Cell

A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing

Richard E. Green,[1,*] Anna-Sapfo Malaspinas,[2] Johannes Krause,[1] Adrian W. Briggs,[1] Philip L.F. Johnson,[3] Caroline Uhler,[4] Matthias Meyer,[1] Jeffrey M. Good,[1] Tomislav Maricic,[1] Udo Stenzel,[1] Kay Prüfer,[1] Michael Siebauer,[1] Hernán A. Burbano,[1] Michael Ronan,[5] Jonathan M. Rothberg,[6] Michael Egholm,[5] Pavao Rudan,[7] Dejana Brajković,[8] Željko Kućan,[7] Ivan Gušić,[7] Mårten Wikström,[9] Liisa Laakkonen,[10] Janet Kelso,[1] Montgomery Slatkin,[2] and Svante Pääbo[1]

[1]Max-Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany
[2]Department of Integrative Biology
[3]Biophysics Graduate Group
[4]Department of Statistics
University of California, Berkeley, CA 94720, USA
[5]454 Life Sciences, Branford, CT 06405, USA
[6]The Rothberg Institute for Childhood Diseases, Guilford, CT 06437, USA
[7]Croatian Academy of Sciences and Arts, Zrinski trg 11, HR-10000 Zagreb, Croatia
[8]Croatian Academy of Sciences and Arts, Institute for Quaternary Paleontology and Geology, Ante Kovačića 5, HR-10000 Zagreb, Croatia
[9]Helsinki Bioenergentics Group, Program for Structural Biology and Biophysics, Institute of Biotechnology
[10]Division of Biochemistry, Department of Biological and Environmental Sciences, Faculty of Biosciences, University of Helsinki
FIN-00014 Helsinki, Finland
*Correspondence: green@eva.mpg.de
DOI 10.1016/j.cell.2008.06.021

SUMMARY

A complete mitochondrial (mt) genome sequence was reconstructed from a 38,000 year-old Neandertal individual with 8341 mtDNA sequences identified among 4.8 Gb of DNA generated from ~0.3 g of bone. Analysis of the assembled sequence unequivocally establishes that the Neandertal mtDNA falls outside the variation of extant human mtDNAs, and allows an estimate of the divergence date between the two mtDNA lineages of 660,000 ± 140,000 years. Of the 13 proteins encoded in the mtDNA, subunit 2 of cytochrome *c* oxidase of the mitochondrial electron transport chain has experienced the largest number of amino acid substitutions in human ancestors since the separation from Neandertals. There is evidence that purifying selection in the Neandertal mtDNA was reduced compared with other primate lineages, suggesting that the effective population size of Neandertals was small.

INTRODUCTION

Although it is well established that Neandertals are the hominid form most closely related to present-day humans, their exact relationship with modern humans remains a topic of debate (Hublin and Pääbo, 2006; Soficaru et al., 2006; Harvati et al., 2007). Molecular genetic data first spoke to this issue in 1997, when a 379 base pair section of the hypervariable region I (*HVRI*) of the mitochondrial genome (mtDNA) was determined from the Neandertal-type specimen found in 1856 in Neander Valley, near Düsseldorf, Germany (Krings et al., 1997). Since then, a total of 15 complete or partial Neandertal *HVRI* sequences, as well as two *HVRII* sequences (Krings et al., 1999; Krings et al., 2000), have been described. Phylogenetic analyses of these suggest that Neandertal mtDNA falls outside the variation of modern human mtDNA. Since the mtDNA genome is maternally inherited without recombination, these results indicate that Neandertals made no lasting contribution to the modern human mtDNA gene pool (Krings et al., 1997; Currat and Excoffier, 2004; Serre et al., 2004).

High-throughput *454* sequencing techniques have recently been applied to ancient DNA (Green et al., 2006; Poinar et al., 2006; Stiller et al., 2006). These methods open new possibilities for the retrieval of ancient DNA that has hitherto relied either on the cloning of random molecules in bacteria (Higuchi et al., 1984; Pääbo, 1985; Noonan et al., 2005, 2006) or on the PCR amplification of individual DNA sequences of interest (Pääbo and Wilson, 1988; Pääbo et al., 2004). The main benefit of the *454* sequencing technique is the sheer volume of sequence data that make it practical to undertake genome-scale ancient DNA sequencing projects. This is particularly feasible for mitochondrial genomes (Gilbert et al., 2007), given their smaller size relative to the nuclear genome and their abundance in cells, where, typically, several hundred mtDNAs per nuclear genome exist.

The *454* sequence data from ancient DNA have also allowed an increased understanding of DNA diagenesis (i.e., how DNA

is modified during deposition in a burial context). In particular, they have allowed a quantitative model of how DNA degradation and chemical modification occurs, and how the effects of these processes interact with the molecular manipulations used to generate sequencing libraries (Briggs et al., 2007). Notably, although it was previously known that a high rate of cytosine deamination occurs in ancient DNA (Hofreiter et al., 2001), it has become clear that this is particularly prevalent in the ends of the ancient molecules, presumably because these are often single stranded (Briggs et al., 2007). Deamination of cytosine residues results in uracil residues that are read as thymine by DNA polymerases, leading to a high rate of C-to-T transitions. A high rate of G-to-A transitions observed near the 3′ ends of sequence reads is thought to be caused by deaminated cytosine residues on the complementary strands used as templates during the fill-in reaction to create blunt ends when sequencing libraries are constructed (Briggs et al., 2007).

By *454* sequencing, we have generated 34.9-fold coverage of the Neandertal mtDNA genome from a Neandertal bone (Vindija bone 33.16) excavated in 1980 from Vindija Cave, Croatia (Malez and Ullrich, 1982). It has been dated to 38,310 ± 2130 years before present (Serre et al., 2004). Previously, the mtDNA *HVRI* sequence of this bone has been determined (Serre et al., 2004), as well as 2414 bp of mtDNA sequences by *454* sequencing (Green et al., 2006). Here, we present its complete mtDNA sequence, as well as the insights it allows into recent human and Neandertal mtDNA evolution.

RESULTS

DNA Sequence Determination

Three DNA extracts, each from 100–200 mg of a Neandertal bone (Vindija 33.16) were prepared in our cleanroom facility where several precautions against DNA contamination are implemented (Experimental Procedures). These include complete separation from other parts of the laboratories, direct delivery of all equipment and reagents to the facility, positive pressure generated with filtered air that excludes particles larger than 0.2 μm, and UV irradiation and bleach treatment of all surfaces. The bone surface was removed prior to extraction. However, the interior of bones is also often contaminated with modern human DNA, presumably due to past washing and other treatments of Neandertal bones. Thus, we analyzed each extract for contamination by extant human mtDNA by PCR with primers flanking positions in the *HVRI* that distinguish extant humans from Neandertals (Green et al., 2006), and amplify both types of mtDNA with similar efficiencies. Following amplification, we cloned the PCR product and sequenced 103–112 clones to determine the ratio of Neandertal to extant human mtDNA. The contamination rate in the three extracts ranged from 0%–0.9% (see Figure S1 available online).

From these extracts, we generated a total of nine *454* libraries in the cleanroom facility with *454* adapters with a Neandertal-specific sequence key that is unique to this project, and thus unequivocally identifies each sequence determined as derived from the extract of a Neandertal bone (Briggs et al., 2007). This allows detection of any contamination that may be introduced in subsequent handling and sequencing steps outside the cleanroom. To maximize the library and sequence yield, we incorporate two modifications to the standard *454* protocol that reduce the need to perform titration runs of libraries (Meyer et al., 2008) and allow more molecules to be retrieved during library preparation (Maricic and Pääbo, unpublished results). From these libraries, we generated a total of 39 million sequence reads by 147 runs on the GS FLX sequencing platform. Bases were called with the standard *454* signal threshold and filtering criteria with minor modifications tailored for short, ancient sequence reads (Meyer et al., 2008).

Neandertal sequences were identified within each run as described previously (Green et al., 2006), with the chief criterion being sequence similarity to a primate genome. mtDNA sequences were identified from these with further criteria (see Experimental Procedures), including similarity to the human mtDNA at least as great as to any nuclear DNA sequence. While the total fraction of sequences that are identified as Neandertal varied across libraries and library pools, the ratio of putative Neandertal nuclear DNA sequences to mtDNA sequences varied little among these libraries, and averaged 171. This corresponds to ~2100 mtDNA molecules per cell. In total, 8341 mtDNA sequences were identified. They are of an average length of 69.4 bp (SD = 26.4), with the shortest fragment identified being 30 bp (limited by the length cut-off in the analysis pipeline), and the longest fragment being 278 bp (limited by the flow cycles performed on the GS FLX instrument).

mtDNA Genome Assembly

Ancient DNA sequences present a challenge for DNA sequence assembly since they are typically short and exhibit high rates of nucleotide misincorporation. A further complication is that pyrosequencing (Ronaghi et al., 1998), for example, as performed on the GS FLX platform, calls long polymers of the same base with reduced accuracy. With these issues in mind, we designed an assembly procedure for ancient DNA. In short, each sequence identified as mtDNA was aligned over its entire length to the human reference mtDNA sequence (UCSC build hg18). These alignments were then merged, and each alignment column was examined to determine the majority base, yielding an assembled mtDNA sequence. Homopolymer lengths at positions where the reference human carries ≥5 identical bases were determined by analysis of the raw signal distributions as described in the Supplemental Data.

Following this examination, some problematic regions remained in the assembly. These include four regions of a total length of 20 bp, where no sequence coverage existed, eight other regions amounting to a total of 117 bp covered by only single reads, nine positions where no majority base existed due to low coverage, and 31 homopolymers for which the data were not sufficient to determine their length. These regions were amplified by PCR from a Vindija 33.16 bone extract in two-step multiplex PCRs (Krause et al., 2006), cloned, and sequenced by Sanger technology in order to complete the assembly (Experimental Procedures).

We then reapplied our mtDNA fragment detection pipeline to all Neandertal DNA sequences determined from the bone, but compared them to the assembled Neandertal mtDNA instead of the reference human mtDNA. This resulted in the detection