



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

V. Fedotovskaya<sup>1</sup>, M. Sadovsky<sup>1,2</sup>, A. Kolesnikova<sup>1,3</sup>, T. Shpagina<sup>1</sup>; SibFU<sup>1</sup>, ICM SB RAS<sup>2</sup>,  
Laboratory of Genomics & Biotechnology, Federal Research Center SB RAS<sup>3</sup>

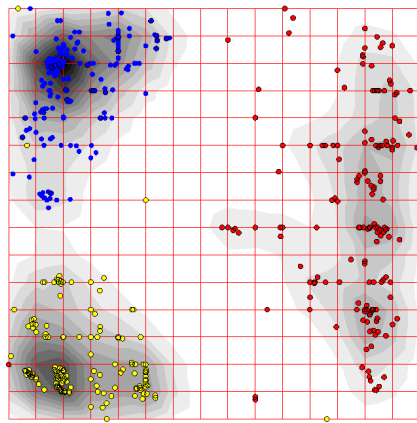


Figure 1: Distribution of ATP synthase genes in amino acid frequency space.

**Goal:** We studied the relations between triplet composition of the family of mitochondrial ATP synthase genes (*atp6*, *atp8* and *atp9*) and their function. In fact, the frequency dictionary of mRNA sequences with  $t = 3$  matches 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. Besides, the distribution of individual genes families (*atp6*, *atp8* or *atp9*, separately) was studied in the nucleotides triplets frequency space.

**Results:** 223 fungal mitochondrial genomes were studied. 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  $t = 3$ . Fig. 1 shows the distribution of three genes families in amino acids frequency space: genes of *atp6* is shown in blue, *atp8* is shown in red and *atp9* is shown in yellow. Fig. 2 shows the distributions of (left to right) *atp6*, *atp8* and *atp9* genes

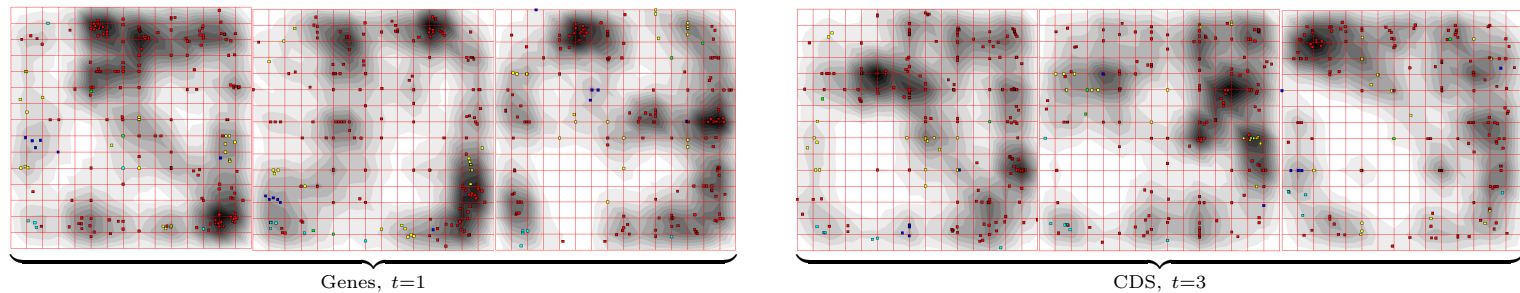


Figure 2: Gallery of individual distributions of three *atp* families in nucleotide triplet frequency space.

in the nucleotide triplets frequency space; the step and sequence type are shown in the figure legend. Gray scaled background represents the local density of the points on the elastic map, in all figures.

It should be stressed that color in Fig. 2 marks up the subphylums: *Agaricomycotina* is shown in blue, *Saccharomycotina* is shown in red, and *Pezizomycotina* is shown in yellow.

**Conclusion:** Previously, it was found that the joint distribution of the genes from all three families in the space of trinucleotides frequencies unambiguously follows the function: the genes gather into clusters, and the clusters almost perfectly correspond to a gene family (in other words, *atp6*, *atp8* and *atp9*) comprise separate clusters.

Analysis of the distribution of the genes from the same family (say, *atp9*) shows very weak correlation to taxonomy: the genes are split into a set of clusters, while the clusters usually comprise the genes of various species belonging to diverse subphylums.

Clustering of all three families of genes in amino acids frequency space yields perfect separation. Nonetheless, the *atp9* family yields a split into two very apparent subclusters. Further study of the distribution may reveal the key amino acids (and relevant synonymic codons) determining the cluster pattern.