SUPPLEMENTARY INFORMATION

Description of Denisova Cave and the manual phalanx

Located near the border of the Altai Krai and the Altai Republic, in the northern part of the Altai Mountains, Denisova Cave (51°40' North; 84°68' East) is on the right side of the Anui River, in a Silurian sandstone cliff, about 28m above the level of the river. The southwest facing entrance is 2m high and 7m broad, opening into a large main chamber that is 33m long, 11m broad and 10 m high.

The cave preserves cultural horizons indicating episodic human occupation for the last 125,000 years. The lowermost horizons of the central chamber were dated to 171±43ky to 282±56ky[1] ago by radiothermoluminescence, but these high ages seem unlikely based on the archaeological and faunal assemblage. The high standard deviations also advise caution. A more probable age for the bottom of the sequence based on the microfauna would be OIS 5e, about 125ky ago (pers. com. Alexander K. Agadzhinyan).

Archaeologically, layer 11 from which the sample derives, is characterized as a transitional industry where both Upper Palaeolithic and Middle Palaeolithic elements are present. Besides a rich lithic assemblage, numerous bone implements and adornments were also recovered[2,3]. The association of these generally Upper Palaeolithic elements with foliate bifaces, sidescrapers, notches and Mousterian as well as Levallois points is hard to interpret. One possible interpretation would be post-depositional mixing, but the presence of both Upper and Middle Palaeolithic elements at other sites in the region indicates that this mixture is a characteristic of the Early Upper Palaeolithic in the Altai. At Kara-Bom, all these Middle Palaeolithic elements also occur in the Early Upper Palaeolithic OH 5 and 6 horizons, while refitting studies by N. Zwyns (Zwyns, pers. comm.) indicate no vertical movement between these horizons and the underlying Middle Palaeolithic level MP1. Similarly, bifacial and Levallois elements also occur in the Early Upper Palaeolithic of Ust-Karakol, where no Middle Palaeolithic occupational horizons are present[1].

The Densiova phalanx is from Level 11.2 of the 2008 excavation in the so-called Eastern Gallery, a small diverticle of the main chamber. It is a proximal epiphysis of a juvenile manual phalanx, preserving the proximal articular surface and the bone surrounding it. Three ^{14}C dates are available for Level 11, all from the Main

Chamber[3]. The lowermost subhorizon 11.3 yielded an age of 48,650±2380 years ago (KIA 25285, uncal. BP, like all following dates), subhorizon 11.2 an infinite age of >37,235 years ago (SOAN-2504), while a sample from the interface of Level 11 and the overlying Level 10 was dated to 29,200±360 years ago (AA-35321), giving a minimum age for the horizon. The stratigraphy of the Main Chamber and of the Eastern Gallery, from which the phalanx derives, can not be connected physically, but are correlated based on the geology and the archaeological material.

DNA extraction and PEC mtDNA enrichment

DNA was extracted as described[4] from 30mg of bone except that DNA elution from the silica particles was done using TE buffer with 0.05% Tween 20. 25µl of extract were used to produce a Illumina/Solexa library using a modified 454 library preparation protocol[5], except that Solexa p5 and p7 adapter were used (Meyer, et. al, manuscript in preparation). For that, two truncated p5 and p7 adapters (Table S1) were prepared by hybridizing complementary oligonucleotides. The truncated adapters were used in blunt-end ligation. Sequencing libraries with full-length adapter sequences were generated by amplifying the truncated library with 5'-tailed PCR primers in a 100µl reaction containing 1X Amplitaq Gold buffer (Applied Biosystems), 2.5mM MgCl2 (Applied Biosystems), 250µM each of each dNTP (New England Biosystems), 200µg/ml bovine serum albumen (BSA) (Sigma-Aldrich), 300nM each emPCR primer (Sigma Aldrich) and 0.05units/µl Amplitaq Gold DNA polymerase (Applied Biosystems). Libraries were amplified for 10 cycles, purified with QIAquick spin columns (Qiagen) and quantified using a Q-PCR[6]. Aliquots of the products were used as template in an additional 8-10 cycle amplification step. Both adapters were modified to allow detection of contamination from other Illumina libraries by a 4-bp-tag sequence at their 3'-ends (5'-GTCT-3'). The p7 adapter carried an additional internal barcode (5'-CTATACG-3'). The combination of this tag and barcode was used exclusively for this particular fossil.

PEC was performed as described[7] using the four 144-plex and 143-plex mixes previously used to retrieve complete Neandertal mtDNAs and a Pleistocene human mtDNA[8]. These primers are 5'-biotinylated and carry a twelve-base 5'-spacer sequence (5'-CAAGGACATCCG-3') followed by a primer site.

Table S1. Oligonucleotide sequences used in library preparation (PTO bonds indicated by *)

Name	Sequence
Adapter P5 oligo 1	5'-A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCT*g*t*c*t-3'
Adapter P5 oligo 2	5'-a*g*a*c*AGATCGGAA*G*A*G*C-3'
Adapter P7 oligo 1	5'-Bio-G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCT*g*t*c*t-3'
Adapter P7 oligo 2	identical to Adapter P5 oligo 2
PCR primer P5	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT-3'
PCR primer P7	5'-CAAGCAGAAGACGGCATACGAGATctatacgGTGACTGGAGTTCAGACGTGT-3'

Illumina/Solexa sequencing

The *lllumina/Solexa* library was amplified in a 100µl reaction containing 50µl Phusion™ High-Fidelity Master Mix, 500nM of each the p5 and p7 *Illumina* primer and 10µl PEC product template. Annealing temperature was 60°C and a total of 10 cycles of PCR was performed. The amplified products were spin column purified and quantified on an Agilent 2100 Bioanalyzer DNA 1000 chip. The PEC product was than diluted and sequenced according to *Illumina* GA2 protocols on a paired-end run with a total of 76 cycles each direction.

The sequencing run was analyzed starting from raw images using the Illumina Genome Analyzer pipeline 1.3.2. The first five cycles were used for cluster identification. After standard base calling, reads of the PhiX 174 control lane were aligned to the corresponding reference sequence to obtain a training data set for the base caller Ibis[9]. Raw sequences obtained from Ibis for the two paired end reads of each sequencing cluster were merged (including adapter removal) requiring at least 11nt overlap between reads. For bases in the overlap, quality scores were summed up. In cases where different bases were called, the base with the higher quality score was chosen.

Assembly procedure

The 1,178,300 merged sequences were used as input for a custom iterative mapping assembler[7]. In the first assembly round, sequences were aligned to the revised Cambridge Reference Sequence (NC_012920.1[10]). The mapper uses a position-specific scoring matrix designed to capture the most relevant features of nucleotide misincorporations affecting ancient DNA sequences. In total, 93,349 (12.4%) of the merged sequences aligned to the reference mtDNA. Since several amplification steps were performed, the aligned sequences were filtered for uniqueness by grouping

sequences with the same direction, start and end coordinates. From each such cluster a consensus sequence was made by taking, for each position, the base with the highest quality score. That resulted in a total of 30,443 distinct sequences. These were then used to call a most likely assembled sequence, after that a second round of alignment to this most likely sequence was performed, after which a new most likely sequence was called. This process was iterated until convergence on the final most likely sequence after 3 rounds. When the mapping assembly was repeated using the Neandertal Vi33.16 mtDNA[11] as a reference sequence, the most likely sequence called after three rounds of iterative assembly was identical to the one called using the rCRS as a reference sequence. The assembly covers all positions in the mtDNA at least twice with an average coverage of 156.4 for each position. At all 67 positions covered less than six times, overlapping sequences all carried the same base. Among the 16,569 positions in the circular mtDNA genome, the average fraction of sequences that agree with the assembled mtDNA was 99.1% (lowest 70%, highest 100%). All 8 positions with support lower than 80% had minor bases consistent with cytosine deamination[7]. Between rCRS nucleotide positions 5,895 and 5,900 a homopolymer of 9 to 20 cytosine residues occurs. We interpret this as an *in vivo* length heteroplasmy (Suppl. Fig 2) which are frequent in some mammalian species[12,13].

To check the presence of potential animal contamination in our sequencing data we mapped all reads retrieved from the PEC experiment to all available mammal mtDNAs in *Genbank* using the mapper *bwa*[14]. We separated the mammalian mtDNAs into families and found that 99.95% (82,091) of the reads that mapped best to one mammalian family had their best hit to hominids. Larger amounts of animal contamination can therefore be excluded.

Shot-gun sequencing of the Denisova mtDNA

In order to replicate the Denisova mtDNA, we extracted DNA as described above from an additional 10mg of bone from the phalanx. 25µl of extract were used to produce a Illumina/Solexa library using a modified protocol as described above. The *lllumina/Solexa* library was amplified in a 100µl reaction containing 4 Units of *Taq* Gold polymerase, 2.5mM MgCl2, 1 mg/ml BSA, 250µM of each dNTP and 500nM of the p5 and p7 *Illumina* primer for 10 cycles. The product was spin column purified and further amplified in a 100µl reaction containing 50µl Phusion™ High-Fidelity

Master Mix, 500nM of the p5 and p7 *Illumina* primer and 10µl template. Annealing temperature was 60°C and a total of 8 cycles of PCR was performed. The amplified products were spin column purified and quantified on an Agilent 2100 Bioanalyzer DNA 1000 chip, diluted and sequenced according to *Illumina* GA2 protocols on a paired-end run, 76 cycles in each direction. The sequencing run was analyzed as described above, resulting in 9,908 unique mtDNA fragments that align to the human mtDNA reference sequence. The average fragment length of the mapped reads was 77.3bp. The mtDNA sequence assembled was identical to the one retrieved by PEC enrichment. The average coverage was 46.2-fold. A direct comparison of shot-gun reads and reads from the PEC enrichment for positions 16,195 -16,399 of the hypervariable region of the mtDNA is shown in Suppl. Figure 4.

Phylogenetic analysis

The 54 modern human mtDNAs, 6 Neandertal mtDNAs, the Kostenki early modern human mtDNA, chimpanzee, bonobo and the Denisova mtDNA were aligned using the software Muscle[15]. A phylogenetic tree was estimated in a Bayesian framework using MrBayes 3.1.2[16], with a GTR+I+Γ model of substitutions and default parameters for the MCMC, with 5,000,000 generations sampling every 1,000 generations and a burn-in of 1,000,000 generations. We observed stationarity (using Tracer 1.4[17]) after 1,000,000 generations. A consensus tree from all 4,000 trees was calculated using TreeAnnotater V.1.4.8[17] (Fig 3). Pairwise nucleotide difference between mtDNAs were calculated using MEGA 4.1[18] (Fig 2).

Table S2. Average pairwise nucleotide differences among 54 humans mtDNAs, from the humans to 6 Neandertal mtDNAs, to the Denisova mtDNA, and to the chimpanzee mtDNA.

	Average nucleotide differences	Maximum differences	Minimum differences
Among 54 humans (Green et al. 2008)	59.7	106	1
54 humans to 6 Neandertals (Briggs et al. 2009)	201.6	220	185
54 humans to Denisova (FM673705)	384.9	396	372
54 humans to chimpanzee (X93335)	1461.5	1474	1448

Estimation of divergence times

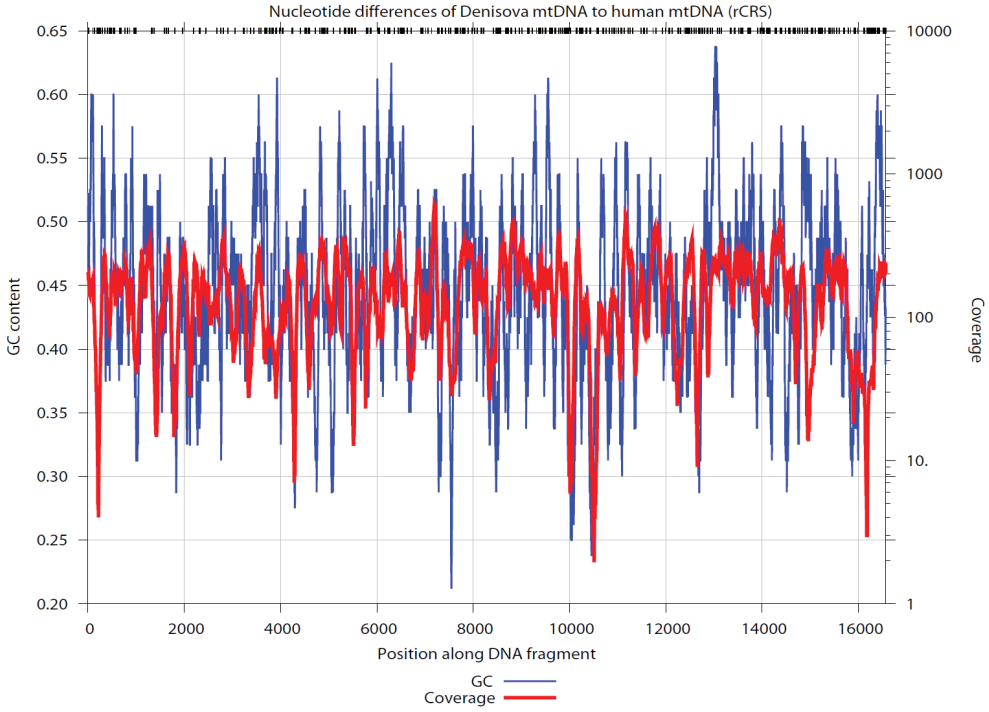
We estimated divergence times between major lineages of hominin mtDNAs (Table S2) using the Bayesian framework employed by BEAST v1.5.3[17]. In addition to the Denisova sequence, we chose a single mtDNA genome to represent Neandertals (Vindija 33.16; NC_011137), modern humans (rCRS; AC_000021), chimpanzees (NC_001643), and bonobos (NC_001644). Using this five-taxon alignment, we partition the data into three nested subsets: all positions (N=16,586 aligned positions), third codon positions (N=3,575), and four-fold degenerate third codon positions (N=1,827). For all third codon positions, we considered non-overlapping codons from the 12 genes encoded on the light strand of the mitochondrial genome (3,575 codons). Each data partition was analyzed using two molecular clock models: a strict molecular clock or a relaxed uncorrelated lognormal molecular clock[19]. For each data partition and clock model, we provide two alternative clock calibrations by assuming that the mtDNA divergence time between chimpanzees and humans follows a normal distribution with a mean of 6 or 7 million years (+/- 500,000 years). Both clock calibrations incorporated the approximate fossil ages of the Vindija 33.16 Neandertal (38,310 years) and of Denisova (40,000 years). For each strict molecular clock analysis, we performed two independent runs of 20,000,000 generations of the Markov chain Monte Carlo (MCMC) with the first 2,000,000 generations of each run discarded as burn-in. For analyses using a relaxed molecular clock, we ran the MCMC for two independent runs of 40,000,000 generations with the first 4,000,000 generations discarded as burn-in. For each data partition, we chose the best-fit model of sequence evolution using DT-ModSel[18] and PAUP*[19].

For each analysis, the mean divergence time of Denisova mtDNA was consistently at least ~2X older than the divergence of Neandertals and modern humans (Table S2). Divergence times were very similar between alternative molecular clock models but were consistently slightly older (<10%) when estimated with the relaxed molecular clock. Assuming a relaxed molecular clock, the standard deviation of rates across branches was near zero, and thus consistent with little substitution rate heterogeneity among branches. Due to uncertainty regarding the age of the Denisova mtDNA, we also re-estimated divergence times assuming that the Denisova mtDNA derives from a contemporary individual (*i.e.* as above but with no Denisova tip-calibration and assuming a strict molecular clock with a human-chimp divergence of 6my). The estimated divergence time between the Denisova mtDNA

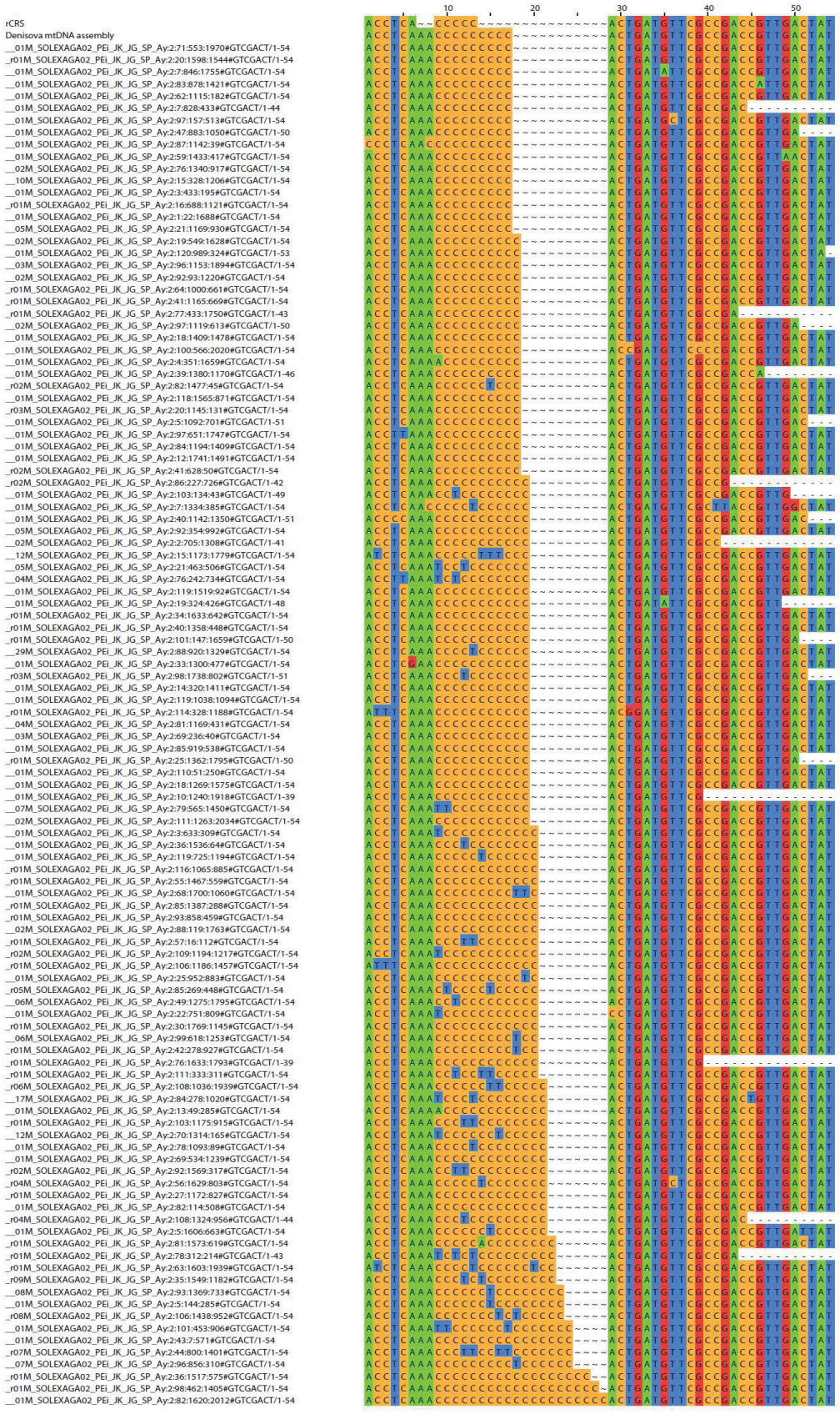
and the mtDNA common ancestors of modern humans and Neandertals changed by <2% for the three data partitions. Thus, our estimated divergence times are largely dominated by the prior assumption of the time of the human-chimpanzee split and are not sensitive to the date of the Denisova fossil.

Protein evolution

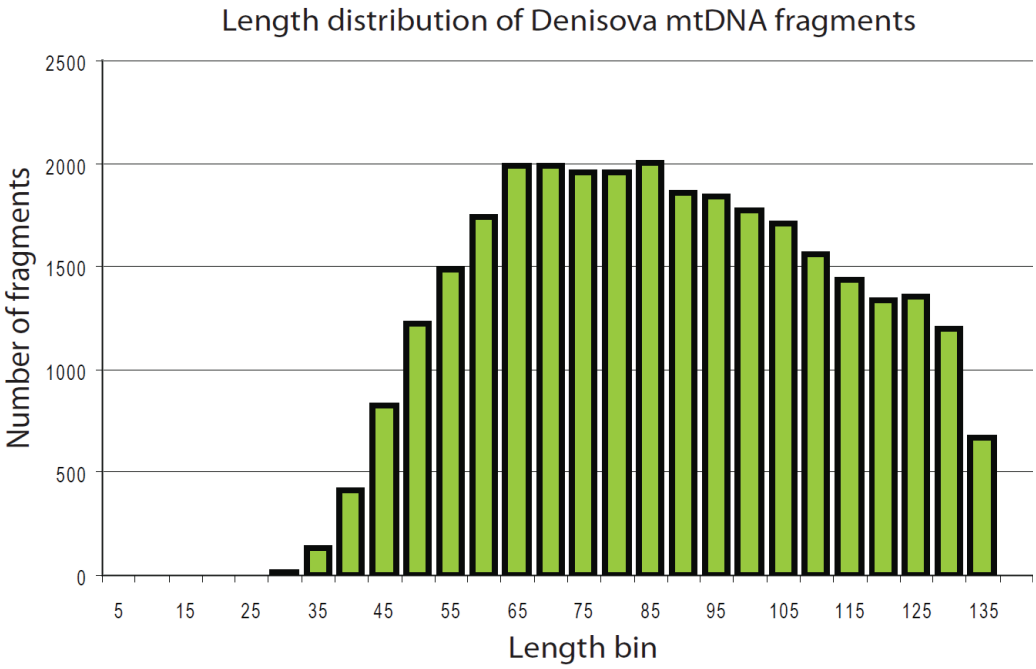
We constructed a concatenated alignment of the 12 genes coded on the light strand of the mitochondrial genome. We excluded all overlapping codons and *ND6* which is encoded on the heavy strand and differs from the other genes in base and codon composition[22]. We aligned the Denisova sequence, two Neandertals (FM865411 and AM948965), two modern humans (AF346985 and AC_000021), one chimpanzee (NC_001643), and one bonobo (NC_001644) and estimated branch-specific rates of *dN* and *dS* using maximum likelihood as implemented in *codeml* in the PAML program suite v4.3[23]. We then repeated this analysis with the Denisova sequence removed (Suppl. Figure 5). Substitution patterns across the two unrooted phylogenies were highly similar with the exception of the localization of a subset of synonymous substitutions on the modern human and Neandertal lineages, respectively. The reconstructed ancestral sequence of the common ancestor of modern humans and Neandertals differs by 36 synonymous and 5 non-synonymous substitutions between the two analyses, resulting in a net increase in ~12 synonymous substitutions assigned to the Neandertal branch and a corresponding reduction in *dN/dS* along this lineage. Briggs *et al*. (2009) suggested that an observed increase in *dN/dS* in Neandertals was consistent with reduced purifying selection on mtDNA protein-coding regions due to a small long-term effective population size of Neandertals but urged caution because the observation was based on relatively few substitutions. Assuming that the ancestral reconstructions are more accurate with the inclusion of the Denisova sequence, it now appears that the higher *dN/dS* along the Neandertal lineage[7,11] was largely the result of incorrect assignment of ancestral states for the common ancestor of modern humans and Neandertals and does not reflect Neandertal population biology *per se*. The inference of a low effective population size in Neandertals based on the direct observation of low population levels of mtDNA genetic diversity are unaffected by these results[7].



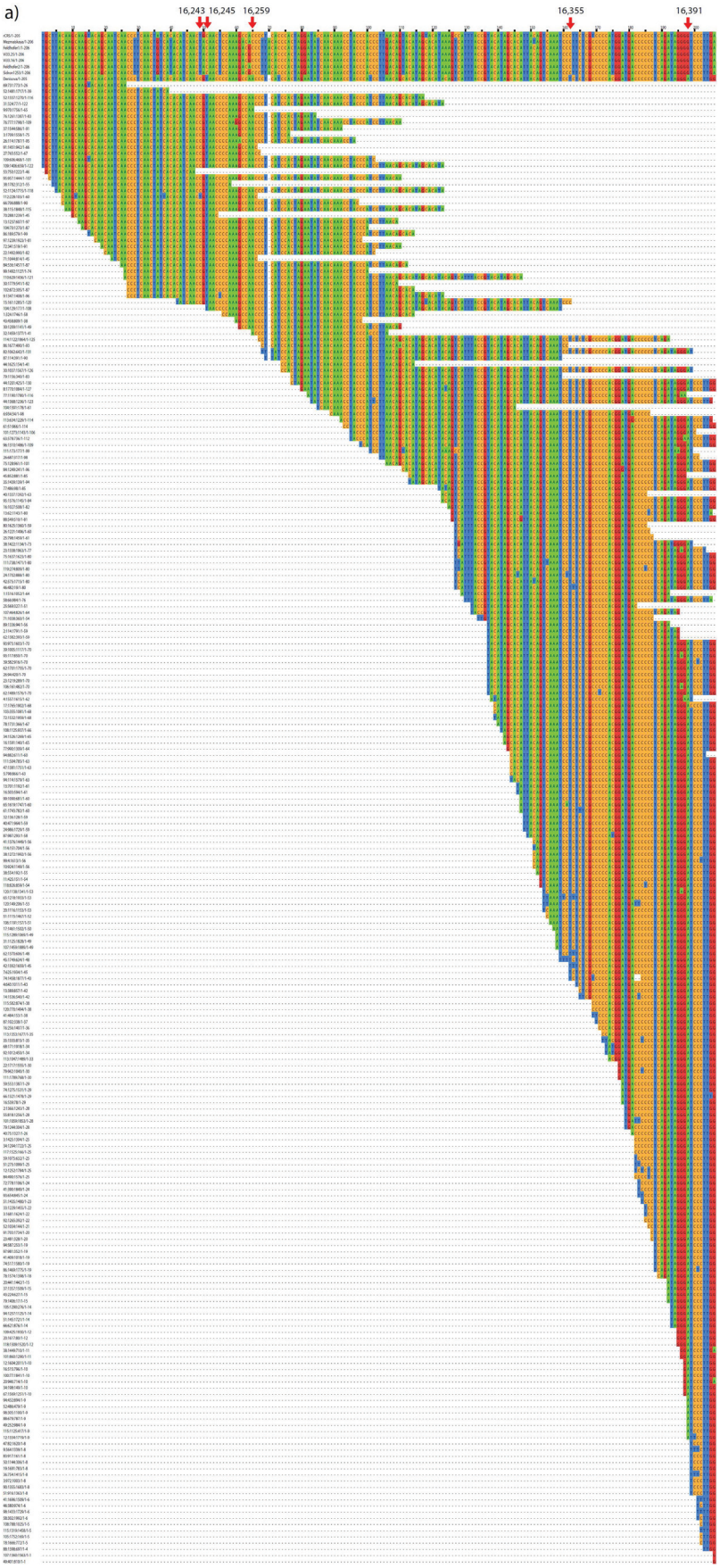
Suppl. Figure 1. Sequence coverage of the Denisova mtDNA (red, after clustering unique molecules) and GC content (blue) for the complete mtDNA. Above, along the second x-axis, the nucleotide differences along the Denisova mtDNA to a present-day human mtDNA (rCRS) are shown.

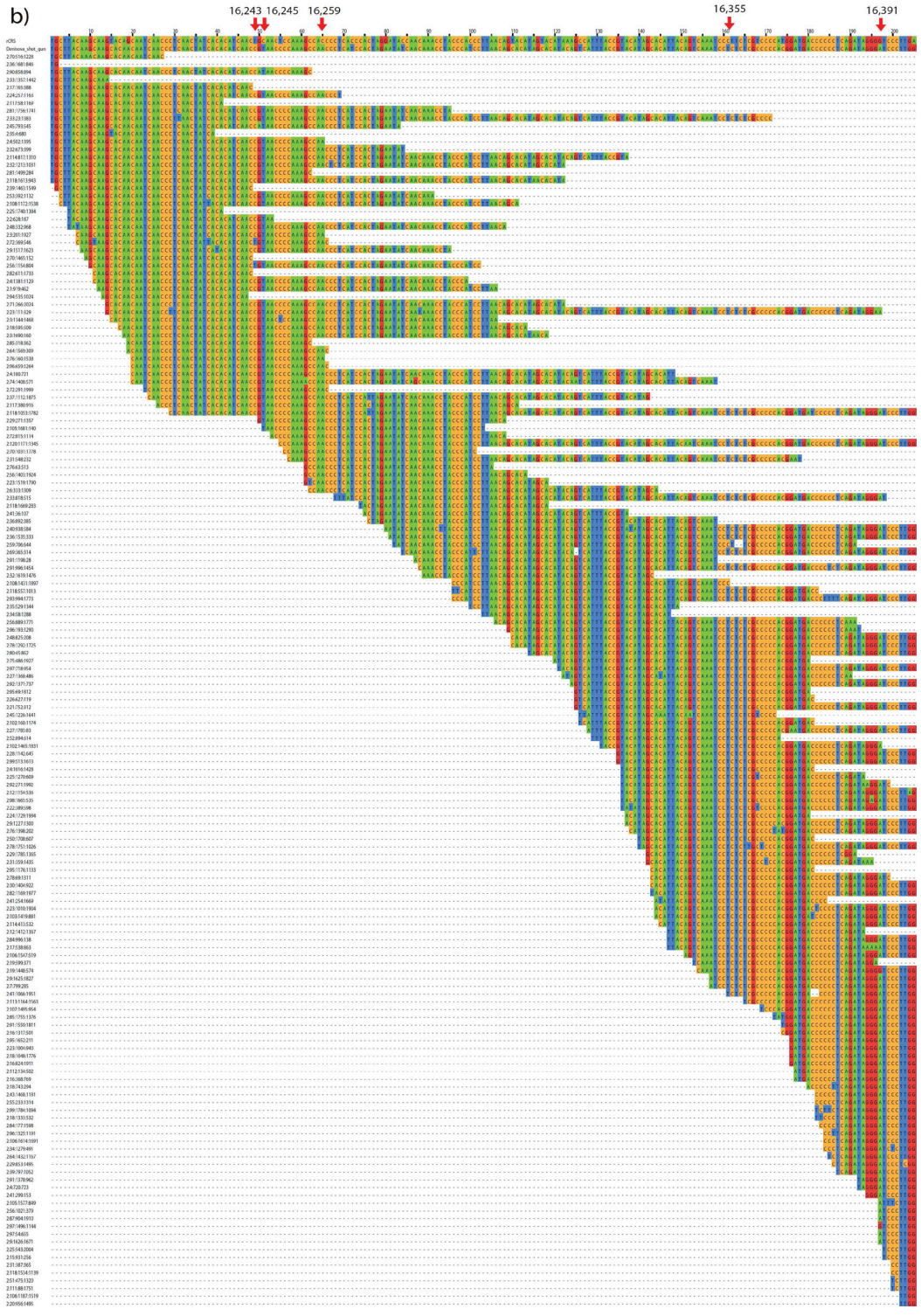


Suppl. Figure 2. Fragments overlapping rCRS nucleotide positions 5,889 to 5,925 of the Denisova mtDNA.

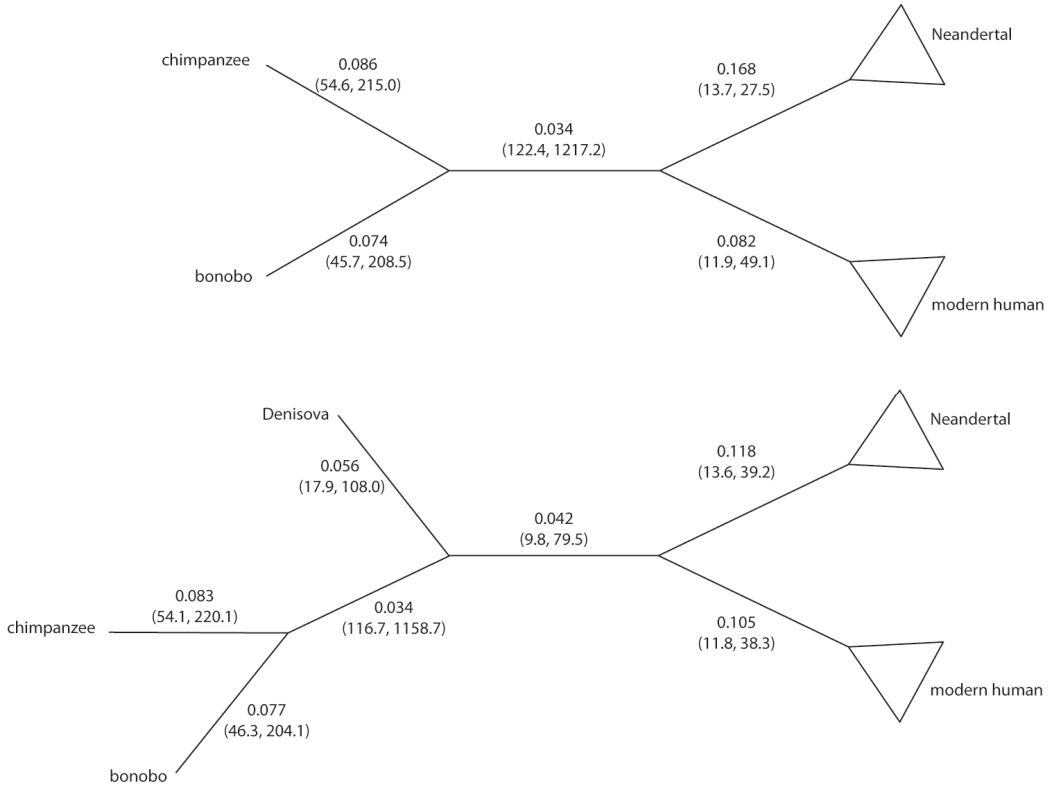


Suppl. Figure 3. Length distribution of Denisova mtDNA fragments.


a)
16,243
16,245
16,259
16,355
16,391



Suppl. Figure 4. Fragments overlapping rCRS nucleotide positions 16,195 to 16,399 of the Denisova mtDNA aligned to the modern human rCRS and six complete Neandertal mtDNAs, above (a) for fragments isolated by PEC; and below (b) for fragments sequenced directly from a library. Red arrows indicate positions that are informative with regard to modern human mtDNA contamination.



Suppl. Figure 5. Rates of hominin mtDNA protein evolution estimated without (top panel) or with the Denisova mtDNA. For each branch, the per site rate of protein evolution (*dN/dS*) and the estimated number of non-synonymous and synonymous substitutions (in parentheses) are given.

Table S2. TMRCA between major mtDNA lineages of hominids.

calibration	partition	sites	model[1]	Sub. Rate[2] (95% HPD)	H-N TMRCA[3] (95% HPD)	HN–D TMRCA[4] (95% HPD)	Rel. TMRCA[5]
strict clock							
H-C 6MYA	all sites	16,586	HKY+I	1.31E-08 (1.06E-08 - 1.58E-08)	525,700 (408,400 - 646,800)	1,021,600 (804,700 - 1,228,700)	1.9
H-C 6MYA	3rd codon	3,575	TrN+I	3.90E-08 (3.05E-08 - 4.82E-08)	371,300 (267,200 - 479,900)	869,100 (661,800 - 1,083,400)	2.3
H-C 6MYA	4-fold	1,827	TrN	3.43E-08 (2.72E-08 - 4.17E-08)	465,700 (321,200 - 618,000	1,040,900 (779,300 - 1,313,500)	2.2
H-C 7MYA	all sites	16,586	HKY+I	1.12E-08 (9.27E-09 - 1.32E-08)	612,400 (485,500 - 745,200)	1,193,300 (973,300 - 1,422,700)	1.9
H-C 7MYA	3rd codon	3,575	TrN+I	3.31E-08 (2.64E-08 - 4.03E-08)	432,600 (317,400 - 556,800)	1,018,200 (790,900 - 1,258,500)	2.4
H-C 7MYA	4-fold	1,827	TrN	2.91E-08 (2.36E-08 - 3.47E-08)	543,400 (378,500 - 713,100)	1,218,500 (927,400 - 1,517,800)	2.2
relaxed clock							
H-C 6MYA	all sites	16,586	HKY+I	1.31E-08 (1.05E-08 - 1.59E-08)	534,400 (353,000 - 720,900)	1,037,800 (723,400 - 1,352,900)	1.9
H-C 6MYA	3rd codon	3,575	TrN+I	3.87E-08 (2.96E-08 - 4.80E-08)	386,500 (198,800 - 587,100)	923,700 (534,700 - 1,371,300)	2.4
H-C 6MYA	4-fold	1,827	TrN	3.40E-08 (2.65E-08 - 4.18E-08)	487,500 (219,700 - 778,800)	1,129,100 (585,100 - 1,763,800)	2.3
H-C 7MYA	all sites	16,586	HKY+I	1.11E-08 (9.08E-09 - 1.32E-08)	623,500 (409,000 - 852,300)	1,216,300 (835,600 - 1,600,100)	2.0
H-C 7MYA	3rd codon	3,575	TrN+I	3.29E-08 (2.60E-08 - 4.03E-08)	453,000 (232,100 - 715,100)	1,090,00 (611,100 - 1,687,800)	2.4
H-C 7MYA	4-fold	1,827	TrN	2.89E-08 (2.31E-08 - 3.50E-08)	569,300 (262,000 - 911,900)	1,323,800 (685,200 - 2,063,700)	2.3

1. Best-fit model of sequence evolution. HKY = Hasegawa-Kishino-Yano model[24] , TrN= Tamura-Nei model[25], I = invariant sites correction for rate heterogeneity.
2. Substitution rate per site per year.
3. Average time in years to the most recent common mtDNA ancestor of Neandertals and modern humans.
4. Average time in years to the most recent common mtDNA ancestor of Denisova mtDNA and the common ancestor of Neandertals and modern humans.
5. Relative average TMRCA (HN–A TMRCA / H-N TMRCA).

Supplementary References

1 Derevianko, A. P. in *The Paleolithic of Siberia - New Discoveries and Interpretation* (ed A. P. Derevianko) (University of Illinois Press, 1998).

2 Derevianko, A. P. & Shunkov, M. V. in *Paleoenvironment and Palaeolithic Human Occupation of Gorny Altai* (ed A. P. Derevianko, Shunkov, M. V.) (Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, 2003).

3 Derevianko, A., Shunkov, M. & Volkov, P. A Paleolithic bracelet from Denisova cave. *Archaeology, Ethnology and Anthropology of Eurasia* **34**, 13-25 (2008).

4 Rohland, N. & Hofreiter, M. Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343-352 (2007).

5 Maricic, T. & Paabo, S. Optimization of 454 sequencing library preparation from small amounts of DNA permits sequence determination of both DNA strands. *Biotechniques* **46**, 51-52, 54-57 (2009).

6 Meyer, M. *et al.* From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res* **36**, (2008).

7 Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321 (2009).

8 Krause, J. *et al.* A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Curr Biol* 20, 231-236 (2010).

9 Kircher, M., Stenzel, U. & Kelso, J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* **10**, R83 (2009).

10 Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147 (1999).

11 Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426 (2008).

12 Lee, H. Y., Chung, U., Yoo, J. E., Park, M. J. & Shin, K. J. Quantitative and qualitative profiling of mitochondrial DNA length heteroplasmy. *Electrophoresis* **25**, 28-34 (2004).

13 Petri, B., von Haeseler, A. & Paabo, S. Extreme sequence heteroplasmy in bat mitochondrial DNA. *Biol Chem* **377**, 661-667 (1996).

14 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

15 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

16 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).

17 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214 (2007).

18 Kumar, S., Tamura, K. & Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150-163 (2004).

19 Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88 (2006).

20 Minin, V., Abdo, Z., Joyce, P. & Sullivan, J. Performance-based selection of likelihood models for phylogeny estimation. *Systematic biology* **52**, 674-683 (2003).
21 PAUP* beta version. Phylogenetic analysis using parsimony (*and other methods) (Sinauer Associated, Sunderland, MA, 2002).
22 Yang, Z., Nielsen, R. & Hasegawa, M. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol Biol Evol* **15**, 1600-1611 (1998).
23 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007).
24 Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**, 160-174 (1985).
25 Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512-526 (1993).