.44	0.016	2.25%
	R5661+R5662+old data	1.92	0.021	2.15%

Table S12.
Autosomal contamination rate estimates using a maximum likelihood method.

Sample	Library	Point Est	lower CI	upper CI
Vindija 33.19	A9368	0.49%	0.17%	0.83%
	A9369	0.05%	0.00%	0.34%
	A9401	0.25%	0.00%	0.75%
	A9402	0.22%	0.00%	0.62%
	A9403	0.00%	0.00%	0.34%
	A9404	0.23%	0.00%	0.80%
	B8744	0.20%	0.16%	0.25%
	B8747	0.19%	0.14%	0.25%
	R5473	0.27%	0.20%	0.33%
	All libraries combined	0.21%	0.18%	0.23%
Mezmaiskaya 1	R5661 + R5662	2.08%	0.77%	2.70%
	R5661 + R5662 + old data	2.13%	1.69%	2.60%

Table S13.
Autosomal contamination estimates in percentage using divergent sites between humans and Denisovan + Neandertals. We calculated the 95% CI from the binomial random sampling. Nr and Ns represent the total number of reads and total number of informative sites, respectively.

Sample	Library	Point Est	Lower CI	Upper CI	Nr	Ns
Vindija 33.19	A9368	0.59	0.07	2.11	339	201
	A9369	0.28	0.01	1.56	354	220
	A9401	0.74	0.09	2.65	270	180
	A9402	1.52	0.42	3.85	263	182
	A9403	2.09	0.68	4.81	239	174
	A9404	0.88	0.11	3.16	226	161
	B8744	0.36	0.12	0.84	1384	341
	B8747	0.60	0.24	1.24	1162	343
	R5473	0.42	0.14	0.97	1200	350
	All libraries	0.29	0.17	0.48	5459	363
Mezmaiskaya 1	R5661 + R5662	3.40	1.38	6.88	206	141
	old data	3.94	1.29	8.95	127	104
	R5661+R5662+old	3.30	1.66	5.83	333	201

Table S14.

Average autosomal heterozygosity per 10,000 sites with and without HBD tracts longer than 2.5 cM. We also reported the ratio of heterozygosity between each sample and Vindija, as well as the scan π parameter for the detection of HBD tracts. Figure S15 shows the distribution of heterozygosity given by the 22 autosomes.

SAMPLE	Het	Het. Ratio Sample / Vindija	Het1 (no HBD)	Het1. Ratio Sample / Vindija	scan π
Vindija33.19	1.62	1.0	1.78	1.0	0.950
Altai	1.58	1.0	1.99	1.1	0.950
Denisova	1.83	1.1	1.89	1.1	0.990
S_Karitiana-2	4.97	3.1	5.71	3.2	0.800
S_Pima-2	5.35	3.3	5.80	3.3	0.875
S_Quechua-1	5.93	3.7	5.98	3.4	0.875
S_Papuan-13	5.66	3.5	5.69	3.2	0.975
S_Dusun-1	6.14	3.8	6.25	3.5	0.990
S_Bougainville-2	6.17	3.8	6.19	3.5	0.950
S_Dai-1	6.51	4.0	6.53	3.7	0.825
S_Han-1	6.56	4.1	6.56	3.7	0.800
S_Japanese-2	6.59	4.1	6.60	3.7	0.800
S_French-2	6.98	4.3	6.99	3.9	0.875
S_Finnish-1	7.02	4.3	7.03	3.9	0.800
S_Sardinian-2	7.03	4.4	7.06	4.0	0.800
S_Khomani-San-1	8.81	5.5	9.10	5.1	0.990
S_Mbuti-2	9.25	5.7	9.38	5.3	0.990
S_Yoruba-1	9.42	5.8	9.43	5.3	0.825

Table S15.
Average autosomal pairwise nucleotide differences per 1,000 sites between the genomes of the Denisovan, Altai and Vindija.

	Vindija	Altai
Denisova	1.181	1.146
Altai	0.339	

Table S16.
State of Vindija 33.15 and Vindija 33.19 heterozygous calls in other archaic genomes.

Sample1	Sample2	heterozygous	homozygous	%heterozygous
Vi 33.15	Altai	294	1593	15.58%
	Denisova	88	3216	2.66%
	Sidron	625	1277	32.86%
	Vi33.19	3631	37	98.99%
Vi 33.19	Altai	332	1866	15.10%
	Denisova	97	3803	2.49%
	Sidron	604	1285	31.97%
	Vi33.15	3631	41	98.88%

Table S17.

Matching of Vindija 33.19 heterozygous sites in low-coverage Neandertal data. A match corresponds to two randomly sampled sequences showing both alleles. Columns under "only transversions" exclude heterozygous sites that are A/G or C/T.

	All Sites			Only transversions		
Sample	**match**	**no match**	**%match**	**match**	**no match**	**%match**
Mez. 1	25294	131923	16.1%	7322	45328	13.9%
Vi 33.16	2916	13550	17.7%	829	4507	15.5%
Vi 33.25	2756	10625	20.6%	849	3591	19.1%
Vi 33.26	2272	9124	19.9%	703	3137	18.3%
Vi 33.19-sample	8677	9482	47.8%	2742	3304	45.4%

Table S18.
Heterozygosity for different archaic genomes in regions retained by the mappability filter map35_100 and regions unique of mappability filter map35_50.

	map35_50-map35_100	map35_100
Vindija33.19	0,00019	0,000178
Altai	0,000252	0,000173
Denisova	0,000288	0,000199

Table S19.
Branch shortening compared to a modern African genome in kya. Radiocarbon column gives reported calibrated radiocarbon dates (Vindija date: OxA-32278, see S1).

(%H-C div x 10^-3) kya	all		GQ>60		Intergenic		transversions		Radiocarbon
Vindija33.19	(4.4)	56.7±6.6	(5.0)	64.7±6.4	(4.5)	58.1±8.7	(4.0)	51.8±12	>45.5
Altai	(9.5)	123±6.4	(9.8)	127.2±6.2	(9.4)	122.4±8.7	(9.4)	122.4±12.4	
Denisova	(6.5)	84.5±7.9	(6.9)	90.1±7.7	(6.9)	90.1±9.6	(5.5)	72±12	
Ust'-Ishim	(2.9)	37.6±5.7	(3.7)	48.6±5.5	(4.0)	51.4±7.9	(3.3)	43.4±9.8	45.0±1.9 (*44*)
Loschbour	(0.0)	-1.3±6	(1.2)	15.1±6	(1.3)	17.4±8.2	(1.0)	12.6±10.2	8.2-7.9 (*45*)
Stuttgart (LBK)	(-3.0)	-38.6±6	(1.1)	14±6.1	(1.3)	17.1±7.9	(0.8)	10.1±10.1	~7000 (*45*)

Table S20.

Split times between archaic hominins and an African genome (Mbuti). Human-chimpanzee divergence (HC div) is assumed to be 13 million years. Columns with "+bs" give estimates after correcting for branch shortening based on transversions from Table S19.

A-B	F(A|B) %	% HC div (A-B)	% HC div (A-B+bs)	split A-B	split+bs A-B
	F(B|A) %	% HC div (B-A)	% HC div (B-A+bs)	split B-A	split+bs B-A
Altai-Vindija	27.1	0.6	1.0	78.80	130.5
	35.9	0.2	1.1	22.10	144.5
Denisova-Vindija	12.3	2.9	3.3	378.60	430.4
	12.9	2.5	3.1	327.50	399.4
Altai-Denisova	13.2	2.5	3.0	320.10	392.1
	16.4	2.4	3.4	315.10	437.5
Vindija-Mbuti	17.6	4.0	4.0	521.60	521.6
	10.2	3.7	4.1	483.00	534.7
Altai-Mbuti	17.6	4.0	4.0	525.20	525.2
	13.1	3.8	4.7	489.20	611.7
Denisova-Mbuti	17.2	4.2	4.2	546.40	546.4
	8.3	4.3	4.9	562.00	634.0

Table S21.

Split times between Mezmaskaya 1 and other individuals. Human-chimpanzee divergence (HC div) is assumed to be 13 million years. Columns with "+bs" give estimates after correcting for branch shortening based on transversions from Table S19.

Population B	Mez1 treatment	F(A|B)%	% HC div	%HC div+bs	split (ky)	split+bs (kya)
Vindija33.19	genotypes	31.9	0.2	0.6	30.8	82.5
Altai	genotypes	35.4	0.3	1.2	34.2	156.7
Denisova	genotypes	12.7	2.5	3.1	331.4	403.4
Mbuti	genotypes	17.8	3.9	3.9	512.1	512.1
Vindija33.19	random read	31.9	0.2	0.6	43.0	82.5
Altai	random read	35.4	0.3	1.2	33.9	156.3
Denisova	random read	12.8	2.5	3.1	330.9	402.8
Mbuti	random read	17.9	3.9	3.9	507.7	507.7
Vindija33.19	random deaminated read	30.5	0.4	0.8	47.2	99
Altai	random deaminated read	33.9	0.3	1.3	41.6	164.1
Denisova	random deaminated read	12.5	4.2	3.1	549.3	409.1
Mbuti	random deaminated read	17.2	2.6	4.2	337.1	549.3

Table S22.

Split times between archaic hominins and ancient Eurasian sapiens. Columns with "+bs" give estimates after correcting for branch shortening based on transversions from Table S19.

A-B	F(A|B) %	% HC div (A-B)	% HC div (A-B+bs)	split A-B	split+bs A-B
	F(B|A) %	% HC div (B-A)	% HC div (B-A+bs)	split B-A	split+bs B-A
Ust-Ishim-Vindija	10.8	3.2	3.6	413.30	465.1
	20.5	3.6	3.9	461.70	505.1
Ust-Ishim-Altai	13.5	3.6	4.5	464.90	587.3
	20.2	3.7	4.0	478.50	521.9
Ust-Ishim-Denisova	8.5	4.2	4.8	548.40	620.4
	18.4	4.6	4.9	599.40	642.8
Ust-Ishim-Mbuti	27.6	0.6	0.6	77.40	77.4
	28.2	0.7	1.0	92.60	136.1
Loschbour-Vindija	10.7	3.2	3.6	420.60	472.4
	19.2	3.7	3.8	475.40	488.1
Loschbour-Altai	13.5	3.6	4.5	466.00	588.5
	19.0	3.8	3.9	491.10	503.7
Loschbour-Denisova	8.4	4.3	4.8	554.70	626.6
	17.4	4.6	4.7	597.00	609.7
Loschbour-Mbuti	27.7	0.6	0.6	76.00	76.0
	26.6	1.0	1.1	131.50	144.1

Table S23.
Archaic-allele sharing differences among Africans.

Comparison	D%	|Z|
D(San,Yoruba,Vindija,chimp)	-0.53%	1.13
D(Mbuti,San,Vindija,chimp)	-0.35%	0.74
D(Mbuti,Yoruba,Vindija,chimp)	-0.92%	2.13
D(San,Yoruba,Altai,chimp)	-0.37%	0.83
D(Mbuti,San,Altai,chimp)	-0.43%	0.98
D(Mbuti,Yoruba,Altai,chimp)	-0.83%	2.10
D(San,Yoruba,Denisova,chimp)	-0.54%	1.32
D(Mbuti,San,Denisova,chimp)	-0.25%	0.68
D(Mbuti,Yoruba,Denisova,chimp)	-0.82%	2.10

Table S24.
Archaic-allele sharing differences between Africans and non-Africans. Column header gives the third individual in the comparison; chimpanzee was used as outgroup.

	Vindija		Altai		Denisova	
Comparison	**D%**	**|Z|**	**D%**	**|Z|**	**D%**	**|Z|**
Yoruba,French	-5.11%	8.5	-4.60%	8.2	-1.60%	2.8
San,French	-5.18%	7.7	-4.56%	7.6	-2.00%	3.8
Mbuti,French	-5.72%	9.5	-5.16%	9.2	-2.32%	4.8
San,Han	-6.95%	11.1	-6.14%	10.5	-3.09%	7.0
Yoruba,Han	-7.06%	14.2	-6.34%	12.9	-2.80%	5.8
San,Papuan	-7.52%	11.1	-6.83%	11.1	-7.17%	12.8
Mbuti,Han	-7.52%	14.3	-6.75%	13.2	-3.44%	8.1
Yoruba,Papuan	-7.72%	11.8	-7.12%	11.3	-7.32%	12.5
Mbuti,Papuan	-8.14%	13.4	-7.50%	12.4	-7.70%	14.3

Table S25.
African-allele sharing differences among the archaic humans.

Comparison	D%	|Z|
D(Vindija,Altai,Yoruba,chimp)	0.52%	0.48
D(Vindija,Altai,Mbuti,chimp)	0.06%	0.06
D(Vindija,Altai,San,chimp)	-0.33%	0.31
D(Altai,Denisova,Yoruba,chimp)	6.46%	11.66
D(Altai,Denisova,Mbuti,chimp)	6.43%	11.77
D(Altai,Denisova,San,chimp)	6.61%	12.59
D(Vindija,Denisova,Yoruba,chimp)	6.47%	12.70
D(Vindija,Denisova,Mbuti,chimp)	6.37%	11.79
D(Vindija,Denisova,San,chimp)	6.47%	12.09

Table S26.
African-allele sharing differences among the archaic humans on chromosome 21.

Comparison	D%	|Z|
D(Sidron1253,Altai,Yoruba,chimp)	4.12%	0.62
D(Vindija33.15,Altai,Yoruba,chimp)	-2.11%	0.29
D(Sidron1253,Altai,San,chimp)	-2.51%	0.27
D(Vindija33.19,Altai,Yoruba,chimp)	-2.96%	0.40
D(Sidron1253,Altai,Mbuti,chimp)	-4.63%	0.82
D(Vindija33.15,Altai,Mbuti,chimp)	-6.08%	0.86
D(Vindija33.19,Altai,Mbuti,chimp)	-6.79%	0.96
D(Vindija33.15,Altai,San,chimp)	-7.07%	0.68
D(Vindija33.19,Altai,San,chimp)	-9.61%	1.04

Table S27.
Parameters for simulating different admixture scenarios.

Scenario	Parameter	Simulated values
Neandertal to modern human	Time of Admixture (kya)	40,45,50,55,60,65,70,75
	Percentage (Yorubans) (%)	0, 0.02, 0.04, 0.06, 0.08, 0.1
	Percentage (CEPH Eur.) (%)	0.75, 1, 1.25, 1.5, 1.75, 2. 2.25, 2.5, 2.75, 3
Modern human to Neandertal	Time of Admixture (kya)	150,170,200,250,300,350,400
	Percentage (%)	0,3,5,7,10,15
Superarchaic to Denisova	Split time superarchaic (kya)	750,1000,1500,2000
	Percentage Admixture (%)	0,1,2,3,4,5,6,8,10

Table S28.
Estimated parameters for the YRI and CEU allele-frequency stratified D(Altai,Denisova,YRI/CEU, Outgroup). Ranges were omitted when the best estimates did not vary after rounding. Abbreviations: N=Neandertal, MH=Modern human, Den=Denisova, *T*=admixture time, α =admixture percentage, Div_{SA}=superarchaic population split time.

Pop	*LL*	$T_{N\rightarrow MH}$ (kya)	$\alpha_{N\rightarrow MH}$ (%)	$T_{MH\rightarrow N}$ (kya)	$\alpha_{MH\rightarrow N}$ (%)	Div_{SA} (Mya)	$\alpha_{SA\rightarrow Den}$ (%)
YRI	-1026	43 (40-43)	0.015 (0.014-0.016)	329 (328-329)	5.0	0.91	6
	-1033	43 (43-46)	0.017 (0.017-0.019)	-	-	1.0	6
	-1089	42 (40-46)	0.016 (0.015-0.017)	306 (305-308)	15	-	-
CEU	-1107	49	1.0 (1.0-1.1)	150 (150-400)	4.5 (4.1-7.3)	2.0 (1.4-2.0)	1.8 (1.6-4.4)
	-1131	49	1.1	-	-	1.5	4.9
	-1416	49	1.0	242 (241-242)	15	-	-

Table S29.
Estimated parameters to fit stratified D(Vindija,Altai,YRI,Outgroup).

Scenario	*LL*	$T_{N\rightarrow MH}$ (kya)	$\alpha_{N\rightarrow MH}$ (%)	$T_{MH\rightarrow N}$ (kya)	$\alpha_{MH\rightarrow N}$ (%)
Vindija→MH, MH→Vindija	-827	48	0.017	120	0.11
Vindija→MH	-833	48	0.017	-	-

Table S30.
Denisova-derived allele-sharing compared between Vindija and Altai.

Comparison	**D%**	**|Z|**
D(Vindija,Altai,Denisova,chimp)	-6.2%	5.2
D(Vindija,Altai,Denisova,bonobo)	-6.3%	5.3
D(Vindija,Altai,Denisova,gorilla)	-6.0%	4.9
D(Vindija,Altai,Denisova,orang)	-6.9%	6.5
D(Vindija,Altai,Denisova,rhesus)	-7.4%	7.5

Table S31.
Relationship of the Altai and Vindija Neandertals to the Mezmaiskaya 1 individual

Comparison				D%	|Z|
Altai	Vindija33.19	Mez1_snpAD	San	-33.39	-37.17
Altai	Vindija33.19	Mez1_snpAD	Mbuti	-33.56	-37.90
Altai	Vindija33.19	Mez1_snpAD	Yoruba	-33.31	-37.34
Altai	Vindija33.19	Mez1_snpAD	Chimp	-33.96	-39.18
Altai	Vindija33.19	Mez1_snpAD	Bonobo	-34.14	-39.82
Altai	Vindija33.19	Mez1_snpAD	Gorilla	-33.93	-38.89
Altai	Vindija33.19	Mez1_snpAD	Orangutan	-33.40	-38.28
Altai	Vindija33.19	Mez1_snpAD	Rhesus	-33.21	-38.66
Altai	Vindija33.19	Mez1_random	San	-33.87	-37.21
Altai	Vindija33.19	Mez1_random	Mbuti	-34.11	-38.18
Altai	Vindija33.19	Mez1_random	Yoruba	-33.86	-37.43
Altai	Vindija33.19	Mez1_random	Chimp	-34.44	-39.47
Altai	Vindija33.19	Mez1_random	Bonobo	-34.55	-39.96
Altai	Vindija33.19	Mez1_random	Gorilla	-34.37	-39.07
Altai	Vindija33.19	Mez1_random	Orangutan	-33.83	-38.38
Altai	Vindija33.19	Mez1_random	Rhesus	-33.67	-39.13
Altai	Vindija33.19	Mez1_random_deam	San	-35.27	-33.47
Altai	Vindija33.19	Mez1_random_deam	Mbuti	-35.47	-33.77
Altai	Vindija33.19	Mez1_random_deam	Yoruba	-35.32	-33.75
Altai	Vindija33.19	Mez1_random_deam	Chimp	-35.29	-34.03
Altai	Vindija33.19	Mez1_random_deam	Bonobo	-35.47	-34.22
Altai	Vindija33.19	Mez1_random_deam	Gorilla	-35.01	-33.13
Altai	Vindija33.19	Mez1_random_deam	Orangutan	-34.80	-32.47
Altai	Vindija33.19	Mez1_random_deam	Rhesus	-34.56	-32.56

Table S32.
Africans form a clade compared to Neandertals

Comparison				D%	|Z|
Altai	Mez1_snpAD	San	Yoruba	2.76	2.77
Altai	Mez1_snpAD	San	Mbuti	-0.54	-0.73
Altai	Mez1_snpAD	Yoruba	Mbuti	-3.42	-3.47
Vindija33.19	Mez1_snpAD	San	Yoruba	3.21	3.63
Vindija33.19	Mez1_snpAD	San	Mbuti	0.30	0.38
Vindija33.19	Mez1_snpAD	Yoruba	Mbuti	-3.03	-3.97
Altai	Mez1_random	San	Yoruba	3.28	3.47
Altai	Mez1_random	San	Mbuti	-0.17	-0.24
Altai	Mez1_random	Yoruba	Mbuti	-3.59	-3.88
Vindija33.19	Mez1_random	San	Yoruba	4.14	5.06
Vindija33.19	Mez1_random	San	Mbuti	0.96	1.27
Vindija33.19	Mez1_random	Yoruba	Mbuti	-3.35	-4.66
Altai	Mez1_random_deam	San	Yoruba	0.61	0.50
Altai	Mez1_random_deam	San	Mbuti	-1.67	-1.63
Altai	Mez1_random_deam	Yoruba	Mbuti	-2.38	-2.15
Vindija33.19	Mez1_random_deam	San	Yoruba	0.60	0.47
Vindija33.19	Mez1_random_deam	San	Mbuti	-1.33	-1.13
Vindija33.19	Mez1_random_deam	Yoruba	Mbuti	-2.00	-1.77

Table S33.
The proximity of Altai and Mezmaiskaya 1 to the introgressed Neandertal(s) in modern humans

Comparison				D%	|Z|
Altai	Mez1_snpAD	French	San	-11.04	-13.38
Altai	Mez1_snpAD	French	Mbuti	-11.87	-14.85
Altai	Mez1_snpAD	French	Yoruba	-9.27	-9.95
Altai	Mez1_snpAD	Sardinian	San	-9.69	-11.33
Altai	Mez1_snpAD	Sardinian	Mbuti	-10.47	-13.36
Altai	Mez1_snpAD	Sardinian	Yoruba	-7.83	-7.56
Altai	Mez1_snpAD	Han	San	-10.67	-11.44
Altai	Mez1_snpAD	Han	Mbuti	-11.51	-12.38
Altai	Mez1_snpAD	Han	Yoruba	-8.93	-10.47
Altai	Mez1_snpAD	Dai	San	-9.89	-11.40
Altai	Mez1_snpAD	Dai	Mbuti	-10.73	-13.69
Altai	Mez1_snpAD	Dai	Yoruba	-8.09	-7.44
Altai	Mez1_snpAD	Karitiana	San	-10.46	-10.89
Altai	Mez1_snpAD	Karitiana	Mbuti	-11.42	-12.61
Altai	Mez1_snpAD	Karitiana	Yoruba	-8.71	-7.10
Altai	Mez1_snpAD	Mixe	San	-11.24	-11.78
Altai	Mez1_snpAD	Mixe	Mbuti	-12.22	-13.83
Altai	Mez1_snpAD	Mixe	Yoruba	-9.59	-9.01
Altai	Mez1_snpAD	Australian	San	-8.05	-8.97
Altai	Mez1_snpAD	Australian	Mbuti	-8.78	-10.39
Altai	Mez1_snpAD	Australian	Yoruba	-6.02	-6.31
Altai	Mez1_snpAD	Papuan	San	-8.01	-9.30
Altai	Mez1_snpAD	Papuan	Mbuti	-8.76	-11.62
Altai	Mez1_snpAD	Papuan	Yoruba	-6.02	-5.98
Altai	Mez1_random	French	San	-12.58	-15.87
Altai	Mez1_random	French	Mbuti	-13.02	-17.27
Altai	Mez1_random	French	Yoruba	-10.33	-11.41
Altai	Mez1_random	Sardinian	San	-11.44	-13.79
Altai	Mez1_random	Sardinian	Mbuti	-11.86	-15.90
Altai	Mez1_random	Sardinian	Yoruba	-9.14	-9.16
Altai	Mez1_random	Han	San	-12.58	-14.21
Altai	Mez1_random	Han	Mbuti	-13.08	-14.97
Altai	Mez1_random	Han	Yoruba	-10.44	-12.53
Altai	Mez1_random	Dai	San	-11.95	-13.97
Altai	Mez1_random	Dai	Mbuti	-12.44	-16.16
Altai	Mez1_random	Dai	Yoruba	-9.79	-9.48

Altai	Mez1_random	Karitiana	San	-12.44	-13.13
Altai	Mez1_random	Karitiana	Mbuti	-13.05	-15.04
Altai	Mez1_random	Karitiana	Yoruba	-10.30	-8.83
Altai	Mez1_random	Mixe	San	-13.06	-14.29
Altai	Mez1_random	Mixe	Mbuti	-13.67	-16.48
Altai	Mez1_random	Mixe	Yoruba	-10.99	-10.95
Altai	Mez1_random	Australian	San	-9.86	-11.40
Altai	Mez1_random	Australian	Mbuti	-10.27	-12.67
Altai	Mez1_random	Australian	Yoruba	-7.46	-8.20
Altai	Mez1_random	Papuan	San	-9.87	-11.68
Altai	Mez1_random	Papuan	Mbuti	-10.29	-14.03
Altai	Mez1_random	Papuan	Yoruba	-7.48	-7.70
Altai	Mez1_random_deam	French	San	-4.71	-3.75
Altai	Mez1_random_deam	French	Mbuti	-6.56	-5.42
Altai	Mez1_random_deam	French	Yoruba	-4.57	-3.66
Altai	Mez1_random_deam	Sardinian	San	-2.33	-1.92
Altai	Mez1_random_deam	Sardinian	Mbuti	-4.13	-3.62
Altai	Mez1_random_deam	Sardinian	Yoruba	-2.01	-1.68
Altai	Mez1_random_deam	Han	San	-3.45	-2.62
Altai	Mez1_random_deam	Han	Mbuti	-5.21	-4.03
Altai	Mez1_random_deam	Han	Yoruba	-3.19	-2.69
Altai	Mez1_random_deam	Dai	San	-3.12	-2.65
Altai	Mez1_random_deam	Dai	Mbuti	-4.84	-4.25
Altai	Mez1_random_deam	Dai	Yoruba	-2.78	-2.34
Altai	Mez1_random_deam	Karitiana	San	-3.97	-2.99
Altai	Mez1_random_deam	Karitiana	Mbuti	-5.83	-4.37
Altai	Mez1_random_deam	Karitiana	Yoruba	-3.75	-2.57
Altai	Mez1_random_deam	Mixe	San	-5.03	-3.86
Altai	Mez1_random_deam	Mixe	Mbuti	-6.83	-5.36
Altai	Mez1_random_deam	Mixe	Yoruba	-4.87	-3.70
Altai	Mez1_random_deam	Australian	San	-2.79	-2.26
Altai	Mez1_random_deam	Australian	Mbuti	-4.53	-3.82
Altai	Mez1_random_deam	Australian	Yoruba	-2.54	-2.01
Altai	Mez1_random_deam	Papuan	San	-2.26	-1.99
Altai	Mez1_random_deam	Papuan	Mbuti	-3.96	-3.69
Altai	Mez1_random_deam	Papuan	Yoruba	-1.88	-1.60

Table S34.
The proximity of Vindija and Mezmaiskaya 1 to the introgressed Neandertal(s) in modern humans

Comparison				D%	|Z|
Vindija33.19	Mez1_snpAD	French	San	-9.96	-11.43
Vindija33.19	Mez1_snpAD	French	Mbuti	-9.92	-12.17
Vindija33.19	Mez1_snpAD	French	Yoruba	-7.49	-8.58
Vindija33.19	Mez1_snpAD	Sardinian	San	-8.81	-9.55
Vindija33.19	Mez1_snpAD	Sardinian	Mbuti	-8.77	-9.93
Vindija33.19	Mez1_snpAD	Sardinian	Yoruba	-6.34	-7.07
Vindija33.19	Mez1_snpAD	Han	San	-8.66	-9.15
Vindija33.19	Mez1_snpAD	Han	Mbuti	-8.65	-9.22
Vindija33.19	Mez1_snpAD	Han	Yoruba	-6.21	-6.60
Vindija33.19	Mez1_snpAD	Dai	San	-9.28	-9.79
Vindija33.19	Mez1_snpAD	Dai	Mbuti	-9.26	-10.15
Vindija33.19	Mez1_snpAD	Dai	Yoruba	-6.80	-7.88
Vindija33.19	Mez1_snpAD	Karitiana	San	-8.39	-8.43
Vindija33.19	Mez1_snpAD	Karitiana	Mbuti	-8.45	-9.08
Vindija33.19	Mez1_snpAD	Karitiana	Yoruba	-5.91	-6.15
Vindija33.19	Mez1_snpAD	Mixe	San	-9.74	-9.78
Vindija33.19	Mez1_snpAD	Mixe	Mbuti	-9.81	-10.03
Vindija33.19	Mez1_snpAD	Mixe	Yoruba	-7.42	-7.37
Vindija33.19	Mez1_snpAD	Australian	San	-7.30	-7.73
Vindija33.19	Mez1_snpAD	Australian	Mbuti	-7.23	-8.07
Vindija33.19	Mez1_snpAD	Australian	Yoruba	-4.69	-5.10
Vindija33.19	Mez1_snpAD	Papuan	San	-7.27	-7.58
Vindija33.19	Mez1_snpAD	Papuan	Mbuti	-7.21	-8.01
Vindija33.19	Mez1_snpAD	Papuan	Yoruba	-4.69	-5.31
Vindija33.19	Mez1_random	French	San	-12.13	-14.84
Vindija33.19	Mez1_random	French	Mbuti	-11.58	-14.87
Vindija33.19	Mez1_random	French	Yoruba	-8.94	-10.90
Vindija33.19	Mez1_random	Sardinian	San	-11.21	-12.81
Vindija33.19	Mez1_random	Sardinian	Mbuti	-10.63	-12.23
Vindija33.19	Mez1_random	Sardinian	Yoruba	-7.99	-9.18
Vindija33.19	Mez1_random	Han	San	-11.34	-12.66
Vindija33.19	Mez1_random	Han	Mbuti	-10.81	-11.94
Vindija33.19	Mez1_random	Han	Yoruba	-8.19	-9.19
Vindija33.19	Mez1_random	Dai	San	-12.14	-13.51
Vindija33.19	Mez1_random	Dai	Mbuti	-11.62	-13.11
Vindija33.19	Mez1_random	Dai	Yoruba	-9.02	-10.79

Vindija33.19	Mez1_random	Karitiana	San	-11.16	-11.66
Vindija33.19	Mez1_random	Karitiana	Mbuti	-10.70	-11.77
Vindija33.19	Mez1_random	Karitiana	Yoruba	-7.97	-8.70
Vindija33.19	Mez1_random	Mixe	San	-12.39	-13.27
Vindija33.19	Mez1_random	Mixe	Mbuti	-11.96	-13.01
Vindija33.19	Mez1_random	Mixe	Yoruba	-9.32	-10.00
Vindija33.19	Mez1_random	Australian	San	-9.75	-10.90
Vindija33.19	Mez1_random	Australian	Mbuti	-9.20	-10.62
Vindija33.19	Mez1_random	Australian	Yoruba	-6.46	-7.33
Vindija33.19	Mez1_random	Papuan	San	-9.72	-10.54
Vindija33.19	Mez1_random	Papuan	Mbuti	-9.12	-10.32
Vindija33.19	Mez1_random	Papuan	Yoruba	-6.39	-7.39
Vindija33.19	Mez1_random_deam	French	San	-0.70	-0.48
Vindija33.19	Mez1_random_deam	French	Mbuti	-2.04	-1.57
Vindija33.19	Mez1_random_deam	French	Yoruba	-0.28	-0.21
Vindija33.19	Mez1_random_deam	Sardinian	San	0.09	0.06
Vindija33.19	Mez1_random_deam	Sardinian	Mbuti	-1.18	-0.93
Vindija33.19	Mez1_random_deam	Sardinian	Yoruba	0.61	0.49
Vindija33.19	Mez1_random_deam	Han	San	1.92	1.37
Vindija33.19	Mez1_random_deam	Han	Mbuti	0.72	0.56
Vindija33.19	Mez1_random_deam	Han	Yoruba	2.64	2.02
Vindija33.19	Mez1_random_deam	Dai	San	0.13	0.09
Vindija33.19	Mez1_random_deam	Dai	Mbuti	-1.11	-0.85
Vindija33.19	Mez1_random_deam	Dai	Yoruba	0.79	0.63
Vindija33.19	Mez1_random_deam	Karitiana	San	1.14	0.78
Vindija33.19	Mez1_random_deam	Karitiana	Mbuti	-0.13	-0.09
Vindija33.19	Mez1_random_deam	Karitiana	Yoruba	1.86	1.34
Vindija33.19	Mez1_random_deam	Mixe	San	-0.92	-0.63
Vindija33.19	Mez1_random_deam	Mixe	Mbuti	-2.24	-1.58
Vindija33.19	Mez1_random_deam	Mixe	Yoruba	-0.48	-0.34
Vindija33.19	Mez1_random_deam	Australian	San	-0.63	-0.44
Vindija33.19	Mez1_random_deam	Australian	Mbuti	-1.85	-1.40
Vindija33.19	Mez1_random_deam	Australian	Yoruba	-0.20	-0.14
Vindija33.19	Mez1_random_deam	Papuan	San	0.90	0.61
Vindija33.19	Mez1_random_deam	Papuan	Mbuti	-0.32	-0.24
Vindija33.19	Mez1_random_deam	Papuan	Yoruba	1.63	1.19

Table S35.
The proportion of present-day human contamination in Mezmaiskaya 1 data

Comparisons									alpha	std. err	Z (null=0)
Altai	Mez1_snpAD	Han	Mbuti	:	Altai	French	Han	Mbuti	3.32	0.29	11.26
Altai	Mez1_snpAD	French	Mbuti	:	Altai	Han	French	Mbuti	3.33	0.25	13.18
Altai	Mez1_random	Han	Mbuti	:	Altai	French	Han	Mbuti	4.05	0.30	13.63
Altai	Mez1_random	French	Mbuti	:	Altai	Han	French	Mbuti	3.92	0.25	15.51
Altai	Mez1_random_deam	Han	Mbuti	:	Altai	French	Han	Mbuti	1.26	0.32	3.91
Altai	Mez1_random_deam	French	Mbuti	:	Altai	Han	French	Mbuti	1.56	0.30	5.26
Vindija33.19	Mez1_snpAD	Han	Mbuti	:	Vindija33.19	French	Han	Mbuti	1.98	0.20	9.82
Vindija33.19	Mez1_snpAD	French	Mbuti	:	Vindija33.19	Han	French	Mbuti	2.20	0.18	12.36
Vindija33.19	Mez1_random	Han	Mbuti	:	Vindija33.19	French	Han	Mbuti	2.72	0.21	13.12
Vindija33.19	Mez1_random	French	Mbuti	:	Vindija33.19	Han	French	Mbuti	2.82	0.18	15.38
Vindija33.19	Mez1_random_deam	Han	Mbuti	:	Vindija33.19	French	Han	Mbuti	-0.14	0.24	-0.60
Vindija33.19	Mez1_random_deam	French	Mbuti	:	Vindija33.19	Han	French	Mbuti	0.36	0.23	1.58

Table S36.
Maximum likelihood demographic models for each archaic combination. Demographic parameters are Neandertal-Denisovan split time (N-D), Altai-Vindija Neandertal split time (NA-NV), split time between the sequenced Neandertal and the introgressing Neandertal popultion (N-Intr N), and the same for Denisovan (D-Intr D).

Archaics	**pop**	**N-D**	**N-Intr N**	**D-Intr D**
neand_altai.den_altai	EAS	400000	200000	350000
neand_altai.den_altai	EUR	400000	200000	350000
neand_altai.den_altai	SAS	400000	200000	350000
neand_altai.den_altai	MEL	400000	200000	350000
neand_vindija.den_altai	EAS	400000	150000	350000
neand_vindija.den_altai	EUR	400000	150000	350000
neand_vindija.den_altai	SAS	400000	150000	350000
neand_vindija.den_altai	MEL	400000	150000	350000
Archaics	**pop**	**NA-NV**	**NV-Intr NV**	**NA-Intr NA**
neand_vindija.neand_altai	EAS	150000	100000	50000*
neand_vindija.neand_altai	EUR	150000	100000	50000*
neand_vindija.neand_altai	SAS	150000	100000	50000*
neand_vindija.neand_altai	MEL	150000	100000	50000*

Table S37.
Maximum likelihood estimates for the proportion of S* sequence from each archaic, for all archaic pairs and all populations, with CI generated via jackknife (+/- 2se).

archaics	pop	Prop. N	Jackknife CI N	Prop. D	Jackknife CI D
neand_altai.den_altai	EAS	0.496	0.494-0.499	0.010	0.010-0.011
neand_altai.den_altai	EUR	0.458	0.457-0.459	0.000	0.000-0.000
neand_altai.den_altai	SAS	0.442	0.441-0.443	0.011	0.011-0.011
neand_altai.den_altai	MEL	0.234	0.233-0.236	0.301	0.299-0.302
neand_vindija.den_altai	EAS	0.521	0.518-0.523	0.011	0.011-0.011
neand_vindija.den_altai	EUR	0.482	0.482-0.483	0.000	0.000-0.000
neand_vindija.den_altai	SAS	0.463	0.462-0.464	0.012	0.012-0.013
neand_vindija.den_altai	MEL	0.258	0.256-0.259	0.295	0.294-0.297
archaics	**pop**	**Prop. NV**	**Jackknife CI NV**	**Prop. NA**	**Jackknife CI NA**
neand_vindija.neand_altai	EAS	0.462	0.460-0.465	0.000	0.000-0.000
neand_vindija.neand_altai	EUR	0.431	0.430-0.432	0.000	0.000-0.000
neand_vindija.neand_altai	SAS	0.408	0.407-0.409	0.001	0.001-0.001
neand_vindija.neand_altai	MEL	0.272	0.271-0.273	0.000	0.000-0.000

Table S38.
The path to download the call sets used for this analysis along with their name is listed.

Source	Variant call set name
http://cdna.eva.mpg.de/neandertal/Vindija/VCF/	ALTAI_SNP
http://cdna.eva.mpg.de/neandertal/Vindija/VCF/indels/	ALTAI_INDEL
http://cdna.eva.mpg.de/neandertal/Vindija/VCF/	DENISOVA_SNP
http://cdna.eva.mpg.de/neandertal/Vindija/VCF/indels/	DESIVOA_INDEL
http://cdna.eva.mpg.de/neandertal/Vindija/VCF/	VINDIJA_SNP
http://cdna.eva.mpg.de/neandertal/Vindija/VCF/indels/	VINDIJA_INDEL
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/integrated_sv_map/	SUDMANT
Mills_and_1000G_gold_standard.indels.b37.sites.vcf	MILLS
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/	1000G

Table S39.
Individuals used for this analysis are shown with their, name, location, sex and coverage.

Type	Name	Location	Sex	Coverage
Neandertal	Vindija	Croatia	F	30X
Neandertal	Altai (*2*)	Altai cave	F	52X
Denisova	Denisova(*1*)	Altai cave	F	31X
Present day human	NA19017(*48*)	Africa, Luhya in Webuye	F	30-40X
Present day human	HG01565 (*48*)	South American from Peru	M	30-40X
Present day human	HG00096(*48*)	British in England	M	30-40X
Present day human	NA18525(*48*)	Han Chinese	F	30-40X
Chimpanzee	Carl	?	M	40X
Orangutan (Pongo Abelii)	Buschi	Osnabrück Zoo	M	40X

Table S40.

Number of shared indels between Denisova, both Neandertals, Vindija Neandertal and Altai Neandertal. The number of indels shared with specific 1000 genomes super populations are shown below. For Denisova specific we counted variants private to Denisova. For Altai and Vindija we counted the variants that were present in the given individual and absent in Denisova (had to be homozygous for the reference allele). The numbers in brackets for Altai and Vindija are indels private to that individual.

Denisova specific	All populations	27595
	Africa	1487
	America	20
	East Asia	357
	Europe	22
	South East Asia	958
Neandertal specific	All populations	48778
	Africa	1033
	America	20
	East Asia	554
	Europe	101
	South East Asia	1096
Altai	All populations	41137 (4434)
	Africa	866 (215)
	America	16 (5)
	East Asia	419 (51)
	Europe	82 (9)
	South East Asia	826 (79)
Vindija	All populations	44344 (7641)
	Africa	818 (167)
	America	15 (4)
	East Asia	503 (135)
	Europe	92 (19)
	South East Asia	1017 (270)

Table S41.

Vindija Neanderthal shows evidence for copy number variation in 40 hominin-specific loci. AMH: anatomically modern human; ANC: ancient DNA of modern humans; NDL: Neanderthal; DNS: Denisovan. The format of copy number count is defined as CNV-genotype:#individuals.

lineage	locus	length	annotation	Copy number count				
				AMH	ANC	NDL,Vindija	NDL,Altai	DNS
AMH-archaic shared	chr1:143873214-143985618	112404	*FAM72C,FAM72D*;exonic	6:7,7:46,8:167, 9:3,10:1	12:1,8:1,9:1	8	8	7
AMH-archaic shared	chr1:143991800-144095775	103975	intergenic	5:12,6:209,7:3	6:1,7:2	7	6	7
AMH-archaic shared	chr1:145610541-145626462	15921	*POLR3C,RNF115*; exonic	4:1,5:143,6:80	5:1,7:2	6	6	7
AMH-archaic shared	chr1:149344367-149435118	90751	*FCGR1C*;exonic	4:7,5:41,6:171, 7:4,8:1	6:1,7:1,10:1	5	4	5
AMH-archaic shared	chr1:234911364-234956952	45588	intergenic	3:1,4:8,5:17, 6:67,7:77,8:36, 9:16,10:1,11:1	6:1,8:2	9	11	5
AMH-archaic shared	chr2:87731136-87806717	75581	*LINC00152*;exonic	2:1,3:1,4:216, 5:5,6:1	4:1,5:2	5	4	4
AMH-archaic shared	chr5:70300054-70314195	14141	*NAIP*;exonic	2:23,3:147,4:38, 5:11,6:4,7:1	3:2,4:1	4	4	3
AMH-archaic shared	chr5:70368978-70389542	20564	*LOC647859*;exonic	2:47,3:92,4:73, 5:11,6:1	3:3	6	5	4
AMH-archaic shared	chr5:150201232-150223428	22196	intergenic	0:20,1:86,2:118	1:1,2:2	0	0	0
AMH-archaic shared	chr6:35754346-35767153	12807	*CLPS,CLPSL1*; exonic	2:76,3:98,4:39, 5:10,6:1	2:1,3:2	5	4	4
AMH-archaic shared	chr7:143876792-144074868	198076	*ARHGEF35,ARHGEF5, CTAGE4,CTAGE8,OR2A1, OR2A42,OR2A7*;exonic	2:2,3:8,4:39,5:50, 6:59,7:38,8:15,9:8, 10:4,12:1	3:1,5:1,6:1	5	3	4
AMH-archaic shared	chr8:32678973-32692337	13364	intergenic	0:212,1:9,2:3	0:3	0	0	0
AMH-archaic shared	chr10:127572984-127623735	50751	*FANK1*;exonic	13:2,14:11,15:35, 16:42,17:50,18:38, 19:27,20:16,21:2, 22:1	15:1,20:1, 23:1	12	23	21
AMH-NDL shared	chr4:34779463-34830156	50693	intergenic	0:7,1:63,2:154	2:3	0	0	2
AMH-NDL shared	chr7:155405636-155415734	10098	intergenic	1:1,2:223	2:3	0	0	2
AMH-NDL shared	chr9:104713896-104726456	12560	intergenic	0:14,1:54,2:156	0:1,2:2	0	0	2
AMH-Vindija shared	chr1:26114093-26124568	10475	intergenic	2:224	2:2,3:1	3	2	2
AMH-Vindija shared	chr4:42756981-42778724	21743	intergenic	0:21,1:60,2:143	1:2,2:1	1	2	2
AMH-Vindija shared	chr8:9054892-9065882	10990	*LOC101929128*;exonic	1:6,2:218	2:3	1	2	2
AMH-Vindija shared	chr12:133504687-133683440	178753	*ZNF140,ZNF26,ZNF605, ZNF84*;exonic	1:1,2:223	2:1,3:2	1	2	2
AMH-Vindija shared	chr15:86511206-86521514	10308	intergenic	1:4,2:220	1:1,2:2	0	2	2
AMH-Vindija shared	chr17:15042081-15059270	17189	intergenic	0:2,1:19,2:203	2:3	0	2	2
AMH-Vindija shared	chr19:47263341-47278112	14771	intergenic	1:1,2:223	2:2,3:1	3	2	2
Vindija specific	chr1:4683611-4712905	29294	intergenic	2:224	2:3	3	2	2
Vindija specific	chr1:190383841-190394813	10972	*BRINP3*;intronic	2:224	2:3	1	2	2
Vindija specific	chr1:218695738-218706809	11071	*MIR548F3*;intronic	2:224	2:3	1	2	2
Vindija specific	chr2:156186257-156219423	33166	intergenic	2:224	2:3	1	2	2
Vindija specific	chr7:105847233-105887724	40491	intergenic	2:224	2:3	3	2	2
Vindija specific	chr8:29474655-29488505	13850	intergenic	2:224	2:3	0	2	2
Vindija specific	chr8:38418242-38428862	10620	intergenic	2:224	2:3	3	2	2
Vindija specific	chr9:79399194-79416216	17022	*PCA3*;exonic	2:224	2:3	1	2	2
Vindija specific	chr10:50453159-50479480	26321	intergenic	2:224	2:3	1	2	2

Vindija specific	chr13:26563041-26619600	56559	*ATP8A2*;exonic	2:224	2:3	3	2	2
Vindija specific	chr15:98395133-98409027	13894	*LINC00923*;intronic	2:224	2:3	1	2	2
Vindija specific	chr20:25357877-25402419	44542	*ABHD12*,*GINS1*;exonic	2:224	2:3	3	2	2
NDL specific	chr4:171467570-171480017	12447	intergenic	2:224	2:3	1	0	2
NDL specific	chr5:41581332-41595266	13934	intergenic	2:224	2:3	1	1	2
NDL specific	chr8:35506699-35521832	15133	*UNC5D*;intronic	2:224	2:3	0	1	2
NDL specific	chr13:100776502-100796519	20018	*PCCA*;intronic	2:224	2:3	1	1	2
NDL-DNS shared	chr2:56318746-56328868	10122	intergenic	2:224	2:3	1	2	1

Table S42.

Depletion of deletions in exonic sequences among the 197 hominin-specific CNVs (Chi-squared test, p=0.0002).

	With exonic sequences	Without exonic sequences
Biallelic deletion	16	103
Biallelic duplication	17	30
Multi-allelic CNV	13	18

Table S43.

Lineage-specific duplications inferred in Prüfer et al. 2014 (*2*) and their genotype counts in each lineage. AMH: anatomically modern human; ANC: ancient DNA of modern humans; NDL: Neanderthal; DNS: Denisovan. The format of copy number count is defined as CNV-genotype:#individuals.

Locus	Annotation	Lineage identified as in (*2*)	Genotype count					
			AMH	ANC	Vindija NDL	Altai NDL	DNS	Non-human primates
chr3:12639069-12641393	*RAF1*	Altai NDL	2:222,3:2	2:3	2	3	2	2:85,3:1
chr6:95473793-95532866	intergenic	Altai NDL	1:1,2:223	2:3	2	4	2	2:86
chr12:122079832-122087495	*ORAI1*	Altai NDL	2:224	2:3	2	4	2	2:86
chr12:132295389-132391442	*MMP17,ULK1*	Altai NDL	1:1,2:223	2:3	2	4	2	1:33,2:53
chr19:9284044-9291195	intergenic	Altai NDL	2:209,3:3,5:1, 6:4,7:3,8:3,9:1	2:2,3:1	6	6	2	2:82,3:4
chr20:281880-290717	intergenic	Altai NDL	2:224	2:3	3	5	2	2:84,3:2
chr1:161272681-161274838	*MPZ*	DNS	2:224	2:3	2	2	3	2:86
chr2:48781187-48787915	intergenic	DNS	2:224	2:3	2	2	4	1:2;2:84
chr4:68542692-68577288	*UBA6,LOC550112*	DNS	2:224	2:3	2	2	3	2:86
chr4:68579206-68581585	*LOC550112*	DNS	2:224	2:3	2	2	4	2:86
chr7:140872574-140879065	*LOC100131199*	DNS	2:224	2:3	2	2	6	2:86
chr11:39901956-39909545	intergenic	DNS	1:1,2:223	2:3	2	2	10	1:33,2:53
chr12:49894191-49897733	*SPATS2*	DNS	2:224	2:3	2	2	5	2:86
chr19:55302094-55315197	*KIR3DP1,KIR2DL4*	DNS	0:2,1:8,2:199, 3:13,4:2	2:3	1	2	5	1:3,2:70,3:7, 4:6
chr1:108924526-108990191	intergenic	AMH	2:4,3:45,4:175	3:2,4:1	2	2	2	2:81,3:5
chr2:87417089-87420544	intergenic	AMH	2:1,3:61,4:104, 5:51,6:7	5:1,6:2	2	2	2	0:19,1:8,2:59
chr16:30200098-30206185	*CORO1A, LOC606724,BOLA2*	AMH	4:9,5:32,6:90, 7:78,8:13,9:2	7:3	2	2	2	1:6,2:62,3:18

Table S44.

Exon-deletions along the Neanderthal-Denisovan lineage reported in Prüfer et al. 2014 (*2*) and their genotype counts in each lineage. AMH: anatomically modern human; ANC: ancient DNA of modern humans; NDL: Neanderthal; DNS: Denisovan. The format of copy number count is defined as CNV-genotype:#individuals.

Locus	Annotation	Lineage-identified as in (*2*)	length	Genotype counts					
				AMH	ANC	Vindija NDL	Altai NDL	DNS	Non-human primates
chr1:152573137-152573562	*LCE3C,LCEB*	DNS	425	0:5,1:82,2:96, 3:40,4:1	1:2,2:1	3	3	1	2:14,3:58,4:14
chr10:114112072-114117451	*GUCY2GP*	DNS	5379	0:53,1:113,2:58	0:1,2:2	2	2	0	2:86
chr12:19401869-19407358	*PLEKHA5*	DNS	5489	2:224	2:3	2	2	0	1:22,2:64
chr5:159618047-159625750	*FABP6*	DNS	7703	2:224	2:3	2	2	0	2:71,3:15
chr1:213000404-213017911	*C1orf227*	Altai NDL	17507	1:10,2:214	1:1,2:2	1	1	2	1:2,2:84
chr10:123735993-123750736	*TACC2*	Altai NDL	14743	2:224	2:3	1	1	2	1:1,2:81,3:4
chr11:128682442-128683785	*FLI1*	Altai NDL	4025	1:59,2:165	2:3	2	1	2	1:1,2:81,3:4
chr12:27646876-27657357	*C12orf70*	Altai NDL	10481	1:16,2:208	2:3	1	1	2	1:1,2:85
chr7:1654105-1656328	*TFAMP1*	Altai NDL	2223	2:10,3:214	3:3	1	1	3	2:18,3:66,4:2
chr8:144632032-144638482	*GSDMD*	Altai NDL	6450	0:11,1:68,2:145	2:3	1	0	2	0:1,1:41,2:43, 3:1
chr8:38774633-38778670	*PLEKHA2*	Altai NDL	4037	1:1,2:223	2:3	1	1	2	2:52,3:34
chr11:3239173-3240043	*MRGPRG*	Altai NDL/DNS	870	1:31,2:191,3:2	2:2,3:1	0	0	0	0:38,1:7,2:37, 3:4
chr11:3239561-3244361	*C11orf36*	Altai NDL/DNS	4800	1:32,2:192	2:2,3:1	0	0	0	1:29,2:55,3:2
chr22:24373116-24374043	*LOC391322*	Altai NDL/DNS	927	1:45,2:100,3:70, 4:8,5:1	1:1,2:1 ,4:1	2	1	1	1:12,2:26,3:10 ,4:29,5:8,6:1
chr22:24376138-24384284	*GSTT1*	Altai NDL/DNS	8146	0:63,1:112,2:49	0:2,2:1	0	0	0	2:83,3:3
chr22:24365191-24401108	*GSTTP2*	Altai NDL/DNS	35917	1:60,2:113,3:50, 4:1	1:1,2:1 ,3:1	1	1	1	3:61,4:24,6:1
chr5:147549295-147554961	*SPINK14*	Altai NDL/DNS	5666	1:133,2:91	1:1,2:2	1	1	1	1:1,2:85

Table S45.

Independent SNP-trait associations (r^2<0.8) involving Vindija-specific potentially introgressed alleles.

chr	Position (hg19)	GWAS p-value	Trait	rs number-risk allele	potentially introduced allele	Altai genotype	Vindija genotype
chr22	26189657	3.00E-07	Trans fatty acid levels (*107*)	rs5752209-C	G	C/C	G/G
chr9	30889007	2.00E-06	Post bronchodilator FEV1/FVC ratio (*108*)	rs78589086-C	C	T/T	C/C
chr6	32797773	4.00E-07	Thionamide-induced agranulocytosis in Graves' disease (*109*)	rs2228391-C	C	T/T	C/T
chr6	32953280	6.00E-07	IgG glycosylation (*110*)	rs3097645-C	G	C/C	G/G
chr15	86984240	2.00E-06	Schizophrenia (*111*)	rs16977195-?	G	A/A	G/G
chr14	98840443	2.00E-06	Response to antipsychotic therapy (extrapyramidal side effects) (*112*)	rs1459148-?	T	C/C	T/T
chr2	118835841	2.00E-12	LDL cholesterol (*106*)	rs10490626-A	A	G/G	A/A
chr4	123399491	4.00E-06	Rheumatoid arthritis (*113*)	rs45475795-G	G	A/A	G/G
chr8	134778342	7.00E-07	Eating disorders (purging via substances) (*114*)	rs74566133-C	T	C/C	T/T
chr1	217718132	6.00E-06	Visceral adipose tissue adjusted for BMI (*115*)	rs2059397-G	G	C/C	G/C
chr2	223049021	1.00E-06	Vitamin D levels (*116*)	rs6730714-A	A	G/G	A/A

Table S46.
Independent SNP-trait associations (r^2<0.8) involving Altai-specific potentially introgressed alleles.

chr	Position (hg19)	GWAS p-value	Trait	rs number-risk allele	Potentially introduced allele	Altai genotype	Vindija genotype
chr1	3651031	2.00E-06	Visceral adipose tissue/subcutaneous adipose tissue ratio (*115*)	rs12562437-T	T	T/T	C/C
chr17	6945087	5.00E-15	Type 2 diabetes (*10*)	rs75493593-?	T	T/T	G/G
chr6	19236109	2.00E-06	3-hydroxypropylmercapturic acid levels in smokers (*117*)	rs6931743-?	A	A/G	G/G
chr11	24091143	2.00E-06	Thrombin-antithrombin complex levels in ischemic stroke (*118*)	rs16912059-?	A	A/G	G/G
chr11	102738075	8.00E-10	Post bronchodilator FEV1/FVC ratio (*108*)	rs17368582-T	C	C/C	T/T

Table S47.

Independent SNP-trait associations (r^2<0.8) involving potentially introgressed alleles that are shared by both Vindija and Altai.

chr	position (hg19)	GWAS p-value	Trait	rs number-risk allele	Potentially introduced allele	Altai genotype	Vindija genotype
chr17	800593	3.00E-08	Colorectal cancer	rs12603526-C	C	C/C	C/C
chr20	3737495	7.00E-06	Survival in colon cancer	rs658495-?	G	G/G	G/G
chr10	3804257	2.00E-06	Amyotrophic lateral sclerosis (sporadic)	rs10508264-?	A	A/A	A/A
chr22	17597462	2.00E-06	Heschl's gyrus morphology	rs971768-A	A	A/A	A/A
chr9	18109235	7.00E-07	Heart failure	rs2210327-?	T	T/T	T/T
chr11	20562519	5.00E-06	Response to treatment for acute lymphoblastic leukemia	rs7128311-C	C	C/C	C/C
chr3	21955198	2.00E-06	Attention deficit hyperactivity disorder	rs11719664-?	T	T/T	T/T
chr9	25452812	4.00E-06	RR interval (heart rate)	rs13300284-A	A	A/A	A/A
chr9	27209469	2.00E-65	Endothelial growth factor levels	rs2273720-C	C	C/C	C/C
chr9	30902950	3.00E-07	Post bronchodilator FEV1/FVC ratio	rs139088314-T	T	T/T	T/T
chr22	32783904	3.00E-06	IgG glycosylation	rs12530-C	C	C/C	C/C
chr22	34078057	8.00E-06	Alcohol dependence (age at onset)	rs5754638-T	T	T/T	T/T
chr21	37013621	5.00E-06	Obesity-related traits	rs12483148-G	C	C/C	C/C
chr13	38737821	3.00E-08	Low vWF levels	rs17057285-?	C	C/C	C/C
chr12	40528432	4.00E-25	Crohn's disease	rs12422544-G	C	C/C	C/C
chr12	40528432	3.00E-06	Ulcerative colitis	rs12422544-G	C	C/C	C/C
chr13	40833012	1.00E-07	Crohn's disease	rs17061048-T	A	A/A	A/A
chr13	40833012	5.00E-09	Inflammatory bowel disease	rs17061048-A	A	A/A	A/A
chr19	41309150	5.00E-06	Post bronchodilator FEV1	rs117391664-A	A	A/A	A/A
chr10	44524675	1.00E-07	HIV-associated dementia	rs17154929-T	T	T/T	T/T
chr3	46235201	3.00E-17	Celiac disease	rs13098911-A	T	T/T	T/T
chr18	53050646	2.00E-06	Schizophrenia	rs72926932-A	C	C/C	C/C
chr1	54320337	5.00E-06	Nonalcoholic fatty liver disease	rs11206226-A	G	G/G	G/G

chr12	56740682	1.00E-13	Height	rs2066807-C	G	G/G	G/G
chr12	56740682	5.00E-12	Psoriasis	rs2066807-G	G	G/G	G/G
chr5	58082723	9.00E-06	IgG glycosylation	rs10065350-C	T	T/T	T/T
chr18	58826022	7.00E-07	Cannabis use (age at onset)	rs142981069-G	G	G/G	G/G
chr14	62968395	1.00E-07	Estradiol plasma levels (breast cancer)	rs4902141-?	C	C/C	C/C
chr13	63634350	2.00E-07	Bone mineral density	rs9317284-?	T	T/T	T/T
chr10	63915972	1.00E-06	Response to haloperidol in psychosis	rs7912580-A	A	A/A	A/A
chr2	65752713	8.00E-06	Electroencephalogram traits	rs702885-?	A	A/A	A/A
chr10	71580120	6.00E-07	Parkinson's disease	rs17497526-G	C	C/C	C/C
chr6	75158266	2.00E-06	Metabolite levels (MHPG)	rs72960926-A	A	A/A	A/A
chr14	75741751	6.00E-11	Crohn's disease	rs1569328-A	T	T/T	T/T
chr14	75741751	3.00E-09	Inflammatory bowel disease	rs1569328-G	T	T/T	T/T
chr8	76319166	5.00E-06	Diisocyanate-induced asthma	rs117628011-A	T	T/T	T/T
chr16	77007437	2.00E-06	Educational attainment	rs1995082-T	G	G/G	G/G
chr16	77328895	4.00E-06	Adverse response to chemotherapy (neutropenia/leucopenia) (all antimicrotubule drugs)	rs12935229-A	T	T/T	T/T
chr5	79581768	9.00E-06	Alzheimer's disease in APOE e4- carriers	rs71636213-G	G	T/G	G/G
chr17	80524184	8.00E-06	3-hydroxypropylmercapturic acid levels in smokers	rs78461346-?	A	A/A	A/A
chr10	81185314	3.00E-06	Serum dimethylarginine levels (asymmetric/symetric ratio)	rs181715218-T	C	C/C	C/C
chr15	82193846	4.00E-06	Post-traumatic stress disorder	rs77963519-?	A	A/A	A/A
chr15	87977476	4.00E-06	Coronary artery calcification	rs2679073-A	G	G/G	G/G
chr14	95848294	8.00E-06	Obesity-related traits	rs12432260-A	T	T/T	T/T
chr9	99002185	2.00E-06	Smoking initiation	rs8190560-?	T	T/T	T/T
chr4	100395414	8.00E-06	Eating disorders	rs148915469-C	C	C/C	C/C
chr12	102506044	4.00E-46	Height	rs2271266-T	C	C/T	C/C
chr11	102680949	2.00E-07	Post bronchodilator FEV1/FVC ratio	rs470530-A	A	A/A	A/A
chr11	102720344	4.00E-10	Post bronchodilator FEV1/FVC ratio	rs17361668-A	A	A/A	A/A
chr12	103011894	4.00E-10	Mammographic density (dense area)	rs703556-A	G	G/G	G/G

chr11	116438851	9.00E-06	Clozapine-induced cytotoxicity	rs17492380-?	A	A/G	A/A
chr12	120880434	1.00E-06	Insulin resistance/response	rs17431357-C	T	T/T	T/T
chr12	129300694	2.00E-11	Systemic lupus erythematosus	rs1385374-A	T	T/T	T/T
chr8	134616136	1.00E-06	Response to protease inhibitor treatment in hepatitis c (peak serum total bilirubin levels)	rs2978048-?	G	T/G	G/G
chr5	149036976	4.00E-06	Major depressive disorder	rs17710780-T	C	C/C	C/C
chr5	150585867	7.00E-06	Bulimia nervosa	rs7724774-G	A	A/A	A/A
chr2	157096776	1.00E-09	Menarche (age at onset)	rs17188434-C	C	C/C	C/C
chr6	167548547	7.00E-09	Triglycerides	rs62436827-G	G	G/G	G/G
chr5	168386089	1.00E-08	Inflammatory skin disease	rs12188351-?	A	A/A	A/A
chr1	169099037	1.00E-31	QT interval	rs10919070-C	C	C/C	C/C
chr2	173311553	9.00E-23	Prostate cancer	rs12621278-?	G	G/G	G/G
chr2	179641975	2.00E-06	QT interval	rs12476289-A	T	T/T	T/T
chr4	182568250	2.00E-06	Age at smoking initiation in chronic obstructive pulmonary disease	rs17278117-A	G	A/G	G/G
chr3	191695113	1.00E-08	Severe influenza A (H1N1) infection	rs4261353-?	C	C/C	C/C
chr1	205458163	5.00E-07	Plasma omega-6 polyunsaturated fatty acid levels (gamma-linolenic acid)	rs6680582-A	A	A/A	A/A
chr1	205483595	4.00E-06	IgG glycosylation	rs12127944-C	G	G/G	G/G
chr1	208994936	7.00E-06	Educational attainment	rs17013497-T	T	T/T	T/T
chr1	209964080	9.00E-22	Nonsyndromic cleft lip with or without cleft palate	rs2235371-C	T	T/T	T/T
chr1	216504269	6.00E-06	IgG glycosylation	rs7532570-G	G	G/G	G/G

Table S48.

Allele frequencies of the potentially introgressed alleles matching specifically either the Vindija or Altai Neandertal (EAS= East Asians; EUR=Europeans) and overlap with the haplotypes detected in the introgression maps reported in S12.

chr	position	rs number	EAS allele frequency	EUR allele frequency	Potentially introgressed allele	Matching Neanderthal	Overlap with maps
chr22	26189657	rs5752209	0.1151	0.0119	G	Vindija	yes (EAS and EUR)
chr17	6945087	rs75493593	0.1002	0.0169	T	Altai	yes (EAS and EUR)
chr15	86984240	rs16977195	0.1349	0.0527	G	Vindija	yes (EAS and EUR)
chr14	98840443	rs1459148	0	0.0189	T	Vindija	yes (EUR)
chr9	30889007	rs78589086	0.0139	0	C	Vindija	yes (EAS)
chr11	24091143	rs16912059	0.004	0.0378	A	Altai	yes (EUR)
chr11	102738075	rs17368582	0.004	0.1223	C	Altai	yes (EAS and EUR)
chr8	134778342	rs74566133	0.001	0.0298	T	Vindija	no
chr6	19236109	rs6931743	0.001	0.0249	A	Altai	no
chr6	32797773	rs2228391	0.0962	0	C	Vindija	no
chr6	32953280	rs3097645	0.1954	0.0934	G	Vindija	yes (EAS and EUR)
chr4	123399491	rs45475795	0	0.0805	G	Vindija	no
chr1	3651031	rs12562437	0.2123	0.0288	T	Altai	no
chr1	217718132	rs2059397	0	0.0268	G	Vindija	yes (EUR)
chr2	118835841	rs10490626	0	0.0795	A	Vindija	yes (EUR)
chr2	223049021	rs6730714	0.0188	0.0746	A	Vindija	yes (EAS and EUR)

Table S49.

SNP-trait associations involving Denisova-specific alleles.

chr	Position (hg19)	GWAS p-value	Trait	rs number-risk allele	Potentially introduced allele	Altai genotype	Vindija genotype	Denisova genotype
chr22	48923459	5.00E-06	Bipolar disorder (body mass index interaction)	rs80088139-?	G	C/C	C/C	G/G
chr6	70074232	4.00E-06	Response to statin therapy	rs3757057-T	T	A/A	A/A	T/T
chr5	172191052	8.00E-09	Vertical cup-disc ratio	rs17658229-C	C	T/T	T/T	C/C

Table S50.
Probability of the Denisova-specific haplotypes to be shared due to incomplete lineage sorting.

chr	Position	GWAS trait	GWAS p-value	Length of haplotype (bp)	Recombination rate (cM/Mb)	p-value
22	48923459	Bipolar disorder (body mass index interaction)	5E-6	20282	4.369	2.786 e-14
5	172191052	Vertical cup-disc ratio	8E-9	27060	0.4959	0.0325
6	70074232	Response to statin therapy	4E-6	1	NA	NA

Table S51.
Frequencies of Denisova-specific trait-associated alleles.

chr	Position	GWAS trait	East Asians	Europeans	Americans	South Asians
22	48923459	Bipolar disorder (body mass index interaction)	0.1746	0.004	0.049	0.0051
5	172191052	Vertical cup-disc ratio	0.001	0.0567	0.0159	0.0102
6	70074232	Response to statin therapy	0.0635	0.0258	0.1124	0.0204

Table S52.
Hominin-specific copy number variants (Table_S52.xlsx)

References and Notes

1. M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andrés, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012). doi:10.1126/science.1224344 Medline

2. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014). doi:10.1038/nature12886 Medline

3. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Y. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010). doi:10.1126/science.1188021 Medline

4. K. Harris, R. Nielsen, The genetic cost of Neanderthal introgression. *Genetics* **203**, 881–891 (2016). doi:10.1534/genetics.116.186890 Medline

5. I. Juric, S. Aeschbacher, G. Coop, The strength of selection against Neanderthal introgression. *PLOS Genet.* **12**, e1006340 (2016). doi:10.1371/journal.pgen.1006340 Medline

6. S. Sankararaman, S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, D. Reich, The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014). doi:10.1038/nature12961 Medline

7. C. N. Simonti, B. Vernot, L. Bastarache, E. Bottinger, D. S. Carrell, R. L. Chisholm, D. R. Crosslin, S. J. Hebbring, G. P. Jarvik, I. J. Kullo, R. Li, J. Pathak, M. D. Ritchie, D. M. Roden, S. S. Verma, G. Tromp, J. D. Prato, W. S. Bush, J. M. Akey, J. C. Denny, J. A. Capra, The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016). doi:10.1126/science.aad2149 Medline

8. B. Vernot, J. M. Akey, Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014). doi:10.1126/science.1245938 Medline

9. F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sánchez, Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015). doi:10.1038/nrg3936 Medline

10. A. L. Williams, S. B. Jacobs, H. Moreno-Macías, A. Huerta-Chagoya, C. Churchhouse, C. Márquez-Luna, H. García-Ortíz, M. J. Gómez-Vázquez, N. P. Burtt, C. A. Aguilar-Salinas, C. González-Villalpando, J. C. Florez, L. Orozco, C. A. Haiman, T. Tusié-Luna, D. Altshuler; SIGMA Type 2 Diabetes Consortium, Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* **506**, 97–101 (2014). Medline

11. M. Dannemann, A. M. Andrés, J. Kelso, Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human Toll-like receptors. *Am. J. Hum. Genet.* **98**, 22–33 (2016). doi:10.1016/j.ajhg.2015.11.015 Medline

12. H. Quach, M. Rotival, J. Pothlichet, Y. E. Loh, M. Dannemann, N. Zidane, G. Laval, E. Patin, C. Harmant, M. Lopez, M. Deschamps, N. Naffakh, D. Duffy, A. Coen, G. Leroux-Roels, F. Clément, A. Boland, J.-F. Deleuze, J. Kelso, M. L. Albert, L. Quintana-Murci, Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell* **167**, 643–656.e17 (2016). doi:10.1016/j.cell.2016.09.024 Medline

13. M. Kuhlwilm, I. Gronau, M. J. Hubisz, C. de Filippo, J. Prado-Martinez, M. Kircher, Q. Fu, H. A. Burbano, C. Lalueza-Fox, M. de la Rasilla, A. Rosas, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, T. Marques-Bonet, A. M. Andrés, B. Viola, S. Pääbo, M. Meyer, A. Siepel, S. Castellano, Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530**, 429–433 (2016). doi:10.1038/nature16544 Medline

14. S. Castellano, G. Parra, F. A. Sánchez-Quinto, F. Racimo, M. Kuhlwilm, M. Kircher, S. Sawyer, Q. Fu, A. Heinze, B. Nickel, J. Dabney, M. Siebauer, L. White, H. A. Burbano, G. Renaud, U. Stenzel, C. Lalueza-Fox, M. de la Rasilla, A. Rosas, P. Rudan, D. Brajković, Ž. Kucan, I. Gušic, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Meyer, J. Kelso, A. M. Andrés, S. Pääbo, Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6666–6671 (2014). doi:10.1073/pnas.1405138111 Medline

15. T. Devièse, I. Karavanić, D. Comeskey, C. Kubiak, P. Korlević, M. Hajdinjak, S. Radović, N. Procopio, M. Buckley, S. Pääbo, T. Higham, Direct dating of Neanderthal remains from the site of Vindija Cave and implications for the Middle to Upper Paleolithic transition. *Proc. Natl. Acad. Sci. U.S.A.* 201709235 (2017). doi:10.1073/pnas.1709235114 Medline

16. A. D. Greenwood, S. Pääbo, Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Mol. Ecol.* **8**, 133–137 (1999). doi:10.1046/j.1365-294X.1999.00507.x Medline

17. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi:10.1073/pnas.1314445110 Medline

18. M. T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013). doi:10.1038/nprot.2013.038 Medline

19. M. T. Gansauge, M. Meyer, Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* **24**, 1543–1549 (2014). doi:10.1101/gr.174201.114 Medline

20. Materials and methods are available as supplementary materials.

21. A. W. Briggs, U. Stenzel, M. Meyer, J. Krause, M. Kircher, S. Pääbo, Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010). doi:10.1093/nar/gkp1163 Medline

22. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi:10.1038/nature18964 Medline

23. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). doi:10.1534/genetics.112.145037 Medline

24. D. Reich, N. Patterson, M. Kircher, F. Delfin, M. R. Nandineni, I. Pugach, A. M.-S. Ko, Y.-C. Ko, T. A. Jinam, M. E. Phipps, N. Saitou, A. Wollstein, M. Kayser, S. Pääbo, M. Stoneking, Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011). doi:10.1016/j.ajhg.2011.09.005 Medline

25. J. D. Wall, M. A. Yang, F. Jay, S. K. Kim, E. Y. Durand, L. S. Stevison, C. Gignoux, A. Woerner, M. F. Hammer, M. Slatkin, Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**, 199–209 (2013). doi:10.1534/genetics.112.148213 Medline

26. A. W. Briggs, J. M. Good, R. E. Green, J. Krause, T. Maricic, U. Stenzel, S. Pääbo, Primer extension capture: Targeted sequence retrieval from heavily degraded DNA sources. *J. Vis. Exp.* **3** 1573 (2009). Medline

27. N. Rohland, M. Hofreiter, Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343–352 (2007). doi:10.2144/000112383 Medline

28. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, t5448 (2010). doi:10.1101/pdb.prot5448 Medline

29. T. Maricic, M. Whitten, S. Pääbo, Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLOS ONE* **5**, e14004 (2010). doi:10.1371/journal.pone.0014004 Medline

30. R. E. Green, A.-S. Malaspinas, J. Krause, A. W. Briggs, P. L. F. Johnson, C. Uhler, M. Meyer, J. M. Good, T. Maricic, U. Stenzel, K. Prüfer, M. Siebauer, H. A. Burbano, M. Ronan, J. M. Rothberg, M. Egholm, P. Rudan, D. Brajković, Z. Kućan, I. Gusić, M. Wikström, L. Laakkonen, J. Kelso, M. Slatkin, S. Pääbo, A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008). doi:10.1016/j.cell.2008.06.021 Medline

31. M. Malez, H. Ullrich, Neuere paläanthropologische Untersuchungen am Material aus der Höhle Vindija (Kroatien, Jugoslawien). *Palaeontol. Jugoslav.* **29**, 1–44 (1982).

32. L. V. Golovanova, J. F. Hoffecker, V. M. Kharitonov, G. P. Romanova, Mezmaiskaya cave: A Neanderthal occupation in the Northern Caucasus. *Curr. Anthropol.* **40**, 77–86 (1999). doi:10.1086/515805

33. A. R. Skinner, B. A. B. Blackwell, S. Martin, A. Ortega, J. I. B. Blickstein, L. V. Golovanova, V. B. Doronichev, ESR dating at Mezmaiskaya Cave, Russia. *Appl. Radiat. Isot.* **62**, 219–224 (2005). doi:10.1016/j.apradiso.2004.08.008 Medline

34. P. Korlević, T. Gerber, M.-T. Gansauge, M. Hajdinjak, S. Nagel, A. Aximu-Petri, M. Meyer, Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015). doi:10.2144/000114320 Medline

35. J. Dabney, M. Meyer, Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94 (2012). doi:10.2144/000113809 Medline

36. M. T. Gansauge, T. Gerber, I. Glocke, P. Korlevic, L. Lippik, S. Nagel, L. M. Riehl, A. Schmidt, M. Meyer, Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017). Medline

37. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012). doi:10.1093/nar/gkr771 Medline

38. M. Kircher, U. Stenzel, J. Kelso, Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* **10**, R83 (2009). doi:10.1186/gb-2009-10-8-r83 Medline

39. G. Renaud, M. Kircher, U. Stenzel, J. Kelso, freeIbis: An efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208–1209 (2013). doi:10.1093/bioinformatics/btt117 Medline

40. G. Renaud, U. Stenzel, J. Kelso, leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014). doi:10.1093/nar/gku699 Medline

41. G. Renaud, U. Stenzel, T. Maricic, V. Wiebe, J. Kelso, deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**, 770–772 (2015). doi:10.1093/bioinformatics/btu719 Medline

42. 1000 Genomes Phase 2 decoy sequences: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz.

43. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324 Medline

44. https://bitbucket.org/ustenzel/network-aware-bwa.

45. https://bitbucket.org/ustenzel/biohazard-tools.

46. Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. C. Salazar-García, Y. V. Kuzmin, S. G. Keates, P. A. Kosintsev, D. I. Razhev, M. P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T. F. G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T. B. Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014). doi:10.1038/nature13810 Medline

47. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi:10.1038/nature13673 Medline

48. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi:10.1101/gr.107524.110 Medline

49. U. Varshney, J. H. van de Sande, Specificities and kinetics of uracil excision from uracil-containing DNA oligomers by *Escherichia coli* uracil DNA glycosylase. *Biochemistry* **30**, 4055–4061 (1991). doi:10.1021/bi00230a033 Medline

50. A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis; 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). Medline

51. R. Nielsen, J. S. Paul, A. Albrechtsen, Y. S. Song, Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **12**, 443–451 (2011). doi:10.1038/nrg2986 Medline

52. R. Li, Y. Li, X. Fang, H. Yang, J. Wang, K. Kristiansen, J. Wang, SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **19**, 1124–1132 (2009). doi:10.1101/gr.088013.108 Medline

53. M. J. D. Powell, The BOBYQA algorithm for bound constrained optimization without derivatives. *Technical Report DAMTP 2009/NA06* (2009).

54. S. G. Johnson, The NLopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt.

55. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999). doi:10.1093/nar/27.2.573 Medline

56. K. R. Rosenbloom, J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. A. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn, W. J. Kent, The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43** (D1), D670–D681 (2015). doi:10.1093/nar/gku1177 Medline

57. M. Meyer, J.-L. Arsuaga, C. de Filippo, S. Nagel, A. Aximu-Petri, B. Nickel, I. Martínez, A. Gracia, J. M. Bermúdez de Castro, E. Carbonell, B. Viola, J. Kelso, K. Prüfer, S. Pääbo, Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016). doi:10.1038/nature17405 Medline

58. http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/simpleRepeat.txt.gz.

59. A. G. Hinch, A. Tandon, N. Patterson, Y. Song, N. Rohland, C. D. Palmer, G. K. Chen, K. Wang, S. G. Buxbaum, E. L. Akylbekova, M. C. Aldrich, C. B. Ambrosone, C. Amos, E. V. Bandera, S. I. Berndt, L. Bernstein, W. J. Blot, C. H. Bock, E. Boerwinkle, Q. Cai, N. Caporaso, G. Casey, L. A. Cupples, S. L. Deming, W. R. Diver, J. Divers, M. Fornage, E. M. Gillanders, J. Glessner, C. C. Harris, J. J. Hu, S. A. Ingles, W. Isaacs, E. M. John, W. H. Kao, B. Keating, R. A. Kittles, L. N. Kolonel, E. Larkin, L. Le Marchand, L. H. McNeill, R. C. Millikan, A. Murphy, S. Musani, C. Neslund-Dudas, S. Nyante, G. J. Papanicolaou, M. F. Press, B. M. Psaty, A. P. Reiner, S. S. Rich, J. L. Rodriguez-Gil, J. I. Rotter, B. A. Rybicki, A. G. Schwartz, L. B. Signorello, M. Spitz, S. S. Strom, M. J. Thun, M. A. Tucker, Z. Wang, J. K. Wiencke, J. S. Witte, M. Wrensch, X. Wu, Y. Yamamura, K. A. Zanetti, W. Zheng, R. G. Ziegler, X. Zhu, S. Redline, J. N. Hirschhorn, B. E. Henderson, H. A. Taylor Jr., A. L. Price, H. Hakonarson, S. J. Chanock, C. A. Haiman, J. G. Wilson, D. Reich, S. R. Myers, The landscape of recombination in African Americans. *Nature* **476**, 170–175 (2011). doi:10.1038/nature10336 Medline

60. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011). doi:10.1038/ng.806 Medline

61. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting FST: The impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013). doi:10.1101/gr.154831.113 Medline

62. A. W. Briggs, J. M. Good, R. E. Green, J. Krause, T. Maricic, U. Stenzel, C. Lalueza-Fox, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, R. Schmitz, V. B. Doronichev, L. V. Golovanova, M. de la Rasilla, J. Fortea, A. Rosas, S. Pääbo, Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321 (2009). doi:10.1126/science.1174462 Medline

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

64. R. R. Hudson, Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002). doi:10.1093/bioinformatics/18.2.337 Medline

65. J. N. Fenner, Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005). doi:10.1002/ajpa.20188 Medline

66. K. E. Langergraber, K. Prüfer, C. Rowney, C. Boesch, C. Crockford, K. Fawcett, E. Inoue, M. Inoue-Muruyama, J. C. Mitani, M. N. Muller, M. M. Robbins, G. Schubert, T. S. Stoinski, B. Viola, D. Watts, R. M. Wittig, R. W. Wrangham, K. Zuberbühler, S. Pääbo, L. Vigilant, Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15716–15721 (2012). doi:10.1073/pnas.1211740109 Medline

67. A. Scally, R. Durbin, Revising the human mutation rate: Implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753 (2012). doi:10.1038/nrg3295 Medline

68. O. Venn, I. Turner, I. Mathieson, N. de Groot, R. Bontrop, G. McVean, Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. *Science* **344**, 1272–1275 (2014). doi:10.1126/science.344.6189.1272 Medline

69. A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, P. Flicek, S. B. Gabriel, R. A. Gibbs, E. D. Green, M. E. Hurles, B. M. Knoppers, J. O. Korbel, E. S. Lander, C. Lee, H. Lehrach, E. R. Mardis, G. T. Marth, G. A. McVean, D. A. Nickerson, J. P. Schmidt, S. T. Sherry, J. Wang, R. K. Wilson, R. A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J. G. Reid, Y. Zhu, J. Wang, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G. Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X. Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, E. S. Lander, D. M. Altshuler, S. B. Gabriel, N. Gupta, N. Gharani, L. H. Toji, N. P. Gerry, A. M. Resch, P. Flicek, J. Barker, L. Clarke, L. Gil, S. E. Hunt, G. Kelman, E. Kulesha, R. Leinonen, W. M. McLaren, R. Radhakrishnan, A. Roa, D. Smirnov, R. E. Smith, I. Streeter, A. Thormann, I. Toneva, B. Vaughan, X. Zheng-Bradley, D. R. Bentley, R. Grocock, S. Humphray, T. James, Z. Kingsbury, H. Lehrach, R. Sudbrak, M. W. Albrecht, V. S. Amstislavskiy, T. A. Borodina, M. Lienhard, F. Mertes, M. Sultan, B. Timmermann, M.-L. Yaspo, E. R. Mardis, R. K. Wilson, L. Fulton, R. Fulton, S. T. Sherry, V. Ananiev, Z. Belaia, D. Beloslyudtsev, N. Bouk, C. Chen, D. Church, R. Cohen, C. Cook, J. Garner, T. Hefferon, M. Kimelman, C. Liu, J. Lopez, P. Meric, C. O'Sullivan, Y. Ostapchuk, L. Phan, S.

Ponomarov, V. Schneider, E. Shekhtman, K. Sirotkin, D. Slotta, H. Zhang, G. A. McVean, R. M. Durbin, S. Balasubramaniam, J. Burton, P. Danecek, T. M. Keane, A. Kolb-Kokocinski, S. McCarthy, J. Stalker, M. Quail, J. P. Schmidt, C. J. Davies, J. Gollub, T. Webster, B. Wong, Y. Zhan, A. Auton, C. L. Campbell, Y. Kong, A. Marcketta, R. A. Gibbs, F. Yu, L. Antunes, M. Bainbridge, D. Muzny, A. Sabo, Z. Huang, J. Wang, L. J. M. Coin, L. Fang, X. Guo, X. Jin, G. Li, Q. Li, Y. Li, Z. Li, H. Lin, B. Liu, R. Luo, H. Shao, Y. Xie, C. Ye, C. Yu, F. Zhang, H. Zheng, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G. T. Marth, E. P. Garrison, D. Kural, W.-P. Lee, W. Fung Leong, M. Stromberg, A. N. Ward, J. Wu, M. Zhang, M. J. Daly, M. A. DePristo, R. E. Handsaker, D. M. Altshuler, E. Banks, G. Bhatia, G. del Angel, S. B. Gabriel, G. Genovese, N. Gupta, H. Li, S. Kashin, E. S. Lander, S. A. McCarroll, J. C. Nemesh, R. E. Poplin, S. C. Yoon, J. Lihm, V. Makarov, A. G. Clark, S. Gottipati, A. Keinan, J. L. Rodriguez-Flores, J. O. Korbel, T. Rausch, M. H. Fritz, A. M. Stütz, P. Flicek, K. Beal, L. Clarke, A. Datta, J. Herrero, W. M. McLaren, G. R. S. Ritchie, R. E. Smith, D. Zerbino, X. Zheng-Bradley, P. C. Sabeti, I. Shlyakhter, S. F. Schaffner, J. Vitti, D. N. Cooper, E. V. Ball, P. D. Stenson, D. R. Bentley, B. Barnes, M. Bauer, R. Keira Cheetham, A. Cox, M. Eberle, S. Humphray, S. Kahn, L. Murray, J. Peden, R. Shaw, E. E. Kenny, M. A. Batzer, M. K. Konkel, J. A. Walker, D. G. MacArthur, M. Lek, R. Sudbrak, V. S. Amstislavskiy, R. Herwig, E. R. Mardis, L. Ding, D. C. Koboldt, D. Larson, K. Ye, S. Gravel, A. Swaroop, E. Chew, T. Lappalainen, Y. Erlich, M. Gymrek, T. Frederick Willems, J. T. Simpson, M. D. Shriver, J. A. Rosenfeld, C. D. Bustamante, S. B. Montgomery, F. M. De La Vega, J. K. Byrnes, A. W. Carroll, M. K. DeGorter, P. Lacroute, B. K. Maples, A. R. Martin, A. Moreno-Estrada, S. S. Shringarpure, F. Zakharia, E. Halperin, Y. Baran, C. Lee, E. Cerveira, J. Hwang, A. Malhotra, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, F. C. L. Hyland, D. W. Craig, A. Christoforides, N. Homer, T. Izatt, A. A. Kurdoglu, S. A. Sinari, K. Squire, S. T. Sherry, C. Xiao, J. Sebat, D. Antaki, M. Gujral, A. Noor, K. Ye, E. G. Burchard, R. D. Hernandez, C. R. Gignoux, D. Haussler, S. J. Katzman, W. James Kent, B. Howie, A. Ruiz-Linares, E. T. Dermitzakis, S. E. Devine, G. R. Abecasis, H. Min Kang, J. M. Kidd, T. Blackwell, S. Caron, W. Chen, S. Emery, L. Fritsche, C. Fuchsberger, G. Jun, B. Li, R. Lyons, C. Scheller, C. Sidore, S. Song, E. Sliwerska, D. Taliun, A. Tan, R. Welch, M. Kate Wing, X. Zhan, P. Awadalla, A. Hodgkinson, Y. Li, X. Shi, A. Quitadamo, G. Lunter, G. A. McVean, J. L. Marchini, S. Myers, C. Churchhouse, O. Delaneau, A. Gupta-Hinch, W. Kretzschmar, Z. Iqbal, I. Mathieson, A. Menelaou, A. Rimmer, D. K. Xifara, T. K. Oleksyk, Y. Fu, X. Liu, M. Xiong, L. Jorde, D. Witherspoon, J. Xing, E. E. Eichler, B. L. Browning, S. R. Browning, F. Hormozdiari, P. H. Sudmant, E. Khurana, R. M. Durbin, M. E. Hurles, C. Tyler-Smith, C. A. Albers, Q. Ayub, S. Balasubramaniam, Y. Chen, V. Colonna, P. Danecek, L. Jostins, T. M. Keane, S. McCarthy, K. Walter, Y. Xue, M. B. Gerstein, A. Abyzov, S. Balasubramanian, J. Chen, D. Clarke, Y. Fu, A. O. Harmanci, M. Jin, D. Lee, J. Liu, X. Jasmine Mu, J. Zhang, Y. Zhang, Y. Li, R. Luo, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G. T. Marth, E. P. Garrison, D. Kural, W.-P. Lee, A. N. Ward, J. Wu, M. Zhang, S. A. McCarroll, R. E. Handsaker, D. M. Altshuler, E. Banks, G. del Angel, G. Genovese, C. Hartl, H. Li, S. Kashin, J. C. Nemesh, K. Shakir, S. C. Yoon, J. Lihm, V. Makarov, J. Degenhardt, J. O. Korbel, M. H. Fritz, S. Meiers, B. Raeder, T. Rausch, A. M. Stütz, P. Flicek, F. Paolo Casale, L. Clarke, R. E. Smith, O. Stegle, X. Zheng-Bradley, D. R. Bentley, B. Barnes, R. Keira Cheetham, M. Eberle, S. Humphray, S. Kahn, L. Murray, R. Shaw, E.-W. Lameijer, M. A. Batzer, M. K. Konkel, J. A. Walker, L. Ding, I. Hall, K. Ye, P. Lacroute, C. Lee, E. Cerveira, A. Malhotra, J. Hwang, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, D. W.

Craig, N. Homer, D. Church, C. Xiao, J. Sebat, D. Antaki, V. Bafna, J. Michaelson, K. Ye, S. E. Devine, E. J. Gardner, G. R. Abecasis, J. M. Kidd, R. E. Mills, G. Dayama, S. Emery, G. Jun, X. Shi, A. Quitadamo, G. Lunter, G. A. McVean, K. Chen, X. Fan, Z. Chong, T. Chen, D. Witherspoon, J. Xing, E. E. Eichler, M. J. Chaisson, F. Hormozdiari, J. Huddleston, M. Malig, B. J. Nelson, P. H. Sudmant, N. F. Parrish, E. Khurana, M. E. Hurles, B. Blackburne, S. J. Lindsay, Z. Ning, K. Walter, Y. Zhang, M. B. Gerstein, A. Abyzov, J. Chen, D. Clarke, H. Lam, X. Jasmine Mu, C. Sisu, J. Zhang, Y. Zhang, R. A. Gibbs, F. Yu, M. Bainbridge, D. Challis, U. S. Evani, C. Kovar, J. Lu, D. Muzny, U. Nagaswamy, J. G. Reid, A. Sabo, J. Yu, X. Guo, W. Li, Y. Li, R. Wu, G. T. Marth, E. P. Garrison, W. Fung Leong, A. N. Ward, G. del Angel, M. A. DePristo, S. B. Gabriel, N. Gupta, C. Hartl, R. E. Poplin, A. G. Clark, J. L. Rodriguez-Flores, P. Flicek, L. Clarke, R. E. Smith, X. Zheng-Bradley, D. G. MacArthur, E. R. Mardis, R. Fulton, D. C. Koboldt, S. Gravel, C. D. Bustamante, D. W. Craig, A. Christoforides, N. Homer, T. Izatt, S. T. Sherry, C. Xiao, E. T. Dermitzakis, G. R. Abecasis, H. Min Kang, G. A. McVean, M. B. Gerstein, S. Balasubramanian, L. Habegger, H. Yu, P. Flicek, L. Clarke, F. Cunningham, I. Dunham, D. Zerbino, X. Zheng-Bradley, K. Lage, J. Berg Jespersen, H. Horn, S. B. Montgomery, M. K. DeGorter, E. Khurana, C. Tyler-Smith, Y. Chen, V. Colonna, Y. Xue, M. B. Gerstein, S. Balasubramanian, Y. Fu, D. Kim, A. Auton, A. Marcketta, R. Desalle, A. Narechania, M. A. Wilson Sayres, E. P. Garrison, R. E. Handsaker, S. Kashin, S. A. McCarroll, J. L. Rodriguez-Flores, P. Flicek, L. Clarke, X. Zheng-Bradley, Y. Erlich, M. Gymrek, T. Frederick Willems, C. D. Bustamante, F. L. Mendez, G. David Poznik, P. A. Underhill, C. Lee, E. Cerveira, A. Malhotra, M. Romanovitch, C. Zhang, G. R. Abecasis, L. Coin, H. Shao, D. Mittelman, C. Tyler-Smith, Q. Ayub, R. Banerjee, M. Cerezo, Y. Chen, T. W. Fitzgerald, S. Louzada, A. Massaia, S. McCarthy, G. R. Ritchie, Y. Xue, F. Yang, R. A. Gibbs, C. Kovar, D. Kalra, W. Hale, D. Muzny, J. G. Reid, J. Wang, X. Dan, X. Guo, G. Li, Y. Li, C. Ye, X. Zheng, D. M. Altshuler, P. Flicek, L. Clarke, X. Zheng-Bradley, D. R. Bentley, A. Cox, S. Humphray, S. Kahn, R. Sudbrak, M. W. Albrecht, M. Lienhard, D. Larson, D. W. Craig, T. Izatt, A. A. Kurdoglu, S. T. Sherry, C. Xiao, D. Haussler, G. R. Abecasis, G. A. McVean, R. M. Durbin, S. Balasubramaniam, T. M. Keane, S. McCarthy, J. Stalker, A. Chakravarti, B. M. Knoppers, G. R. Abecasis, K. C. Barnes, C. Beiswanger, E. G. Burchard, C. D. Bustamante, H. Cai, H. Cao, R. M. Durbin, N. P. Gerry, N. Gharani, R. A. Gibbs, C. R. Gignoux, S. Gravel, B. Henn, D. Jones, L. Jorde, J. S. Kaye, A. Keinan, A. Kent, A. Kerasidou, Y. Li, R. Mathias, G. A. McVean, A. Moreno-Estrada, P. N. Ossorio, M. Parker, A. M. Resch, C. N. Rotimi, C. D. Royal, K. Sandoval, Y. Su, R. Sudbrak, Z. Tian, S. Tishkoff, L. H. Toji, C. Tyler-Smith, M. Via, Y. Wang, H. Yang, L. Yang, J. Zhu, W. Bodmer, G. Bedoya, A. Ruiz-Linares, Z. Cai, Y. Gao, J. Chu, L. Peltonen, A. Garcia-Montero, A. Orfao, J. Dutil, J. C. Martinez-Cruzado, T. K. Oleksyk, K. C. Barnes, R. A. Mathias, A. Hennis, H. Watson, C. McKenzie, F. Qadri, R. LaRocque, P. C. Sabeti, J. Zhu, X. Deng, P. C. Sabeti, D. Asogun, O. Folarin, C. Happi, O. Omoniwa, M. Stremlau, R. Tariyal, M. Jallow, F. Sisay Joof, T. Corrah, K. Rockett, D. Kwiatkowski, J. Kooner, T. Tịnh Hiê'n, S. J. Dunstan, N. Thuy Hang, R. Fonnie, R. Garry, L. Kanneh, L. Moses, P. C. Sabeti, J. Schieffelin, D. S. Grant, C. Gallo, G. Poletti, D. Saleheen, A. Rasheed, L. D. Brooks, A. L. Felsenfeld, J. E. McEwen, Y. Vaydylevich, E. D. Green, A. Duncanson, M. Dunn, J. A. Schloss, J. Wang, H. Yang, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. Min Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis; 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi:10.1038/nature15393 Medline

70. B. Vernot, S. Tucci, J. Kelso, J. G. Schraiber, A. B. Wolf, R. M. Gittelman, M. Dannemann, S. Grote, R. C. McCoy, H. Norton, L. B. Scheinfeldt, D. A. Merriwether, G. Koki, J. S. Friedlaender, J. Wakefield, S. Pääbo, J. M. Akey, Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016). Medline

71. D. Reich, R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. F. Johnson, T. Maricic, J. M. Good, T. Marques-Bonet, C. Alkan, Q. Fu, S. Mallick, H. Li, M. Meyer, E. E. Eichler, M. Stoneking, M. Richards, S. Talamo, M. V. Shunkov, A. P. Derevianko, J.-J. Hublin, J. Kelso, M. Slatkin, S. Pääbo, Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010). doi:10.1038/nature09710 Medline

72. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011). doi:10.1093/molbev/msr048 Medline

73. F. M. T. A. Busing, E. Meijer, R. Van Der Leeden, Delete-m jackknife for unequal m. *Stat. Comput.* **9**, 3–8 (1999). doi:10.1023/A:1008800423698

74. P. Qin, M. Stoneking, Denisovan ancestry in East Eurasian and Native American populations. *Mol. Biol. Evol.* **32**, 2665–2674 (2015). Medline

75. M. Gallego Llorente, E. R. Jones, A. Eriksson, V. Siska, K. W. Arthur, J. W. Arthur, M. C. Curtis, J. T. Stock, M. Coltorti, P. Pieruccini, S. Stretton, F. Brock, T. Higham, Y. Park, M. Hofreiter, D. G. Bradley, J. Bhak, R. Pinhasi, A. Manica, Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015). doi:10.1126/science.aad2879 Medline

76. F. Sánchez-Quinto, L. R. Botigué, S. Civit, C. Arenas, M. C. Ávila-Arcos, C. D. Bustamante, D. Comas, C. Lalueza-Fox, North African populations carry the signature of admixture with Neandertals. *PLOS ONE* **7**, e47765 (2012). doi:10.1371/journal.pone.0047765 Medline

77. J. K. Pickrell, N. Patterson, P.-R. Loh, M. Lipson, B. Berger, M. Stoneking, B. Pakendorf, D. Reich, Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2632–2637 (2014). doi:10.1073/pnas.1313787111 Medline

78. S. Sankararaman, S. Mallick, N. Patterson, D. Reich, The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr. Biol.* **26**, 1241–1247 (2016). doi:10.1016/j.cub.2016.03.037 Medline

79. B. Y. Kim, K. E. Lohmueller, Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. *Am. J. Hum. Genet.* **96**, 454–461 (2015). doi:10.1016/j.ajhg.2014.12.029 Medline

80. B. Vernot, J. M. Akey, Complex history of admixture between modern humans and Neandertals. *Am. J. Hum. Genet.* **96**, 448–453 (2015). doi:10.1016/j.ajhg.2015.01.006 Medline

81. Q. Fu, C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R.

Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Pääbo, D. Reich, The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016). Medline

82. T. C. S. and Analysis Consortium; Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005). doi:10.1038/nature04072 Medline

83. R. A. Gibbs, J. Rogers, M. G. Katze, R. Bumgarner, G. M. Weinstock, E. R. Mardis, K. A. Remington, R. L. Strausberg, J. C. Venter, R. K. Wilson, M. A. Batzer, C. D. Bustamante, E. E. Eichler, M. W. Hahn, R. C. Hardison, K. D. Makova, W. Miller, A. Milosavljevic, R. E. Palermo, A. Siepel, J. M. Sikela, T. Attaway, S. Bell, K. E. Bernard, C. J. Buhay, M. N. Chandrabose, M. Dao, C. Davis, K. D. Delehaunty, Y. Ding, H. H. Dinh, S. Dugan-Rocha, L. A. Fulton, R. A. Gabisi, T. T. Garner, J. Godfrey, A. C. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S. N. Jhangiani, V. Joshi, Z. M. Khan, E. F. Kirkness, A. Cree, R. G. Fowler, S. Lee, L. R. Lewis, Z. Li, Y. S. Liu, S. M. Moore, D. Muzny, L. V. Nazareth, D. N. Ngo, G. O. Okwuonu, G. Pai, D. Parker, H. A. Paul, C. Pfannkoch, C. S. Pohl, Y.-H. Rogers, S. J. Ruiz, A. Sabo, J. Santibanez, B. W. Schneider, S. M. Smith, E. Sodergren, A. F. Svatek, T. R. Utterback, S. Vattathil, W. Warren, C. S. White, A. T. Chinwalla, Y. Feng, A. L. Halpern, L. W. Hillier, X. Huang, P. Minx, J. O. Nelson, K. H. Pepin, X. Qin, G. G. Sutton, E. Venter, B. P. Walenz, J. W. Wallis, K. C. Worley, S.-P. Yang, S. M. Jones, M. A. Marra, M. Rocchi, J. E. Schein, R. Baertsch, L. Clarke, M. Csürös, J. Glasscock, R. A. Harris, P. Havlak, A. R. Jackson, H. Jiang, Y. Liu, D. N. Messina, Y. Shen, H. X.-Z. Song, T. Wylie, L. Zhang, E. Birney, K. Han, M. K. Konkel, J. Lee, A. F. A. Smit, B. Ullmer, H. Wang, J. Xing, R. Burhans, Z. Cheng, J. E. Karro, J. Ma, B. Raney, X. She, M. J. Cox, J. P. Demuth, L. J. Dumas, S.-G. Han, J. Hopkins, A. Karimpour-Fard, Y. H. Kim, J. R. Pollack, T. Vinar, C. Addo-Quaye, J. Degenhardt, A. Denby, M. J. Hubisz, A. Indap, C. Kosiol, B. T. Lahn, H. A. Lawson, A. Marklein, R. Nielsen, E. J. Vallender, A. G. Clark, B. Ferguson, R. D. Hernandez, K. Hirani, H. Kehrer-Sawatzki, J. Kolb, S. Patil, L.-L. Pu, Y. Ren, D. G. Smith, D. A. Wheeler, I. Schenck, E. V. Ball, R. Chen, D. N. Cooper, B. Giardine, F. Hsu, W. J. Kent, A. Lesk, D. L. Nelson, W. E. O'brien, K. Prüfer, P. D. Stenson, J. C. Wallace, H. Ke, X.-M. Liu, P. Wang, A. P. Xiang, F. Yang, G. P. Barber, D. Haussler, D. Karolchik, A. D. Kern, R. M. Kuhn, K. E. Smith, A. S. Zwieg; Rhesus Macaque Genome Sequencing and Analysis Consortium, Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007). doi:10.1126/science.1139247 Medline

84. D. P. Locke, L. W. Hillier, W. C. Warren, K. C. Worley, L. V. Nazareth, D. M. Muzny, S.-P. Yang, Z. Wang, A. T. Chinwalla, P. Minx, M. Mitreva, L. Cook, K. D. Delehaunty, C. Fronick, H. Schmidt, L. A. Fulton, R. S. Fulton, J. O. Nelson, V. Magrini, C. Pohl, T. A. Graves, C. Markovic, A. Cree, H. H. Dinh, J. Hume, C. L. Kovar, G. R. Fowler, G. Lunter, S. Meader, A. Heger, C. P. Ponting, T. Marques-Bonet, C. Alkan, L. Chen, Z. Cheng, J. M. Kidd, E. E. Eichler, S. White, S. Searle, A. J. Vilella, Y. Chen, P. Flicek, J. Ma, B. Raney, B. Suh, R. Burhans, J. Herrero, D. Haussler, R. Faria, O. Fernando, F. Darré, D. Farré, E. Gazave, M. Oliva, A. Navarro, R. Roberto, O. Capozzi, N.

Archidiacono, G. Della Valle, S. Purgato, M. Rocchi, M. K. Konkel, J. A. Walker, B. Ullmer, M. A. Batzer, A. F. A. Smit, R. Hubley, C. Casola, D. R. Schrider, M. W. Hahn, V. Quesada, X. S. Puente, G. R. Ordoñez, C. López-Otín, T. Vinar, B. Brejova, A. Ratan, R. S. Harris, W. Miller, C. Kosiol, H. A. Lawson, V. Taliwal, A. L. Martins, A. Siepel, A. Roychoudhury, X. Ma, J. Degenhardt, C. D. Bustamante, R. N. Gutenkunst, T. Mailund, J. Y. Dutheil, A. Hobolth, M. H. Schierup, O. A. Ryder, Y. Yoshinaga, P. J. de Jong, G. M. Weinstock, J. Rogers, E. R. Mardis, R. A. Gibbs, R. K. Wilson, Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529–533 (2011). doi:10.1038/nature09687 Medline

85. A. Scally, J. Y. Dutheil, L. W. Hillier, G. E. Jordan, I. Goodhead, J. Herrero, A. Hobolth, T. Lappalainen, T. Mailund, T. Marques-Bonet, S. McCarthy, S. H. Montgomery, P. C. Schwalie, Y. A. Tang, M. C. Ward, Y. Xue, B. Yngvadottir, C. Alkan, L. N. Andersen, Q. Ayub, E. V. Ball, K. Beal, B. J. Bradley, Y. Chen, C. M. Clee, S. Fitzgerald, T. A. Graves, Y. Gu, P. Heath, A. Heger, E. Karakoc, A. Kolb-Kokocinski, G. K. Laird, G. Lunter, S. Meader, M. Mort, J. C. Mullikin, K. Munch, T. D. O'Connor, A. D. Phillips, J. Prado-Martinez, A. S. Rogers, S. Sajjadian, D. Schmidt, K. Shaw, J. T. Simpson, P. D. Stenson, D. J. Turner, L. Vigilant, A. J. Vilella, W. Whitener, B. Zhu, D. N. Cooper, P. de Jong, E. T. Dermitzakis, E. E. Eichler, P. Flicek, N. Goldman, N. I. Mundy, Z. Ning, D. T. Odom, C. P. Ponting, M. A. Quail, O. A. Ryder, S. M. Searle, W. C. Warren, R. K. Wilson, M. H. Schierup, J. Rogers, C. Tyler-Smith, R. Durbin, Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175 (2012). doi:10.1038/nature10842 Medline

86. K. Prüfer, K. Munch, I. Hellmann, K. Akagi, J. R. Miller, B. Walenz, S. Koren, G. Sutton, C. Kodira, R. Winer, J. R. Knight, J. C. Mullikin, S. J. Meader, C. P. Ponting, G. Lunter, S. Higashino, A. Hobolth, J. Dutheil, E. Karakoç, C. Alkan, S. Sajjadian, C. R. Catacchio, M. Ventura, T. Marques-Bonet, E. E. Eichler, C. André, R. Atencia, L. Mugisha, J. Junhold, N. Patterson, M. Siebauer, J. M. Good, A. Fischer, S. E. Ptak, M. Lachmann, D. E. Symer, T. Mailund, M. H. Schierup, A. M. Andrés, J. Kelso, S. Pääbo, The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012). Medline

87. A. Kong, G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson, A. Jonasdottir, G. B. Walters, A. Jonasdottir, A. Gylfason, K. T. Kristinsson, S. A. Gudjonsson, M. L. Frigge, A. Helgason, U. Thorsteinsdottir, K. Stefansson, Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010). doi:10.1038/nature09525 Medline

88. P. R. Staab, S. Zhu, D. Metzler, G. Lunter, scrm: Efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics* **31**, 1680–1682 (2015). doi:10.1093/bioinformatics/btu861 Medline

89. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017). doi:10.1038/ng.3748 Medline

90. S. Sankararaman, N. Patterson, H. Li, S. Pääbo, D. Reich, The date of interbreeding between Neandertals and modern humans. *PLOS Genet.* **8**, e1002947 (2012). doi:10.1371/journal.pgen.1002947 Medline

91. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson, N. Rohland, I. Lazaridis, B. Nickel, B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich, S. Pääbo, An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015). doi:10.1038/nature14558 Medline

92. P. H. Sudmant, J. Huddleston, C. R. Catacchio, M. Malig, L. W. Hillier, C. Baker, K. Mohajeri, I. Kondova, R. E. Bontrop, S. Persengiev, F. Antonacci, M. Ventura, J. Prado-Martinez, T. Marques-Bonet, E. E. Eichler; Great Ape Genome Project, Evolution and diversity of copy number variation in the great ape lineage. *Genome Res.* **23**, 1373–1382 (2013). doi:10.1101/gr.158543.113 Medline

93. P. H. Sudmant, S. Mallick, B. J. Nelson, F. Hormozdiari, N. Krumm, J. Huddleston, B. P. Coe, C. Baker, S. Nordenfelt, M. Bamshad, L. B. Jorde, O. L. Posukh, H. Sahakyan, W. S. Watkins, L. Yepiskoposyan, M. S. Abdullah, C. M. Bravi, C. Capelli, T. Hervig, J. T. S. Wee, C. Tyler-Smith, G. van Driem, I. G. Romero, A. R. Jha, S. Karachanak-Yankova, D. Toncheva, D. Comas, B. Henn, T. Kivisild, A. Ruiz-Linares, A. Sajantila, E. Metspalu, J. Parik, R. Villems, E. B. Starikovskaya, G. Ayodo, C. M. Beall, A. Di Rienzo, M. F. Hammer, R. Khusainova, E. Khusnutdinova, W. Klitz, C. Winkler, D. Labuda, M. Metspalu, S. A. Tishkoff, S. Dryomov, R. Sukernik, N. Patterson, D. Reich, E. E. Eichler, Global diversity, population stratification, and selection of human copy-number variation. *Science* **349**, aab3761 (2015). doi:10.1126/science.aab3761 Medline

94. P. H. Sudmant, T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. H. Fritz, M. K. Konkel, A. Malhotra, A. M. Stütz, X. Shi, F. P. Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. J. Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E.-W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler, J. O. Korbel; 1000 Genomes Project Consortium, An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015). doi:10.1038/nature15394 Medline

95. M. Y. Dennis, X. Nuttle, P. H. Sudmant, F. Antonacci, T. A. Graves, M. Nefedov, J. A. Rosenfeld, S. Sajjadian, M. Malig, H. Kotkiewicz, C. J. Curry, S. Shafer, L. G. Shaffer, P. J. de Jong, R. K. Wilson, E. E. Eichler, Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* **149**, 912–922 (2012). doi:10.1016/j.cell.2012.03.033 Medline

96. M. Florio, M. Albert, E. Taverna, T. Namba, H. Brandl, E. Lewitus, C. Haffner, A. Sykes, F. K. Wong, J. Peters, E. Guhr, S. Klemroth, K. Prüfer, J. Kelso, R. Naumann, I. Nüsslein, A. Dahl, R. Lachmann, S. Pääbo, W. B. Huttner, Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **347**, 1465–1470 (2015). doi:10.1126/science.aaa1975 Medline

97. X. Nuttle, G. Giannuzzi, M. H. Duyzend, J. G. Schraiber, I. Narvaiza, P. H. Sudmant, O. Penn, G. Chiatante, M. Malig, J. Huddleston, C. Benner, F. Camponeschi, S. Ciofi-

Baffoni, H. A. F. Stessman, M. C. N. Marchetto, L. Denman, L. Harshman, C. Baker, A. Raja, K. Penewit, N. Janke, W. J. Tang, M. Ventura, L. Banci, F. Antonacci, J. M. Akey, C. T. Amemiya, F. H. Gage, A. Reymond, E. E. Eichler, Emergence of a *Homo sapiens*-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature* **536**, 205–209 (2016). doi:10.1038/nature19075 Medline

98. F. Hach, F. Hormozdiari, C. Alkan, F. Hormozdiari, I. Birol, E. E. Eichler, S. C. Sahinalp, mrsFAST: A cache-oblivious algorithm for short-read mapping. *Nat. Methods* **7**, 576–577 (2010). doi:10.1038/nmeth0810-576 Medline

99. T. Fiskerstrand, D. H'mida-Ben Brahim, S. Johansson, A. M'zahem, B. I. Haukanes, N. Drouot, J. Zimmermann, A. J. Cole, C. Vedeler, C. Bredrup, M. Assoum, M. Tazir, T. Klockgether, A. Hamri, V. M. Steen, H. Boman, L. A. Bindoff, M. Koenig, P. M. Knappskog, Mutations in ABHD12 cause the neurodegenerative disease PHARC: An inborn error of endocannabinoid metabolism. *Am. J. Hum. Genet.* **87**, 410–417 (2010). doi:10.1016/j.ajhg.2010.08.002 Medline

100. W. M. Knowlton, D. D. McKemy, TRPM8: From cold to cancer, peppermint to pain. *Curr. Pharm. Biotechnol.* **12**, 68–77 (2011). doi:10.2174/138920111793937961 Medline

101. M. Ugarte, C. Pérez-Cerdá, P. Rodríguez-Pombo, L. R. Desviat, B. Pérez, E. Richard, S. Muro, E. Campeau, T. Ohura, R. A. Gravel, Overview of mutations in the PCCA and PCCB genes causing propionic acidemia. *Hum. Mutat.* **14**, 275–282 (1999). doi:10.1002/(SICI)1098-1004(199910)14:4<275:AID-HUMU1>3.0.CO;2-N Medline

102. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011). doi:10.1093/bioinformatics/btr330 Medline

103. D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, H. Parkinson, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42** (D1), D1001–D1006 (2014). doi:10.1093/nar/gkt1229 Medline

104. R. M. Kuhn, D. Haussler, W. J. Kent, The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013). doi:10.1093/bib/bbs038 Medline

105. E. Huerta-Sánchez, X. Jin, Z. Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, J. Huasang, J. Luosang, Z. X. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, R. Nielsen, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014). doi:10.1038/nature13408 Medline

106. C. J. Willer, E. M. Schmidt, S. Sengupta, G. M. Peloso, S. Gustafsson, S. Kanoni, A. Ganna, J. Chen, M. L. Buchkovich, S. Mora, J. S. Beckmann, J. L. Bragg-Gresham, H. Y. Chang, A. Demirkan, H. M. Den Hertog, R. Do, L. A. Donnelly, G. B. Ehret, T. Esko, M. F. Feitosa, T. Ferreira, K. Fischer, P. Fontanillas, R. M. Fraser, D. F. Freitag, D. Gurdasani, K. Heikkilä, E. Hyppönen, A. Isaacs, A. U. Jackson, Å. Johansson, T. Johnson, M. Kaakinen, J. Kettunen, M. E. Kleber, X. Li, J. Luan, L. P. Lyytikäinen, P. K. E. Magnusson, M. Mangino, E. Mihailov, M. E. Montasser, M. Müller-Nurasyid, I. M. Nolte, J. R. O'Connell, C. D. Palmer, M. Perola, A. K. Petersen, S. Sanna, R. Saxena, S. K. Service, S. Shah, D. Shungin, C. Sidore, C. Song, R. J. Strawbridge, I. Surakka, T.

Tanaka, T. M. Teslovich, G. Thorleifsson, E. G. Van den Herik, B. F. Voight, K. A. Volcik, L. L. Waite, A. Wong, Y. Wu, W. Zhang, D. Absher, G. Asiki, I. Barroso, L. F. Been, J. L. Bolton, L. L. Bonnycastle, P. Brambilla, M. S. Burnett, G. Cesana, M. Dimitriou, A. S. F. Doney, A. Döring, P. Elliott, S. E. Epstein, G. Ingi Eyjolfsson, B. Gigante, M. O. Goodarzi, H. Grallert, M. L. Gravito, C. J. Groves, G. Hallmans, A. L. Hartikainen, C. Hayward, D. Hernandez, A. A. Hicks, H. Holm, Y. J. Hung, T. Illig, M. R. Jones, P. Kaleebu, J. J. P. Kastelein, K. T. Khaw, E. Kim, N. Klopp, P. Komulainen, M. Kumari, C. Langenberg, T. Lehtimäki, S. Y. Lin, J. Lindström, R. J. F. Loos, F. Mach, W. L. McArdle, C. Meisinger, B. D. Mitchell, G. Müller, R. Nagaraja, N. Narisu, T. V. M. Nieminen, R. N. Nsubuga, I. Olafsson, K. K. Ong, A. Palotie, T. Papamarkou, C. Pomilla, A. Pouta, D. J. Rader, M. P. Reilly, P. M. Ridker, F. Rivadeneira, I. Rudan, A. Ruokonen, N. Samani, H. Scharnagl, J. Seeley, K. Silander, A. Stančáková, K. Stirrups, A. J. Swift, L. Tiret, A. G. Uitterlinden, L. J. van Pelt, S. Vedantam, N. Wainwright, C. Wijmenga, S. H. Wild, G. Willemsen, T. Wilsgaard, J. F. Wilson, E. H. Young, J. H. Zhao, L. S. Adair, D. Arveiler, T. L. Assimes, S. Bandinelli, F. Bennett, M. Bochud, B. O. Boehm, D. I. Boomsma, I. B. Borecki, S. R. Bornstein, P. Bovet, M. Burnier, H. Campbell, A. Chakravarti, J. C. Chambers, Y. I. Chen, F. S. Collins, R. S. Cooper, J. Danesh, G. Dedoussis, U. de Faire, A. B. Feranil, J. Ferrières, L. Ferrucci, N. B. Freimer, C. Gieger, L. C. Groop, V. Gudnason, U. Gyllensten, A. Hamsten, T. B. Harris, A. Hingorani, J. N. Hirschhorn, A. Hofman, G. K. Hovingh, C. A. Hsiung, S. E. Humphries, S. C. Hunt, K. Hveem, C. Iribarren, M. R. Järvelin, A. Jula, M. Kähönen, J. Kaprio, A. Kesäniemi, M. Kivimaki, J. S. Kooner, P. J. Koudstaal, R. M. Krauss, D. Kuh, J. Kuusisto, K. O. Kyvik, M. Laakso, T. A. Lakka, L. Lind, C. M. Lindgren, N. G. Martin, W. März, M. I. McCarthy, C. A. McKenzie, P. Meneton, A. Metspalu, L. Moilanen, A. D. Morris, P. B. Munroe, I. Njølstad, N. L. Pedersen, C. Power, P. P. Pramstaller, J. F. Price, B. M. Psaty, T. Quertermous, R. Rauramaa, D. Saleheen, V. Salomaa, D. K. Sanghera, J. Saramies, P. E. H. Schwarz, W. H. Sheu, A. R. Shuldiner, A. Siegbahn, T. D. Spector, K. Stefansson, D. P. Strachan, B. O. Tayo, E. Tremoli, J. Tuomilehto, M. Uusitupa, C. M. van Duijn, P. Vollenweider, L. Wallentin, N. J. Wareham, J. B. Whitfield, B. H. R. Wolffenbuttel, J. M. Ordovas, E. Boerwinkle, C. N. A. Palmer, U. Thorsteinsdottir, D. I. Chasman, J. I. Rotter, P. W. Franks, S. Ripatti, L. A. Cupples, M. S. Sandhu, S. S. Rich, M. Boehnke, P. Deloukas, S. Kathiresan, K. L. Mohlke, E. Ingelsson, G. R. Abecasis; Global Lipids Genetics Consortium, Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013). doi:10.1038/ng.2797 Medline

107. D. Mozaffarian, E. K. Kabagambe, C. O. Johnson, R. N. Lemaitre, A. Manichaikul, Q. Sun, M. Foy, L. Wang, H. Wiener, M. R. Irvin, S. S. Rich, H. Wu, M. K. Jensen, D. I. Chasman, A. Y. Chu, M. Fornage, L. Steffen, I. B. King, B. McKnight, B. M. Psaty, L. Djoussé, I. Y.-D. Chen, J. H. Wu, D. S. Siscovick, P. M. Ridker, M. Y. Tsai, E. B. Rimm, F. B. Hu, D. K. Arnett, Genetic loci associated with circulating phospholipid trans fatty acids: A meta-analysis of genome-wide association studies from the CHARGE Consortium. *Am. J. Clin. Nutr.* **101**, 398–406 (2015). doi:10.3945/ajcn.114.094557 Medline

108. S. M. Lutz, M. H. Cho, K. Young, C. P. Hersh, P. J. Castaldi, M.-L. McDonald, E. Regan, M. Mattheisen, D. L. DeMeo, M. Parker, M. Foreman, B. J. Make, R. L. Jensen, R. Casaburi, D. A. Lomas, S. P. Bhatt, P. Bakke, A. Gulsvik, J. D. Crapo, T. H. Beaty, N. M. Laird, C. Lange, J. E. Hokanson, E. K. Silverman; ECLIPSE Investigators; COPDGene Investigators, A genome-wide association study identifies risk loci for spirometric

measures among smokers of European and African ancestry. *BMC Genet.* **16**, 138 (2015). doi:10.1186/s12863-015-0299-4 Medline

109. P. L. Chen, S.-R. Shih, P.-W. Wang, Y.-C. Lin, C.-C. Chu, J.-H. Lin, S.-C. Chen, C.-C. Chang, T.-S. Huang, K. S. Tsai, F.-Y. Tseng, C.-Y. Wang, J.-Y. Lu, W.-Y. Chiu, C.-C. Chang, Y.-H. Chen, Y.-T. Chen, C. S.-J. Fann, W.-S. Yang, T.-C. Chang, Genetic determinants of antithyroid drug-induced agranulocytosis by human leukocyte antigen genotyping and genome-wide association study. *Nat. Commun.* **6**, 7633 (2015). doi:10.1038/ncomms8633 Medline

110. G. Lauc, J. E. Huffman, M. Pučić, L. Zgaga, B. Adamczyk, A. Mužinić, M. Novokmet, O. Polašek, O. Gornik, J. Krištić, T. Keser, V. Vitart, B. Scheijen, H.-W. Uh, M. Molokhia, A. L. Patrick, P. McKeigue, I. Kolčić, I. K. Lukić, O. Swann, F. N. van Leeuwen, L. R. Ruhaak, J. J. Houwing-Duistermaat, P. E. Slagboom, M. Beekman, A. J. M. de Craen, A. M. Deelder, Q. Zeng, W. Wang, N. D. Hastie, U. Gyllensten, J. F. Wilson, M. Wuhrer, A. F. Wright, P. M. Rudd, C. Hayward, Y. Aulchenko, H. Campbell, I. Rudan, Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLOS Genet.* **9**, e1003225 (2013). doi:10.1371/journal.pgen.1003225 Medline

111. P. F. Sullivan, D. Lin, J.-Y. Tzeng, E. van den Oord, D. Perkins, T. S. Stroup, M. Wagner, S. Lee, F. A. Wright, F. Zou, W. Liu, A. M. Downing, J. Lieberman, S. L. Close, Genomewide association for schizophrenia in the CATIE study: Results of stage 1. *Mol. Psychiatry* **13**, 570–584 (2008). doi:10.1038/mp.2008.25 Medline

112. K. Aberg, D. E. Adkins, J. Bukszár, B. T. Webb, S. N. Caroff, D. D. Miller, J. Sebat, S. Stroup, A. H. Fanous, V. I. Vladimirov, J. L. McClay, J. A. Lieberman, P. F. Sullivan, E. J. C. G. van den Oord, Genomewide association study of movement-related adverse antipsychotic effects. *Biol. Psychiatry* **67**, 279–282 (2010). doi:10.1016/j.biopsych.2009.08.036 Medline

113. Y. Okada, D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, S. Yoshida, R. R. Graham, A. Manoharan, W. Ortmann, T. Bhangale, J. C. Denny, R. J. Carroll, A. E. Eyler, J. D. Greenberg, J. M. Kremer, D. A. Pappas, L. Jiang, J. Yin, L. Ye, D.-F. Su, J. Yang, G. Xie, E. Keystone, H.-J. Westra, T. Esko, A. Metspalu, X. Zhou, N. Gupta, D. Mirel, E. A. Stahl, D. Diogo, J. Cui, K. Liao, M. H. Guo, K. Myouzen, T. Kawaguchi, M. J. H. Coenen, P. L. C. M. van Riel, M. A. F. J. van de Laar, H.-J. Guchelaar, T. W. J. Huizinga, P. Dieudé, X. Mariette, S. L. Bridges Jr., A. Zhernakova, R. E. M. Toes, P. P. Tak, C. Miceli-Richard, S.-Y. Bang, H.-S. Lee, J. Martin, M. A. Gonzalez-Gay, L. Rodriguez-Rodriguez, S. Rantapää-Dahlqvist, L. Ärlestig, H. K. Choi, Y. Kamatani, P. Galan, M. Lathrop, S. Eyre, J. Bowes, A. Barton, N. de Vries, L. W. Moreland, L. A. Criswell, E. W. Karlson, A. Taniguchi, R. Yamada, M. Kubo, J. S. Liu, S.-C. Bae, J. Worthington, L. Padyukov, L. Klareskog, P. K. Gregersen, S. Raychaudhuri, B. E. Stranger, P. L. De Jager, L. Franke, P. M. Visscher, M. A. Brown, H. Yamanaka, T. Mimori, A. Takahashi, H. Xu, T. W. Behrens, K. A. Siminovitch, S. Momohara, F. Matsuda, K. Yamamoto, R. M. Plenge; RACI consortium; GARNET consortium, Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014). doi:10.1038/nature12873 Medline

114. T. D. Wade, S. Gordon, S. Medland, C. M. Bulik, A. C. Heath, G. W. Montgomery, N. G. Martin, Genetic variants associated with disordered eating. *Int. J. Eat. Disord.* **46**, 594–608 (2013). doi:10.1002/eat.22133 Medline

115. C. S. Fox, Y. Liu, C. C. White, M. Feitosa, A. V. Smith, N. Heard-Costa, K. Lohman, A. D. Johnson, M. C. Foster, D. M. Greenawalt, P. Griffin, J. Ding, A. B. Newman, F. Tylavsky, I. Miljkovic, S. B. Kritchevsky, L. Launer, M. Garcia, G. Eiriksdottir, J. J. Carr, V. Gudnason, T. B. Harris, L. A. Cupples, I. B. Borecki; GIANT Consortium; MAGIC Consortium; GLGC Consortium, Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLOS Genet.* **8**, e1002695 (2012). doi:10.1371/journal.pgen.1002695 Medline

116. D. Anderson, B. J. Holt, C. E. Pennell, P. G. Holt, P. H. Hart, J. M. Blackwell, Genome-wide association study of vitamin D levels in children: Replication in the Western Australian Pregnancy Cohort (Raine) study. *Genes Immun.* **15**, 578–583 (2014). doi:10.1038/gene.2014.52 Medline

117. S. L. Park, S. G. Carmella, M. Chen, Y. Patel, D. O. Stram, C. A. Haiman, L. Le Marchand, S. S. Hecht, Mercapturic acids derived from the toxicants acrolein and crotonaldehyde in the urine of cigarette smokers from five ethnic groups with differing risks for lung cancer. *PLOS ONE* **10**, e0124841 (2015). doi:10.1371/journal.pone.0124841 Medline

118. S. R. Williams, F.-C. Hsu, K. L. Keene, W.-M. Chen, S. Nelson, A. M. Southerland, E. B. Madden, B. Coull, S. M. Gogarten, K. L. Furie, G. Dzhivhuho, J. L. Rowles, P. Mehndiratta, R. Malik, J. Dupuis, H. Lin, S. Seshadri, S. S. Rich, M. M. Sale, B. B. Worrall; METASTROKE, The Genomics and Randomized Trials Network (GARNET) Collaborative Research Group; METASTROKE The Genomics and Randomized Trials Network GARNET Collaborative Research Group, Shared genetic susceptibility of vascular-related biomarkers with ischemic and recurrent stroke. *Neurology* **86**, 351–359 (2016). doi:10.1212/WNL.0000000000002319 Medline