RECONSTRUCTION OF AUTHENTIC COMPLETE MITOCHONDRIAL GENOME SEQUENCE FROM ANCIENT DNA: PERSPECTIVES FOR EVOLUTIONARY, HISTORICAL AND FORENSIC IDENTIFICATIONS

Evgeny I. Rogaev[1,2,3,4¶], **Anastasia P. Grigorenko**[2,3], **Yuri K. Moliaka**[1,2]

[1]University of Massachusetts Medical School, 303 Belmont St., Worcester, MA, 01604, USA; [2] Laboratory of Molecular Brain Genetics, Research Center of Mental Health, RAMS, Zagorodnoe sh., 2, 113152; [3]Lomonosov Moscow State University, faculty of bioengineering and bioinformatics, 119992, [4]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia
[¶] To whom correspondence should be addressed.
E-mail: Evgeny.Rogaev@umassmed.edu

There are many advantages of using mitochondrial DNA (mtDNA) in evolutionary and forensic analysis: relatively high intra- and inter-species diversity, one-lineage (maternal) inheritance and high copy presentation in somatic cells. The small (~400bp) variable sequences in the control region (HV1 and HV2) and the cytochrome C gene are commonly used for individual, population and inter-species comparisons. There are some limitations in the application of the mtDNA: 1) insufficient information about the selected short mtDNA sequences; postmortem DNA degradation, contamination by exogenous DNA, accumulation of the "artificial mutations" in ancient specimens; existence of nuclear mt pseudogenes (numt) ; "heteroplasmy" observed in mtDNA which may be real but tissue specific or reflect the presence of numts. These problems may be overcome and highly informative mitochondrial genome sequence may be reconstructed from very old specimens. As an example we present the reconstruction of the sequence of the complete mitochondrial genome of a woolly mammoth extracted from permafrost-preserved remains, which is to date the oldest mitochondrial genome recovered from ancient specimens (the Pleistocene epoch). We demonstrate that well preserved mitochondrial genome fragments, as long as ~1,700 base pairs, can be retrieved from pre-Holocene remains (~33,000 years old). We developed several approaches to control possible postmortem DNA mutations and sequencing errors (which we identified in a number of reported sequences from ancient or old specimens) by analyses of (1) comparative divergence in evolutionary conserved versus non-conserved sites and sequences; (2) the distribution and ratio of synonymous to nonsynonymous substitutions; (3) rates of transitions (A->G/T->C or more common C->T/G->A), corresponding to the type I and type II mutations in ancient DNA. Several lines of evidence have been obtained to demonstrate that our experimental approaches provided no nucleotide errors and that the mammoth genome sequence (16,842 base pairs) determined in this study is authentic. The phylogeny of Elephantinae (closely related species *Mammuthus primigenius (M. primigenius), Elephas maximus (E. maximus)* and *Loxodonta africana (L. africana))* has not been previously resolved by analysis of short mtDNA fragments. To determine the evolutionary history of the woolly mammoth, we sequenced complete mitochondrial genomes of Asian and African elephants (*L. africana* and *E. maximus)*. Phylogenetic reconstruction of the Elephantinae clade based on complete mt genome sequence suggests that *M. primigenius* and *E. maximus* are sister species that diverged soon after their common ancestor split from the *L. africana* lineage. We also reconstructed the phylogeny of these species only using individual protein (13 protein coding genes), 2 rRNAs and the D-loop control region (22 tRNA genes are too short and contain too few substitutions). Remarkably, the majority, but not all of trees reconstructed with the sequence of

individual genes supported the topology recovered using the complete genome. These data further illuminate the importance of comparative analysis of long range mt genome sequences rather the selected short fragments.

To reconstruct mitochondrial genome sequences from multiple PCR products stringent precautions have to be undertaken to work with forensic or ancient DNA. The possibility of cross-contamination has to be eliminated a priori by separation in procedures of analysis of DNA extracted from ancient or criminological samples and DNA from living subjects used for comparison[1,2]. To determine the complete mitochondrial genome sequence from a limited amount of biological material in forensic casework the multiplex PCR can be applied. The multiplex human mtDNA PCR assay designed to avoid numt amplification will be discussed. Search for "unique" individual mt genome signatures in forensic mtDNA will be exemplified by analysis of historical specimens of the Romanov (Nicholas II) family.

1. Rogaev E., Moliaka I., Malyarchuk B., Kondrashov F., Derenko M., Chumakov I., Grigorenko A. Complete Mitochondrial Genome and Phylogeny of Pleistocene Mammoth *Mammuthus primigenius. PlosBiology, V.4. p.403-409, 2006*

2. Rogaev E.I. Analysis of mitochondrial DNA of putative remains of Nicholas II (skeleton N4) and his nephew. In book: Materials of Government Commission to study issues related to investigation and re-burying of remains of Russian Empire Nicholas II and members of his family (Russ) (selected documents), p. 171-182., Moscow, 1998.