DOI 10.18699/BioGenEvo-2018-43

The reasons for mtDNA structural instability: evolutionary physico-chemical retrospective

V.N. Timonina[1]*, D.A. Knorre[2], K.Yu. Popadin[1, 3], K.V. Gunbin[1, 4]
[1] *School of Life Science, Immanuel Kant Federal Baltic University, Kaliningrad, Russia*
[2] *A.N. Belozersky Institute of Physico-Chemical Biology, MSU, Moscow, Russia*
[3] *Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*
[4] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** e-mail: valeratimonina@gmail.com*

Key words: mtDNA, non-B-DNA conformations, dinucleotide properties, molecular evolution

Motivation and Aim: It is well known that unevenness of physico-chemical DNA properties along the human mtDNA is associated with the unevenness in mtDNA deletion breakpoints and SNPs and, ultimately with longevity.
Methods and Algorithms: More than 1000 completely sequenced mtDNAs of Mammalia and more than 24000 human mtDNAs were analysed. To uncover the basis of mtDNA structural instability (SI) we analyzed both intraspecial (human) and interspecial (Mammalians) mtDNA variations with (1) the non-B-DNA conformations (cruciform, triplex, hairpins, G-quadruplex, Z-DNA, etc.) and (2) the dinucleotide properties of mtDNA enriched with variations. Our analysis was based on (1) various available software tools for non-B-DNA identification (various EMBOSS package programs; R packages triplex and pqsfinder; SIST and triplexator software) and (2) DiProDB dinucleotide properties, respectively. Random forest approach implemented in R randomForest and Boruta libraries and various sliding window lengths were used to dissect mtDNA dinucleotide properties unevenness; phylogenetic comparative methods implemented in R ape and caper libraries were used for linking Mammalian species longevity and variations in mtDNA properties or between ancestry of human mtDNAs (based on reconstructed maximum likelihood phylogeny) and their properties.
Results and conclusions: During the analysis, we identified various important regularities in intraspecial and interspecial evolution of mtDNA properties.
For example, interspecial analysis allowed us to find positive relations between quadruplex frequencies, CpG island frequencies and species longevity; the relations of DNA bend, free energy and enthalpy with species longevity. All these relations associated with mtDNA replication optimization. Various intriguing relations between di-/terta-nucleotide under- or overrepresentations were found, for example, we found a tendency for CG and TT dinucleotides loss in long-lived mammals that can be related with SI minimization due to high mutational pressure on such nucleotide patterns.
The results of intraspecial analysis demonstrates nonrandom physical causes of SI. For example, it was found that DNA properties associated with 5' and 3' breakpoint hotspots are quite different: the properties of 5' breakpoints as well as SNP hotspots relates with the rigidity of DNA and with protein-DNA interactions while 3' breakpoints associates with the number of imperfect repeats. These facts allowed us to conclude that mtDNA deletion breakpoints and SNP fixations is initiated by protein-DNA interactions while the termination of breakpoints is defined by complementary interaction between the single stranded DNA by imperfect repeats.
Acknowledgements: This study was supported by the 5 Top 100 Project at the I. Kant Baltic Federal University.