

'DNA Fossils', ancient mtDNA preserved in the nuclear genome: witnesses of human evolution milestones

Fragments of mitochondrial DNA (mtDNA) are known to get inserted into the DNA of the cell nucleus (i.e., our main genome) and to form NUMTs, i.e. dysfunctional sequences, nuclear pseudogenes of the mtDNA. Hundreds of NUMTs have been inserted into the human genome over the period of about 50 million years of our evolution. Once in the nucleus, fragments of mtDNA enjoy a safer environment and better repair, so the rate of mutation accumulation decreases. As a result mtDNA sequence remain fairly less changed compared to their mtDNA counterparts (hence the term 'DNA fossils'). Moreover, the mutations that accumulate in NUMT can be distinguished from the mutations accumulated previously in the mitochondrion because NUMTs are dysfunctional sequences and tend to accumulate more non-neutral mutations. These unique features allow one to use NUMT's as a glimpse into our species' distant past.

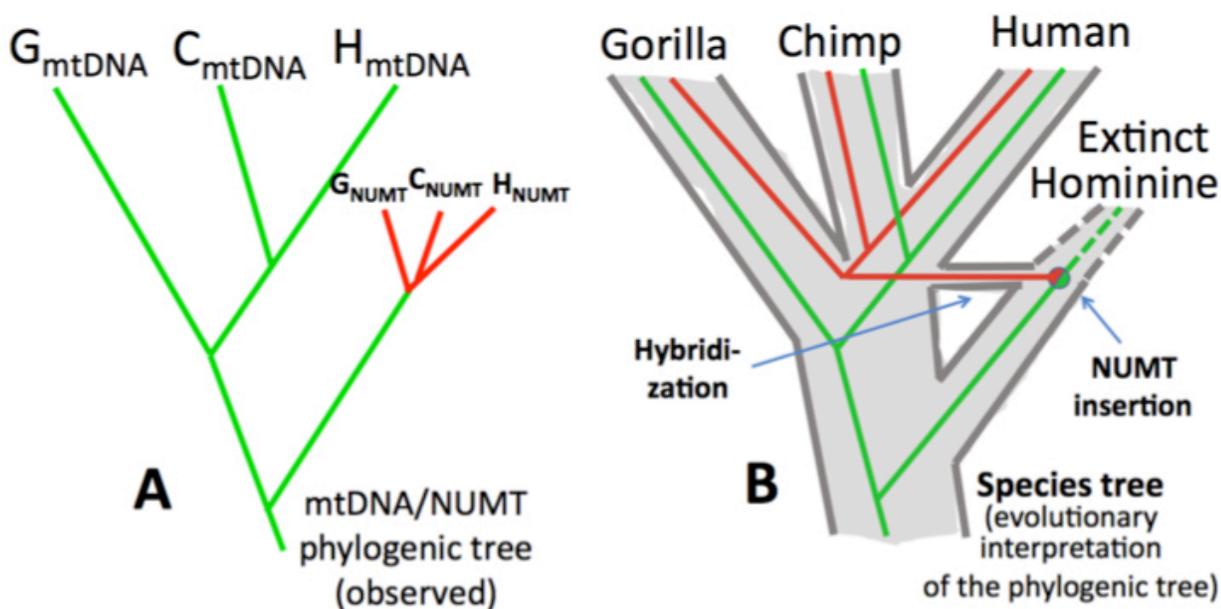


Fig. 1. (A). Phylogeny of a family of 3 closely related pseudogenes of mtDNA (NUMTs), found in humans, chimps and gorillas embedded into phylogeny of the great ape mtDNA. The NUMT phylogeny is rather unusual: the three sequences share a disproportionately long common stem that separates them from mtDNAs of the three species.

(B) The following hybridization scenario could explain the observed phylogeny of NUMT: NUMT is formed in one of the individuals of a different Hominine species, which later goes extinct. Branch color: Green: mitochondrial branches, red – NUMT branches. Note that the branch leading to the three diverging NUMT5 branches is recognized as mitochondrial (green). This is because mutations that happened along this branch are mostly neutral, and thus should have accumulated in a living mitochondrion, which would not tolerate detrimental mutations. In contrast, branches

leading to the individual NUMTs in the three species is red because mutations along these branches are mostly non-synonymous, as expected for a dysfunctional sequence like NUMT.

Evidence of ancient interspecies recombination between our direct ancestors. Because NUMT sequences are close “relatives” of the mtDNA, they can be put on the genealogical/ phylogenetic mtDNA trees, such as trees connecting humans with great apes, who possess NUMTs of their own. The tree in Figure 1A shows the phylogeny of a NUMT we call NUMT5 as it is located on chromosome 5 in human, chimp, and gorilla nuclear genomes. NUMT5 was inserted into the nuclear genome about 10 Million years ago. This tree is remarkable in that NUMT’s common ancestor resides on a long *mitochondrial* branch (green color). The length of this mitochondrial branch common to the three NUMT5s corresponds to about 4 million years of evolution. This implies that individuals whose mtDNA later became NUMT, belonged to a separate Hominine species, that had been evolving for about 4 My independently from our direct ancestors. Because mtDNA is subject to rapid genetic drift, an independent mtDNA lineage cannot survive for any time period close to 4 My unless it is isolated in a separate population. After 4 My of isolation, such a population ought to become a separate species. The fact that a NUMT that has derived from this mtDNA now resides in Human, Chimp and Gorilla genomes implies that this hypothetical Hominine interbred (hybridized) with the ancestral Human/Chimp/Gorilla population before its complete split. This hominine later presumably went extinct (Fig. 1B).

Upon hybridization with the human /chimp/gorilla ancestral population, NUMT5 spreads and gets fixed in the population of our direct ancestors, perhaps due to positive selection. Of note, although NUMT sequence itself is dysfunctional, its insertion into nuclear genome can be highly selective, as it may disrupt a gene, or modify gene expression e.g. by altering chromatin structure. In fact, there is evidence that a gene adjacent to a NUMT with genealogy similar to NUMT5 indeed was under intense selection at about the time of NUMT integration. Note that the divergence between species in this hybridization is expected to be much greater than that in well documented hybridizations between humans and Neanderthals or Denisovans. It may appear bizarre that such a distant interspecies interbreeding could be a part of human evolutionary history. In corroboration of this scenario, however, hybridization and NUMT transfers over ~4My have been reported among other groups, including nonhuman primates, e.g. baboons and colobines.

The hypothesis that human evolution involved hybridization between highly diverged species has been intensely debated. The importance of hybridization is that it may provide extra phenotypic variation and thus potentially initiate “leaps” in evolution. It also provides a potential explanation to the apparent mosaicism of fossil morphology at the emergence of the Hominin (Human, Chimp, Gorilla) lineage, i.e. situation when a fossil specimen appears to be more human-like or more ape-like in different aspects.

NUMTs as a measure of mutation rate in evolution. Insertion of a NUMT into the nuclear genome

can be viewed as another type of mutation, i.e. a special kind of genomic insertion. Because NUMTs can be represented as branches of phylogenetic trees, the time of insertion, which corresponds to the point on the branch where mitochondrial pattern of mutations switches to the pseudogenic pattern of mutations (see Fig 1 caption for details; the methodology of this analysis is described in (3)). Thus time of insertion of a NUMT into the nuclear genome can be determined with some certainty (shown schematically in Figure 2, insert). Note that this is a unique advantage of NUMTs, where the actual mutational events are linked to the reliable evolution clock - mtDNA. For comparison, the timing of a point mutation can be determined only very approximately. For an illustration, all we can say about a point mutation is that happened, for example, after separation between Human and Chimp, that is, within an interval of about 6 My (till the time of the earliest branch off the human lineage, such as Homo Heidelbergensis mtDNA).

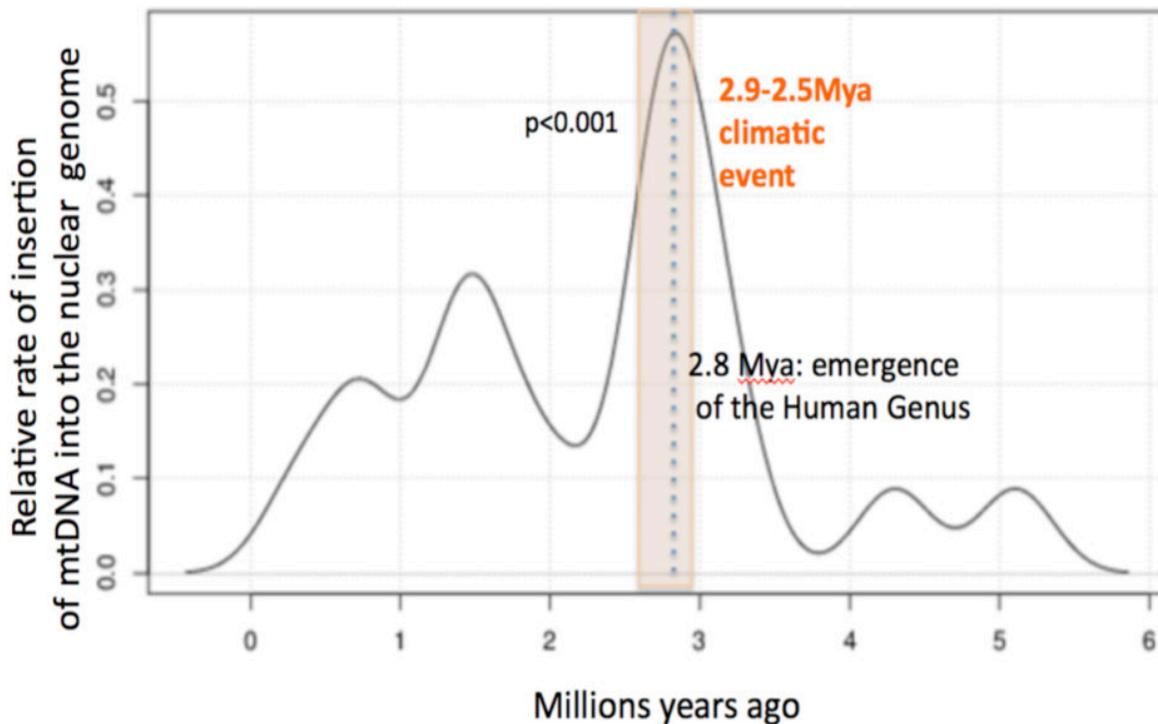


Fig. 2. A smoothed distribution of the relative rates of the NUMT integration into the nuclear genome within the last 6 My (22 NUMTs included). Insert: determining the time of NUMT integration is based on the position where mitochondrial pattern of mutations in the NUMT branch (green) changes to the pseudogenic pattern (red). Statistical significance of the peak at 2.8 Mya has been determined by a numerical simulation.

An interesting question then is whether NUMTs (and, by analogy, other mutations) are being inserted uniformly over time, or preferentially during certain “critical” periods of rapid evolution, in

accordance with the well known “punctuated equilibrium” evolution hypothesis. To explore this, we determined the times of insertion a set of human NUMTs that have been integrated into the nuclear genome over the last 6 million years. To our surprise, the NUMT integration times appear to be distributed non-uniformly between present time and 6 My ago. In particular, there is a peak of insertion frequency detectable around 2.8 My ago. Interestingly, 2.8 My ago closely corresponds to the time of emergence of the genus *Homo*, and to a well-documented period of major climate change ca. 2.9–2.5 Ma. It is thus tempting to hypothesize that integration of NUMTs is somehow related to the speciation process. NUMTs could be either “riders”, i.e., their insertion could be facilitated, for example, by the overall higher genome rearrangement activity during speciation, of a similar phenomenon. Alternatively, they could be “drivers”, i.e. they may more readily get fixed in the population due to positive selection associated with speciation and/ or with the population response to environmental changes/major climate changes.

Zoe Fleischmann, Samantha Brecht, Konstantin Gunbin, Konstantin Khrapko
Northeastern University, Boston

Publications

[Integration of mtDNA pseudogenes into the nuclear genome coincides with speciation of the human genus. A hypothesis.](#)

Gunbin K, Peshkin L, Popadin K, Annis S, Ackermann RR, Khrapko K
Mitochondrion. 2017 May

[Data on the time of integration of the human mitochondrial pseudogenes \(NUMTs\) into the nuclear genome.](#)

Gunbin K, Peshkin L, Popadin K, Annis S, Ackermann RR, Khrapko K
Data Brief. 2017 May 17