

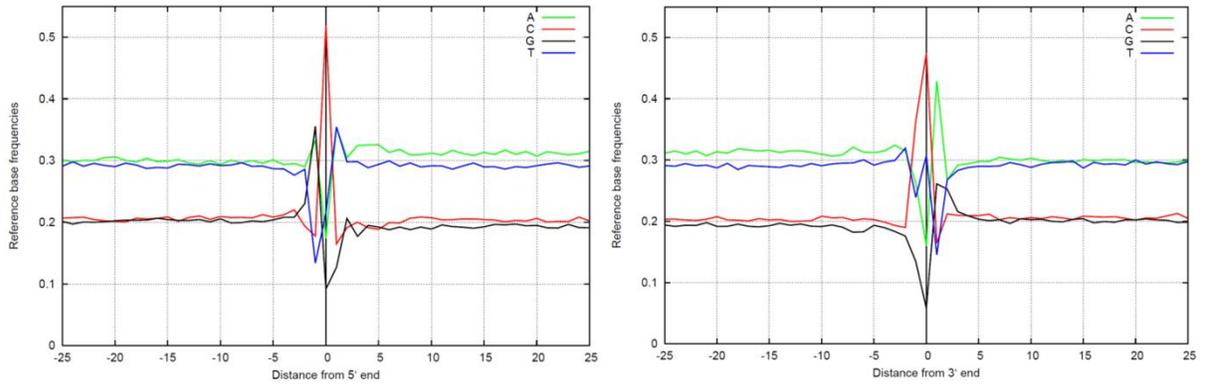
Supplementary Information

Selective enrichment of damaged DNA molecules for ancient genome sequencing

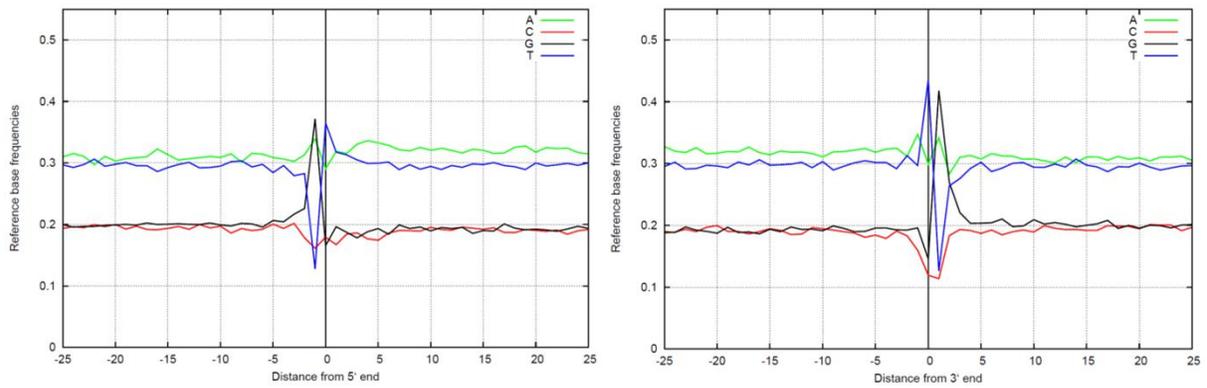
Marie-Theres Gansauge¹ and Matthias Meyer¹

¹*Max Planck Institute for Evolutionary Anthropology, Evolutionary Genetics Department,
Deutscher Platz 6, D-04103 Leipzig, Germany*

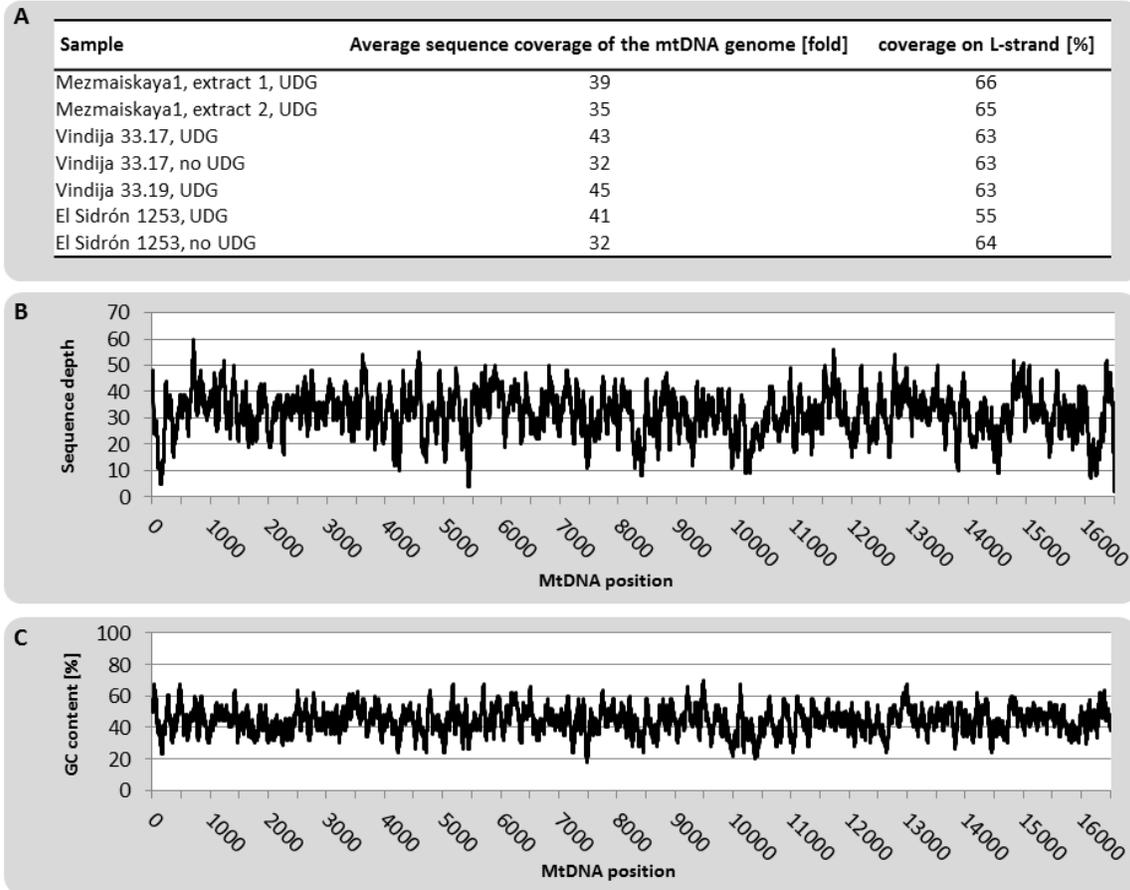
Supernatant fraction



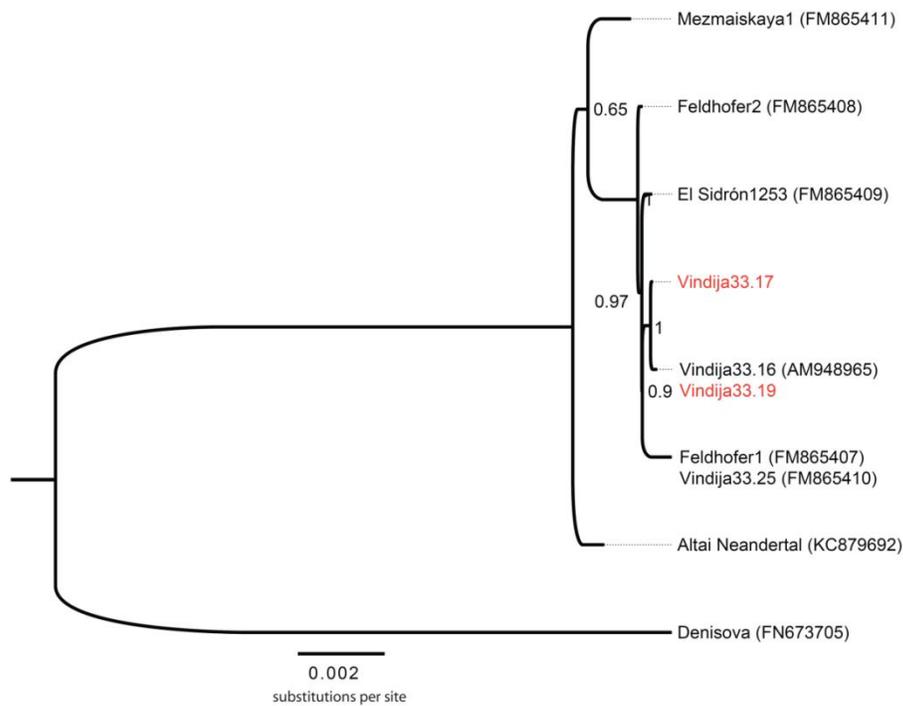
Bead fraction



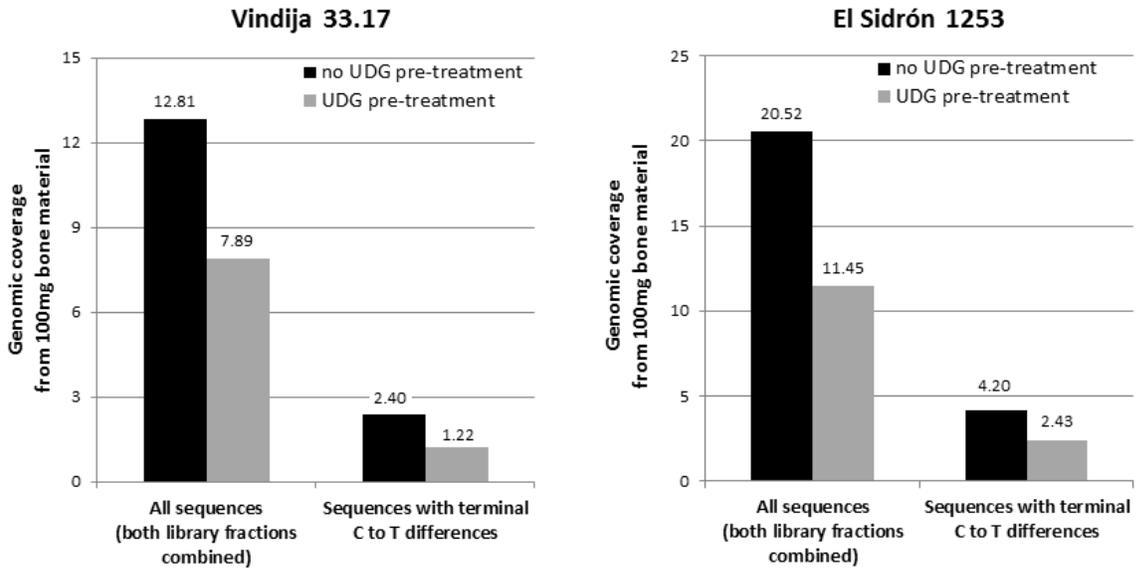
Supplementary Figure 1: Base composition of the reference genome around the start and end positions of aligned sequences from the supernatant and bead library fractions of Vi33.17 (no UDG pretreatment).



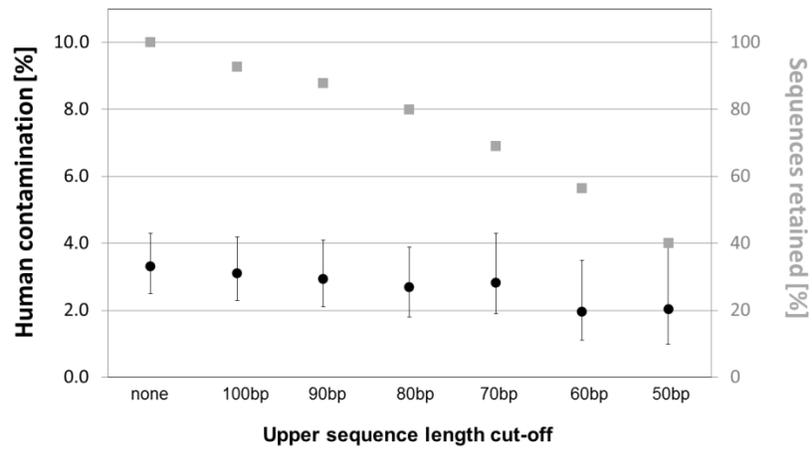
Supplementary Figure 2: Overview of the mitochondrial genome assemblies. (A) Average sequence coverage obtained from sequences with terminal C to T substitutions. All sequences were re-aligned to the consensus sequence of the respective Neanderthal individual. We also show that coverage is skewed towards the cytosine-rich light strand. (B) As shown for the U selected library that produced the lowest average coverage of the mtDNA genome (El Sidrón, no UDG pre-treatment), all positions of the mitochondrial genome are covered by 3 or more sequence reads, despite substantial variation in local GC-content. (C) The GC-content of the mitochondrial genome is shown in a 50bp sliding window.



Supplementary Figure 3: Mid-point rooted phylogenetic tree of mtDNA genome sequences of Vindija 33.17, Vindija 33.19 (highlighted in red), 7 published Neanderthals (Briggs et al. 2009; Prüfer 2013), and the Denisovan phalanx (Briggs et al. 2010). Sequences were aligned using MAFFT (Katoh and Standley 2013). The General Time Reversible sequence evolution model was applied with a substitution site model (GTR+Gamma) determined by the best-fit model approach of Modeltest and PAUP* (Posada and Crandall 1998). The tree was reconstructed in a Bayesian statistical framework using MrBayes (Ronquist and Huelsenbeck 2003), running 20,000,000 iterations of the Markov Chain Monte Carlo and discarding the first 2,000,000 iterations as burn-in. Genbank accession numbers are provided in brackets.



Supplementary Figure 4: Predicted number of nuclear genomes that would be recovered in DNA libraries prepared from 100 mg of bone powder from two Neanderthal samples depending on the library preparation strategy chosen. Sequences shorter than 35bp were discarded.



Supplementary Figure 5: Further reduction of human contamination that can be achieved by applying an upper length cutoff on the mtDNA sequences generated from the second Mezmaiskaya DNA extract (in addition to U selection and filtering for sequences with terminal C to T differences). Bars denote 95% confidence intervals. The fraction of sequences retained after length filtering is denoted by gray rectangles. The lower sequence length cutoff is 35bp.

Supplementary References

- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z et al. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**(5938): 318-321.
- Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**(6): e87.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol.Biol.Evol.* **30**(4): 772-780.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**(9): 817-818.
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C et al. 2013. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**(7481): 43-49.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**(12): 1572-1574.