

УДК 575.8

О ПРИРОДЕ ПРОТЯЖЕННЫХ НУКЛЕОТИДНЫХ ПОВТОРОВ В ГЕНОМАХ ПРОКАРИОТИЧЕСКИХ ОРГАНИЗМОВ

Л.Е. Михеева, Е.А. Карбышева, С.В. Шестаков

(кафедра генетики; e-mail: shestakovgen@mail.ru)

Структурные гены, кодирующие белки и РНК, составляют в геномах прокариот около 85–95% генома; остальная часть генома относится к некодирующим нуклеотидным последовательностям (Rogozin et al., 2002). В прокариотических геномах большинство генов являются уникальными (единичными), но некоторые участки генома могут быть дублированы или мультиплицированы. Доля совершенных повторов в геномах прокариот составляет у различных видов от 0,7 до 13% генома (Орлов, 2004). Многие повторы связаны с мультикопийностью рибосомных оперонов (Турова, 2003).

Определение количества, размеров и распределения геномных повторов используется для оценки филогенетического сходства и различия геномов прокариот и позволяет анализировать особенности геномной организации в связи с экофизиологическими характеристиками организмов (Михеева и др., 2005; Henz et al., 2005). Так, например, на основе полногеномного анализа показана адаптивная роль мультипликации генов, кодирующих у цианобактерий белки светособирающих комплексов (Bibby et al., 2003) и белки, индуцируемые в условиях повышенной освещенности (Bhaya et al., 2002).

Прокариотические геномы содержат также повторяющиеся последовательности, представленные инсерционными (IS) и транспозоноподобными элементами (Mahillon et al., 1999). Наличие копий IS-элементов различного типа характерно, например, для галофильных и алкалофильных бацилл, обитающих в экстремальных экологических условиях (Takaki et al., 2004). Известно, что IS-элементы и транспозоны содержат гены, кодирующие белки транспозиции, и могут участвовать в геномных перестройках, приводящих к мутациям и/или к изменениям в регуляции экспрессии генов (Mahillon et al., 1999; Schneider, Lenski, 2004), что создает условия для изменения вектора отбора и скоростей эволюции (Frost et al., 2005).

В целях выяснения природы и распределения повторяющихся последовательностей в геномах различных таксономических групп прокариот мы исследовали геномы 64 свободноживущих микроорганизмов (52 вида бактерий и 12 видов архей), используя данные по полногеномному секвенированию.

Материалы и методы

Геномный анализ проводили на основе базы данных GenBank (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>). Для поиска повторов в полных геномах использовали компьютерные программы, разработанные в Институте цитологии и генетики СО РАН (Orlov et al., 2003, 2004, <http://www.mgs.bionet.nsc.ru/mgs/programs/lzcomposer/>). Для каждого генома определяли количество неслучайных внутригеномных прямых и инвертированных нуклеотидных повторов длиной 20 и более пар нуклеотидов (пн), долю генома, занятого повторяющимися последовательностями, и распределение повторов на молекулярно-генетических картах. Данные о количестве, локализации и номенклатуре IS-элементов в геномах прокариот получены из базы данных ISfinder (Siguier et al., 2006; <http://www-is.biotoul.fr>).

Результаты и обсуждение

Для анализа были выбраны полностью секвенированные геномы свободноживущих микроорганизмов, принадлежащих к основным систематическим группам прокариот, обитающих в различных экологических нишах и различающихся по типам метаболизма.

В табл. 1 представлены данные о количестве и природе повторяющихся последовательностей в геномах проанализированных прокариотических организмов. Доля генома, занятого неслучайными повторами длиной 20 пн, в бактериальных геномах редко превышает 10%. Археобактерии в этом отношении близки к бактериям за исключением некоторых представителей, у которых это значение достигает 13–14%. Минимальное содержание повторяющихся последовательностей наблюдается у *Prochlorococcus marinus* 1375 (0,6% генома) и *Thermoplasma acidophilum* (0,7%), характеризующихся малыми размерами геномов.

Основное внимание было уделено совершенным протяженным повторам (500 и более пн), которые могут принадлежать мультиплицированным генам. Определение локализации таких повторов на геномных картах показало, что они представлены генетическими элементами различной природы.

В геномах большинства бактерий значительная часть протяженных повторов связана с кластерами

генов, кодирующих рибосомные РНК (р-РНК) и являющихся эволюционно консервативной частью геномов. Совершенные повторы в кластерах генов р-РНК могут быть единичными и множественными в зависимости от количества таких кластеров в данном геноме и протяженности полностью гомологичных участков внутри кластеров. В изученных геномах *Bacillus*, *Lactobacillus*, *Oceanobacillus*, *Agrobacterium*, *Corynebacterium*, *Zymomonas*, *Pseudomonas* протяженные повторы в кластерах генов р-РНК составляют от 50 до 100% всех протяженных внутригеномных повторов (данные выделены жирным шрифтом в последнем столбце табл. 1). Как правило, в геномах бактерий, содержащих более одного протяженного кластера р-РНК (занимающего в сред-

нем от 5 до 6 тпн), локализуются и максимальные по протяженности совершенные нуклеотидные повторы. Среди 52 проанализированных бактериальных геномов в 39 случаях максимальные по протяженности внутригеномные повторы находятся в области генов р-РНК. Если в геномах большинства изученных архей гены р-РНК не собраны в кластеры (табл. 1), то в двух геномах (*Methanococcus*, *Methanosarcina*) наблюдаются объединение и мультипликация генов р-РНК, в результате чего именно в этих областях обнаруживаются максимальные по протяженности внутригеномные повторы. Особенности структуры и распределения многокопийных кластеров генов рибосомной РНК в геномах прокариот детально рассмотрены в обзоре (Турова, 2003).

Таблица 1

**Анализ внутригеномных повторяющихся последовательностей
у свободноживущих представителей различных систематических групп прокариот**

Систематическая группа	Организм	Размер генома, Мпн	Доля генома, занятого повторами	Общее число внутригеномных повторов (П, И)*		Число генов в геноме, кодирующих ферменты сайт-специфической рекомбинации		Число внутригеномных повторов 500 пн	
				20 пн	в том числе 500 пн	всего	в том числе транс-позазы	в генах транспозаз (П, И)*	в генах риб-РНК ***
1	2	3	4	5	6	7	8	9	10
<i>Crenarchaeota</i>	<i>Aeropyrum pemix</i> K1	1,67	0,013	254 (172,82)	1 (0,1)	1	0	—	0 (1)
	<i>Sulfolobus acidocaldarius</i> DSM639	2,23	0,014	109 (82,27)	1 (1,0)	7	5	0	0 (1)
	<i>Sulfolobus tokodaii</i> str. 7	2,69	0,056	1672 (955,717)	10 (1,9)	26	25	5 (1,4)	0 (1)
	<i>Sulfolobus solfataricus</i> P2	2,99	0,137	3621 (1945,1676)	70 (31,39)	236	235	50 (20,30)	0 (1)
<i>Euryarchaeota</i>	<i>Picrophilus torridus</i> DSM9790	1,55	0,014	224 (170,54)	0	6	3	0	0 (1)
	<i>Thermoplasma acidophilum</i> DSM1728	1,56	0,007	109 (82,27)	1 (1,0)	4	3	0	0 (1)
	<i>Thermoplasma volcanium</i> GSS1	1,58	0,024	268 (166,102)	1 (0,1)	49	38	1 (0,1)	0 (1)
	<i>Methanococcus maripalidus</i> S2	1,66	0,028	218 (150,68)	7 (7,0)	4	0	0	4 (3)
	<i>Pyrococcus abissi</i> GE5	1,77	0,008	165 (115,50)	1 (0,1)	2	0	—	0 (1)
	<i>Pyrococcus furiosus</i> DSM3638	1,91	0,027	403 (303,110)	22 (12,10)	30	27	19 (10,9)	0 (1)
	<i>Archaeoglobus fulgidus</i> DSM4304	2,18	0,028	461 (290,171)	10 (8,2)	27	27	8 (6,2)	0 (1)
	<i>Methanosarcina mazei</i>	4,10	0,072	1871 (1475,396)	68 (36,32)	97	96	50 (26,24)	6 (3)
	<i>Candidatus Pelagibacter ubique</i>	1,33	0,016	164 (146,18)	1 (1,0)	1	0	—	0 (1)

Продолжение табл. 1

1	2	3	4	5	6	7	8	9	10
Alphaproteobacteria	<i>Zymomonas mobilis</i>	2,06	0,021	200 (168,32)	4 (4,0)	6	3	0	3 (3)
	<i>Gluconobacter oxydans</i> 621H	2,92	0,043	226 (102,124)	64 (22,42)	77	69	49 (21,28)	9 (4)
	<i>Rhodobacter sphaeroides</i>	3,19	0,031	591 (428,169)	1 (1,0)	12	3	0	0 (1)
	<i>Caulobacter crescentus</i> CB15	4,02	0,049	1353 (753,600)	20 (11,9)	38	34	14 (6,8)	2 (2)
	<i>Rhodopseudomonas palustris</i> CGA009	5,46	0,040	1431 (779,652)	10 (6,4)	18	7	6 (2,4)	1 (2)
	<i>Agrobacterium tumefaciens</i> C58	2,075	0,021	288 (173,115)	5 (5,0)	16	13	2 (2,0)	3 (2)
		2,84	0,014	287 (149,138)	1 (0,1)	7	2	0	1 (2)
	<i>Sinorhizobium meliloti</i> 1021	3,65	0,045	1102 (597,505)	41 (21,20)	79	70	37 (19,18)	2 (3)
	<i>Mesorhizobium loti</i> MAFF 303099	7,04	0,041	1891 (1205,686)	18 (9,9)	131	90	13 (4,9)	1 (2)
Chlorobi	<i>Chlorobium tepidum</i> TLS	2,16	0,027	360 (251,109)	9 (2,7)	10	7	1 (0,1)	2 (2)
	<i>Chlorobium chlorochromatii</i> CaD3	2,57	0,074	1938 (1627,311)	18 (12,6)	19	17	12 (6,6)	0 (1)
Betaproteobacteria	<i>Nitrosomonas europaea</i> ATCC19718	2,81	0,063	557 (357,200)	87 (50,37)	128	117	67 (35,32)	0 (1)
	<i>Dechloromonas aromatica</i> RCB	4,50	0,063	741 (540,201)	25 (20,5)	36	17	15 (10,5)	3 (4)
	<i>Azoarcus</i> EbN1	4,30	0,078	1158 (714,444)	98 (53,45)	196	167	83 (45,38)	4 (4)
	<i>Rhialstonia solanacearum</i> GM11000	3,71	0,049	767 (411,456)	31 (21,10)	79	64	23 (16,7)	4 (4)
		2,09**	0,088	887 (710,177)	26 (15,11)	66	63	22 (12,10)	0 (1)
Gammaproteobacteria	<i>Idiomarina loihiensis</i> L2TR	2,84	0,025	176 (119,57)	29 (19,10)	35	28	18 (8,10)	10 (4)
	<i>Methylococcus capsulatus</i> Bath	3,3	0,031	281 (215,66)	33 (20,13)	47	39	26 (15,11)	1 (2)
	<i>Nitrosococcus oceani</i> ATCC19707	3,48	0,028	579 (355,224)	27 (12,15)	76	64	19 (10,9)	2 (2)
	<i>Shewanella oneidensis</i> MR-1	4,97	0,068	1516 (914,602)	139 (63,76)	204	184	108 (47,61)	11 (9)
	<i>Photobacterium profundum</i> SS9	4,09	0,079	1472 (957,515)	142 (90,52)	111	101	61 (31,30)	50 (14)
		2,24	0,121	1256 (749,519)	147 (76,71)	147	103	56 (25,31)	0 (1)
	<i>Pseudomonas putida</i> KT2440	6,18	0,052	2189 (1421,768)	64 (47,17)	87	64	27 (18,9)	29 (7)
	<i>Pseudomonas fluorescens</i> Pf-5	7,07	0,046	2440 (1525,915)	12 (9,3)	15	2	0	7 (5)
Deltaproteobacteria	<i>Pelobacter carbinolicus</i>	3,66	0,037	857 (611,246)	20 (13,7)	29	20	4(1,3) + 8Fis (5,3)	2 (2)
	<i>Geobacter sulfurreducens</i>	3,81	0,029	587 (451,136)	22 (11,11)	44	36	17 (7,10)	1 (2)

Продолжение табл. 1

1	2	3	4	5	6	7	8	9	10
Actino- bacteria	<i>Corynebacterium glutamicum</i> ATCC 13032	3,31	0,038	560 (336,224)	29 (15,14)	27	20	10 (3,7)	16 (6)
	<i>Symbiobacterium thermophilum</i> IAM14863	3,57	0,080	1417 (1055,362)	83 (51,32)	52	49	20 (8,12)	21 (6)
Thermo- togae	<i>Thermotoga maritima</i> MSB8	1,86	0,02	336 (207,129)	2 (2,0)	12	11	0	0 (1)
Deinococcus— Thermus	<i>Thermus thermophilus</i> HB 8	1,85	0,068	502 (273,229)	10 (3,7)	10	7	3 (2,1)	2 (4)
	<i>Deinococcus radiodurans</i> R1	2,65	0,045	779 (404,375)	24 (10,14)	35	34	13 (7,6)	8 (3)
		0,41	0,013	43 (31,12)	0	3	2	0	0 (0)
Cyanobacteria	<i>Prochlorococcus marinus</i> 1375	1,75	0,006	99 (82,17)	0	1	0	—	0 (1)
	<i>Prochlorococcus marinus</i> MIT9313	2,41	0,030	489 (357,132)	10 (7,3)	6	0	—	1 (2)
	<i>Synechococcus</i> WH8102	2,43	0,030	461 (390,71)	12 (11,1)	19	0	—	3 (2)
	<i>Thermosynechococcus</i> BP-1	2,59	0,058	948 (623,325)	48 (34,14)	81	81	31 (19,13)	0 (1)
	<i>Synechococcus elongatus</i> PCC6301	2,70	0,013	145 (92,53)	2 (0,2)	5	3	0	1 (2)
	<i>Synechococcus elongatus</i> PCC7942	2,70	0,013	163 (104,59)	3 (1,2)	3	1	0	1 (2)
	<i>Synechocystis</i> sp. PCC6803	3,57	0,036	758 (378,380)	35 (19,16)	99	97	30 (15,15)	1 (2)
	<i>Gloeobacter violaceus</i> PCC7421	4,66	0,039	1206 (768,631)	24 (13,11)	63	60	13 (6,7)	0 (1)
	<i>Anabaena variabilis</i> ATCC29413	6,73	0,056	3566 (2114,1452)	46 (23,23)	70	55	29 (13,16) + 8Hyp	5 (4)
	<i>Nostoc</i> sp. PCC7120	6,41	0,049	2892 (1808,1084)	44 (21,23)	75	72	27 (12,15)	6 (4)
Plancto- mycetes	<i>Rhodopirellula baltica</i> SH1	7,15	0,034	1837 (1232,605)	51 (31,20)	75	59	32 (19,13)	0 (1)
Chloro- flexi	<i>Dehalococcoides ethenogenes</i> 195	1,47	0,095	42 (27,15)	10 (8,2)	24	5	0	0 (1)
Firmicutes	<i>Streptococcus thermophilus</i> CNRZ1066	1,80	0,058	505 (324,171)	24 (16,8)	102	95	13 (9,4)	7 (6)
	<i>Streptococcus thermophilus</i> LMG18311	1,80	0,062	570 (352,218)	24 (13,11)	97	91	15 (7,8)	8 (6)
	<i>Lactobacillus acidophilus</i> NCFM	1,99	0,032	289 (198,91)	19 (11,8)	33	20	8 (4,4)	11 (4)
	<i>Streptococcus pneumoniae</i> TIGR4	2,14	0,087	1520 (855,665)	28 (13,15)	40	33	7 (3,4)	4 (4)
	<i>Lactococcus lactis</i> IL1403	2,37	0,064	610 (301,309)	45 (31,14)	59	50	35 (25,10)	10 (6)
	<i>Thermoanaerobacter tengcongensis</i> MB4	2,69	0,101	1585 (1073,512)	81 (53,27)	63	49	41 (21,20)	4 (4)
	<i>Oceanobacillus iheyensis</i> HTE831	3,63	0,031	701 (508,193)	23 (18,5)	14	11	5 (3,2)	15 (7)

Окончание табл. 1

1	2	3	4	5	6	7	8	9	10
	<i>Bacillus halodurans</i> C-125	4,20	0,053	853 (597,254)	102 (72,30)	112	106	74 (42,32)	24 (8)
	<i>Bacillus subtilis</i> 168	4,21	0,022	411 (304,107)	31 (30,1)	3	0	—	31 (10)
	<i>Bacillus cereus</i> ATCC10987	5,22	0,036	1270 (943,327)	36 (32,4)	43	23	2 (1,1)	33 (12)
	<i>Bacillus cereus</i> ATCC14579	5,41	0,040	1269 (928,341)	41 (36,5)	41	25	9 (7,2)	32 (13)

Примечание. * — П — прямые повторы, И — инвертированные повторы; ** — данные для мегаплазмиды; *** — в скобках приведено число кластеров р-РНК в геноме.

Кроме повторов, относящихся к генам р-РНК, в геномах многих бактерий встречаются дубликации или множественные повторы, связанные как с межгенными участками, так и с генами, кодирующими специфические белки. Количество таких повторов для разных геномов варьирует от 1 (*Aeropyrum*, *Zygomonas* и др.) до 36 (*Thermoanaerobacter*). В табл. 2 приведены сведения о протяженных повторах, связанных с некоторыми генами фототрофных бактерий. В их число входят гены, кодирующие белки фотосистем, фикобилисом, систем хемотаксиса и связывания ионов металлов. Увеличение числа копий таких генов, очевидно, связано с обеспечением светозависимых и других существенных ме-

таболических процессов, определяющих адаптивную устойчивость и экологическую приспособленность организмов. Примером такого протяженного повтора (1254 пн) является дубликация кластера светоиндуцибельных *hli*-генов в геноме светоустойчивого штамма *Prochlorococcus marinus* MED4 (Михеева и др., 2005). Среди всех изученных нами геномов максимальный по протяженности совершенный нуклеотидный повтор, не связанный с кластерами генов р-РНК, обнаруживается в хромосоме *Dehalococcoides* (21,1 тпн — блок из 21 гена). Протяженные дубликации находятся также в мегаплазмиде *Rhialstonia* (30,2 тпн — блок из 30 генов) и в одной из плазмид *Synechocystis* (23,6 тпн — блок из 27 генов).

Таблица 2

Повторяющиеся последовательности (500пн), связанные с генами, кодирующими белки у фототрофных бактерий

Бактерия	Повторяющиеся гены	Функции повторяющихся генов	Количество/размер повторов (пн)
<i>Prochlorococcus marinus</i> MIT9313	0959, 0960, 0961 // 0956, 0957, 0958	Кинезин, белок базального тела флагеллы, гипотетический белок	2/1936, 1429
	0982, 0983 // 0962, 0963	Кинезин, гипотетический белок	1/811
	<i>psbA-2</i> // <i>psbA-1</i>	D1 белок фотосистемы II	2/538, 613
<i>Synechococcus</i> WH8102	<i>psbD-2</i> // <i>psbD-1</i>	D2 белок фотосистемы II	1/1043
	<i>psbA-4</i> // <i>psbA-3</i>	Белок фотосистемы I	1/830
	Внутригенные повторы в гене 0953 (гомолог <i>vsb</i> -гена)	Экзобелок-сидерофор?	7/632, 734, 1485, 1910, 1365, 1385, 977
<i>Synechococcus elongatus</i> 6301	<i>psbD-II</i> // <i>psbD-I</i>	D2 белок фотосистемы II	1/910
<i>Synechococcus elongatus</i> PCC7942	1636 // 0655	D2 белок фотосистемы II	1/1042
	1051, 1052 // 1047, 1048	α и β -субъединицы фикоцианина	1/1093
<i>Chlorobium tepidum</i> TLS	<i>dskB-2</i> , <i>dsrA-2</i> , 2247 // <i>dskB-1</i> , <i>dsrA-1</i> , 0854	2 субъединицы сульфитредуктазы и Fe-S-связывающий белок	1/3403
<i>Chlorobium chlorochromatii</i>	0664 // 0657	Регулятор транскрипции	1/765
	Внутригенные повторы в генах 0738, 1242 (гомологи <i>vsb</i> -генов)	Экзобелки-сидерофоры?	2/663, 671
<i>Rhodospseudomonas palustris</i> CGA009	<i>tuf A-2</i> // <i>tuf A-1</i>	Фактор элонгации трансляции Tu	1/1191
	<i>che A-3</i> // <i>che A-2</i>	Гистидинпротеинкиназа хемотаксиса	2/851, 683
<i>Rhodobacter sphaeroides</i>	<i>tuf A-2</i> // <i>tuf A-1</i>	Фактор элонгации трансляции Tu	1/695

Третий, и самый многочисленный, тип протяженных повторов выявлен в генах, кодирующих ферменты транспозиции (транспозазы). На гены транспозаз (и гены близкие по гомологии, но нечетко аннотированные) приходится от 50 до 90% протяженных повторов от их общего количества у многих архей и бактерий (их видовые названия выделены в табл. 1 жирным шрифтом). Это явление, по-видимому, лежит в основе положительной корреляции между числом генов, кодирующих транспозазы, и суммарной протяженностью всех внутригеномных повторов для 80 проанализированных нами геномов свободноживущих прокариот (Михеева и др., 2006).

Гены транспозаз обнаруживаются в геномах архей и бактерий во всех систематических группах, но по числу генов, кодирующих ферменты транспозиции, между видами наблюдаются большие различия. Анализ наличия и распределения генов транспозаз у большого числа свободноживущих ви-

дов прокариот не выявил зависимости этих показателей от таксономической принадлежности: представители даже одного рода (например, *Pyrococcus*, *Sulfolobus*, *Bacillus*, *Pseudomonas*) могут существенно различаться между собой, хотя геномы разных штаммов одного вида (например, штаммы *Bacillus cereus*, *Streptococcus thermophilus*) обладают высокой степенью сходства по этому признаку. Прямой зависимости между количеством генов транспозаз и размерами геномов не выявлено, но отмечена тенденция к снижению числа генов транспозаз в редуцированных геномах (кроме *Thermoplasma volcanium*). Эти данные согласуются с тем, что количество генов транспозаз в геномах облигатных внутриклеточных паразитирующих бактерий значительно меньше, чем в геномах факультативных эндосимбионтов (Bordenstein, Reznikoff, 2005).

Гены, кодирующие транспозазы, в геномах прокариот находятся в составе простейших мигрирующих элементов (IS-элементов) с размером от

Таблица 3

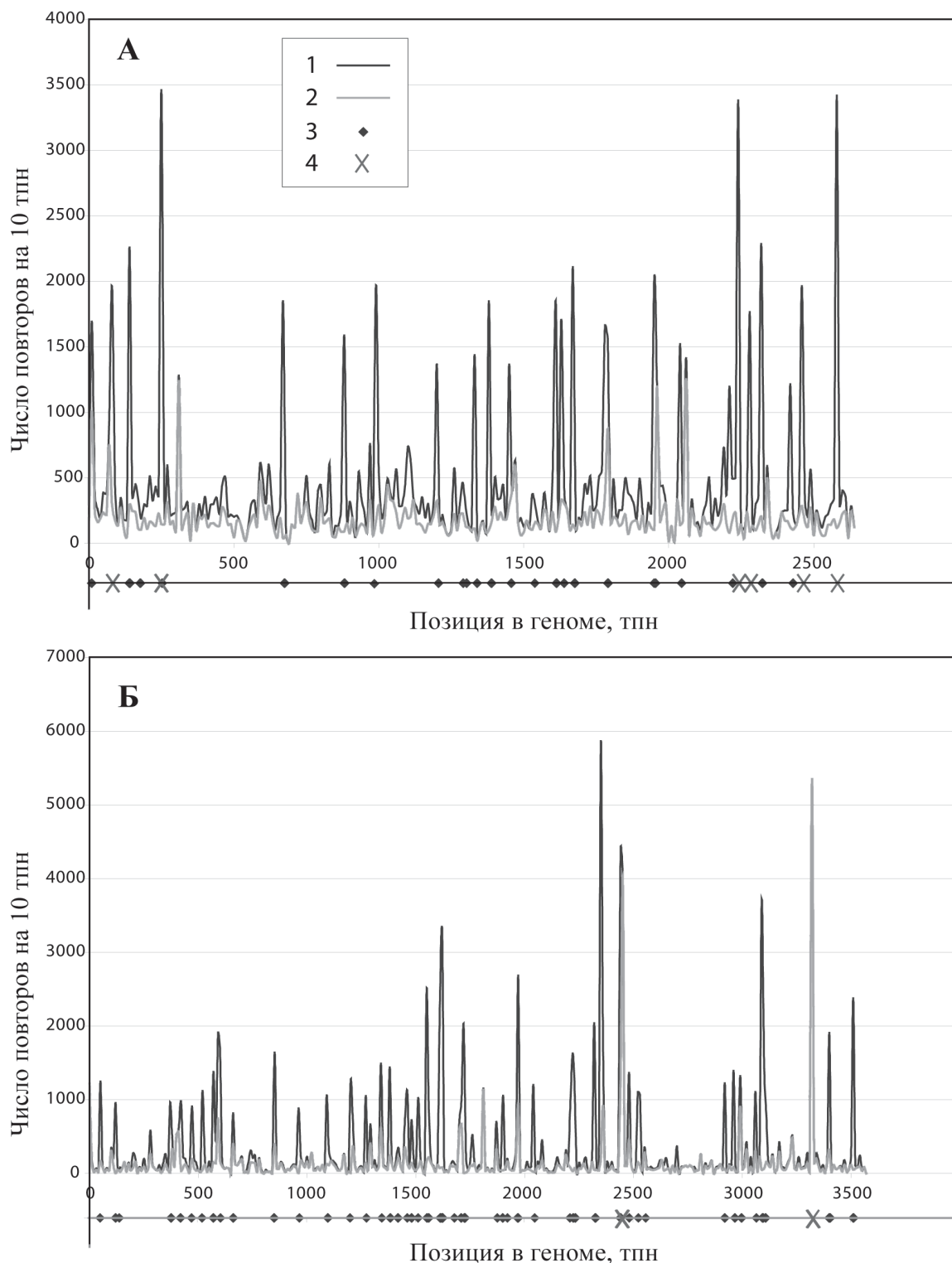
Протяженные нуклеотидные повторы, связанные с мультипликацией IS-элементов в геномах некоторых прокариот

Организм	Транспозазы			IS-элементы			Повторы 500 пн	
	номер по COG*	длина кодирующей рамки**, пн	количество генов в геноме**	тип (семе-йство)***	длина, ** пн	количество копий в геноме**	длина, пн	количество
<i>Archaeoglobus fulgidus</i>	3316	(282, 432)	(2p)	IS6	(282, 432)	(2p)	—	—
	2801	816—900	7	IS481	963	7	545—780	3
	3335 + 3415	549 + 486	6	IS630	1081—1083	3	1083	1
	5558	144 + 969	12	ISNCY1	1211—1214	6	1214	4
<i>Deinococcus radiodurans</i> , (хромосома 1)	3385	1202—1250	4	IS4	1322	4	1323	3
	Неаннотировано	866, 983	2		1206—1240	2		
	3293	812	1	IS5	910	1	—	—
	1943 + 0675	423 + 1226	16	IS605	1730—1742	8	1743	6
	3415, 3335	485, 518	5	IS630	971—1070	5	915	2
	3385	707—770	2				834	1
<i>Synechocystis</i> sp. PCC6803	1943	437	1				—	—
	3677 + 1662	425 + 416	6 + (1p)	IS1	801—802	3 + (1p)	—	—
	2801	(659)	(1p)	IS3	(659)	(1p)	—	—
	3385	1016	6 + (8p)	IS4	1174	6 + (8p)	645—1175	9
	5659	1076, 1019	2		1416, 1267	2		
	3293	779—785	9 + (6p)	IS5	865—871	9 + (6p)	745—878	3
	3293 + 3293	428 + 386/323	4 + (8p)		907—908	2 + (4p)		
	3328	(197)	(1p)	IS256	(197)	(1p)	—	—
	0675	(1202)	(1p)	IS602	(1202)	(1p)	—	—
	3335	848	9 + (10p)	IS630	946—947	9 + (10p)	575—955	18
	3335 + 3415	494 + 359	20 + (4p)			10 + (2p)		

Примечание. * — COG — классификация ортологических генов (по данным GenBank); ** — цифра в скобках обозначает длину или количество редуцированных копий (p) генов транспозаз или IS-элементов, знак “+” в столбцах 2 и 3 обозначает, что IS-элементы содержат две рамки считывания; *** — по Singer et al., 2006.

0,75 тпн до 2,5 тпн или локализованы внутри транспозонов, более характерных для представителей патогенных бактерий. В геномах, проанализированных в данной работе, единичные транспозоны обнаружены у *Nitrosomonas* (транспозон Tn3-семейства, содержащий гены, контролирующие устойчи-

вость к ртути (<http://cmr.tigr.org/CMR/Search.shtml>), *Oceanobacillus* (транспозон Tn3-семейства, содержащий кластер генов, контролирующих прорастание спор (Takaki et al., 2004)), у фитопатогена *Rhizoctonia* (аналог Tn4371, с генами, контролирующими процесс конъюгации (Salanoubat et al., 2002)).



Распределение внутригеномных повторов и IS-элементов на молекулярно-генетических картах:

А — хромосома 1 *Deinococcus radiodurans* R1; Б — хромосома *Synechocystis* sp. PCC6803. 1 — число позиций, занятых повторами, на 10 тпн, 2 — число позиций, занятых одноразовыми повторами, на 10 тпн, 3 — положение IS-элементов в геномах, 4 — положение кластеров генов р-РНК

Данные, приведенные в табл. 3, показывают, что максимальные размеры протяженных повторов, содержащих гены транспозаз, совпадают с размерами интактных IS-элементов, которые могут быть представлены в геномах прокариот многочисленными гомологичными копиями. На рисунке сопоставлено распределение неслучайных совершенных повторов в геномах и распределение IS-элементов вдоль генных карт хромосом *Deinococcus radiodurans* R1 и *Synechocystis* sp. PCC6803. Совпадение большинства максимумов повторяющихся последовательностей с положением IS-элементов подтверждает вывод о том, что именно мультипликация мобильных IS-элементов вносит наибольший вклад во внутригеномные повторы у многих прокариот.

В эту же категорию протяженных повторов можно включить интроны II группы, кодирующие каталитические РНК, функционирующие как мобильные элементы, участвующие в транспозиции и инсерции ДНК-копий (Belfort et al., 2002). Эти гены представлены в ряде геномов несколькими копиями (*Bacillus halodurans* — 5 генов, *Oceanobacillus* — 5, *Symbiobacterium* — 26, *Pseudomonas putida* — 8, *Thermosynechococcus* — 9, *Pelobacter* — 5, *Dechloromonas* — 3 и др.) и, следовательно, тоже вносят существенный вклад в общее число протяженных внутригеномных повторов.

Таким образом, значительная доля протяженных внутригеномных повторов у прокариот обусловлена умножением числа копий IS-элементов или интронов II группы в результате репликативной транспозиции самих мобильных элементов. Такие повторы можно рассматривать как особый тип повторяющихся последовательностей в отношении их происхождения и возможных функций в геномных перестройках (Achaz et al., 2002).

Следует отметить, что гены, кодирующие другие виды ферментов сайт-специфической рекомбинации, в частности различного типа интегразы, как правило, не обнаруживаются в составе протяженных повторов.

Проведенный анализ показал, что протяженные нуклеотидные повторы в геномах свободноживущих прокариот можно подразделить на три основные группы и наиболее многочисленными являются повторы генетических элементов, способных к транспозиции. По современным представлениям, мобильные элементы генома прокариот не являются “паразитической ДНК”, распространяющейся между различными организмами в результате горизонтального переноса. Такие элементы мо-

гут быть источником “благоприятных” адаптивных мутационных и рекомбинационных изменений генома, увеличивающих генетическое разнообразие популяций как у прокариот, так и у дрожжей (Schneider, Lenski, 2004). Концевые повторы в IS-элементах и интронах II группы могут участвовать в регуляции транскрипции близко расположенных генов (Bonen, Vogel, 2000; Belfort et al., 2002), и, следовательно, наличие внутригеномных повторов является предпосылкой для рекомбинационной реорганизации геномов, в результате которой активируется или репрессируется действие генов, ответственных за фенотипическое проявление признаков, имеющих адаптивное значение. Производными IS-элементов являются также более короткие повторяющиеся межгенные последовательности, такие как RUP-элементы в геномах штаммов *Streptococcus pneumoniae*, которые, вероятно, остаются способными к мобилизации родственными транспозазами в транспозиции (Oggioni, Claverys, 1999).

Транспозиции мобильных элементов и связанные с ними мутации и внутригеномные перестройки, приводящие к изменениям в регуляции генной активности, способны повышать степень пластичности прокариотического генома, ускорять процессы адаптации к изменяющимся условиям внешней среды и, соответственно, способствовать дивергенции, обеспечивающей видообразование (Frost et al., 2005). Результаты анализа повторяющихся последовательностей в геномах прокариот свидетельствуют об очевидной тенденции к увеличению фракции геномной повторяющейся ДНК за счет “самореплицирующихся” мигрирующих элементов генома. Можно предположить, что этот механизм, способствующий процессам усложнения и дивергенции геномов, был развит и эволюционно закреплен в процессе становления эукариотического генома. Сведения о количестве и локализации повторов, связанных с мобильными элементами и системами транспозиции, могут служить показателем рекомбинационного потенциала, определяющего внутригеномную мобильность и способность клеток к интеграции “чужеродных” генов, привносимых через горизонтальный перенос, являющийся одним из ключевых факторов эволюции.

* * *

Исследования поддержаны грантами программы “Ведущие научные школы” НШ-4202.2006.4 и программы Президиума РАН “Происхождение и эволюция биосферы”.

СПИСОК ЛИТЕРАТУРЫ

Михеева Л.Е., Карбышева Е.А., Шестаков С.В. 2006. Эволюционная геномика прокариот: протяженные повторяющиеся последовательности, связанные с IS-элементами и генами, кодирующими транс-

позазы // Открытое образование. № 3 (Мат-лы междунар. науч. конф. “Новые информационные технологии в медицине, биологии, фармакологии и экологии”). М. С. 28—30.

Михеева Л.Е., Орлов Ю.Л., Колчанов Н.А., Шестаков С.В. 2005. Филогенетический анализ совершенных повторов в геномах одноклеточных цианобактерий // Экологическая генетика. **3**. 15–21.

Орлов Ю.Л. 2004. Анализ повторов в полных бактериальных геномах // Тр. III съезда ВОГиС “Генетика в XXI веке: современное состояние и перспективы развития”. Москва, 6–12 июня 2004. М. Т. 2. Р. 364.

Турова Т.П. 2003. Мультикопийность рибосомных оперонов прокариот и ее влияние на проведение филогенетического анализа // Микробиология. **72**. № 1. 437–452.

Achaz G., Rocha E.P., Netter P., Coissac E. 2002. Origin and fate of repeats in bacteria // Nucleic Acid Research. **30**. 2987–2994.

Belfort M., Derbyshire V., Cousineau B., Lambowitz A. 2002. Mobile introns: pathway and proteins // Mobile DNA II / Eds. N. Craig, R. Craigie, M. Gellert, A. Lambowitz. Washington. P. 761–783.

Bhaya D., Dufresne A., Vaillot D., Grossman A. 2002. Analysis of the *hli* family in marine and freshwater cyanobacteria // FEMS Microbiol Lett. **215**. 209–219.

Bibby T.S., Mary I., Nield J., Partensky F., Barber J. 2003. Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem // Nature. **424**. 1051–1054.

Bonen L., Vogel J. 2001. The ins and outs of group II introns // Trends Genet. **17**. 322–331.

Bordenstein S.R., Reznikoff W.S. 2005. Mobile DNA in obligate intracellular bacteria // Nature Rev. (Microbiol.). **3**. 688–699.

Frost L.S., Leplae R., Summers A.O., Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution // Nat. Rev. Microbiol. **3**. 722–732.

Henz S.R., Huson D.H., Auch A.F., Nieselt-Struwe K., Schuster S.C. 2005.

Whole-genome prokaryotic phylogeny // Bioinformatics. **21**. 2329–2335.

Mahillon J., Leonard C., Chandler M. 1999. IS elements as constituents of bacterial genomes // Res. Microbiol. **150**. 675–687.

Oggioni M., Claverys J. 1999. Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae* // Microbiol. **145**. 2647–2653.

Orlov Y.L., Gusev V.D., Miroshnichenko L.A. 2003. LZcomposer: Decomposition of genomic sequences by repeat fragments // Biofizika (Mosk). **48**. Suppl. 1. 7–16.

Orlov Y.L., Potapov V.N., Poplavsky A.S. 2004. Computer analysis of genomic sequence complexity: new applications // Proceedings of BGRS'2004. Vol. 1. Novosibirsk. P. 153–157.

Rogozin I.B., Makarova K.S., Natale D.A., Spiridonov A.N., Tatusov R.L., Wolf Y.I., Yin J., Koonin E.V. 2002. Congruent evolution of different classes of non-coding DNA in prokaryotic genomes // Nucleic Acids Res. **30**. 4264–4271.

Salanoubat M., Genin S., Artiguenave F. et al. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum* // Nature. **415** (6871). 497–502.

Schneider D., Lenski R.I. 2004. Dynamics insertion sequence elements during experimental evolution of bacteria // Res. Microbiol. **155**. 319–327.

Siguiet P., Perochon J., Lestrade L., Mahillon J., Chandler M. 2006. ISfinder: the reference centre for bacterial insertions sequences // Nucleic Acid Research. **34**. 32–36.

Takaki Y., Matsuki A., Chee G.J., Takami H. 2004. Identification and distribution of new insertion sequences in the genome of the extremely halotolerant and alkaliphilic *Oceanobacillus iheyensis* HTE831 // DNA Res. **11**. 233–245.

ON THE NATURE OF LENGTHY NUCLEOTIDE REPEATED SEQUENCES IN GENOMES OF PROKARYOTIC ORGANISMS

L.E. Mikheeva, E.A. Karbysheva, S.V. Shestakov

The whole-genomic analysis *in silico* of 64 free-living prokaryotic species has been performed to determine the number, length, distribution and location of direct and inverted intragenomic repeated sequences (LRS). Three main types of lengthy (> 500 bp) repeated sequences were revealed: a) associated with ribosomal RNA genes; b) with copies of protein coding genes; c) with IS-elements and genes encoding putative transposases. Lengthy repeated sequences related to transposases comprise 50–95% of total number of LRS depending on species. Intragenomic LRS associated with transposases and IS-elements can reflect a recombinational potential of different prokaryotic species determining both the ability to adaptive gene rearrangements and cell capacity for integration of genes acquired through horizontal transfer ways.