

Recent human evolution by a complete Neandertal mitochondrial genome sequence (to appear in Cell)

This paper emerged from a joint work of various collaborators. In the following, I will describe in which parts of the project I was involved.

Together with Anna-Sapfo Malaspinas, a first year Ph.D. student in the Department of Integrative Biology, we were responsible for most phylogenetic and population genetic analysis related to the Neandertal genome project. So we received the assembly of the Neandertal genome project from our collaborators and, assuming this was correct, we did the following analysis.

Annotation:

To guide annotation we aligned the Neandertal mtDNA sequence to the Cambridge human reference sequence. We found 13 protein-coding genes, 22 tRNAs and 2 rRNAs, as well as a control region. For the 13 protein-coding genes the start and stop codons were shared with at least the chimpanzee or the human mitochondrial genome and no frameshift, deletion or insertion was found. In addition, the 3-dimensional structure of the tRNAs did not differ significantly from the corresponding structure in humans.

Phylogenetic tree:

We aligned the assembled Neandertal mtDNA sequence to mtDNA genomes of 54 modern humans sampled from all major language groups worldwide and to the mtDNA genome of one chimpanzee. The substitution pattern between Neandertal and chimpanzee was very similar to the substitution pattern between human and chimpanzee advocating for low damage.

To reconstruct the phylogenetic tree we used different partition schemes and tree reconstruction methods. As partitions we used the whole genomes, a subset of the whole genome (concatenation of proteins, then tRNAs and finally rRNAs), and a partition of the whole genome into proteins, rRNAs, tRNAs, and non-coding base pairs. We used neighbor-joining, maximum parsimony, maximum likelihood and a Bayesian approach to reconstruct the tree for the various partition schemes. We tested for the model of substitution and found GTR+G and TN93+G to be the best fitting models. Due to limitations by the programs used in our study, we chose the GTR+G model for all further analysis. We found that the Neandertal mitochondrial genome falls outside the human diversity for all tree reconstruction methods and partitions of the data.

Divergence times:

To further explore the evolutionary relationship between Neandertal and modern human mtDNA, we used the topology found above to date the divergence events. In particular, we used only a subset of 10 humans from the previous 54 humans to restrict ourselves to a topology with high support. We tested the molecular clock hypothesis and could not reject the null hypothesis of a constant mutation rate across the tree. The Neanderthal-human divergence time was estimated by a Bayesian approach assuming that the chimpanzees diverged 6-8 million years ago. The posterior distribution resulting from this Bayesian analysis showed a Neanderthal-human divergence time of 660,000 +/- 140,000 years, which is consistent with fossil evidence. In addition, we also showed that an increase of the amount of data would not reduce the confidence interval of the time estimate.

Natural selection:

To test for selection on each of the genes, we took a population genetics approach with a MacDonald Kreitman test using first the Neandertal and then the chimpanzee as outgroup. We found no gene departing significantly from neutrality after Bonferroni correction when using the Neandertal as outgroup, although the test was significant for the protein COX2 without Bonferroni correction. Interestingly, using the chimpanzee as outgroup did not lead to the same genes under positive selection. Finally, we also tested for selection using a phylogenetic approach. With this approach, we did not detect a burst of positive selection in the human lineage for any of the genes.