

Supplementary Information

Supplementary Methods

Fossil assay for biomolecular preservation

In order to evaluate the preservation of DNA in Neandertal fossils from different sites in Europe and Asia, we removed about 10 mg of bone material from each tested fossil and analyzed the overall content and degree of racimization for eight amino acids as described⁵¹. The eight amino acids analyzed were: D- and L-alanine, glycine, D and L-aspartic acid, serine, glutamic acid, valine, D- and L-leucine and isoleucine. Fossils showing a overall content of amino acids of more than 20.000 parts per million (ppm) and an aspartic acid racemization (i.e., the stereoisomeric D/L ratio) less than 0.10 have an increased probability for endogenous DNA to be preserved⁵¹⁻⁵⁴ and were used for DNA extraction.

PCR assay for the ratio of modern human versus Neandertal mtDNA

DNA was extracted in a laboratory dedicated to ancient DNA work. We extracted nucleic acids from bone powder as described previously⁵⁵ and eluted the DNA in a final volume of 100 µl 1x TE. We used two different primer pairs (Supp. Table 1) that amplify 63 bp and 119 bp of human and Neandertal mtDNA from the HVR1 with equal efficiencies. All previously published Neandertal HVR1 sequences carry combinations of substitutions in these fragments that are not found in combination in modern humans and thus allow us to discriminate modern human from Neandertal mtDNA. To conserve the extract we used 5 µl of Neandertal DNA for a two-step multiplex PCR⁵⁶ in a total volume of 20 µl containing both primer pairs. In total 60 cycles of PCR were carried out, 27 in the first and 33 in the second step. The final concentrations of reagents in the amplification reactions were as described⁵⁶ except that the annealing temperature was 56°C. Amplification products of the correct size were cloned using the TOPO TA cloning Kit (Invitrogen) and a minimum of 100 clones were sequenced

on an ABI3730 capillary sequencer (Applied Biosystems). For all reactions no-template PCR controls and blank extraction controls were carried along. All control reactions were negative. For the 10 products obtained, using different Neandertal extracts, we estimated the ratio of Neandertal DNA clone sequences to contaminating modern human DNA sequences (Fig. 1).

PCR confirmation of mitochondrial 454 sequences

To confirm the 39 substitutions found in the 41 sequences from 454 with the closest hit to modern human mtDNA, 29 primer pairs, varying in length between 51 and 98bp (including primers), were designed based on modern human mitochondrial sequence data. Five μ l of the Vi-80 Neandertal DNA extract were used in triplicates for a two-step multiplex PCR with the same concentrations as described above, but using 29 primer pairs in the multiplex primer mix and an annealing temperature of 52°C. We obtained 75 PCR products for 25 primer pairs. For each primer pair three no-template controls were carried out. Out of 87 controls, 21 were positive, a frequent finding when short human mtDNA fragments are amplified by PCR. None of the four primer pairs that gave no PCR product using the Neandertal extract yielded any positive no-template control suggesting that these primer pairs are of low sensitivity or functionality. For each primer pair, one product was cloned as described and at least six clones of each product sequenced. All products obtained from the negative controls were similarly analyzed and all clone sequences from these were found to be different from the 454 Neandertal sequences. The consensus sequences from Vi-80 products were compared to the corresponding 454 fragments. For two primer pairs, multiple substitutions in the clones sequenced suggested large amounts of nucleotide misincorporations. Subsequent cloning and sequencing of the second and third products for these two primer pairs resulted in interpretable consensus sequences.

	primer	Sequence	product length
a) contamination test	16219FWD	ACAAGCAAGCACAGCAATCA	
	16245RVS	GGGGTGTCTTTGGAGTTG	63bp
	16022FWD	CTAAGATTCTAATTTAACTATTCTCT	
	16095RVS	GTGGCTGGCAGTAATGTACG	119bp
b) Confirmation of mtDNA sequences from 454	384FWD	CTAACCAGATTTCAAATTTTATC	
	463RVS	GAGGGGAAAATAATGTGTTAG	80bp
	1199FWD	GAGGAGCCTGTTCTGTAATC	
	1260_RVS	TATAGGCTGAGCAAGAGGTG	60bp
	1382_FWD	AACTACGATAGCCCTTATGA	
	1442_RVS	CTCTTAGTTTACTGCTAAATCCA	61bp
	2668FWD	TTAACCAGTGAAATTGACCT	
	2742_RVS	TAAAGCTCCATAGGGTCTTC	75bp
	3370_FWD	CTTACCGAACGAAAATTCTA	
	3449_RVS	GGTTGTAGTAGCCCGTAGG	80bp
	3920_FWD	CCGAACTAGTCTCAGGCTT	
	3988_RVS	ATTCGGCTATGAAGAATAGG	69bp
	4577_FWD	CATGCTAGCTTTTATTCCAG	
	4627_RVS	GTGGAACGAGGGTTTATTT	51bp
	5145FWD	CACGACCCTACTACTATCTCG	
	5208RVS	GGATGGAATTAAGGGTGTTA	64bp
	5476FWD	CTCCTACCTATCTCCCCTTT	
	5546RVS	GAAGGCTCTTGGTCTGTATT	71bp
	5799FWD	CAATTCAATATGAAAATCACCT	
	5868RVS	GGACTGTAAATCTAAAGACAGG	70bp
	5940_FWD	AAAGACATTGGAACACTATACCT	
	5997_RVS	CTGTGCCTAGGACTCCA	58bp
	7470FWD	CCCAAAGCTGGTTTCAAG	
	7554RVS	TTTGACAAAAGTTATGAAATGG	85bp
	7834_FWD	CCTTTACATAACAGACGAGGT	
	7891_RVS	GTGGCCAATTGATTTGAT	58bp
	8923_FWD	ACACCTACACCCCTTATCC	
	8977_RVS	TGAGTAGGCTGATGGTTTC	55bp
	8965_FWD	ATCAGCCTACTCATTCAACC	
	9021_RVS	AGTAATGTTAGCGTTAGGC	57bp
	9846_FWD	TTTCTCACTATCTGCTTCA	
	9903RVS	GTGATGTTTGGATGTAAAGTG	58bp
	9922FWD	GCCTGATACTGGCATTIT	
	9973RVS	TGGAGACATACAGAAATAGTCAA	52bp
	10080_FWD	ATAATCAACACCCTCCTAGC	
	10139_RVS	GAGTTGTGGTAGTCAAAATGTAA	60bp
10643_FWD	TGTGCCTATTGCCATACTAG		
10740_RVS	GTACGTAGTCTAGGCCATATGT	98bp	
10782_FWD	CAATTATATTACTACCACTGACATG		
10845_RVS	GTGGTTGTGTTGATTCAAAT	64bp	
12124_FWD	CGGGTTTTCTCTTGTAAT		
12188_RVS	AGCCTCTGTTGTCAGATTC	65bp	

12607_FWD	CCTGTAGCATTGTTTCGTTAC	
12671_RVS	GGGTCTGAGTTTATATATCACAGT	65bp
13175FWD	TTAGGCGCTATCACCACT	
13233RVS	TTGATGTCATTTTGTGTAAGG	59bp
13779_FWD	AATCCCCCTCTACCTAAAAAC	
13831_RVS	GAAGTCCTAGGAAAGTGACAG	53bp
13855FWD	CCTAACCAACAACTTAAAATAAA	
13919rvs	ATCCGAGTATGTTGGAGAAA	65bp
14671FWD	GCAAAATTAACCCCTAATA	
14811RVS	GGAGGTCGATGAATGAGT	51bp
14910FWD	TACTACCAGACGCCTCA	
14968RVS	TTTACGTCTCGAGTGATGTG	59bp
15204FWD	ATCCCATACATTGGGACA	
15253RVS	GAGTAGCCTCCTCAGATTCA	50bp
15595FWD	CGTCCCTAACAACTAGGAG	
15651RVS	GCTAGGATGAGGATGGATAG	57bp

Supplementary Table 1: a) Primer sequences that were used for the PCR assay to test the relative rate of modern human contamination on different Neandertal fossils, **b)** Primer sequences used to amplify mtDNA fragments where the 454 sequences revealed differences to the consensus sequence of 311 modern human mt genomes

Sample preparation and sequencing on the 454 platform

The ancient DNA was processed into a single-stranded library suitable for sequencing on the GS-20 instrument using a protocol similar to that described recently by Margulies et al.⁵⁷.

The initial step in the previously described process is the random cleavage of genomic DNA into approximately 200 to 600 base pair fragments, compatible size-wise with emulsion-based amplification (emPCR) and sequencing. This step was omitted, since ancient DNA is typically severely fragmented and already of a length suitable for sequencing. In addition, by omitting the fragmentation step a bias in favour of the ancient DNA fragments was achieved, as high molecular weight, contaminating modern DNA is not amenable as such for sequencing on the platform. The ancient DNA was therefore directly treated with T4 polynucleotide kinase (1units/ μ l) and T4 DNA polymerase (0.3units/ μ l in the presence of 1X NEB Buffer 4, 0.1mg/ml BSA, 1mM ATP, 400 μ M dNTP's (NEB, Beverly, MA) under

standard conditions to generate ends suitable for blunt-end ligation. The blunt-ended DNA was purified using a MinElute column (Qiagen, Valencia, CA), whereafter 20pmoles of unphosphorylated, A and B adapters, needed for subsequent emulsion PCR and sequencing primer annealing, were ligated to each end of the fragments using the Quick Ligation kit (NEB). Following ligation, the DNA was purified using a MinElute column and immobilized onto M-270 streptavidin beads (Dynal, Carlsbad, CA) through a biotin moiety present on the B adapter. The immobilized fragments were subjected to a fill-in reaction using Bst polymerase (0.48 units/ μ l) in the presence of 1X Thermopol Buffer and 400 μ M dNTPs (NEB) to generate intact DNA sequencing templates flanked by adapters at both ends. Finally, the non-biotinylated strand was eluted off the beads using mild alkaline conditions (125 mM NaOH) and concentrated into a final volume of 15 μ l TE using a MinElute column. The eluted material represent the final sequencing-ready DNA library, enriched in A-B adapted fragments (A-A and B-B fragments either fail to bind or fail to be released from the capture beads and will be present in negligible amounts in the final library). The isolation of single-stranded DNA at the final library preparation step also ensures that the identity of the single ancient DNA molecule used as template in the emulsion PCR is retained; sequencing from the A adapter produces sequence data corresponding to the exact strand of the original ancient DNA template molecule.

Emulsion PCR was performed as described previously⁵⁷. The optimal amount of DNA library input per emulsion was determined empirically through a series of small scale titrations. It was determined that 1/950 of the final library was optimal per single 600,000 bead emulsion. A total of 32 individual emulsions were performed to generate enough beads for a full 70x75 PicoTiterPlate. After completion of the emulsion PCR 735,000 template-carrying beads were selectively isolated and all were loaded onto a 70x75 PicoTiterPlate™. The sequencing run was performed as previously described for a total of 42 cycles of

nucleotide flows (flow order TACG), using a small amount of adenovirus DNA library for signal threshold determination. Signal processing was performed using 454 Life Sciences' software release 1.0.52. Base-called reads were trimmed and filtered for quality and converted into FASTA format. A total of 197,725 quality filtered reads were thus retrieved for further analysis. A second round of emPCR and sequencing was also performed with comparable results, generating an additional 171,905 quality filtered reads.

The processing of ancient DNA was performed on the 454 GS20 platform utilizing the reagents and kits for library preparation, emPCR, and sequencing that accompany this platform.

For blunt-end ligation of ancient DNA, the following adapter designs were used:

UA3A(TCAG)

5' – CCATCT CAT CCC TGC GTG TCC CAT CTG TTC CCT CCC TGT CTC AG –3'

UA3Aprime_12-mer

5' - CTGAGACAGGGA – 3'

UA3B(TCAG)

/5' - Bio/CCTATC CCC TGT GTG CCT TGC CTA TCC CCT GTT GCG TGT CTCAG –
3'

UA3Bprime_12-mer

5' - CTGAGACACGCA – 3'

Clustering 454 sequences to remove repeats

For each of the two 454 runs, we first clustered the 454 sequences amongst themselves as the emulsion PCR and parallel sequencing can sometimes produce artefact repeat sequences. The clustering was done by aligning each 454 sequence against all other 454 sequences from the same run. Then, any pair of sequences found to be $\geq 95\%$ identical over 95% or more of their

length was clustered. In this way, single linkage clusters were made such that each sequence in a cluster was nearly identical over almost its entire length to at least one other sequence in the cluster. In subsequent analyses, the single best representative in terms of alignment score was used. This clustering protocol was motivated by the consideration that given the depth of sampling we have, it is unlikely that any Neandertal genomic fragment would be sampled multiple times, beginning and ending at or near the exact same genomic location. Therefore, the repeat sequences likely do not derive from independently sampled molecules. Indeed, many of the repeat sequences come from neighbouring wells of the 454 pico-titer plate and thus may be two signals read from the same source. Finally, since these repeat sequences likely derive from a single source molecule from the extract, if differences exist among them, the version with the highest similarity to a database sequence is likely to be the correct one since any deviation is more likely to reduce the alignment score than it is to increase the alignment score. Reassuringly, after clustering, there were very few overlapping alignments against the human genome, as expected given the depth of sampling. The sequence clusters produced by this method are given in Supplementary File 1.

Database searching of 454 sequences

The database of sequences produced by each 454 run was used as a query to search five databases: GenBank env (sequences from environmental sampling), GenBank nt (all GenBank nucleotide sequences, minus STS and EST sequences), the mouse genome (UCSC mm7), the chimpanzee genome (UCSC panTro1), and the human genome (UCSC hg18). The program blastall was used with the following options: -p blastn -K 100 -b 10 -v 10 -U F -I T -e 0.001 -F F -a 1. The blast reports were then parsed for alignment information and each hit was stored in a MySQL table. Then, for each sequence cluster (see above) the best single alignment from all its aligning sequences against all databases was found and analyzed. To

compare alignments within databases, the e-value was used. For comparison across databases, the total number of aligned bases was used, since e-value comparisons across databases of different sizes are not valid. A summary table (Supplementary File 2) was produced that contains fields for all relevant alignment information, including the alignment itself. To find the taxonomic distribution of best aligning sequences, we used the *gi_taxid_nucl.dmp* table from NCBI (ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid.nucl.dmp.gz).

For the mitochondrial alignments, a master alignment of complete mtDNA sequences from 311 humans from mtDB⁵⁸, 3 chimpanzees (GIs: 1262390, 5835121, 643679), 2 bonobos (GIs: 5835135, 643684), and 1 gorilla (GI: 5835149) was made using the *muscle*⁵⁹ program with the *-diags* parameters. Then, each 454 sequence was aligned to each of these hominoid mtDNA sequences separately. For all 454 sequence clusters that had an alignment longer than 30 nt and $\geq 95\%$ sequence identity to one or more of these hominoid mtDNA sequences, the single best alignment was selected. For each alignment, the mtDNA alignment was as good or better than any nuclear genome alignment. This best alignment was then used to merge the Neandertal sequence into the master alignment by propagating gaps from the pairwise Neandertal-hominoid alignment into the master alignment, or vice-versa. Therefore, this merge does not change any aligned bases in either alignment. To ensure that no 454 repeat sequences were introduced into the mtDNA alignments, we further filtered any alignments that overlapped by more than 20 bases on the master alignment. The result of this protocol is 41 alignment segments of the master hominoid alignment that also contains Neandertal sequence. The master alignment is given as Supplementary File 3. The alignment segments containing Neandertal mtDNA sequence are given before PCR-confirmation filtering [Supplementary File 4] and after [Supplementary File 5].

Generating 3-way human-Neandertal-chimpanzee alignments

To generate three-way alignments between Neandertal, human, and chimpanzee we used the best-in-genome human-chimpanzee pairwise whole genome alignment from UCSC (<http://hgdownload.cse.ucsc.edu/goldenPath/hg18/vsPanTro1/>). For each 454 sequence cluster whose best hit was to the human genome (compared to all other non-primate databases searched) and mapped with a unique best hit within the human genome, we extracted the corresponding region of the chimpanzee genome from the whole genome alignment and the quality scores for these chimpanzee bases. These two pairwise alignments, human-Neandertal and human-chimpanzee, were merged using the human sequence as reference. This alignment merge is a simple propagation of gapped positions from one alignment into the other. It is not a re-alignment and it does not change any aligned bases. From this merge, a set of 15,239 3-way human-Neandertal-chimpanzee alignments is created that includes quality scores for the chimpanzee bases. These 3-way alignments are given in Supplementary File 6. Note that chimp bases with a “Q” character are positions of no alignment and were not analyzed.

Analysis of 3-way alignments

Since best hits shorter than 30 bases are not sufficient evidence to confidently identify the species (personal observation), we restricted our analysis to alignments of at least 30 bases. We also restricted our analysis to alignment columns in which the chimpanzee base has a quality score ≥ 30 . As discussed in the main text, we also excluded 5 bases from each end of the alignment, since, due to the local alignment, these bases must be identical between human and Neandertal. Each alignment column passing these criteria that contained a base character

(A, C, G, or T) for each of the three sequences, was classified according to which sequences had identical bases and which sequences differed.

Neutralizing the effect of sequence error in the Neandertal sequence on branch length estimates

The human-Neandertal-chimpanzee sequences in the 3-way alignments are of very different quality. The human sequence comes from an essentially finished genome (NCBI version 36.1). The chimpanzee sequence comes from the first, draft sequence of the chimp genome (UCSC panTro1). The Neandertal sequence comes from single-pass sequencing reads of ancient molecules. Therefore, the human sequence can be considered correct and the chimpanzee sequence can be filtered by its associated quality score. The Neandertal sequence, however, which contains the bulk of the sequence errors, cannot be filtered. Therefore, we devised a strategy to quantify and neutralize the effect of Neandertal sequence error in the context of our estimate of genome divergence time. Note that this method is used only to more accurately determine the branch lengths for chimpanzee, human, and Neandertal. It is not a method to correct any individual base substitution.

Consider that each non-gapped alignment column contains a Neandertal base (N), a human base (H), and a chimpanzee base (C). Then, there are five possible categories for which of these three bases are the same (see table below, left-most column). For closely related species such as these three, the most populous category will be the first. That is, for most positions, there has been no sequence divergence between chimp, Neandertal, and human. Now, consider the effect of a sequence error in the Neandertal base (N) on the number of alignment columns observed in each of these five categories. For the case where all bases are the same ($N=H=C$), an incorrectly called Neandertal base, N, will be misclassified into the category $N\neq(H=C)$. That is, the human and chimp base will remain the

same, but the Neandertal base will appear different due to the error. In other words, all of the sequence errors from this category will be misclassified as Neandertal-specific divergence. This accounts for the accelerated Neandertal lineage. Likewise, for each category, one can track what fraction of the three possible wrong bases in sequence error will go to which other category (right-most column).

Case	Act	Obs	If sequence error...
N=H=C	a_1	b_1	$\frac{2}{3} \times e \rightarrow N!=(H=C)$
(N=H)!=C	a_2	b_2	$\frac{2}{3} \times e \rightarrow N!=H!=C; \frac{1}{3} \times e \rightarrow (N=C)!=H$
N!=(H=C)	a_3	b_3	$\frac{2}{3} \times e \rightarrow N!=(H=C); \frac{1}{3} \times e \rightarrow N=H=C$
(N=C)!=H	a_4	b_4	$\frac{2}{3} \times e \rightarrow N!=H!=C; \frac{1}{3} \times e \rightarrow (N=H)!=C$
N!=H!=C	a_5	b_5	$\frac{1}{3} \times e \rightarrow N!=H!=C; \frac{1}{3} \times e \rightarrow (N=C)!=H; \frac{1}{3} \times e \rightarrow (N=H)!=C$

Supplementary Table 2 Effect of sequence error in phylogenetic context. For each base, human (H), chimp (C), and Neandertal (N), in the 3-way alignments, an error in calling the Neandertal base, which happens at rate e , can cause misclassification of alignment type. For any base error, there are 3 possibilities for calling the wrong base, since there are 4 total bases. Therefore, one can track the effect of sequence error in this phylogenetic context.

Now, if we let the vector $[a_1, a_2, a_3, a_4, a_5]$ be the actual number in each category (that which we would like to know) and we set $[b_1, b_2, b_3, b_4, b_5]$ to the numbers that we observe in our alignments, then there is a linear relationship between these two vectors. Namely, in each category, the observed number, b_i , is the actual number, a_i , minus the number lost to other categories due to sequencing error, plus the number gained from other categories due to sequence error. Given a rate of sequence error, e , then:

$$b_1 = a_1 + (1/3) ea_3 - ea_1$$

$$b_2 = a_2 + (1/3) ea_4 + (1/3) ea_5 - ea_2$$

$$b_3 = a_3 - (1/3) ea_3 + ea_1$$

$$b_4 = a_4 + (1/3) ea_2 + (1/3) ea_5 - ea_4$$

$$b_5 = a_5 + (2/3) ea_2 + (2/3) ea_4 - (2/3) ea_5$$

This set of equations can be simplified and a 0 coefficient added for each category that does not contribute to each observed number to generate the following matrix equation:

$$\begin{bmatrix} 1-e & 0 & \frac{1}{3}e & 0 & 0 \\ 0 & 1-e & 0 & \frac{1}{3}e & \frac{1}{3}e \\ e & 0 & 1-\frac{e}{3} & 0 & 0 \\ 0 & \frac{1}{3}e & 0 & 1-e & \frac{1}{3}e \\ 0 & \frac{2}{3}e & 0 & \frac{2}{3}e & 1-2\frac{e}{3} \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix}$$

Finally, we have a set of five independent equations with five unknowns. We obtain b_i from the observed number of alignment columns in each category. If we know the error rate, e , we can solve to obtain the a_i . There are several ways to estimate the error rate in our alignment data. One way is to take the difference between the Neandertal-specific branch and the human-specific branch and divide by the total alignment columns. This assumes that the Neandertal and human lineages should be equal and that the excess Neandertal branch length is attributable to error. Another way to estimate e is to empirically find a value that satisfies the requirement of equal human and Neandertal divergence. In practice, these two error rate estimates are nearly identical. The first method gives an error rate of 0.004012 for the Neandertal alignments and 0.00193 for the modern human sequence alignments. The second method gives an error rate of 0.004089 for the Neandertal alignments and 0.00196 for the human alignments. Note that these error rates are not useful estimates for the overall base-calling error in these sequences as the alignment flanks that are 100% identical and were

therefore called correctly, are excluded. This estimate is only for the alignment interiors from which our analysis derives.

Ancestral population size

For the analysis of the ancestral population size and split time, we implemented additional filters to the three way alignments (human, chimpanzee, and Neandertal) described above. We retain only autosomal sequences that do not overlap with an exon or UTR to minimize the effects of selection and to insure that the patterns of mutation, recombination, and inheritance are roughly the same across reads. To insure that we are indeed comparing homologous regions we only use sequences for which we have a unique reciprocal hit. A unique reciprocal hit occurs if, when we blast the chimpanzee sequence against the human genome, there is only one best hit sequence and it overlaps with the human sequence (based on results from blastall run with an e-value cutoff of 0.001, *i.e.* -e 0.001 -b 2 -v 2 -K 100 -F F -U F -m 8). We remove all sequences which overlap with another sequence from the same run of 454 machine, as these overlapping sequences are created primarily due to a technology artifact resulting from an inability to properly quantify the low amounts of DNA used (E. Green and M. Ronan, unpublished results). Since we treat all sequences as unlinked regions in our analysis, for any cluster of sequences which are 30kb or closer to one another we randomly exclude all but one of the sequences. Finally, we used a higher quality score cutoff in chimpanzee (≥ 50) than previous analyses to minimize effects of varying error rates, which the method would detect as variations in substitution rate. We then summarize the data for each sequence by counting the number of substitutions inferred to occur on the chimpanzee lineage and on the human lineage (see main text; Fig. 5). Thus, here too, we do not include data on the number of substitutions on the Neandertal lineage, as this category contains most of the sequencing errors in the Neandertal sequence reads. We also had to

eliminate alignment columns where the base is different in all three sequences, as we could not infer on which lineage the change occurred. This leaves us with 8,498 sequences comprising a total of 508,964 usable bases.

Several methods exist to estimate split time and ancestral effective population size from divergence data⁶⁰⁻⁶³. We chose the Wall method since it has the advantage that it explicitly incorporates recombination, handles variation in mutation rates better, and is more efficient and more flexible in its implementation. This method runs coalescent simulations to obtain the expected number of mutations on each lineage given an ancestral effective population size and population split time under a simple model (Fig. 6a), from which one can then compute the likelihood of the data. Thus, we summarized the alignments by counting, for each read, the number of nucleotide changes that occurred on the human or chimpanzee lineages by using the Neandertal as a reference, thus ignoring bases in the Neandertal not corroborated by either the human or the chimpanzee sequence. We then re-implemented the Wall method for three species as outlined in⁶³, except we restrict our analysis to computing the likelihood of substitutions on the chimpanzee and human lineages only. We run coalescent simulations using *ms*⁶⁴, based on a simple split model between three populations, where humans and Neandertals split at time T_{NH} and chimpanzees and the ancestor of humans and Neandertals split at time $T_{C(HN)}$ ($0 < T_{NH} < T_{C(HN)}$) with instantaneous population size changes at the split times (Fig. 6a). From these coalescent simulations, this method then computes the likelihood of the observed number of substitutions on each lineage, and finally finds the split time and ancestral population size that maximizes this likelihood.

For the coalescent simulations, we scale everything with respect to the human effective population size (N_H) and assume the following values for these parameters:

Diversity level: $\theta = 0.00075$ per bp⁶⁵

Mutation rate: $\mu = 2e-8$ per bp per generation⁶⁶

Generation time: $g = 20$ years⁶⁶

Effective population size in humans: $N_H = \theta / 4\mu = 10,000$ ⁶⁷

Recombination rate: $\rho = 4N_c c = 1.19 \text{ cM/Mb}^{68} * 0.00000001 * 4 * 10000 = 0.00048$ per bp

Number of basepairs: nbps, this parameter comes from the data itself, where total length of the region is used for recombination rate while the number of usable bases is used for the mutation rate

Effective population size in chimpanzee: $N_C = 44,000$ ⁶⁹ hence $N_C/N_H = 4.4$

Effective population size in Neandertal: $N_N = N_H$ hence $N_N/N_H = 1.0$

Ratio of effective population size of ancestor to human, chimpanzee, and Neandertal to human effective population size: $N_O/N_H = 6$ ⁶³

Split time between human and chimpanzee: $T_{C(HN)} = 6,000,000 \text{ years}^{70} / 4N_{Hg} = 7.5$

Split time between human and Neandertal: T_{NH} , unknown but is uniformly distributed between 0 and 1.0

Ratio of effective population size of ancestor of humans and Neandertals to human effective population size: N_A/N_H , unknown but is uniformly distributed between 0 and 2.0

Age of Neandertal individual: $T_{bone} = 38,310 \text{ years}^{54} / 4N_{Hg} = 0.048$ if $T_{NH} \geq 0.048$ otherwise, $T_{bone} = 0$

NumRep: For each parameter combination of N_A/N_H and T_{NH} we run 50,000 replications⁶³

That is, to run *ms* we use the following command: `./ms 3 NumRep -t θ -r ρ nbps -T -I 3 1 1 1 -n 1 N_C/N_H -n 2 0 -en T_{bone} 2 N_N/N_H -ej T_{NH} 3 2 -en T_{NH} 2 N_A/N_H -ej $T_{C(HN)}$ 2 1 -en $T_{C(HN)}$ 1 N_O/N_H .`

We thus computed the likelihood of the observed data for the random parameter sets depicted in Fig. 6b. We then applied a likelihood ratio test to determine a set of parameter values that are statistically equivalent (with $p > 0.05$). The likelihood ratio test, while not strictly appropriate, seems to be a good approximation (⁶³; Supplementary Results). The set

of statistically equivalent parameter sets suggests a line of best fit. We thus fit a line to this set of statistically equivalent parameter sets and estimated a slope of -2.0 and an intercept of

$$1.8. \text{ That is, } 1.8 - 2 * T_{HN} = \frac{N_A}{N_H}.$$

However, this is simply a rewriting of the relationship, that for a sample size of two, the expected divergence time or $T_{mrca} = 2gN_A$, which, in a simple split model, can only occur earlier in time than the split⁷¹; that is $T_{mrca} = 2gN_A + 4gN_H T_{HN}$. Solving this equation for

$$N_A/N_H, \text{ yields: } \frac{T_{mrca}}{2gN_H} - 2 * T_{HN} = \frac{N_A}{N_H} \text{ (1).}$$

Thus, in the main text, we use this equation (1) to estimate N_A given a human-chimpanzee sequence divergence time of 6,500,000 years, a human-Neandertal split time of 400,000 years (J.J. Hublin, pers. comm.), and the error-corrected number of mutations on the human and chimpanzee lineages (see above). The estimate of 6,500,000 years, along with agreeing with previous estimates^{70, 72, 73}, is the estimated sequence divergence time for our data (sequence divergence = 0.012), given $g = 20$ years and $\mu = 2e-8$ per bp per year used above. However, both of these dates are uncertain. A rough upper limit on the divergence time is given by $T_{C(HN)} + 2gN_O = 8,400,000$ years, while a lower limit is perhaps 4,700,000 years⁷³. A rough upper limit for the population split time is given by an estimate of the divergence time from mitochondrial DNA⁷⁴ of 500,000 years, which cannot have been younger than a putative population split, while a lower limit may be 300,000 years⁷⁵. Finally, we also include uncertainty in the number of mutations on the human lineage, by assuming a poisson distribution with mean equivalent to the observed value.

Power simulations

To assess the performance of this method we ran three sets of simulations. For all three sets, we generated simulated data using the same command line of ms as described in the methods above (except we simulate a second human individual, so -I 3 1 1 2), using 30

randomly chosen sets of T_{NH} and N_A/N_H . Thus, our simulations represent an ideal case where we know all parameters exactly except the two we are estimating. We ran the simulations assuming 8,530 sequences comprising a total of 510,275 bases, where the length distribution for the 8,530 sequences mimicked the actual data. Thus, the simulations had somewhat more data than the observed data, but the conclusions we draw from these simulations should not be altered by this. We then run these 30 simulated data sets through the same algorithm as the data. The first set of simulations is to assess the performance of the algorithm in determining the relationship between T_{NH} and N_A/N_H . For this, the values of T_{NH} and N_A/N_H are chosen independently of one another, both in generating the simulated data set, and in the algorithm used to determine the likelihood of the simulated data set. The second set of simulations is to assess the performance of the algorithm in determining the relationship between T_{NH} and N_A/N_H , given the presence of sequencing errors on Neandertal and human contamination. It thus uses the same approach and the same 30 data sets, except we additionally simulate sequencing errors and contamination. Based on the sequencing error rate estimate (see main text), with probability 0.004 an error occurs, where 1/3 of the time the Neandertal base switches between the two alleles already present, and 2/3 of the time the Neandertal base switches to a new allele and the base is no longer usable. Errors that occur on bases that are not polymorphic simply create an apparent substitution that occurs on the Neandertal lineage, which is not analyzed further. Based on the human contamination rate as calculated for mitochondrial DNA (see main text), with probability 0.02, the sequence from the second simulated human individual is incorporated rather than the Neandertal sequence. The final set of simulations is to assess the performance of the algorithm in determining the maximum likelihood estimate of T_{NH} and N_A/N_H , given their linear relationship. To do this, we generate 30 random data sets, where $N_A/N_H = -2.0 T_{NH} + 1.8 + \text{error}$, where the error term is uniformly distributed between -0.2 and 0.2, and the slope and intercept come from values estimated

from the actual Neandertal data. We then use our algorithm to infer the maximum likelihood values of N_A/N_H and T_{NH} , given this same linear relationship.

We ran each algorithm for a minimum of 500 parameter sets. In order to estimate the line of best fit, we required that there were at least 20 parameter sets that were most likely (as assessed by likelihood ratio test).

To quantify the performance of the Wall method we focus on two statistics, the average difference between true and estimated value (ignoring direction of change), as a percentage of

the total range of possible values, *i.e.* $\frac{\sum |X - \hat{X}|}{n * range}$ where X is either N_a/N_h (which has range 0-

2.0=2) or T_{NH} (which has a range of 0-1.0=1), and the maximal difference *i.e.* $\frac{\max |X - \hat{X}|}{range}$.

Comparison of Neandertal sequence to polymorphic sites in humans

We downloaded the HapMap data⁷⁶ (<http://www.hapmap.org/downloads/index.html>) on December 6, 2005 and the Perlegen data⁷⁷ (<http://genome.perlegen.com/browser/download.html>) on May 2, 2006, and used the command line lift-over tool provided by UCSC browser (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) that we downloaded on May 2, 2006 to map all SNPs to Build 36.1 (March 2006). Next we took the three way alignments between human, chimpanzee, and Neandertal and again only use autosomal, non-genic sequences, for which we have a unique reciprocal best hit in the human and chimpanzee genomes and which does not overlap with any other sequence from the same run. We then identify all bi-allelic SNPs that overlap with these regions. Rather than masking bases based on quality scores in chimpanzee, we only use alignment columns for which Neandertal and chimpanzee sequence matches one of the alleles seen in the HapMap/Perlegen data (this criterion thus excludes cases where a base is missing or is an indel). Finally, for the Perlegen data, we restricted our analysis to those SNPs that were discovered by the initial

resequencing study (class A SNPs), so that, unlike the HapMap data, all the SNPs were ascertained in the same way and this ascertainment is not biased towards discovery in a given population.

Supplementary Results

Estimate of effective population size for human data

We applied our method of estimating the effective population size to the human 454 sequence data. This allows us to evaluate the performance of our data processing scheme on data where we have a better idea of the true evolutionary history. This also allows us to evaluate the concern of whether the data we have is truly Neandertal data and not contaminating human sequence. Using equation (1) as before, and using a human-human split time of zero (since humans form one breeding population), we estimate $N_A \sim 12,000$. The effective population size of humans has been estimated to be $10,000^{67}$, which is encompassed by the degree of uncertainty. This suggests our data processing scheme is not introducing large biases.

This line for the human data (intercept: 1.5) is shifted downwards compared to the line for the data from the Neandertal bone (intercept: 1.8). However, given the power of the method, we can not statistically differentiate between the two estimates. Thus, although this method is not powerful enough to reject the hypothesis that we have contaminating human sequence, the results of the Neandertal data predict, for a split time of 0, a higher effective population size for humans than is currently believed most likely.

It should be noted that since the coalescent time of two human sequences can extend further back in time than the split time of humans and Neandertals, the estimate of the effective population size in present-day humans includes effects of the ancestral population size of humans and Neandertal. In particular, if the ancestral population size was smaller than the effective population size of humans (as suggested by Fig. 6b), then the current effective population size of modern day humans would have to be larger than the estimated effective population size of modern day humans.

Power simulations of Wall Method

Since assuming a χ^2 distribution may not be appropriate⁶³, to evaluate the performance of the Wall method and the power we have for our data, we tested the method using simulated data sets generated by the coalescent process. Specifically, we generated 30 simulated data sets given random parameter sets of the ancestral population size and split time, and then ran the Wall method on these simulated data sets. We first asked how accurately the ratio of the ancestral effective population size and human effective population size (N_a/N_h) is estimated from a known split time, by estimating the line of best fit and then computing N_a/N_h from this line given the value of split time used in generating the data (Supplementary Fig. 3). On average, the difference between the predicted value and the estimated value of N_a/N_h is 5% of the total range, while the maximum deviation is 14%. When we use instead the confidence intervals generated by assuming the approximation of the likelihood ratio test for the actual data, we notice that points differ maximally by 12%. Thus, the likelihood ratio test seems like a good rough approximation.

Second we used the simulated data to ask how accurately both the ancestral population size and the split time are co-estimated, by examining the maximum likelihood estimates (Supplementary Fig. 4). To address this question, we simulated an additional 30 data sets, in which we assumed the true value lies within ± 0.2 of the line estimated for the actual Neandertal data. We then asked how well the maximum likelihood estimate predicts the true parameters. For both parameters, on average, the difference between the predicted value and the estimated value is $\sim 25\%$ of the total range, while the maximum deviation is $\sim 80\%$. This suggests we have only a little power to distinguish points along the line, which is also consistent with what we see for the actual data by assuming the approximation of the likelihood ratio test. This is not surprising since both increasing the split time (given a fixed split time between humans and chimpanzees) and increasing the ancestral effective population size lead to an increase in the number of substitutions expected on the human

lineage and a decrease on the chimpanzee lineage⁷⁸. The ability to differentiate between these two alternative hypotheses of increasing split time versus increasing population size lies in the fact that the split time is fixed across all loci, whereas the coalescent time in the ancestor varies across loci such that the amount of variation is also dependent on the effective population size in this ancestor. Hence information about the distribution of the number of substitutions across multiple loci is used⁶⁰. However, presumably since our reads are quite short, this greatly reduces the power to infer the amount of variation across loci. Nevertheless, we do have some power. If we compute the average difference between all points along the line and the true value, and compare this to the difference between the maximum likelihood and the true value, the difference is slightly smaller in the latter case ($p = 0.05$ for T_{NH} and $p = 0.02$ for N_A/N_H , $n = 30$, based on a sign test).

Finally, we wanted to examine the effect that sequencing errors and human contamination have on our estimate. To assess the effect of these two factors on our estimates, we took the same 30 simulated data sets for which we estimated the line of best fit and added sequencing errors at a rate of 0.4% (as estimated based upon the excess substitutions on the Neandertal lineage, see main text) and human contamination at a rate of 2% (as assessed from mitochondrial DNA assay, see main text). We then ran this simulated data with these errors through the Wall method, and compared N_a/N_h from the simulated data + error to the estimate we obtained using the original simulated data (Supplementary Fig. 5). The estimate given error and contamination increases on average by 2% of the range ($p = 0.002$, $n = 30$, based on paired t-test) and differs maximally by 6%. Thus, although sequencing error and human contamination has an effect, this effect is small for the levels at which we think these errors occur.

In conclusion, there are a number of reasons to suspect that our estimate of the effective population size is, if anything, an overestimate. The model of speciation used in the

likelihood calculation assumes no gene flow after the Neandertal-human split. If there was gene flow, this leads to an overestimate in our estimate of ancestral population size⁶³. Likewise, we assume no variation in mutation rate across loci. Although the method we use seems not to be overly sensitive to deviations from this assumption⁶³, assuming no variation when there is indeed variation leads to an overestimate in the ancestral population size⁷⁹. The method also assumes no sequencing errors and that all reads come from a single individual (*i.e.* no contamination), and when we simulate these effects, this leads, on average, to an overestimate (see above).

Supplementary Figure Legends

Supplementary Figure 1. Effect of local alignment on polymorphism and divergence position. Local alignment algorithms, like *blast*, find the single maximally-scoring alignment between two sequences. Because the alignment is local and maximally scoring, extension of the alignment in either direction would decrease the score, either by aligning mismatched positions or introducing a gap. Therefore, the next bases upstream and downstream of the alignment (if any – the alignment may begin or end at the beginning or end of one of the sequences), are necessarily not matching. As a corollary, the bases just inside the alignment are matching. Otherwise, the alignment would have ended before these positions to achieve a higher score. This subtle but important effect impacts how alignments relate to other measures of diversity within the human genome and divergence between human, Neandertals, and chimps. The top panel shows the frequency of SNPs within the human genome present in three populations (CEU=Europeans; CHB=Han Chinese; YRI=Yorubans) in relation to the alignment position of Neandertal sequence. The x-axis is the alignment position, anchored from the ends of the alignment. Position 1 is the first alignment position in the alignment whereas position -1 is the first position outside the alignment. There is a much increased rate of polymorphism just outside the alignments. This indicates that in some instances, the Neandertal-human alignment was not extended to cover all bases because the reference human genome sequence at that position was polymorphic and has the allele that does not match the Neandertal sequence. In other words, some alignments ended because of human polymorphism. The bottom panel shows the number of mismatches and gaps as a function of distance from the beginning (5' end) or end (3' end) of human-Neandertal alignments. As described above, there is no divergence in the flanks of the alignments as a consequence of the local alignment algorithm. Therefore, these flanks are necessarily 100%

identical between Neandertal and human. Therefore, they are excluded from human-Neandertal-chimpanzee comparison.

Supplementary Figure 2. Schematic tree illustrating the number of nucleotide changes inferred to have occurred on hominoid lineages. This comparison is identical to that described in Figure 5, except the input sequence data come from a modern human individual instead of a Neandertal extract. Thus, the divergence estimate is between this modern human and the human reference genome. Each alignment position that passes the quality filtering criteria (see main text) is classified by which sequences share bases. In blue is the distribution of all aligned positions that did not change on any lineage. In brown are the changes that occurred either on the chimpanzee lineage, p , or on the hominid lineage prior to reference human-experimental human (exp) divergence, h . In orange are the changes that are unique to the experimental human. Note the absence of C to T and G to A mismatches present in the Neandertal sequence. In yellow are changes that occurred on the human lineage after reference human-experimental human divergence. The distribution of types of changes in each category is also given. The number of counts in each category, corrected for all base-calling error in the 454 sequencing (see Supp Methods) is also shown.

Supplementary Figure 3. Power to estimate linear relationship between split time and ancestral effective population size.

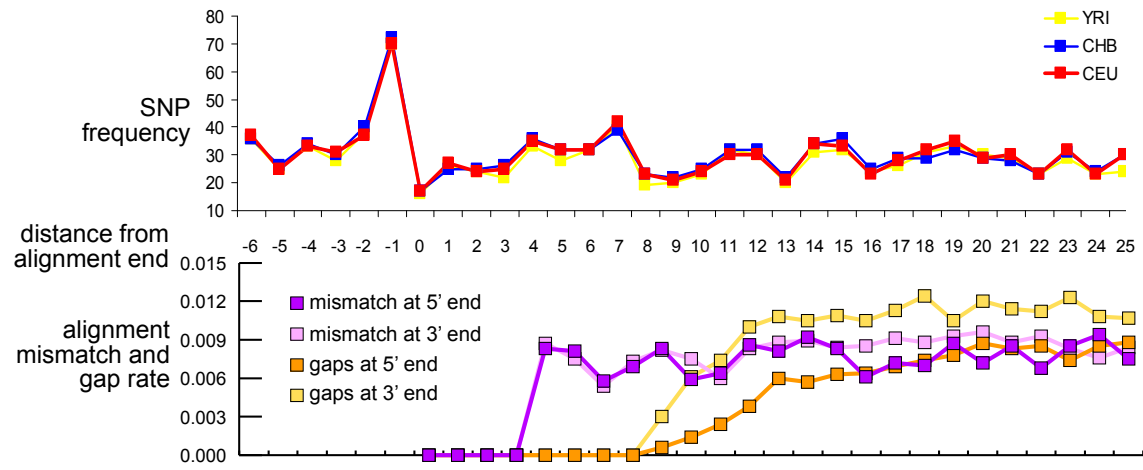
Shown are the 30 parameter sets used to generate 30 simulated data sets (as points) and the lines that were subsequently estimated by applying the Wall method to these 30 simulated data sets. Points and lines are matched based on colour.

Supplementary Figure 4. Power to estimate maximum likelihood estimate of split time and ancestral effective population size given the line of best fit between these two parameters.

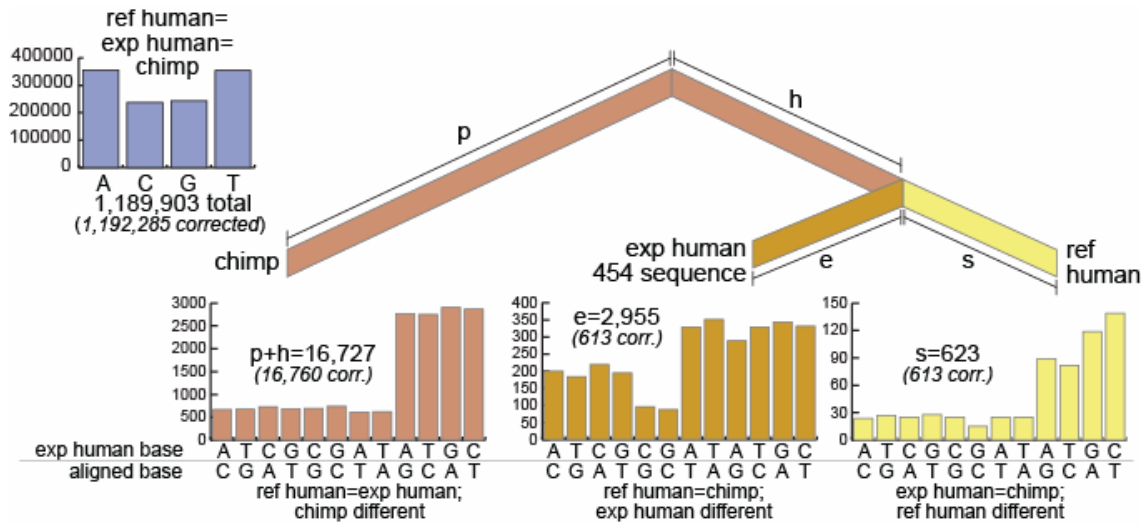
Shown are the 30 parameter sets, which fall within ± 0.2 of the line of best fit inferred for the actual Neandertal data set, used to generate 30 simulated data sets and the estimates of the maximum likelihood that result from applying the Wall method to this simulated data. These two sets of points are matched based on colour and shape, where the true value is indicated by an asterisk.

Supplementary Figure 5. Effect of sequencing error and human contamination on the estimate of the linear relationship between split time and ancestral effective population size.

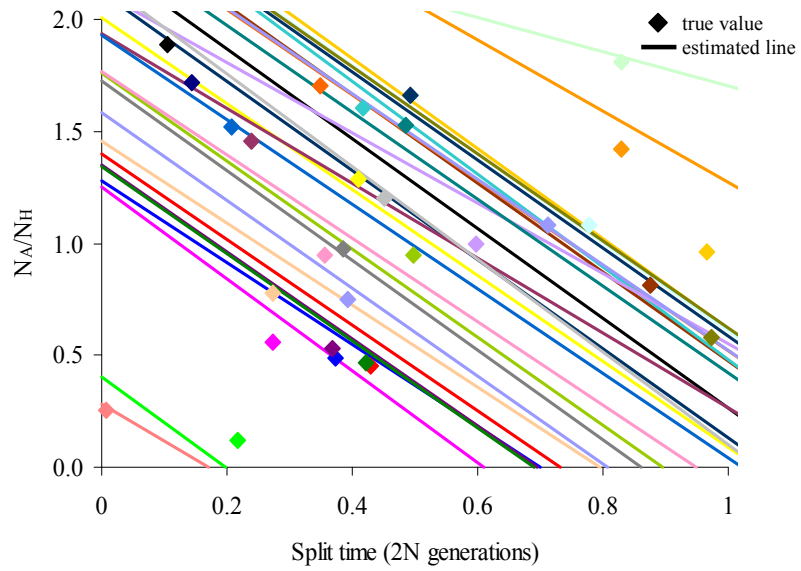
Shown are the 30 parameter sets used to generate 30 simulated data sets (as points) and the lines that were subsequently estimated by applying the Wall method to these 30 simulated data sets with (dashed) and without (solid) errors being added. Points and lines are matched based on colour.



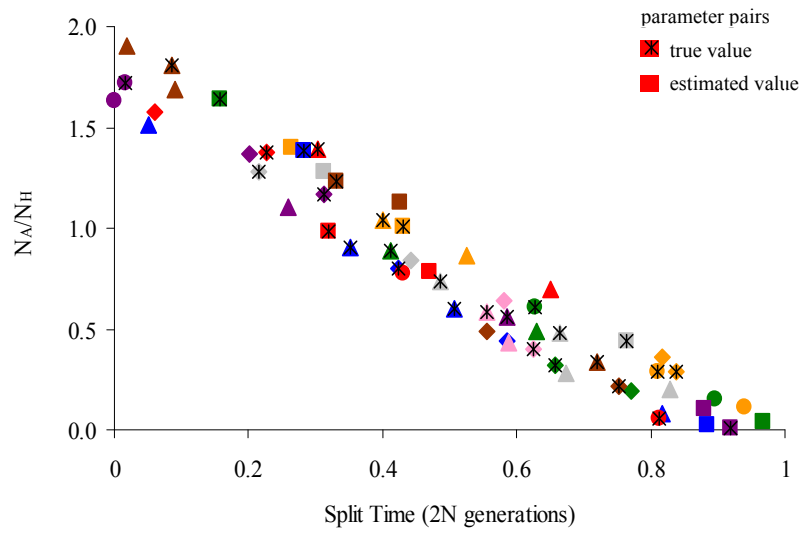
Supplementary Figure 1



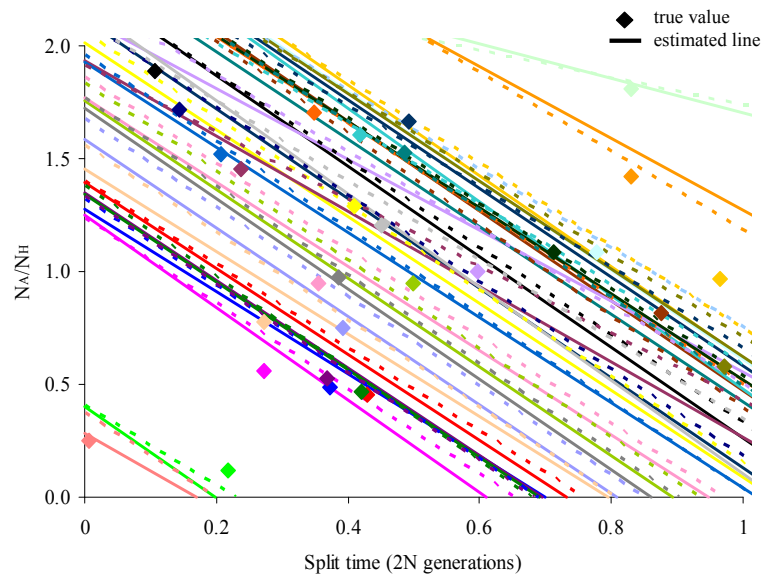
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5

Supplementary References

51. Schmitz, R. W. et al. The Neandertal type site revisited: Interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc. Natl. Acad. Sci. USA* 99, 13342-13347 (2002).
52. Krings, M. et al. A view of Neandertal genetic diversity. *Nat Genet* 26, 144-146 (2000).
53. Poinar, H. N., Höss, M., Bada, J. L. & Pääbo, S. Amino acid racemization and the preservation of ancient DNA. *Science* 272, 864-866 (1996).
54. Serre, D. et al. No evidence of neandertal mtDNA contribution to early modern humans. *PLoS Biology* 2, 313-317 (2004).
55. Krings, M. et al. Neandertal DNA sequences and the origin of modern humans. *Cell* 90, 19-30 (1997).
56. Krause, J. et al. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439, 724-727 (2006).
57. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380 (2005).
58. Ingman, M. & Gyllensten, U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res* 34, D749-751 (2006).
59. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792-1797 (2004).
60. Takahata, N., Satta, Y., & Klein, J. Divergence time and population size in the lineage leading to modern humans. *Theoretical Population Biology* 48, 198-221 (1995).
61. Yang, Z. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics* 162, 1811-1823 (2002).
62. Rannala, B., & Yang, Z. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164, 1645-1656 (2003).
63. Wall, J. D. Estimating ancestral population sizes and divergence times. *Genetics* 163, 395-404 (2003).
64. Hudson, R. R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18, 337-338 (2002).
65. Kruglyak, L., & Nickerson, D.A. Variation is the spice of life. *Nature Genetics* 27, 234-236 (2001).
66. Fischer, A., V. Wiebe, S. Pääbo, & M. Przeworski. Evidence for a complex demographic history of chimpanzees. *Molecular Biology and Evolution* 21, 799-808 (2004).
67. Harpending, H. & Rogers, A. Genetic perspectives on human origins and differentiation. *Annu Rev Genomics Hum Genet* 1, 361-385 (2000).
68. Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., Sigurdardottir, S., Barnard, J. & Hallbeck, B., Masson, G., Shlien, A., Palsson, S. T., Frigge, M. L., Thorgeirsson, T. E., Gulcher, J. R., Stefansson, K. A high-resolution recombination map of the human genome. *Nat Genet* 31, 241-247 (2002).
69. Fischer, A., Pollack, J., Thalmann, O., Nickel, B., & Pääbo, S. Demographic history and genetic differentiation in apes. *Current Biology* 16, 1133-1138 (2006).
70. Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. & Reich, D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441, 1103-1108 (2006).

71. Hudson, R. R. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* 7, 1-44 (1990).
72. Innan, H., & Watanabe, H. The effect of gene flow on the coalescent time in the human-chimpanzee ancestral population. *Molecular Biology and Evolution* 23, 1040-1047 (2006).
73. Kumar, S., Filipowski, A., Swarna, V., Walker, A., & Blair Hedges, S. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc. Natl. Acad. Sci. USA* 102, 18842-18847 (2005).
74. Krings, M., Geisert, H., Schmitz, R. W., Krainitzki, H. & Pääbo, S. DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc Natl Acad Sci U S A* 96, 5581-5585 (1999).
75. Stringer, C. Modern human origins: progress and prospects. *Philos Trans R Soc Lond B Biol Sci* 357, 563-579 (2002).
76. Consortium, T. I. H. A haplotype map of the human genome. *Nature* 437, 1299-1320 (2005).
77. Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A., Cox, D. R. Whole genome patterns of common DNA variation in diverse human populations. *Science* 307, 1072-1079 (2005).
78. Takahata, N. An attempt to estimate the effective size of the ancestral species common to two extant species from which homologous genes are sequenced. *Genetical Research Cambridge* 48, 187-190 (1986).
79. Yang, Z. On the estimation of ancestral population sizes of modern humans. *Genetical Research Cambridge* 69, 111-116 (1997).