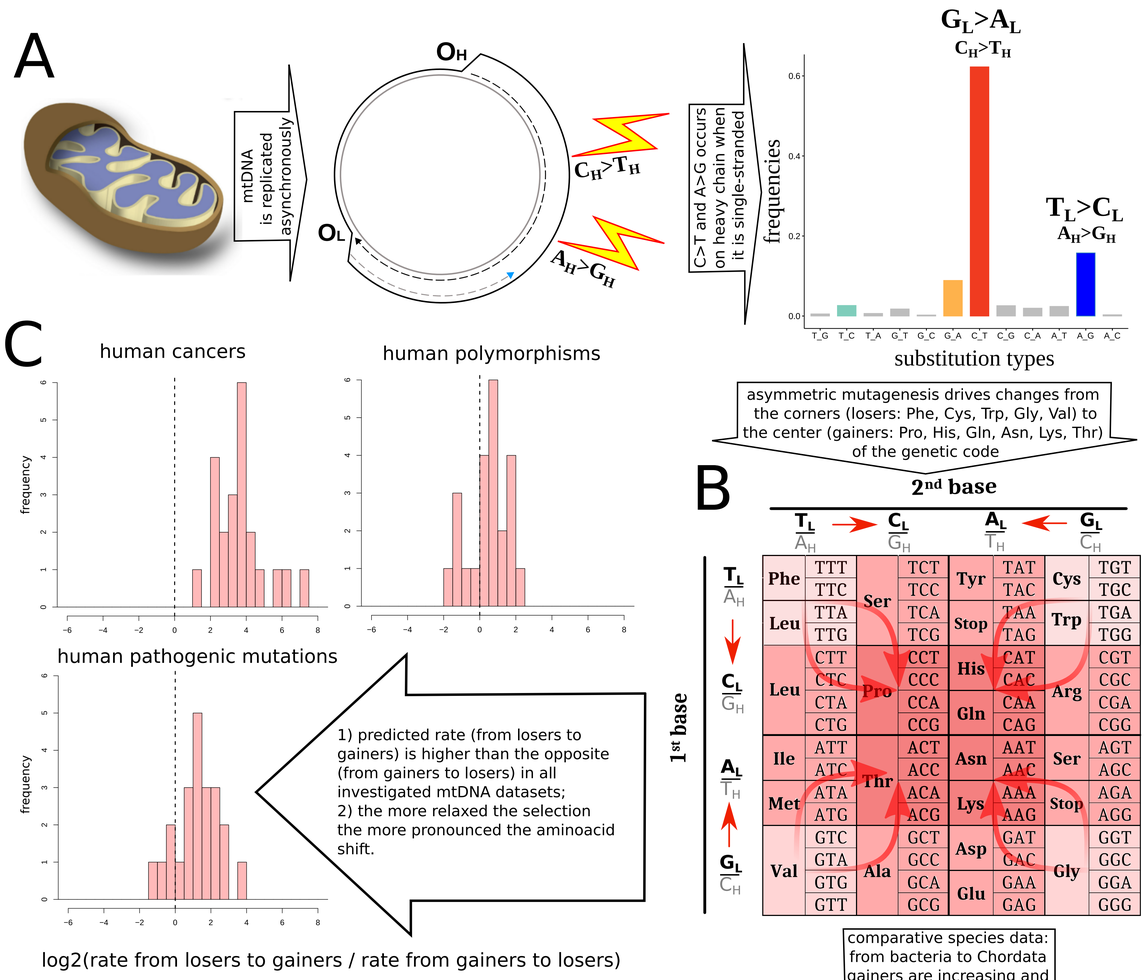
# A billion-year trend of amino acid substitutions in the mitochondrial genome

Alina A Mikhailova[1,2], Alina G Mikhailova[2], Konstantin Gunbin[2], Masashi Tanaka[3], Jacques Fellay[4], Konstantin Popadin[2,4]

1University of Münster, Institute for Evolution and Biodiversity, Münster, Germany; 2 Immanuel Kant Baltic Federal University, Center For Mitochondrial Functional Genomics, Kaliningrad, Russia; 3 Juntendo University Graduate School of Medicine, Department of Neurology, Tokyo, Japan; 4 Ecole Polytechnique Federale de Lausanne (EPFL), school of life sciences, Lausanne, Switzerland

konstantin.popadin@epfl.ch; @KPopadin




| 1st base | 2nd base: TL/AH | CL/GH | AL/TH | GL/CH |
|---|---|---|---|---|
| TL/AH | Phe TTT | Ser TCT | Tyr TAT | Cys TGT |
| | Phe TTC | Ser TCC | Tyr TAC | Cys TGC |
| | Leu TTA | Ser TCA | Stop TAA | Trp TGA |
| | Leu TTG | Ser TCG | Stop TAG | Trp TGG |
| CL/GH | Leu CTT | Pro CCT | His CAT | Arg CGT |
| | Leu CTC | Pro CCC | His CAC | Arg CGC |
| | Leu CTA | Pro CCA | Gln CAA | Arg CGA |
| | Leu CTG | Pro CCG | Gln CAG | Arg CGG |
| AL/TH | Ile ATT | Thr ACT | Asn AAT | Ser AGT |
| | Ile ATC | Thr ACC | Asn AAC | Ser AGC |
| | Met ATA | Thr ACA | Lys AAA | Stop AGA |
| | Met ATG | Thr ACG | Lys AAG | Stop AGG |
| GL/CH | Val GTC | Ala GCT | Asp GAT | Gly GGT |
| | Val GTA | Ala GCC | Asp GAC | Gly GGC |
| | Val GTG | Ala GCA | Glu GAA | Gly GGA |
| | Val GTT | Ala GCG | Glu GAG | Gly GGG |

It has been shown that the rates of reciprocal amino acid substitutions in prokaryotic and eukaryotic organisms are not balanced, leading to the long-term increase (i.e. 'gainers') or decrease (i.e. 'losers') in the frequency of some amino acids (Jordan et al. 2005). However, the evolutionary driving forces establishing this trend (mutagenesis or selection) are still unknown.

Here we focus on the mitochondrial genome, where we can predict the preferential direction of amino acid substitutions, based on the strongly asymmetrical mutational spectrum (an excess of G to A and T to C, light chain notation) (see figures A). According to the structure of the mtDNA genetic code and mtDNA mutational bias, we expect that mtDNA amino acid composition is shifting from the corners of the genetic code towards the center (see figure B), leading to 6 predictions: (i) 27 single-transition nonsynonymous substitutions are non-reciprocal (for example the rate of substitutions from Val to Ala is expected to be higher as compared to the rate from Ala to Val); (ii) Pro, His, Gln, Asn, Lys, and Thr are gainers; (iii) LeuTT, Phe, Cys, Trp, Gly, and Val are losers, (iv) direction of all shifts is opposite for ND6 gene - the only protein-coding gene, coded on the light chain of mtDNA, (v) global amino acid shift is more pronounced in cases when mutagenesis is stronger than selection, (vi) if this trend is long-term and universal it will explain main changes in amino acid composition from the ancestral mtDNA (the common ancestor of mtDNA and alpha-proteo-bacteria) to modern mtDNAs.

First, analysing a collection of 3149 nonsynonymous mtDNA mutations from the human cancers (Pan-Cancer Analysis of Whole Genomes Consortium with whole-genome sequencing data from 2,658 cancers of 38 types), we observed that 89% of them were in the predicted direction (2812 substitutions are from losers to gainers) and only 11% were in the opposite to predicted direction (337 substitutions are from gainers to losers). The observed shift remains significant for the majority of individual amino acid pairs (see figures C).
Second, analysing 295 nonsynonymous pathogenic mtDNA human mutations (MitoMap database), we observed 71% of substitutions in the predicted direction, the trend was significant for the majority of the amino-acid pairs (see graphical abstract).
Third, analysing 87843 nonsynonymous mtDNA substitutions, reconstructed from the human mitochondrial tree, we observed 51% of substitutions in the predicted direction, which, despite the moderate excess, is still significant for the majority of the amino-acid pairs due to high sampling size (see figures C).
It is important to note that the decrease in the percentage of the predicted amino-acid substitutions from cancers (89%), to human pathogenic variants (71%) and human population polymorphisms (51%) reflects the expectedly increasing role of selection over mutagenesis (see the shift of the corresponding distributions from the expected zero in pink histograms, see figures C)).
Fourth, analysing 37'221 nonsynonymous mtDNA polymorphic substitutions observed in 2020 vertebrate species, we confirmed an excess of the predicted substitutions in the whole dataset. Additionally, within mammals, we demonstrated that this excess is stronger in low- versus high- effective population (Ne) sized species (long-lived mammals versus short-lived), which is in line with the accumulation of the slightly-deleterious variants (i.e. excess of the forward substitutions) due to genetic drift.
Fifth, in all analyses described before, as expected, we observed opposite amino-acid shifts in the ND6 gene.
Sixth, comparing the amino acid composition of 10 orthologous genes (all mtDNA coded genes except ND6, ATP8, and ND4L) between alpha-proteo-bacteria, fungi, plants, invertebrates, and five classes of vertebrates (ray-finned fishes, amphibia, reptilia, mammals, birds), we observed a global billion-year trend (see figure D): losers become rarer while gainers become more frequent (graphical abstract) starting from the moment of origin of mtDNA. We explain this trend by the mtDNA mutational bias, which leads to a strong amino acid shift in species with the high mtDNA mutational rate (animals versus plants) and low Ne: Ne(vertebrates) < Ne(invertebrates) < Ne(plants) < Ne(fungi).

Altogether, on different datasets and evolutionary time scales, we unambiguously confirmed the strong effect of the mtDNA mutational bias on the pattern of amino-acid substitutions. It shows a slow and continuous accumulation of slightly-deleterious variants in mtDNA from the moment of endosymbiosis emergence till the current days. This finding allows to (i) predict the direction of amino-acid substitutions in human cancers; (ii) reevaluate pathogenicity of deleterious human mtDNA variants in the light that more than 15% of amino acids in human mtDNA are shaped by mutagenesis, not selection; (iii) provide a new and simple metric (fraction of gainers and losers) approximating the effective population size of mtDNA of different eukaryotic species.

Jordan, I. King, Fyodor A. Kondrashov, Ivan A. Adzhubei, Yuri I. Wolf, Eugene V. Koonin, Alexey S. Kondrashov, and Shamil Sunyaev. 2005. "A Universal Trend of Amino Acid Gain and Loss in Protein Evolution." Nature 433 (7026): 633–38.

Mikhailova et al. Mammalian mitochondrial mutational spectrum as a hallmark of cellular and organismal aging https://www.biorxiv.org/content/10.1101/589168v3

Mikhailova et al. A mitochondrial mutational signature of temperature and longevity in ectothermic and endothermic vertebrates. https://www.biorxiv.org/content/10.1101/2020.07.25.221184v1