



# ГЕНЕТИКА ПОПУЛЯЦИИ И ЭВОЛЮЦИЯ

© <sup>1</sup>Л.Е. Михеева, <sup>2</sup>Ю.Л. Орлов,  
<sup>2</sup>Н.А. Колчанов, <sup>1</sup>С.В. Шестаков

<sup>1</sup> Институт общей генетики  
им. Н.И. Вавилова РАН, Москва

<sup>2</sup> Институт цитологии и генетики  
СО РАН, Новосибирск

✿ Исследование *in silico* структуры, количества, распределения и локализации (внутригенной и межгенной) прямых и инвертированных повторяющихся последовательностей в геномах восьми одноклеточных цианобактерий показало перспективность такого подхода в полногеномном анализе для решения задач молекулярной филогении и экологической геномики. Сравнительный анализ распределения неслучайных повторов позволил: подтвердить высокую степень генетического родства двух штаммов

*Prochlorococcus marinus* (MED4 и SS120), обладающих редуцированными геномами и обитающих в различных по уровням освещенности экинишах; установить филогенетическую близость геномов *Prochlorococcus marinus* MIT9313 и *Synechococcus* WH8102, существенно различающихся по набору белков светособирающих комплексов фотосистем, но занимающих сходные ниши в морских экосистемах; выявить характерные различия в геномной организации между морскими и пресноводными видами одноклеточных цианобактерий.

✿ **Ключевые слова:** эволюционная геномика; цианобактерии; повторяющиеся последовательности; экофизиология

## ФИЛОГЕНЕТИЧЕСКИЙ АНАЛИЗ СОВЕРШЕННЫХ ПОВТОРОВ В ГЕНОМАХ ОДНОКЛЕТОЧНЫХ ЦИАНОБАКТЕРИЙ

### ВВЕДЕНИЕ

Цианобактерии являются таксономически разнообразной группой фотосинтезирующих прокариот, играющих глобальную роль в различных экосистемах. Большое внимание уделяется изучению филогении цианобактерий в связи с проблемами экологической геномики, эволюционной биологии, эндосимбиотического происхождения хлоропластов [6]. На основе классических методов анализа консервативных генов 16S–23S рибосомной РНК построены филогенетические дендрограммы, включающие более 70 видов и штаммов цианобактерий, относящихся к различным морфофизиологическим и экологическим группам [7, 11, 12, 16]. Однако эти дендрограммы часто не согласуются с филогенией по отдельным белкам и с фенотипическими особенностями, определяющими экологически значимые признаки цианобактерий. В частности, цианобактерии, относящиеся к роду *Synechococcus* могут находиться на дендрограмме 16S РНК-генов на отдаленном расстоянии друг от друга, в то время как некоторые из них имеют высокую степень филогенетической близости к оксифотобактериям рода *Prochlorococcus*, отличающимся от других цианобактерий по белкам светособирающих комплексов аппарата фотосинтеза [14].

Сведения о полностью секвенированных геномах различных цианобактерий существенно раздвигают рамки сравнительной эволюционной геномики и позволяют обсуждать вопросы таксономии цианобактерий с учетом их экофизиологических характеристик. В нашей работе для исследования геномов одноклеточных цианобактерий использован новый подход, базирующийся на сравнительном анализе *in silico* протяженных совершенных повторяющихся последовательностей, наличие и распределение которых в геномах разных видов и штаммов может отражать филогенетические связи организмов.

### МАТЕРИАЛЫ И МЕТОДЫ

Исследовали восемь одноклеточных цианобактерий, различающихся по размерам полностью секвенированных геномов, числу генов, среде обитания и морфофизиологическим характеристикам (табл. 1). Три штамма *Prochlorococcus* (MED4, SS120, MIT9313), а также *Synechococcus* WH8102 относятся к формам морского фитопланктона, другие штаммы (BP1, PCC6301, PCC6803) являются пресноводными цианобактериями. Клетки *Prochlorococcus* используют дивинил-хлорофиллы **a/b** в светособирающих комплексах, тогда как другие группы цианобактерий содержат фикобилиновые антенны и хлорофилл **a** в фотосистемах. *Gloeobacter violaceus* PCC7421 относится к бестилакоидным древним цианобактериям. [8]. Для

Таблица 1

## Характеристика полностью секвенированных геномов одноклеточных цианобактерий

Организм	Среда обитания, биологические особенности	Размер генома тнп	Г + Ц состав, %	Число ОРС*	Число генов р-РНК	Число <i>hli</i> -генов**	Число <i>pcb</i> -генов***
<i>Prochlorococcus marinus pastoris</i> MED4	Морской фитопланктон, хлорофилл <b>a/b</b>	1658	31	1713	3	22	1
<i>Prochlorococcus marinus marinus</i> SS120	Морской фитопланктон, хлорофилл <b>a/b</b>	1751	36,4	1884	3	13	8
<i>Prochlorococcus marinus</i> MIT9313	Морской фитопланктон, хлорофилл <b>a/b</b>	2411	51	2265	6	9	2
<i>Synechococcus</i> sp. WH8102	Морской фитопланктон, хлорофилл <b>a</b>	2434	59,5	2522	6	8	0
<i>Thermosynechococcus elongates</i> BP-1	Пресноводная, термофил, хлорофилл <b>a</b>	2594	53,9	2475	3	1	0
<i>Synechococcus elongates</i> PCC 6301	Пресноводная, хлорофилл <b>a</b>	2696	55,5	2525	6	2	0
<i>Synechocystis</i> sp. PCC 6803	Пресноводная, хлорофилл <b>a</b>	357	47	3167	6	5	0
<i>Gloeobacter violaceus</i> PCC 7421	Бестилакоидные клетки	4659	62	4430	3	0	0

Примечание :\* — открытая рамка считывания (ОРС); \*\* — *hli*-гены (high light inducible), кодирующие белки, синтез которых индуцируется при сильной освещенности; \*\*\* — *pcb*-гены, кодирующие светособирающие антенные белки, связанные с хлорофиллами *a/b*.

сравнения были проанализированы геномы бактерий *Rhodospseudomonas palustris* CGA009 и *Bacillus halodurans* C125. Информация о нуклеотидных последовательностях геномов прокариот получена из баз данных GenBank (<http://www.ncbi.nlm.nih.gov/GenBank/>) и CyanoBase (<http://www.kazusa.or.jp/cyano/>).

Сходство геномов оценивалось по общим фрагментам ДНК (совершенным повторам). Действительно, число генов-ортологов в геномах отражает их эволюционную близость. Вариант такой меры — число протяженных совпадающих участков, соответствующих консервативным доменам, генам или кластерам генов при сравнении геномов с помощью программы BLAST [5]. Для поиска повторов в полных геномах использовали компьютерные программы, разработанные в Институте цитологии и генетики СО РАН [9, 10]. Для каждой из исследуемых цианобактерий определены структура, количество, локализация (внутригенная и межгенная) прямых и инвертированных повторов, изучены картины сходства и различия в распределении случайных повторов при попарном сравнении геномных карт.

Для анализа сходства полных геномов использовалась программа MUMmer [4], позволяющая с высокой эффективностью выполнять сравнение последовательностей в геномах близкородственных организмов. В основе метода сравнения лежит принцип разложения геномов на отдельные подпоследовательности, так называемые MUMs (maximal unique matching subsequences), полностью со-

впадающие, но встречающиеся в двух сравниваемых геномах не более одного раза. Размер MUMs составляет от 15 до 100 и более нуклеотидных пар (нп) и соответственно, количество найденных MUMs при попарном сравнении геномов тем больше, чем ниже минимальный заданный размер совпадающих последовательностей [4].

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Одна из существенных характеристик при анализе геномов — насыщенность их случайными прямыми и инвертированными повторами [15]. В табл. 2 приведены данные по количеству и характеристикам внутригеномных совершенных повторов длиной  $\geq 20$  нп для восьми штаммов цианобактерий. В этих геномах присутствуют протяженные случайные совершенные повторы, достигающие в длину более одной тысячи нуклеотидных пар (за исключением штамма SS120).

Общая доля последовательностей генома, занятая протяженными случайными повторами в геномах морских фитопланктонных организмов *Prochlorococcus* и *Synechococcus* меньше, чем у остальных рассматриваемых цианобактерий и в среднем у других бактерий: около 1% против 3–5%. Усредненное значение этого параметра для 122 последовательностей полных бактериальных геномов составляет 4,6% [1]. Прослеживается тенденция увеличения числа повторов в зависимости от размера генома, однако общая доля генома, занятого случай-

ными повторами, не всегда коррелирует с размерами генома. Например, в геноме термофильной цианобактерии *Thermosynechococcus elongatus* BP1 неслучайные протяженные повторы занимают 5,8% и наблюдается наибольшее число мультикопийных повторов (присутствующих в трех и более копиях) по сравнению с другими цианобактериями. Суммарная доля позиций, занятых одинарными (уникальными) повторами, составляет от 19 до 83% (в среднем 56%) от общего числа протяженных повторов; на мультикопийные приходится около 44% (соотношение колонок 5 и 6 в табл. 2).

Морские штаммы *Prochlorococcus* MED4 и SS120 характеризуются высокой степенью сходства геномов, но отличаются от *Prochlorococcus* MIT9313, геном которого по критерию распределения повторов ближе к геному штамма *Synechococcus* WH8102. В табл. 3 приведены

данные детального анализа внутригеномных повторов у *Prochlorococcus* MED4 и SS120, имеющих наименьший размер геномов.

Компьютерный анализ геномов штаммов *Prochlorococcus* MED4 и SS120 показал:

1. В геномах этих штаммов обнаруживается очень мало инвертированных повторов размером  $\geq 50$  нп;
2. Примерно 27–29% всех повторов являются внутригеномными, причем около 50% из них являются тандемно повторяющимися участками низкой сложности в генах, кодирующих сигнальные белки, РНК-связывающие домены, шапероны, факторы инициации трансляции. Большинство близко расположенных повторов, находящихся между соседними открытыми рамками считывания (ОРС) в одном межгеномном участке, являются тандемными повторами коротких участков ДНК;

Таблица 2

Характеристика внутригеномных повторов

Геном штамма	Кол-во повторов (более 20 нп)	Средняя длина повтора (более 20 нп)	Размер максимального повтора, нп	Локализация максимального повтора (дупликация)	Всего занято повторами, нп	Из них одинарных повторов, нп	Доля генома, занятая повторами
	1	2	3	4	5	6	7
MED4	89 (69, 20)	31,6	1257(П)	Кластер 4-х hli-генов	19750	16216	0,012
SS120	99(82, 17)	34,1	303(И)	Гены белков-поринов	10570	7261	0,006
MIT9313	489 (357, 132)	40,3	6182(И)	Кластер генов р-РНК	73174	46620	0,030
WH8102	461 (390, 71)	57,7	3122(И)	Кластер генов р-РНК	72138	40812	0,030
BP-1	948 (623, 325)	66,0	3228(И)	Гены обратной транскриптазы	150160	29596	0,058
6301	145 (92, 53)	40,2	5436(И)	Кластер генов р-РНК	36055	30106	0,013
6803	758 (378, 380)	82,8	5361(И)	Кластер генов р-РНК	127544	49889	0,036
7421	1206 (754, 452)	38,0	1672(П)	Гены транспозаз	183491	117856	0,039
<i>Rhodospseudomonas palustris</i> CGA009	1377 (746, 631)	25,3	5211(П)	Кластер генов р-РНК	220870	160119	0,040
<i>Bacillus halodurans</i> C-125	1224 (768, 456)	113	5283(П)	Кластер генов р-РНК	222251	26401	0,053

Примечание: в колонке 1 цифры в скобках через запятую обозначают количество прямых и инвертированных повторов; в колонке 2 приведены данные о средней длине совершенного повтора для повторов, содержащих 20 и более нп; в колонке 3 буквы в скобках обозначают тип повтора: П — прямой повтор, И — инвертированный повтор; колонка 5 содержит данные об общем числе позиций в геноме, занятых неслучайными совершенными повторами; в колонке 6 приведено число позиций, повторенных не более двух раз (в отличие от мультикопийных, повторенных три-четыре и более раз); в колонке 7 приведена доля (от 0 до 1) позиций в геноме, занятых повторами, которая рассчитывалась как число позиций занятых протяженными повторами (колонка 5), разделенное на размер генома.

3. Около 30% повторов расположены в различных генах: 50% из них у штамма MED4 и более 80% — у штамма SS120 располагаются в гомологичных генах, выполняющих одинаковую или сходную функцию;

4. Остальные 40% повторов находятся либо полностью в некодирующих межгенных участках генома, либо только один из повторов находится между ОРС, а другой — внутри ОРС. Заметное число таких повторов находится в генах тРНК или в непосредственной близости от них. Известно, что именно гены тРНК являются локусами, в которых часто происходит реорганизация генома в результате гомологичной рекомбинации или интеграции фаговой ДНК [13].

Следует отметить, что значительное число генов, содержащих совершенные повторы, у штамма MED4, адаптированного к высокой интенсивности света, представлены «hli»-генами (high light inducible), кодирующими белки, синтез которых индуцируется при сильной освещенности. Участок генома с четырьмя такими генами является дупликацией и образует максимальный прямой повтор в геноме штамма MED4, причем копии расположены на значительном удалении друг от друга, так как

начальные позиции этого повтора в геноме — 1344538 (Hli 6–9) и 772664 (Hli 16–19). Всего в геноме MED4 обнаруживается 22 гена, кодирующих Hli-белки, тогда как в геномах других морских цианобактерий, обитающих в эконишах с более низкой освещенностью, их существенно меньше: в геноме штамма SS120 — 13, у штамма MIT 9313 — 9 и у штамма WH8102 — 8 [2].

Протяженные совершенные повторы в геноме штамма SS120 обнаруживаются в многочисленных «pcb»-генах, кодирующих светособирающие антенные белки, связанные с хлорофиллами a/b. Геном этого штамма содержит 8 различных pcb-генов, геном MIT9313 — всего 2 pcb-гена, в геноме *Synechococcus* WH8102 вообще нет таких генов, так как этот штамм использует фикобилисомы в качестве светособирающего комплекса [3].

Таким образом, анализ повторяющихся последовательностей в геномах штаммов *Prochlorococcus* MED4 и SS120 выявляет наличие гомологичных генов, количество которых связано с адаптацией этих штаммов к условиям высокой (MED4) или низкой (SS120) освещенности. Подобные дупликации, приводящие к появлению в геномах многочисленных функционально гомологичных

Таблица 3

Анализ совершенных внутригеномных повторов у двух морских штаммов *Prochlorococcus*

Характеристика повторов		MED4			SS120		
		Всего	П	И	Всего	П	И
Число повторов (? 20 нп)	20–49 нп	74	54	20	81	65	16
	50–99 нп	6	6	0	13	13	0
	100–194 нп	5	5	0	1	1	0
	200–499 нп	2	2	0	4	3	1
	Более 500 нп	2	2	0	0	0	0
	Общее число	89	69	20	99	82	17
Длина максимальных повторов, нп		—	1257	48	—	239	303
Локализация повторов	Внутригенные (2 копии в одной ОРС)	24	24	0	29	29	0
	Межгенные (2 копии в различных ОРС)	29	21	8	29	16	13
		В том числе дупликация 4-х hli-генов и 12 повторов в 10 парах других гомологичных генов, среди которых еще 6 hli-генов			В том числе 23 повтора в 20 парах гомологичных генов, среди которых 8 pcb-генов		
	2 копии в некодирующих областях	19	13	6	8 + 27*	4 + 27*	4
	Одна копия в ОРС, другая — в некодирующих областях	17	12	5	6	6	0

Примечание: П — прямые повторы, И — инвертированные повторы.

\* — 27 повторов — множественные повторы в двух межгенных участках:

17 — в области между генами Pro0676 и Pro0677 (1587 нп)

10 — в области между генами Pro1460 и Pro1461 (1428 нп).

генов, повидимому, отражают эволюционные пути геномной реорганизации, увеличивающей приспособленность к различным экологическим условиям обитания.

В целях сравнительного геномного анализа была проведена оценка количества и характера совершенных повторов при попарном сравнении цианобактериальных геномов между собой (табл. 4). Определяли размер и положение совершенных повторов длиной 20 и более нуклеотидов в одном из геномов (I) по отношению к другому геному (II). При анализе повторов в попарных межгеномных сравнениях различных цианобактерий было обнаружено, что наиболее длинные повторы оказывались связанными с последовательностями кластеров генов рибосомных РНК. Например, все повторы длиной 100 и более нп локализовались в этих кластерах. Это неудивительно, так как именно критерий сходства последовательностей генов рибосомных РНК используется в систематике для объединения организмов в крупные таксономические группы.

Результаты межгеномного сравнения цианобактерий по критерию количества совершенных неслучайных повторов длиной 20 и более нуклеотидов представлены в табл. 4: приведены данные о количестве прямых и инвертированных повторов за вычетом повторов, попадающих в кластеры генов рибосомных РНК. Результаты попарных геномных сравнений, выявляющих максимальное число повторов, выделены жирным шрифтом. Нересiproкность значений в попарных сравнениях связана с особенностями учета мультикопийных повторов и не изменяет общей картины.

Штаммы MED4 и SS120 похожи друг на друга по количеству взаимно совпадающих последовательностей (см. табл. 4). При сравнении геномов этих двух штаммов с геномами других цианобактерий обнаруживается меньшее число повторов и при этом снижается как максимальная, так и общая длина «совпадающих» участков генома, что коррелирует со степенью филогенетической удаленности различных видов цианобактерий. Наименьшее сходство в картине повторов для большинства цианобактерий наблюдается при их сравнении с эволюционно далекими бактериями — *Bacillus halodurans* C150, обитающей в щелочных условиях, и почвенной бактерией *Rhodopseudomonas palustris* CGA009.

Высокая степень сходства (максимальное число межгеномных повторов) обнаруживается при сравнении геномов штаммов WH8102 и MIT9313, несмотря на то, что в систематике они относятся к разным родам. Высокая степень геномного сходства этих штаммов по данным сравнительного анализа 16S-рибосомной РНК отмечена и другими авторами [12, 13], обращающих внимание на общность экологических ниш, которые занимают в морском фитопланктоне эти два штамма цианобактерий, обладающих разными типами светособирающих комплексов.

Сопоставление расположения повторов, полученных при межгеномных сравнениях, с генетическими картами цианобактерий (из базы данных CyanoBase) показало, что в большинстве случаев сходные совершенные повторы находятся в генах-ортологах. Поэтому высокие значения числа повторов при межгеномных сравнениях отдельных штаммов могут свидетельствовать о близком

Таблица 4

Повторы при попарных межгеномных сравнениях

Сравнение геномов I II	MED4	SS120	9313	8102	BP1	6301	6803	7421	<i>Rhodopseudomonas palustris</i> GA009	<i>Bacillus halodurans</i> C-125
MED4		<b>239</b> (178, 61)	61 (22, 39)	40 (15, 25)	22 (11, 11)	27 (11, 16)	38 (19, 19)	21 (11, 10)	9 (5, 4)	16 (8, 8)
SS120	<b>227</b> (179, 48)		114 (57, 57)	53 (16, 37)	18 (12, 6)	28 (12, 16)	37 (22, 15)	22 (11, 11)	8 (3, 5)	10 (7, 3)
9313	59 (23, 36)	127 (61, 66)		<b>436</b> (228, 208)	59 (29, 30)	104 (49, 55)	45 (21, 24)	54 (26, 28)	21 (7, 14)	9 (6, 3)
8102	45 (17, 28)	64 (19, 45)	<b>447</b> (232, 215)		80 (37, 43)	176 (96, 80)	79 (38, 41)	<b>180</b> (79, 101)	79 (35, 44)	8 (3, 5)
BP1	24 (14, 10)	20 (12, 8)	51 (22, 29)	88 (43, 45)		<b>259</b> (147, 112)	<b>248</b> (117, 131)	78 (31, 47)	15 (6, 9)	19 (9, 10)
6301	29 (17, 12)	31 (12, 19)	91 (43, 48)	176 (96, 80)	<b>247</b> (132, 115)		175 (91, 84)	121 (62, 59)	37 (17, 20)	16 (3, 13)
6803	32 (16, 16)	31 (21, 10)	45 (19, 26)	89 (41, 48)	<b>260</b> (131, 129)	181 (103, 78)		82 (48, 34)	11 (7, 4)	16 (8, 8)
7421	22 (11, 11)	25 (13, 12)	63 (32, 31)	<b>198</b> (101, 97)	80 (37, 43)	149 (77, 72)	79 (45, 34)		<b>213</b> (109, 104)	11 (6, 5)

Примечание: В клетках указано количество совершенных повторов длиной 20 и более нуклеотидных пар, исключая повторы в кластерах генов рибосомных РНК. В скобках приведено количество прямых (первая цифра) и инвертированных (вторая цифра) повторов. Размеры геномов *Rhodopseudomonas palustris* CGA009 — 5, 46 Мнп; размеры геномов *Bacillus halodurans* C-125 — 4, 20 Мнп.



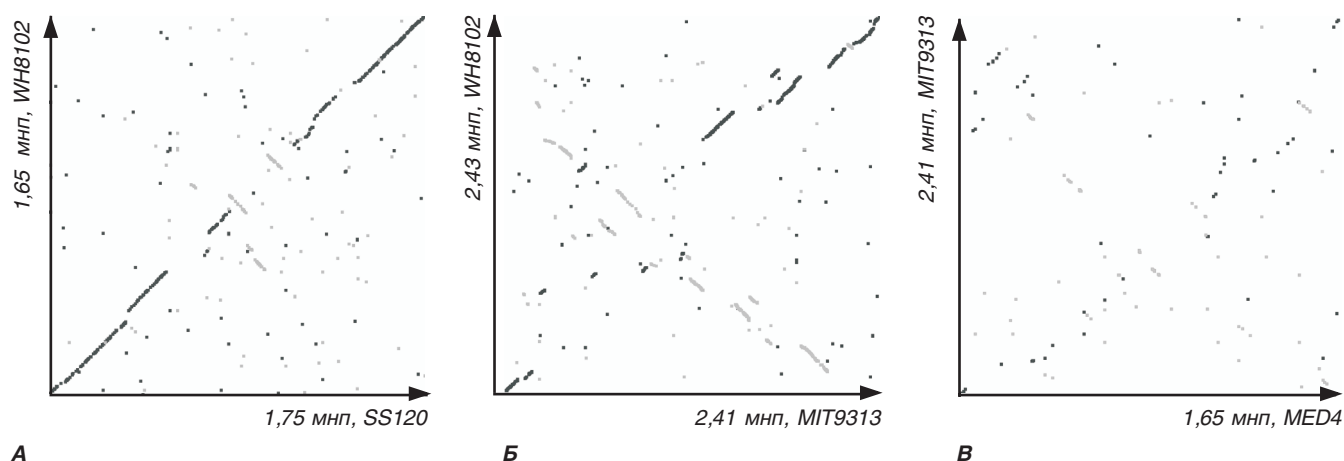


Рис. 1. Диаграммы MUMmer выравниваний геномов некоторых одноклеточных цианобактерий при минимальной длине MUMs-фрагментов равной 20 нп:

каждая точка диаграммы характеризует локализацию отдельных MUMs размером 20 и более нуклеотидных пар в геномах сравниваемых штаммов; темно-серым обозначены повторы, считываемые в обоих геномах в прямом направлении, светло-серым — считываемые по комплементарной нити в одном из геномов (инвертированные повторы).

**А.** Геном *Prochlorococcus* MED4 (ось Y) против генома *Prochlorococcus* SS120 (ось X).

**Б.** Геном *Synechococcus* WH8102 (ось Y) против генома *Prochlorococcus marinus* MIT9313 (ось X).

**В.** Геном *Prochlorococcus* MIT9313 (ось Y) против генома *Prochlorococcus* MED4 (ось X).

генетическом родстве этих организмов. Можно предполагать, что гены-ортологи, в которых в ходе эволюции накапливаются мутации и даже меняются функции в процессе экологической адаптации и дивергенции видов, постепенно «теряют» протяженные совершенные повторы. Напротив, в эволюционно близких видах, «разошедшихся» в относительно недавний период, может сохраниться большее количество сходных участков генома. Чем дальше друг от друга отстоят организмы в эволюционном отношении, тем меньше будет количество и длина межгеномных совершенных повторов. При определении количества совершенных повторов длиной в 50 и более нуклеотидов вообще не было обнаружено таких повторов при сравнении геномов цианобактерий с другими бактериями. При попарном сравнении штаммов цианобактерий друг с другом число таких повторов снижается, оставаясь максимальным в парах MED 4 / SS120 и MIT9313 / WH8102.

Полученные данные хорошо согласуются с результатами проведенного нами полногеномного сравнения различных штаммов цианобактерий с помощью программы MUMmer [4]. На рис. 1 приведены полученные этим методом попарные выравнивания геномов (точечная матрица гомологии) некоторых цианобактерий. Это исследование со всей определенностью свидетельствует о более высокой степени сходства геномов цианобактерий MIT9313 и WH8102 (см. рис. 1Б) и, особенно, MED4 и SS120 (см. рис. 1А) по сравнению со всеми другими вариантами попарных сравнений (в частности, в паре штаммов MIT9313 и MED4 — рис. 1В). На рис. 1А и 1Б обнаруживается не только большое количество совпадающих участков, но и сходство во взаимном расположении

как отдельных генов, так и целых сегментов генома. В наибольшей степени друг другу соответствуют геномы штаммов MED4 и SS120, которые рассматриваются как близкородственные организмы, эволюционно «разошедшиеся» относительно недавно [12].

Представленные в табл. 4 сведения подтверждают, что геном штамма *Prochlorococcus* MIT9313 имеет больше сходных последовательностей с геномом *Synechococcus* WH8102, чем с геномами других штаммов *Prochlorococcus*. В этой паре геномов сходным оказывается и взаимное расположение генов во многих крупных сегментах хромосомы. Эти данные, как и результаты определения степени гомологии 16S-РНК [13] свидетельствуют о филогенетической близости штаммов WH8102 и MIT9313. Некоторые исследователи склоняются к гипотезе о происхождении рода *Prochlorococcus* от древней группы фикобилисом-содержащих цианобактерий. С этой точки зрения штаммы MIT9313 и WH8102 могут оказаться эволюционно более близкими к общему предку, чем к штаммам MED4 и SS120, которые, скорее всего, являются результатом редукционной эволюции древних хлорофилл а/б-содержащих цианобактерий [12, 13, 14].

Сопоставление числа совершенных повторов с размером 20 и более нуклеотидных пар при межгеномных сравнениях (см. табл. 4) свидетельствует об определенном сходстве геномов штаммов пресноводных цианобактерий *Thermosynechococcus* BP1, *Synechococcus* 6301 и *Synechocystis* 6803. Особое место занимает штамм *Gloeobacter* 7421, обладающий геномом большого размера и наибольшим среди анализируемых цианобактерий количеством повторов при сравнении с геномом фо-

тосинтезирующей пурпурной бактерии *Rhodospseudomonas palustris* (но не с *Bacillus halodurans*). Это наблюдение можно рассматривать как косвенное указание на более высокую степень филогенетической близости *Gloeobacter* 7421 к предковому фототрофному организму по сравнению с другими изучаемыми в данной статье цианобактериями. Полученные результаты позволяют обсуждать вопрос о возможной связи молекулярной филогении цианобактерий с их экологическими характеристиками. Количественный анализ совершенных повторов (см. табл. 4) подтверждает выявленные по другим критериям геномного анализа [2] групповые различия между морскими и пресноводными видами цианобактерий. Морские виды, обитающие в более гомогенной окружающей среде, отличаются редуцированными размерами геномов и менее развитыми сигнальными системами по сравнению с пресноводными цианобактериями, обладающими широким диапазоном адаптивных реакций на внешние воздействия [2].

Таким образом, можно заключить, что использованный в работе метод поиска совершенных повторов является эффективным инструментом полногеномного анализа. Наш подход близок к недавно предложенному методу полногеномного анализа [5]. Исследование *in silico* полностью секвенированных геномов цианобактерий позволяет сделать вывод о том, что количество и протяженность прямых и инвертированных повторов в геномах прокариот могут в целом отражать степень генетического родства и/или эволюционной близости исследуемых организмов. Эти сведения могут также дать полезную информацию о корреляции особенностей геномной организации с экофизиологическими характеристиками микроорганизмов.

Работа поддержана грантами «Ведущие научные школы» НШ-1731.2003.4, программой президиума РАН «Происхождение и эволюция биосферы» и РФФИ 05-07-98012.

### Литература

1. Орлов Ю.Л. Анализ повторов в полных бактериальных геномах. // Тр. III съезда ВОГиС. Москва, 6–12 июня 2004: «Генетика в XXI веке: современное состояние и перспективы развития» — 2004. — Т. II. — С. 364.
2. Bhaya D., Dufresne A., Vault D., Grossman A. Analysis of the hli-family in marine and freshwater cyanobacteria // FEMS Microbiol. Lett. — 2002. — Vol. 215(2). — P. 209–219.
3. Bibby T.S., Mary I., Nield J. et al. Lowlightadapted *Prochlorococcus* species possess specific antennae for each photosystem // Nature. — 2003. — Vol. 424(6952). — P. 1051–1054.
4. Delcher A., Kasif S., Fleischmann R. et al. Alignment of whole genomes // Nucleic Acids Res. — 1999. — Vol. 27, N 11. — P. 2369–2376.
5. Программа MUMmer — <http://www.tigr.org/tigrscrits/CMR2/webmum/mumplot>.
6. Henz S.R., Huson D.H., Auch A.F., NieseltStruwe K., Schuster S.C. Wholegenome prokaryotic phylogeny // Bioinformatics. — 2005. — Vol. 21, N 10. — P. 2329–2335.
7. Hess W.R. Genome analysis of marine photosynthetic microbes and their global role // Curr. Opin. Biotechnol. — 2004. — Vol. 15, N 3. — P. 191–198.
8. Honda D., Yokota A., Sugiyama J. Detection of seven major evolutionary lineages in Cyanobacteria based on 16S rRNA gene analysis with new sequences of five marine *Synechococcus* strains // J. Molec. Evol. — 1999. — Vol. 48. — P. 723–739.
9. Nakamura Y., Kaneko T., Sato S. et al. Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids // DNA Res. — 2003. — Vol. 10(4). — P. 137–145.
10. Orlov Y.L., Gusev V.D., Miroshnichenko L.A. LZcomposer: Decomposition of Genomic Sequences by Repeat Fragments. // Biofizika. (Mosk). — 2003. — Vol. 48, Suppl. 1. — P. S7–S16. Интернет-доступная программа Lzcomposer. — <http://www.mgs.bionet.nsc.ru/mgs/programs/lzcomposer/>.
11. Orlov Y.L., Potapov V.N., Poplavsky A.S. Computer analysis of genomic sequence complexity: new applications. // Proceedings of BGRS'2004. — Novosibirsk, Inst. of Cytology & Genetics Press. — 2004. — Vol. 1. — P. 153–157.
12. Partensky F., Hess W.R., Vault D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance // Microbiol. Molec. Biol. Rev. — 1999. — Vol. 63, N 1. — P. 106–127.
13. Rocap G., Distel D.L., Waterbury J.B., Chisholm S.W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal RNA internal transcribed spacer sequences // Appl. Environ. Microbiol. — 2002. — Vol. 68(3). — P. 1180–1191.
14. Rocap G., Larimer F.W., Lamer J. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation // Nature. — 2003. — Vol. 424(6952). — P. 1042–1047.
15. Ting C., Rocap G., King J., Chisholm S. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies // Trends Microbiol. — 2002. — Vol. 10, N 3. — P. 134–142.
16. Van Belkum A. Short sequence repeats in microbial pathogenesis and evolution // Cell Mol. Life Sci. — 1999. — Vol. 56(9–10). — P. 729–734.
17. Wilmotte A. Molecular evolution and taxonomy of the cyanobacteria // Molecular biology of cyanobacteria / Bryant D.A. ed. — Kluwer Acad. Publishers, 1992. — P. 1–26.

✿ SUMMARY: We have fulfilled *in silico* research of number, structure, distribution and location of direct and inverted repeated sequences in eight complete genomes of unicellular cyanobacteria. Analysis of whole genome repeats has shown utility of this approach for purposes of molecular phylogeny and ecological genomics. Comparative analysis of nonrandom repeats patterns has allowed: 1) to confirm the close genetic relationship of two *Prochlorococcus marinus* strains (MED4 and SS120) that have reduced genomes and inhabit the ecotones with different light intensities; 2) to suggest the close phylogenetic relationship of genomes *Prochlorococcus marinus* MIT9313 and *Synechococcus* WH8102 that significantly differ by sets of light-harvesting photosystem; 3) to reveal specific differences in genome organization between marine and freshwater cyanobacteria.

✿ KEY WORDS: ecological genomics; cyanobacteria; repeated sequences; ecophysiology