

Spatial Structure of Chloroplast Genes of Photosynthetic Systems I and II

Maria Yu. Senashova

Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, 50/44 Akademgorodok, Krasnoyarsk, 660036, Russia

Abstract

Spatial structure of chloroplast genes of photosynthetic systems I and II is considered. The spatial structure is understood as the distribution of points corresponding to the frequency dictionaries of genes in the space of triplets frequencies in this work. The photosystems I and II genes are clustered according to their belonging to the forward and reverse strands. Points corresponding to genes in the forward and reverse strands are located at a distance from the main clusters. Any structure wasn't found for the distribution of the genes' GC-content values in the frequency space.

Keywords

Order, distribution, clustering, evolution, triplets

1. Introduction

The sun is the main source of energy on the Earth, and a number of organisms have adapted to use this energy for their needs. Plants, algae and cyanobacteria grow due to their ability to use sunlight to extract electrons from water. This is how photosynthesis and, accordingly, phototrophic nutrition arose. Photosynthetic organisms include phototrophic bacteria and green plants. The study of the structure and function of the photosynthetic system plays an important role, since plants are the main suppliers of oxygen and food. In addition, oil reserves do not last forever, and the question arises about alternative methods of obtaining hydrocarbons. Artificial photosynthesis, the production of organic fuel from carbon dioxide using solar energy, stands apart from the list of these methods [3, 13, 14]. Therefore, a comprehensive study of the process of photosynthesis and photosynthetic systems is very important.

The subcellular location of genes was determined for individual components of the photosynthetic apparatus. The results obtained were presented in the form of contour maps of four energy-transducing thylakoid membranes. Differences were found in the maps for terrestrial plants and red and green algae in the protein subunits encoded in the nucleus from those encoded in the chloroplast [1].

Photosystems I and II (PS I and PS II) were found to be separated in chloroplast membranes (photosystems II in compressed granular membranes and photosystems I in uncompressed stromal membranes) [2, 12].

Many studies are devoted to investigating various aspects of the structure of photosynthetic systems I and II both by mathematical and biological methods. Statistical methods of analysis were used to study the distribution of the complexes of photosystem II in stacked grana thylakoids and formation of the structure of photosystem II in low-light conditions [6, 7].

When studying the structural and functional organization of chloroplasts in higher plants under different lighting conditions, it was found that the structural formations of the grana contributed to the

SibDATA 2021: The 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications 2021, June 25, 2021, Krasnoyarsk, Russia

EMAIL: msen@icm.krasn.ru (M. Senashova)

ORCID: 0000-0002-1023-7103 (M. Senashova)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

centralization and relative increase in the concentration of a specific group of PS II reaction centers [9].

The analysis of the structure of the complexes of photosystems I, photosystems II, cytochrome b6f, and F-ATPase provided the basis for studying the transfer of energy and electrons and the evolutionary forces which formed the photosynthetic apparatus [10].

This study considers the spatial structure of chloroplast genes belonging to photosynthetic systems I and II. Here, the structure is meant to be the distribution of points corresponding to the frequency dictionaries of genes of photosynthetic systems in the 64-dimensional space of triplets. This approach differs from that considering the structure of these genes from the biological point of view.

2. Material and methods

Let us introduce basic concepts. We consider the genes of the photosynthetic systems as symbol sequences of various lengths, consisting of the alphabet symbols $\mathcal{M} = \{A, C, G, T\}$. If the sequence contains symbols differing from the alphabet symbols, then such characters are removed from the sequence, and the number of such symbols reduces the length of the sequence. Each of these sequences is associated with the frequency dictionary of thickness 3. The frequency dictionary W_3 of thickness 3 of the symbol sequence corresponding to DNA is a list of all the triplets $v_1v_2v_3$ of consecutive nucleotides with an indication of the frequencies of these triplets. There can be 64 triplets in total. The frequency f_ω is the ratio of the number of the given word copies n_ω to the total number of all the triplets N , where N is the sum of all n_ω :

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

A frequency dictionary maps a symbolic sequence into the 64-dimensional metric space. The proximity of two genomes is set naturally as the proximity of two points in the Euclidean metric:

$$\rho(W_3^1, W_3^2) = \sqrt{\sum_{\omega=AAA}^{TTT} (f_\omega^1 - f_\omega^2)^2} \quad (2)$$

The preliminary processing is carried out to identify the structure in the set of genes. The given set of the symbolic sequences corresponds to a set of points in the 64-dimensional space of triplets. Each gene has its own point. Each point has a set of parameters: name of the gene, name of the species to which the gene belongs, and whether this strand is forward or reverse. Freely distributed software VidaExpert is used to visualize the distribution of the genes converted into the triplet frequency dictionaries in the 64-dimensional space metric space. The distribution of the points corresponding to the genes in the space of the principal components is studied. The projections on the plane of the space of the first principal components are considered.

3. Results and discussion

The genes of photosynthetic system I and II were isolated from 570 chloroplast genomes currently available in the EMBL-bank. The following genes were found in the studied set of genomes: *psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *psaM*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbG*, *psbH*, *psbI*, *psbJ*, *psbK*, *psaL*, *psbM*, *psbN*, *psbT*, *psbT*.

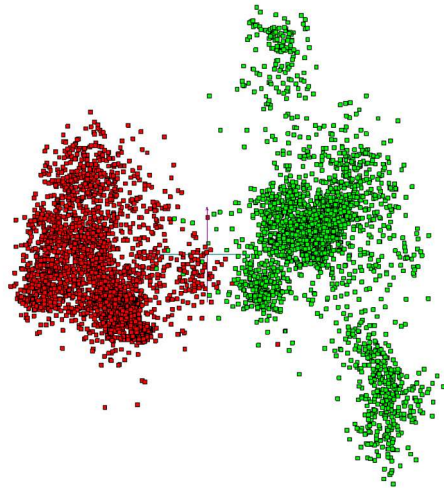


Figure 1: The plane projection of the first two principal components. The genes in the forward strand are shown in red, and the genes in the reverse strand are shown in green

A frequency dictionary $W_{(3,3)}$ was built for each gene. The frequency dictionary $W_{(3,3)}$ is a set of triplet frequencies. The triplets in the gene were taken without crossing symbols. For the genes in the reverse strand, the frequency dictionary was built taking into account that the symbolic sequence related to such genes was inverted. A projection into the space of the first three principal components was built to visualize the spatial structure formed by a set of these points. The set of the points was found to form two large clusters, with one of them including the points related to the forward strand genes, and the other containing the points related to the genes of the reverse strand (Figure 1). This significantly distinguishes the spatial structure of the genes of photosynthetic systems I and II from the spatial structure of the complete genomes of chloroplasts, mitochondria, and bacteria, where such clustering is not observed [4, 8, 11].



Figure 2: The genes *psbF*, *psbJ*, *psbL* and *psbN* a) in the plane of the first and second principal components and b) in the plane of the second and third principal components

In addition, one can see that in the cluster related to the reverse strand, there are groups of points which are quite far from the basic cluster. These groups of points correspond to the genes *psbF*, *psbJ*, *psbL* and *psbN*. In Figure 2., each group of points is highlighted with its own color. The gene *psbF* corresponds to the points of turquoise color, the gene *psbJ* corresponds to crimson color, the gene *psbL* corresponds to light green color and the gene *psbN* is denoted by purple color. The rest of the points are colored according to their belonging to the strands and have a smaller size for clarity. These genes are also present in the direct strand cluster, but there are quite a few of them. The gene *psbF* in higher plants consists of approximately 38 amino acids, the β -subunit of

cytochrome b559. The gene *psbJ* is important for the assembly of PSII and regulates the electron flow to the plastoquinone pool. The gene *psbL* is necessary for the operation of the Qa site, and it prevents the return of an electron from the Qb site to Qa. The gene *psbN* participates in the assembly of the reaction center of photosystem II.

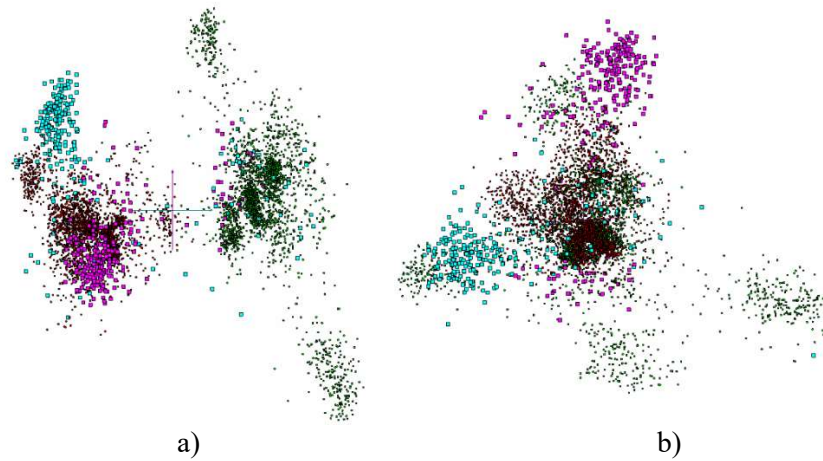


Figure 3: The genes *psaI* and *psbI* a) in the plane of the first and second principal components and b) of the second and third principal components

Two groups of points can also be distinguished, which are far from the basic cluster in the cluster related to the forward strand. These points related to the genes *psaI* (turquoise color) and *psbI* (crimson color) are shown in Figure 3. The rest of the points are colored as is done in Figure 2. These genes are also present in the reverse strand cluster, but their number is small. The gene *psaI* interacts with the gene *psaH*, and binds to the light-harvesting complex of photosystem II. The gene *psbI* is required for the assembly and functioning of photosystem II. Why exactly these genes are far enough from the basic clusters requires additional study.

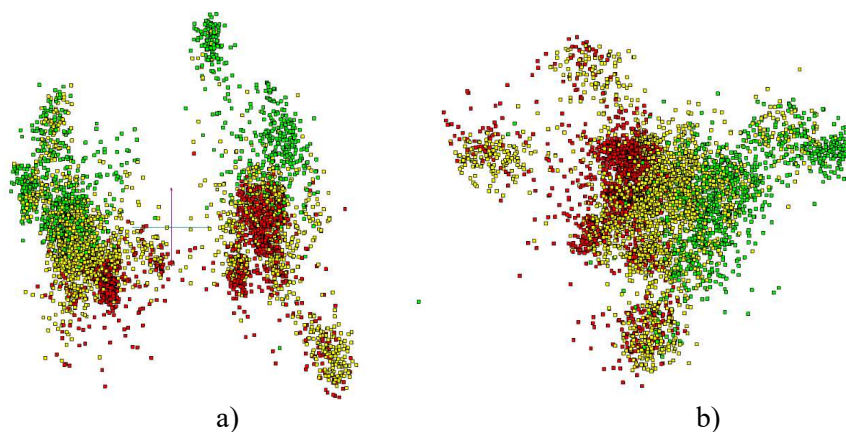


Figure 4: The spatial distribution of the *GC*-content values of the genes of the photosynthetic systems a) in the plane of the first and second and b) second and third principal components

It should be noted that the genes of the same type are very densely distributed, even if they are located in large clusters belonging to the forward and reverse strands. They "adjoin" rather than intersect with genes of other types. That is, clustering occurs precisely according to the type of genes, rather than according to the phylogenetic characteristics of organisms to which they belong.

The *GC*-content is the ratio of the number of nucleotides *C* and *G* to the total number of nucleotides in the gene. The spatial distribution of the values of the gene *GC*-content was considered.

The GC-content values were calculated for each gene. The points corresponding to the genes with the GC-content value lower than the average one are indicated in Fig. 4 in green, the points corresponding to the genes with the average GC-content value are marked in yellow, and the points with the GC-content value greater than the average one are indicated in red. No order is observed in the distribution of the GC-content value of the genes of the photosynthetic systems. This also distinguishes the genes under consideration from the complete genomes of chloroplasts, mitochondria and bacteria, which are characterized by two types of distribution: gradient and centrally symmetric.

4. Conclusion

The spatial structure of chloroplast genes of the photosynthetic systems in the space of triplet frequencies is quite different from similar structures of the previously studied complete genomes of chloroplasts, mitochondria, and bacteria. This concerns the relative position of the genes of the forward and reverse strands and the spatial distribution of the value of the gene GC-content. In addition, the points corresponding to the genes are found to be grouped in the space of the triplet frequencies according to the type of genes rather than according to the type of the corresponding organisms.

5. References

- [1] J. F. Allen et al., A structural phylogenetic map for chloroplast photosynthesis, *Trends in plant science* 16(12) (2011) 645–655. doi:10.1016/j.tplants.2011.10.004.
- [2] J. M. Anderson, A. Melis, Localization of different photosystems in separate regions of chloroplast membranes, *Proceedings of the National Academy of Sciences* 80(3) (1983) 745–749. doi: 10.1073/pnas.80.3.745.
- [3] F. A. Chowdhury et al., A photochemical diode artificial photosynthesis system for unassisted high efficiency overall pure water splitting, *Nature communications*. 9(1) (2018) 1–9. doi:10.1038/s41467-018-04067-1.
- [4] A. N. Gorban, A. Yu. Zinoviyev, T. G. Popova, Four basic symmetry types in the universal 7–cluster structure of microbial genomic sequences, *In Silico Biology* 5(3) (2005) 265–282.
- [5] B. Hankamer, J. Barber, E. J. Boekema, Structure and membrane organization of photosystem II in green plants, *Annual review of plant biology* 48(1) (1997) 641–671. doi:10.1146/annurev.arplant.48.1.641.
- [6] H. Kirchhoff et al., Supramolecular photosystem II organization in grana thylakoid membranes: evidence for a structured arrangement, *Biochemistry* 43(28) (2004) 9204–9213. DOI:10.1021/bi0494626.
- [7] H. Kirchhoff et al., Low-light-induced formation of semicrystalline photosystem II arrays in higher plant chloroplasts, *Biochemistry* 46(39) (2007) 11169–11176. doi:10.1021/bi700748y.
- [8] R. Kosarev, M. Senashova, M. Sadovsky, Intrinsic Structuredness of Mitochondria Genomes, in: L. Nozhenkova, T. Penkova, A. Korobko (Eds.), *The 1st Siberian Scientific Workshop on Data Analysis Technologies with Applications 2020*, volume 2727 of *SibDATA'20*, CEUR, 2020, Krasnoyarsk, Russia, pp. 66–74.
- [9] A. Melis, G. W. Harvey, Regulation of photosystem stoichiometry, chlorophyll a and chlorophyll b content and relation to chloroplast ultrastructure, *Biochimica et Biophysica Acta (BBA)–Bioenergetics* 637(1) (1981) 138–145. doi:10.1016/0005-2728(81)90219-X.
- [10] N. Nelson, A. Ben-Shem, The complex architecture of oxygenic photosynthesis, *Nature Reviews Molecular Cell Biology* 5(12) (2004) 971–982. doi:10.1038/nrm1525.
- [11] M. G. Sadovsky, M. Yu. Senashova, A. V. Malyshev, Amazing symmetrical clustering in chloroplast genomes, *BMC Bioinformatics* 21(Suppl 2) 83 (2020). doi:10.1186/s12859-020-3350-z.
- [12] L. A. Staehelin, C. J. Arntzen, Regulation of chloroplast membrane function: protein phosphorylation changes the spatial organization of membrane components, *The Journal of cell biology* 97(5) (1983) 1327–1337. doi:10.1083/jcb.97.5.1327.

- [13] Y. Wang et al., A quadruple-band metal-nitride nanowire artificial photosynthesis system for high efficiency photocatalytic overall solar water splitting, *Materials Horizons* 6(7) (2019) 1454–1462. doi:10.1039/C9MH00257J.
- [14] S. Zhang et al., An artificial photosynthesis system comprising a covalent triazine framework as an electron relay facilitator for photochemical carbon dioxide reduction, *Journal of Materials Chemistry C* 8(1) (2020) 192–200. doi:10.1039/C9TC05297F.