

Пространственная структура генов фотосинтетической системы хлоропластов водорослей с точки зрения биоинформатики

Сенашова Мария Юрьевна

Институт вычислительного моделирования СО РАН,

Россия, Красноярск, *msen@icm.krasn.ru*

Аннотация. Рассмотрена пространственная структура генов фотосистем I и II хлоропластов водорослей. Под пространственной структурой понимается распределение точек, соответствующих частотным словарям генов, в пространстве частот триплетов. Гены фотосистем образуют два основных кластера, соответствующих прямому и обратному стренду. Не обнаружено группирования точек внутри основных кластеров ни по видам организмов, ни по типам генов, как это характерно для хлоропластов наземных растений и цианобактерий. Распределение по значениям GC-состава неоднородно. Часть надтипов имеют градиентное распределение, часть не обнаруживают выраженного порядка распределения.

Ключевые слова: порядок, кластеризация распределения, эволюция, триплеты

Цитирование: Сенашова М.Ю. Пространственная структура генов фотосинтетической системы хлоропластов водорослей с точки зрения биоинформатики / М.Ю. Сенашова // Информационные и математические технологии в науке и управлении. – 2022. – № 3(27). – С. 34-39. – DOI: 10.38028/ESI.2022.27.3.004.

Введение. Термин «водоросли» является собирательным для огромной группы эукариотических организмов и прокариот, имеющих разное происхождение и не обязательно родственных друг другу. Существует несколько отдельных филогенетических ветвей эукариот, которые эволюционировали независимо друг от друга и содержат фотосинтезирующие организмы. Вопрос о происхождении и эволюции водорослей очень сложен из-за разнообразия этих растений, особенно их субмикроскопического строения и биохимических признаков. Решение этой проблемы затруднено, кроме того, плохой сохранностью большинства водорослей в ископаемом состоянии и отсутствием связующих звеньев между современными отделами растений в виде организмов промежуточного строения.

Не меньший интерес вызывает эволюция фотосинтеза, как с точки зрения изменения структуры фотосистемы в целом, так и ее отдельных единиц. В работах [1-11] рассматриваются различные структурные единицы фотосистем I и II цианобактерий, водорослей и наземных растений с точки зрения биологии и молекулярной генетики, их сходство и различия, возникшие в процессе эволюции. В работах [1-3] анализируются особенности фотосистем I и II красных водорослей, которые являются эволюционным звеном между цианобактериями и растениями. В работах [4-11] сравнивается структура фотосистем I и II цианобактерий, водорослей и высших растений. Показано, что фотосистема имеет высоко консервативное ядро, которое практически не изменяется в процессе эволюции и изменяющуюся часть, которая связана с большим разнообразием среды обитания фотосинтезирующих организмов.

Задачей этой работы было выяснить, можно ли проследить эволюцию фотосинтеза с точки зрения биоинформатики по генетическим текстам. Ранее в [12, 13], были рассмотрены пространственные структуры генов фотосистем I и II цианобактерий и хлоропластов наземных растений. Водоросли являются промежуточным звеном между этими группами организмов. Рассмотрим, чем похожи и чем отличаются структуры, образующиеся в пространстве частот триплетов, для трех этих групп.

1. Материалы и методы. Введём основные понятия. Будем рассматривать гены фотосинтетических систем как символьные последовательности различной длины, состоящие из символов алфавита $M = \{A, C, G, T\}$. Каждой из этих последовательностей мы будем ста-

вить в соответствие частотный словарь толщины 3. Частотный словарь толщины 3 символьной последовательности, соответствующей ДНК – это список всех троек $V_1V_2V_3$ идущих подряд нуклеотидов с указанием частот этих троек; всего может быть 64 триплета. Частота f_ω – это отношение числа копий n_ω данного слова к общему числу всех триплетов N , где N – сумма всех n_ω :

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

Всякий частотный словарь W_3 отображает символьную последовательность в 64-мерное метрическое пространство; близость двух геномов задается естественным образом – например, как близость двух точек в Евклидовой метрике:

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

Для каждого гена был построен частотный словарь $W_{(3,3)}$. Частотный словарь $W_{(3,3)}$ представляет собой множество частот триплетов, причем триплеты в гене брались без пересечения троек символов. Для генов, находящихся в обратном стренде, частотный словарь строился с учетом того, что символьная последовательность, относящаяся к таким генам, инвертировалась. Частотный словарь соответствует точке в 64-мерном пространстве. Таким образом, каждому гену соответствует точка в пространстве частот триплетов. Для визуализации пространственной структуры, образованной множеством этих точек, была построена проекция в пространство первых трех главных компонент. Для выявления структуры в наборе генов проводилась предварительная обработка, которая ставила в соответствие данному набору символьных последовательностей множество точек в 64-мерном пространстве триплетов. Каждому гену соответствует своя точка. Кроме названия гена, с точкой связаны название вида, которому принадлежит ген, к какому стренду ген относится – прямому или обратному, GC-состав гена. По полученному множеству точек в программе *VidaExpert* (<http://bioinfo-out.curie.fr/projects/vidaexpert/>) строился вид данных в пространстве первых трёх главных компонент, вычисленных для 64-мерного пространства триплетов. Рассматривались проекции на плоскости пространства первых главных компонент.

2. Результаты. Были проанализированы 142 генома хлоропластов одноклеточных и многоклеточных водорослей, имеющихся на настоящий момент в EMBL-банке, в том числе одноклеточные водоросли следующих надтипов: Alveolata – 4 вида, Chlorophyta – 57 видов, Cryptophyta – 3 вида, Euglenozoa – 4 вида, Glaucophyta – 1 вид, Haptophyta – 4 вида, Ochrophyta Bacillariophyta – 37 видов, Rhodophyta – 4 вида, Streptophyta – 10 видов. Для многоклеточных водорослей число имеющихся надтипов существенно меньше: Chlorophyta – 4 вида, Ochrophyta Bacillariophyta – 4 вида, Rhodophyta – 7 видов, Streptophyta – 3 вида. К сожалению, EMBL-банк содержит далеко не все надтипы водорослей, а количество видов в имеющихся надтипах сильно отличается. Во всех геномах выделялись гены фотосинтетических систем I и II, для которых анализировалось пространственное распределение точек, соответствующих генам в пространстве частот триплетов, в соответствии со стрендами, принадлежности к определенным видам, типам генов и величине GC-состава.

Ранее для генов наземных растений и цианобактерий было обнаружено четкое разделение соответствующих им точек пространства триплетов на кластеры прямого и обратного стрендов. Аналогичное разделение обнаружено для генов фотосистем одноклеточных и многоклеточных водорослей.

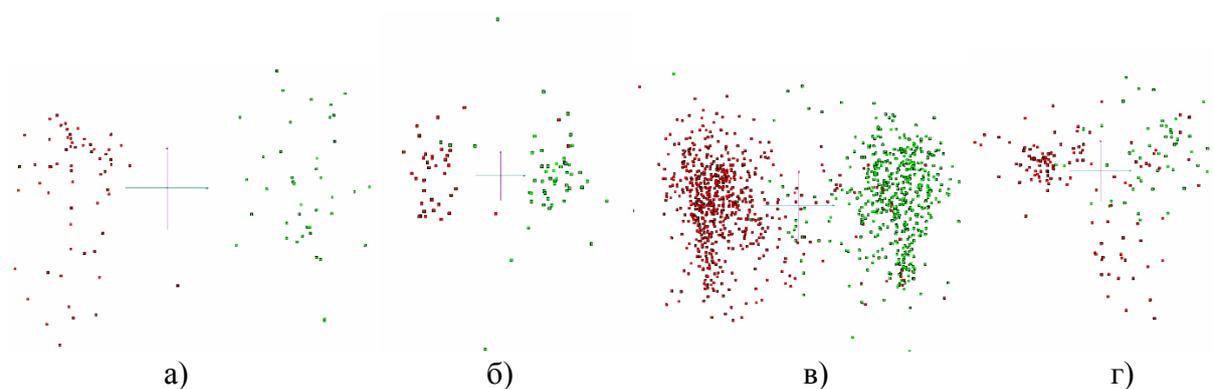


Рис. 1. Проекция в плоскость двух первых главных компонент. Гены, лежащие в прямом стренде, обозначены красным цветом, гены, лежащие в обратном стренде, обозначены зеленым цветом

В качестве примера на рисунке 1а) показано разделения на кластеры, соответствующие прямому и обратному стренду, генов надтипа *Alveolata*. Однако существуют три исключения из такого распределения: гены одноклеточных водорослей надтипа *Chlorophyta* (рисунок 1б), *Cryptophyta* (рисунок 1в), *Euglenozoa* (рисунок 1г).

Гены фотосистемы хлоропластов наземных растений образуют плотные кластеры внутри кластеров, относящихся к прямому и обратному стренду. Гены фотосистемы цианобактерий, напротив, образуют кластеры по видовой принадлежности организмов. Гены фотосистемы как одноклеточных, так и многоклеточных водорослей не группируются ни по одному из этих признаков, как видно из рисунка 2. На рисунке 2а) показаны одинаковым цветом точки, относящиеся к одному и тому же виду из надтипа *Chlorophyta*. Малиновым цветом обозначены точки, относящиеся к виду *Auxenochlorella protothecoides*, бирюзовым – к виду *Botryococcus braunii*, фиолетовым – к виду *Chlorella sorokiniana*. Красными и зелеными квадратами меньшего размера обозначены другие точки прямого и обратного стренда. На рисунке 2б) одинаковым цветом показаны точки, относящиеся к одному и тому же типу гена. Синим цветом обозначены точки, относящиеся к гену *psaA*, оранжевым – к гену *psaB*, розовым – к гену *psaC*.

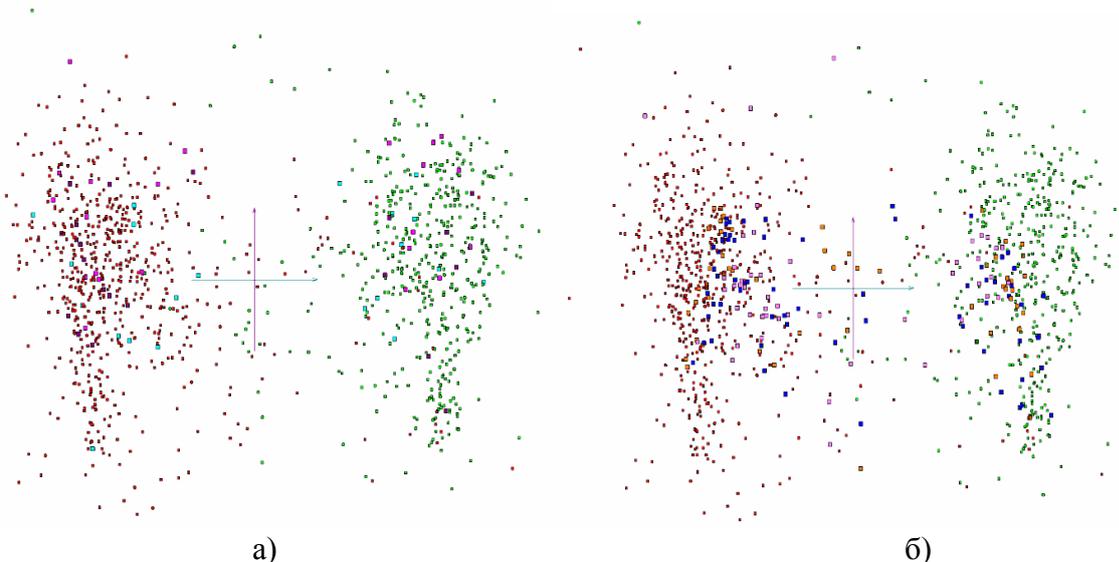


Рис. 2. Распределение точек, относящихся к одному виду (2а) и одному типу гена (2б) в плоскости первой и второй главных компонент

Кластеры прямого и обратного стрендов генов фотосистемы цианобактерий могут быть аппроксимированы вложенными плоскостями [13]. В плоскости 2-ой и 3-ей главных компонент эти плоскости перпендикулярны друг другу (рисунок 3а). Для всех остальных рассмот-

ренных генов фотосистемы (хлоропластов наземных растений и водорослей) такого не наблюдалось. Наиболее похожая структура наблюдается у водорослей надтипа *Rhodophyta* (рисунок 3б). Это очень древняя группа водорослей. Хлоропласты у этих водорослей возникли в результате первичного эндосимбиоза с цианобактериями. При этом у красных водорослей сохраняются фикобилисомы – белковые комплексы, характерные для цианобактерий. К сожалению, в EMBL-банке довольно мало водорослей надтипа *Rhodophyta* (всего 11 видов) и это может давать не точную картину.

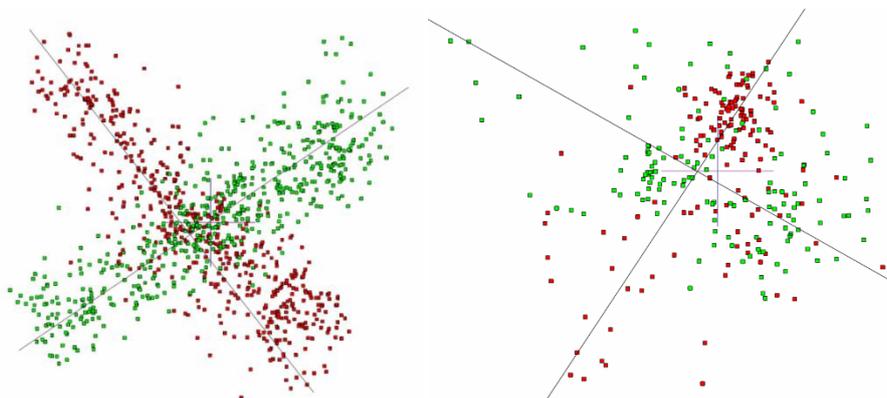


Рис. 3. Структура множества точек, соответствующих прямому и обратному тренду у генов фотосистемы цианобактерий 3а) и водорослей надтипа *Rhodophyta* 3б) в плоскости второй и третьей главных компонент



Рис. 4. Пространственное расположение значений GC - состава генов фотосинтетических систем надтипа *Chlorophyta* 4а) и надтипа *Euglenozoa* 4б) в плоскости первой и второй главных компонент

Для каждого гена вычислялся его GC - состав, то есть отношение количества нуклеотидов G и C к общему числу нуклеотидов в гене. Точки, соответствующие генам со значением GC - состава меньше среднего, обозначены на рис. 4 зеленым цветом, точки, соответствующие генам со средним значением GC - состава, обозначены желтым цветом и точки со значением GC - состава больше среднего обозначены красным цветом. При анализе пространственного расположения значений GC - состава для генов фотосистем хлоропластов наземных растений и цианобактерий было обнаружено, что для цианобактерий характерно градиентное распределение значений GC - состава, а для хлоропластов наземных растений отсутствие какой-либо упорядоченности. Аналогичный анализ был выполнен для генов фотосистем водорослей. Было обнаружено, что градиентное распределение значений GC -

состава наблюдается для надтипов *Alveolata*, *Bacillariophyta*, *Chlorophyta*, *Ochrophyta*, *Streptophyta* (рис. 4а). Для всех остальных исследованных надтипов не наблюдается какой-либо упорядоченности (рис. 4б).

Заключение. Водоросли являются промежуточным звеном эволюции фотосинтеза между цианобактериями и наземными растениями. Водоросли встречаются в основном в водной среде обитания, но широко встречаются и в наземных экосистемах. Отдельные представители обитают в горячих источниках, на скалах, льду, снегу, известковом субстрате. Большое разнообразие среды их обитания обуславливает различия в морфологии и других параметрах организмов. Как было показано выше, нет типичной картины пространственной структуры генов фотосистем хлоропластов водорослей, как это наблюдается для цианобактерий и наземных растений, которые гораздо более однородны по морфологии и выполняемым функциям. Следует обратить внимание, что отсутствие кластеризации как по видам организмов (наблюдаемое у цианобактерий), так и по типам генов (наблюдаемое у наземных растений) может свидетельствовать в пользу того, что водоросли являются промежуточным звеном эволюции между цианобактериями и наземными растениями.

Список источников

1. Vanselow C. et al. Genetic analysis of the Photosystem I subunits from the red alga, *Galdieria sulphuraria*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2009, vol. 1787, no. 1, pp. 46-59.
2. Gardian Z. et al. Organisation of photosystem I and photosystem II in red alga *Cyanidium caldarium*: encounter of cyanobacterial and higher plant concepts. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2007, vol. 1767, no. 6, pp. 725-731.
3. Enami I. et al. Structures and functions of the extrinsic proteins of photosystem II from different species. *Photosynthesis research*, 2008, vol. 98, no. 1, pp. 349-363.
4. Antoshvili M. et al. Structure and function of photosystem I in *Cyanidioschyzon merolae*. *Photosynthesis research*, 2019, vol. 139, no. 1, pp. 499-508.
5. Kato K. et al. Structure of a tetrameric photosystem I from a glaucophyte alga *Cyanophora paradoxa*. *Nature communications*, 2022, vol. 13, no. 1, pp. 1-11.
6. Caspy I. et al. Structure and energy transfer pathways of the *Dunaliella Salina* photosystem I supercomplex. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2020, vol. 1861, no. 10, pp. 148253.
7. Mulo P., Sakurai I., Aro E. M. Strategies for *psbA* gene expression in cyanobacteria, green algae and higher plants: from transcription to PSII repair. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2012, vol. 1817, no. 1, pp. 247-257.
8. Hippler M., Nelson N. The plasticity of photosystem I. *Plant and Cell Physiology*, 2021, vol. 62, no. 7, pp. 1073-1081.
9. Busch A., Hippler M. The structure and function of eukaryotic photosystem I. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2011, vol. 1807, no. 8, pp. 864-877.
10. Alboresi A. et al. Conservation of core complex subunits shaped the structure and function of photosystem I in the secondary endosymbiont alga *Nannochloropsis gaditana*. *New Phytologist*, 2017, vol. 213, no. 2, pp. 714-726.
11. Nugent J.H. A., Purton S., Evans M.C.W. Oxygenic photosynthesis in algae and cyanobacteria: electron transfer in photosystems I and II. *Photosynthesis in algae*, Springer, Dordrecht, 2003, pp. 133-156.
12. Senashova M. Yu. Spatial Structure of Chloroplast Genes of Photosynthetic Systems I And II. *CEUR Workshop Proceedings of the 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications*, 2021, vol. 3047, pp. 115-120.
13. Senashova M. Yu. Unusual Distribution Structure of the Cyanobacteria Photosystem Genes in the Frequency Space of Triplets. *CEUR Workshop Proceedings of the 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications*, 2021, vol. 3047, pp. 121-125

Сенашова Мария Юрьевна. К.т.н., доцент, с.н.с. Института вычислительного моделирования СО РАН. Основные направления исследований: анализ данных, динамика популяций, биоинформатика. *AuthorID: 133140, SPIN-код: 1178-6320, ORCID: 0000-0002-1023-7103, msen@ict.krasn.ru, Россия, г. Красноярск, Академгородок, д. 50, стр. 44.*

UDC 57.015 + 573.2

DOI:10.38028/ESI.2022.27.3.004

Spatial structure of the genes of the algal chloroplasts photosynthetic system from the bioinformatics point of view

Maria Yu. Senashova

Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, Russia, Krasnoyarsk, *msen@icm.krasn.ru*

Annotation. The photosynthetic systems I and II chloroplasts genes spatial structure 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

References

1. Vanselow C. et al. Genetic analysis of the Photosystem I subunits from the red alga, *Galdieria sulphuraria*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2009, vol. 1787, no. 1, pp. 46-59.
2. Gardian Z. et al. Organisation of photosystem I and photosystem II in red alga *Cyanidium caldarium*: encounter of cyanobacterial and higher plant concepts. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2007, vol. 1767, no. 6, pp. 725-731.
3. Enami I. et al. Structures and functions of the extrinsic proteins of photosystem II from different species. *Photosynthesis research*, 2008, vol. 98, no. 1, pp. 349-363.
4. Antoshvili M. et al. Structure and function of photosystem I in *Cyanidioschyzon merolae*. *Photosynthesis research*, 2019, vol. 139, no. 1, pp. 499-508.
5. Kato K. et al. Structure of a tetrameric photosystem I from a glaucophyte alga *Cyanophora paradoxa*. *Nature communications*, 2022, vol. 13, no. 1, pp. 1-11.
6. Caspy I. et al. Structure and energy transfer pathways of the *Dunaliella Salina* photosystem I supercomplex. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2020, vol. 1861, no. 10, pp. 148253.
7. Mulo P., Sakurai I., Aro E. M. Strategies for *psbA* gene expression in cyanobacteria, green algae and higher plants: from transcription to PSII repair. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2012, vol. 1817, no. 1, pp. 247-257.
8. Hippler M., Nelson N. The plasticity of photosystem I. *Plant and Cell Physiology*, 2021, vol. 62, no. 7, pp. 1073-1081.
9. Busch A., Hippler M. The structure and function of eukaryotic photosystem I. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2011, vol. 1807, no. 8, pp. 864-877.
10. Alborese A. et al. Conservation of core complex subunits shaped the structure and function of photosystem I in the secondary endosymbiont alga *Nannochloropsis gaditana*. *New Phytologist*, 2017, vol. 213, no. 2, pp. 714-726.
11. Nugent J.H. A., Purton S., Evans M.C.W. Oxygenic photosynthesis in algae and cyanobacteria: electron transfer in photosystems I and II. *Photosynthesis in algae*, Springer, Dordrecht, 2003, pp. 133-156.
12. Senashova M. Yu. Spatial Structure of Chloroplast Genes of Photosynthetic Systems I And II. *CEUR Workshop Proceedings of the 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications*, 2021, vol. 3047, pp. 115-120.
13. Senashova M. Yu. Unusual Distribution Structure of the Cyanobacteria Photosystem Genes in the Frequency Space of Triplets. *CEUR Workshop Proceedings of the 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications*, 2021, vol. 3047, pp. 121-125.

Senashova Maria Yurievna. *Candidate of Physical and Mathematical Sciences, Associate Professor, Senior Researcher Institute of Computational Modeling SB RAS. Main areas of research: data analysis, population dynamics, bioinformatics. AuthorID: 133140, SPIN: 1178-6320, ORCID: 0000-0002-1023-7103, msen@icm.krasn.ru, Russia, Krasnoyarsk, Akademgorodok, 50, building 44.*

Статья поступила в редакцию 05.08.2022; одобрена после рецензирования 06.09.2022; принята к публикации 16.09.2022.

The article was submitted 08/05/2022; approved after reviewing 09/06/2022; accepted for publication 09/16/2022.