

## THE DISTRIBUTION OF FUNGAL MITOCHONDRIAL *ATP* GENES IN AMINO ACIDS SPACE

Victoria Fedotovskaya\*<sup>1</sup>, Anna Kolesnikova<sup>1,2</sup>,  
Tatiana Shpagina<sup>1</sup> and Michael Sadovsky<sup>1,3</sup>

<sup>1</sup>Siberian Federal University,  
Institute of Fundamental Biology and Biotechnology,  
Svobodny prosp., 79, 660049 Krasnoyarsk, Russia

<sup>2</sup>Laboratory of Genomics and Biotechnology,  
Federal Research Center RAS,  
Akademgorodok, 660036 Krasnoyarsk, Russia

<sup>3</sup>Institute of Computational Modelling of SB RAS,  
Akademgorodok, 660036 Krasnoyarsk, Russia

viktorii.fedotovskaia@gmail.com (\*corresponding author),  
msad@icm.krasn.ru

We studied the relations between triplet composition of the family of mitochondrial *ATP* synthase genes (*atp6*, *atp8* and *atp9*) and their function. In total, 223 fungal mitochondrial genomes were used in this study. All gene sequences were extracted from the genomes in two versions: containing both exons and introns, and sequences without introns corresponding to their mRNA. Each gene sequence was transformed into triplet frequency dictionary, where the reading frame shift was equal either to  $t = 1$  or  $t = 3$ .

In fact, the frequency dictionary of mRNA sequences with  $t = 3$  matches with amino acid codons of an open reading frame, consequently it represents an amino acids frequency dictionary. Thereby, the distribution of the amino acid frequency representing points in 21-dimensional space (20 amino acids plus stop signal) was studied.

It was found that the points are separated into three clusters corresponding to those genes. The distribution in amino acids space was compared to the distribution in triplet space and appeared more accurate. Moreover, we separated the set of dictionaries by  $K$ -means into the classes for  $K = 4$ . Such kind of classification was rather unstable. Also, we checked the structuredness in 64-dimensional metric space on the set of points, which corresponds to each gene separately. It was found that species are spread among the clusters rather equally.

A comparative analysis of the clusters in the 21 and 64-dimensional metric space was carried out. As a result, we conclude that in case of analysis of all three genes (*atp6*, *atp8* and *atp9*), the clusters distinguished by local density are clearly associated with the type of a gene, and in case of a single gene, a larger number of clusters is distinguished.