

The software and database for Vertebrate imperfect mtDNA repeats annotation

V.A. Shamanskiy^{1*}, K.Yu. Popadin^{1,2}, K.V. Gunbin^{1,3}

¹ *School of Life Science, Immanuel Kant Federal Baltic University, Kaliningrad, Russia*

² *Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*

³ *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

* *e-mail: v.a.shamanskiy@gmail.com*

Key words: imperfect (degenerate) repeats, mtDNA, Vertebrata

Motivation and Aim: It is well known that the vast majority of non-B-DNA conformations is tightly associated with pattern compositions especially with various imperfect repeats. It was also known that the number and nature of mtDNA perfect repeats are under strong selection pressure. Thus, the number and nature of various mtDNA repeats is useful for estimation the organismal fitness-related parameters. Despite this fact, there are no any database annotated imperfect mtDNA repeats.

Methods and Algorithms: In order to build the database we selected more than 3800 completely sequenced mtDNAs of Vertebrates and more than 25000 human mtDNAs. We find imperfect repeats in these sequences using our modification of dot-plot analysis: we aligned every 10-bases regions (sliding windows) in mtDNA versus this genome by dynamic programming using various indel-scoring schemes. After that, we linked 10-base imperfect repeats in longer repeats by our hash-algorithm implemented in Python in all cases when neighbor windows marked as a repeat-containing.

In order to estimate the evolutionary relationship between imperfect mtDNA repeats of Vertebrates we mapped these repeats on the 3 alternative multiple mtDNA alignments of Vertebrates generated by MAFFT, PASTA and MISHIMA taking into account predefined phylogenetic tree topology. We mapped human imperfect mtDNA repeats on the alignment of human mtDNA that was reconstructed using BLASR.

Results: In order to discriminate different selection modes associated with various Vertebrate taxa evolution we compared taxa-specific numbers of main repeat types with ones from whole sample using nonparametric U-test. We found that Mammals have a significantly large number of imperfect repeats than the number of such repeats in the whole Vertebrata sample set. Another interesting finding is the excess of direct and symmetrical imperfect repeats in hominids that, at the same time, characterized by the lack of complimentary and inverted degenerate repeats. According to our data Ascidiids are the record-holder in the number of imperfect repeats, while fishes characterized by relatively small number of degenerate repeats. In addition to the taxa specific description of imperfect repeats number we calculated standard statistics on the length of imperfect repeats, degeneration rate, GC richness, dinucleotide composition, etc. There are also data on tandem degenerate repeats and areas of self-replicating.

Conclusion: The software and database for Vertebrate imperfect mtDNA repeats annotation were made. The information in the database allowed us to discriminate taxa-specific selection forces shaping mtDNA repeat composition.

Availability: available upon request.

Acknowledgements: This study was supported by the 5 Top 100 Project at the I. Kant Baltic Federal University.