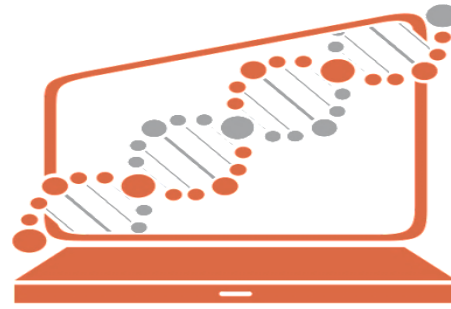




СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
SIBERIAN FEDERAL UNIVERSITY



ЦЕНТР
ГЕНОМНЫХ
ИССЛЕДОВАНИЙ
научно-образовательный центр СФУ



Triplet composition of common mitochondrial and chloroplast genes of plants reveals their differentiation

Fedotovskaya V.*¹, Sadovsky M.^{2, 3, 4}

¹Siberian Federal university, Krasnoyarsk, Russia

²Institute of computational modelling of SD of RAS, Krasnoyarsk, Russia

³V.F. Voino-Yasenetsky Krasnoyarsk state medical university, Krasnoyarsk, Russia

⁴Siberian research & clinic center of FABM of Russia, Krasnoyarsk, Russia



Introduction

The problem of interplay of structure of nucleotide sequences, functions encoded in them, and taxonomy of their bearers still challenges researchers. We studied the relation between triplet composition of genes and taxonomy of the bearers. We compared mitochondrial and chloroplast genes common for both organelles of the same species (Table 1).

Totally, 170 (85 mitochondrial and 85 chloroplast) plant genomes were studied.

Mitochondrial	Chloroplast
<i>atp1</i>	<i>atpA</i>
<i>atp4</i>	<i>atpB</i>
<i>atp6</i>	<i>atpE</i>
<i>atp8</i>	<i>atpF</i>
<i>atp9</i>	<i>atpH</i>
	<i>atpI</i>

Table 1. Extracted genes



Methods

Each gene sequence was transformed into the frequency dictionaries which differ in the reading frame shift $t = 1$. The triplet frequency dictionary represents set of all triplets $w_1 = AAA$ to $w_2 = TTT$ together with their frequency (1).

Three database were studied:

- only mitochondrial genes,
- only chloroplast genes,
- mitochondrial and chloroplast genes simultaneously.

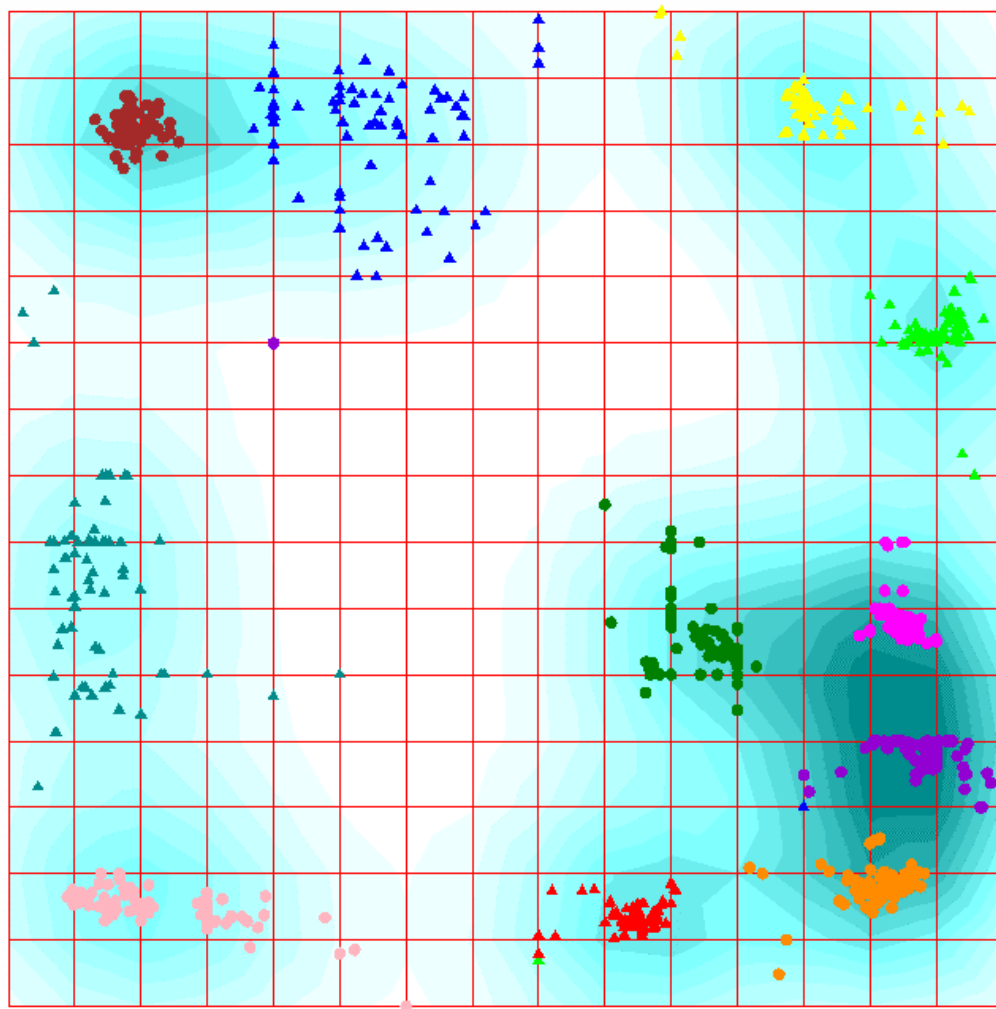
The distribution of the points was clustered with elastic map technique due to ViDaExpert program.

ACGTGCTATACGTACG

Fig. 1. Reading frame shift $t = 1$

$$f_{\omega} = \frac{n_{\omega}}{M} \quad (1)$$

f_{ω} – triplet frequency,
 n_{ω} – number of specific triplet,
 M – total number of triplets counted within a sequence.



Mitochondria:

- ▲ – *atp1*
- ▲ – *atp4*
- ▲ – *atp6*
- ▲ – *atp8*
- ▲ – *atp9*

Chloroplast:

- – *atpA*
- – *atpB*
- – *atpI*
- – *atpE*
- – *atpF*
- – *atpH*

Fig 2. Distribution of the merged set of genes.



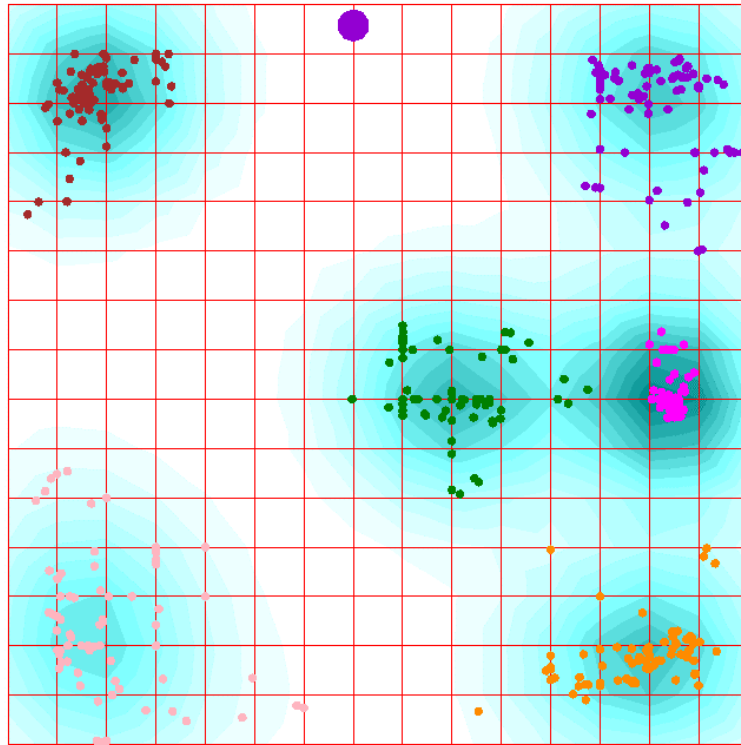


Fig 3. Distribution of the chloroplast genes.

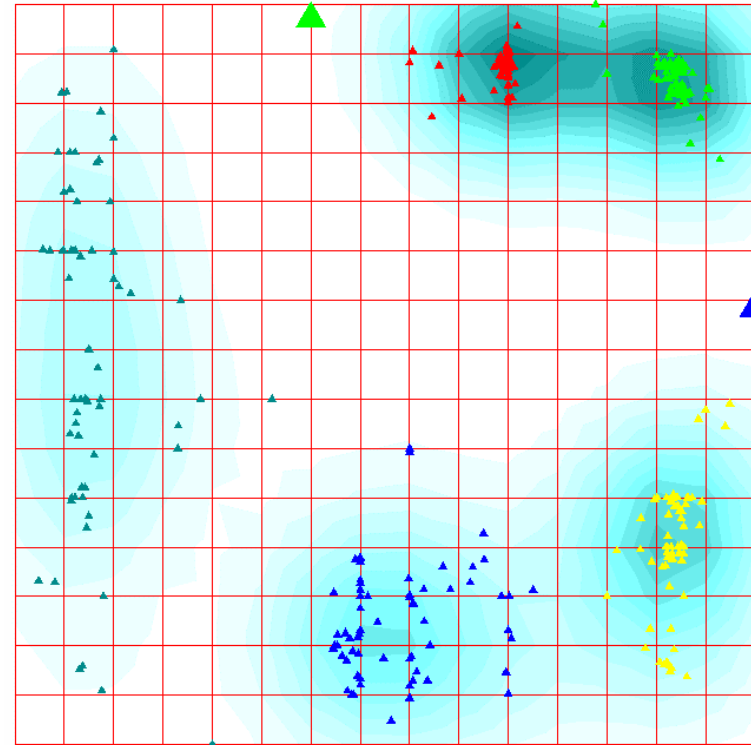


Fig 4. Distribution of the mitochondrial genes.

Mitochondria:

- ▲ – *atp1*
- ▲ – *atp4*
- ▲ – *atp6*
- ▲ – *atp8*
- ▲ – *atp9*

Chloroplast:

- – *atpA*
- – *atpB*
- – *atpI*
- – *atpE*
- – *atpF*
- – *atpH*

Organelle	Gene	Organism
Mitochondria	<i>atp4</i>	<i>Welwitschia mirabilis</i>
Mitochondria	<i>atp6</i>	<i>Ammopiptanthus mongolicus</i>
Chloroplast	<i>atpF</i>	<i>Ammopiptanthus nanus</i>

Table 2. Out-of-distribution genes



Conclusion

The encoded function was found to be the key feature in clustering: all the clusters in all three versions of clustering patterns clearly exhibit distinct separation of the genes encoding the same subunit into a separate cluster. This behaviour was found for all three types of cluster patterns. Thus, an evidence of the prevalence of function over the taxonomy was shown, for ATP synthase genes family of mitochondrial and chloroplast genomes of plants.

