

УДК: 579:252

## **Вариабельность транскрипционного ландшафта межгенной области *oppA-oppB* у бактерий по данным компьютерного предсказания потенциальных стартов синтеза РНК**

**Сухаричева Н.А.<sup>1,2</sup>, Киселев С.С.<sup>1</sup>, Озолин О.Н.<sup>1,2</sup>,  
Масулис И.С.<sup>1,2</sup>**

<sup>1</sup>*Институт биофизики клетки, Российская академия наук, Пущино, Московская область, 142290, Российская Федерация*

<sup>2</sup>*Пуцинский государственный естественно-научный институт, Пущино, Московская область, 142290, Российская Федерация*

**Аннотация.** Трансмембранные белки OppB и OppC, составляя остов бактериальных ABC-транспортёров олигопептидов, остаются совершенно не исследованными с точки зрения регуляции экспрессии кодирующих их генов. С помощью универсального алгоритма поиска потенциальных стартов синтеза РНК (PlatPromU) для геномов 13 микроорганизмов из различных таксономических групп исследована возможность транскрипции *oppB*, инициируемой на промоторах межгенной области *oppA-oppB*. Выявлены сходные паттерны распределения вероятных сайтов инициации транскрипции у энтеробактерий и необычная по насыщенности потенциальными промоторами межгенная область у *Bifidobacterium dentium*. Филогенетический анализ и поиск мотивов, гомологичных нуклеотидному контексту *промоторного островка* свидетельствуют о том, что часть его была привнесена в регуляторную область *oppB Bifidobacterium dentium* из кодирующей области гена *GTNG\_2042* эволюционно удалённого вида *Geobacillus thermodenitrificans*. Эта интеграция могла спровоцировать формирование протяжённого *промоторного островка*, которое в данном случае было направлено не на ассимиляцию нового гена, а на оптимизацию экспрессии собственного гена *oppB Bifidobacterium dentium*.

**Ключевые слова:** бактериальные геномы, ABC-транспортёры, *opp*-опероны, регуляция экспрессии внутренних генов оперонов, промоторы, *промоторные островки*, алгоритмы поиска промоторов

### **ВВЕДЕНИЕ**

АТФ-зависимые транспортные системы осуществляют как импорт, так и экспорт широкого спектра компонентов в прокариотических и эукариотических клетках [1, 2]. У микроорганизмов ABC-транспортёрами (АТФ-Binding Cassette transporters) опосредован активный транспорт сахаров и аминокислот [3], неорганических катионов [4], витаминов [5]. Обеспечивая поступление необходимых бактериям соединений в клетку и поддержание сбалансированного метаболизма, ABC-транспортёры также вносят вклад в развитие вирулентности патогенных микроорганизмов [6, 7].

Две системы, гены которых экспрессируются в составе оперонов *dppABCDF* и *oppABCDF*, специализированы на транспорте олигопептидов. Оперон *oppABCDF* кодирует компоненты мультисубъединичной олигопептидпермеазы (Opp),

интегрированной в клеточную мембрану (рис. 1). Его ортологи имеются как у грамположительных, так и у грамотрицательных бактерий и по современной классификации мембранных транспортных систем относятся к семейству PerT 3.A.1.5 (Peptide/Opine/Nickel Uptake Transporter) [8, 9]. Белки Opp отвечают за поступление в бактериальную клетку от ди- до пентапептидов, которые используются в качестве источников аминокислот и углерода [10].

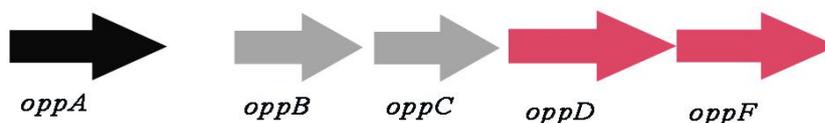


Рис. 1. Схема, иллюстрирующая типичную генетическую организацию оперона *oppABCD*.

Для транспортеров Opp характерна общая схема организации, где OppA – периплазматический субстрат-связывающий компонент, OppD и OppF – АТФ-связывающие белки, а OppB и OppC – субъединицы, формирующие трансмембранный канал [11]. Мутации, инактивирующие *oppA*, влияют на способность бактерий образовывать биоплёнки [12] и вызывают повышенную устойчивость к аминогликозидным антибиотикам [13]. Это свидетельствует об участии транспортной системы Opp в разных физиологических процессах.

Типичная организация оперона *oppABCD* показана на рис. 1, но у грамположительных микроорганизмов, в том числе у стафилококков, обнаружены множественные вариации в порядке расположения генов, возникающие при дупликации этого оперона в геноме. Однако во всех случаях присутствует оперон с последовательным расположением: *oppA*→*oppB*→*oppC*→*oppD*→*oppF* [14]. В геномах многих бактерий гены функционально связанных субъединиц *oppB* и *oppC* (образуют канал в мембране), *oppD* и *oppF* (АТФазы), а также их паралоги сохраняют сцепленность. Молекулярно-филогенетический анализ периплазматических субстрат-связывающих белков OppA и DppA выявляет их явную ко-кластеризацию, что указывает на раннее обособление этих белков, определяющих специфичность ABC-транспортеров к субстратам пептидной природы [15]. У некоторых организмов возможна дупликация *oppA* и *dppA*, приводящая к образованию нескольких паралогов, как это показано для *Borrelia burgdorferi* [16] и *Pseudomonas aeruginosa* [17], хотя гены трансмембранных и АТФ-связывающих белков представлены у этих микроорганизмов единичными копиями. Наличие изоформ первого компонента (OppA или DppA) и устойчивая сцепленность блоков *oppB*→*oppC* и *oppD*→*oppF* допускают возможность независимой регуляции транскрипционной активности, по крайней мере, для *oppB* и следующих за ним генов.

Несмотря на высокую значимость транспорта олигопептидов, особенности транскрипционной регуляции исследованы только для первого гена оперона, *oppA*, исходя из предположения, что оперон транскрибируется в виде единой полицистронной мРНК. Для *Escherichia coli* (*E. coli*) в базе данных RegulonDB [18] на основании компьютерных предсказаний указаны множественные старты инициации транскрипции, включающие 5 кластеров близкорасположенных точек, 5 одиночных стартов для РНК-полимеразы с неизвестным сигма-фактором и одиночный промотор, узнаваемый  $\sigma^{28}$ -РНК-полимеразой [19]. Однако в 3'-нетранслируемой области гена *oppA* предсказывается rho-независимый терминатор транскрипции, который может останавливать синтез РНК, начатый на промоторе, расположенном перед геном *oppA*. Это вызвало интерес к поиску потенциальных промоторов, способных обеспечивать независимую экспрессию гена *oppB* и следующих за ним генов. Поэтому в данной работе с помощью компьютерного поиска был исследован характер распределения

потенциальных точек инициации транскрипции в межгенном участке *oppA-oppB* *E. coli* и проанализирована генетическая организация первых трёх генов оперона *oppABCD* для 59 бактерий других видов. Для 13 геномов с аналогичным порядком расположения генов в опероне, в межгенной области *oppA-oppB* были предсказаны промоторы. При этом у *Bifidobacterium dentium* (*B. dentium*) в межгенном участке была выявлена область с аномально высокой плотностью распределения промотор-подобных сигналов, так называемый *промоторный островок* [20, 21]. Участок, гомологичный этой промоторной области, был обнаружен только в геноме *Geobacillus thermodenitrificans* (*G. thermodenitrificans*). В работе показано, что накопление избыточных сигналов инициации транскрипции не соответствует предположению о постепенном формировании *островка*, свидетельствуя в пользу возможности горизонтального переноса генетического материала из *G. thermodenitrificans* в межгенную область *oppA-oppB* генома *B. dentium*.

## МЕТОДЫ И АЛГОРИТМЫ

### Формирование выборки исследованных геномов

Для анализа генетической организации оперона *oppABCD* использованы геномы 60 видов бактерий (таблица 1). Для 59 из них аннотирован только один оперон с последовательным расположением генов *oppA*→*oppB*→*oppC*→*oppD*→*oppF*. У *Corynebacterium pseudotuberculosis* (*C. pseudotuberculosis*) имеется 6 копий *opp*-оперонов, в каждом из которых были конъюгированные гены (*oppC-oppD* или *oppD-oppF*). Для работы использовали один из трёх оперонов, слитые гены *oppD* и *oppF* которого имели максимальную степень гомологии с ортологами *E. coli*. При составлении выборки не исключали геномы с *opp*-оперонами, имеющими дополнительную копию гена *oppA* на 5'-конце. Спектр представленных видов охватывает 30 семейств из разных таксономических групп эубактерий (альфа-протеобактерии, бета-протеобактерии, гамма-протеобактерии, бациллы, кластридии, актинобактерии, цианобактерии, спирохеты, планктомицеты) и одного представителя архей (*Thermofilum pendens*).

Для *E. coli* K12 использовали геном версии NC\_000913.2, а соответствующая генная карта была взята из базы данных RegulonDB 8.0 [19]. Для всех остальных микроорганизмов нуклеотидные последовательности геномов и соответствующие генные карты были взяты из NCBI (номера доступа указаны в таблице 1) [22].

### Анализ транскрипционного потенциала межгенной области *oppA-oppB*

Для предсказания потенциальных стартов инициации транскрипции с помощью универсального алгоритма поиска промоторов PlatPromU [21, 23] использованы нуклеотидные последовательности геномов *Rhizobium etli* (*Rh. etli*), *Bacillus subtilis* (*B. subtilis*), *Yersinia pseudotuberculosis* (*Y. pseudotuberculosis*), *Clostridium botulinum* (*Cl. botulinum*), *Salmonella enterica* (*S. enterica*), *G. thermodenitrificans*, *E. coli*, *Chlamydomonas reinhardtii* (*Chl. reinhardtii*), *Shigella flexneri* (*Sh. flexneri*), *B. dentium*, *Corynebacterium glutamicum* (*C. glutamicum*), *Corynebacterium pseudotuberculosis* (*C. pseudotuberculosis*) и *Propionibacterium acnes* (*P. acnes*). Пороговые уровни для определения статистически значимых величин показателей промотор-подобия рассчитывали так, как предложено в работе [23]. Приводятся распределения потенциальных точек инициации транскрипции с показателем промотор-подобия, превышающим фоновый уровень на величину трёх стандартных отклонений ( $p$ -value < 0.00139).

## Филогенетический анализ

Филогенетический анализ был выполнен для генов *oppV* и 16S рРНК 60 микроорганизмов, перечисленных в таблице 1. Гены 16S рРНК были выбраны в качестве стандартных филогенетических маркеров [24]. Сравнимые нуклеотидные последовательности выравнивали с использованием алгоритма ClustalW. Для построения филогенетических деревьев в программе MEGA6 [25] было использовано несколько методов: максимального правдоподобия (maximum likelihood), присоединения ближайших соседей (neighbour-joining), минимальной эволюции (minimum evolution), попарной невзвешенной группировки с арифметическим усреднением (UPGMA). В работе показаны результаты, полученные методом максимального правдоподобия с использованием модели замен (substitution model) Тамура-Неи [26]. Поскольку все геномы имеют множественные копии генов 16S рРНК, в набор анализируемых последовательностей включали паралоги максимальной длины.

**Таблица 1.** Список микроорганизмов, гены которых были использованы для анализа протяженности межгеновой области *oppA-oppV* и построения филогенетических деревьев

	Название организма	Идентификатор GenBank	Класс	Семейство	GC-состав (%)	Наличие повторов оперона <i>oppABCDF</i> в геноме организмов
1	<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	NC_009348.1	Gamma proteobacteria	Aeromonadaceae	58.17	–
2	<i>Agrobacterium radiobacter</i> K84	NC_011985.1 (chromosome 1)	Alpha proteobacteria	Rhizobiaceae	59.87	–
3	<i>Aliivibrio salmonicida</i> LFI1238	NC_011312.1	Gamma proteobacteria	Vibrionaceae	38.94	–
4	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168*	NC_000964.3	Bacilli	Bacillaceae	43.50	–
5	<i>Bifidobacterium dentium</i> Bd1	NC_013714.1	Actinobacteria	Bifidobacteriaceae	58.50	–
6	<i>Brucella pinnipedialis</i> B2/94	NC_015857.1	Alpha proteobacteria	Brucellaceae	57.2	–
7	<i>Burkholderia pseudomallei</i> K96243	NC_006351.1 (chromosome 2)	Beta proteobacteria	Burkholderiaceae	68.50	–
8	<i>Chlamydia trachomatis</i> D-LC	NC_017436.1	Chlamydiae	Chlamydiaceae	41.26	–
9	<i>Chlamydophila felis</i> Fe/C-56	NC_007899.1	Chlamydiae	Chlamydiaceae	39.36	–
10	<i>Citrobacter rodentium</i> ICC168	NC_013716.1	Gamma proteobacteria	Enterobacteriaceae	54.53	–
11	<i>Clostridium botulinum</i> A str. ATCC 3502	NC_009495.1	Clostridia	Clostridiaceae	28.19	–
12	<i>Corynebacterium glutamicum</i> ATCC 13032	NC_003450.3	Actinobacteria	Corynebacteriaceae	53.80	–
13	<i>Corynebacterium pseudotuberculosis</i> 3/99-5	NC_016781.1	Actinobacteria	Corynebacteriaceae	52.20	+
14	<i>Coxiella burnetii</i> CbuK Q154	NC_011528.1	Gamma proteobacteria	Coxiellaceae	42.64	–
15	<i>Cronobacter turicensis</i> z3032	NC_013282.2	Gamma proteobacteria	Enterobacteriaceae	57.23	–
16	<i>Dickeya dadantii</i> 3937	NC_014500.1	Gamma proteobacteria	Enterobacteriaceae	56.30	–
17	<i>Edwardsiella tarda</i> EIB202	NC_013508.1	Gamma proteobacteria	Enterobacteriaceae	59.67	–
18	<i>Enterobacter cloacae</i> EcWSU1	NC_016514.1	Gamma proteobacteria	Enterobacteriaceae	54.53	–
19	<i>Enterococcus faecalis</i> OG1RF	NC_017316.1	Bacilli	Enterococcaceae	37.80	–
20	<i>Erwinia amylovora</i> ATCC 49946	NC_013971.1	Gamma proteobacteria	Enterobacteriaceae	53.61	–
21	<i>Escherichia coli</i> str. K-12 substr. MG1655	NC_000913.2	Gamma proteobacteria	Enterobacteriaceae	50.80	–
22	<i>Exiguobacterium antarcticum</i> B7	NC_018665.1	Bacilli	Bacillales Family XII. Incertae Sedis	47.50	–
23	<i>Francisella tularensis</i> subsp. <i>novicida</i> U112	NC_008601.1	Gamma proteobacteria	Francisellaceae	32.50	–
24	<i>Gallibacterium anatis</i> UMN179	NC_015460.1	Gamma proteobacteria	Pasteurellaceae	39.89	–
25	<i>Geobacillus thermodenitrificans</i> NG80-2	NC_009328.1	Bacilli	Bacillaceae	48.85	–
26	<i>Gloeobacter kilauensis</i> JS1	NC_022600.1	Cyanobacteria	Gloeobacteraceae	60.50	–
27	<i>Haemophilus influenzae</i> Rd KW20	NC_000907.1	Gamma proteobacteria	Pasteurellaceae	38.20	–

	Название организма	Идентификатор GenBank	Класс	Семейство	GC-состав (%)	Наличие повторов оперона <i>orrABCD</i> в геноме организмов
28	<i>Janthinobacterium sp.</i> Marseille	NC_009659.1	Betaproteobacteria	Oxalobacteraceae	54.20	–
29	<i>Klebsiella pneumoniae</i> 342	NC_011283.1	Gammaproteobacteria	Enterobacteriaceae	56.88	–
30	<i>Lactobacillus plantarum</i> subsp. <i>plantarum</i> ST-III	NC_014554.1	Bacilli	Lactobacillaceae	44.50	–
31	<i>Leuconostoc citreum</i> KM20	NC_010471.1	Bacilli	Leuconostocaceae	38.88	–
32	<i>Lysinibacillus sphaericus</i> C3-41	NC_010382.1	Bacilli	Bacillaceae	37.15	–
33	<i>Oceanimonas sp.</i> GK1	NC_016745.1	Gammaproteobacteria	Aeromonadaceae	61.08	–
34	<i>Pantoea ananatis</i> LMG20103	NC_013956.2	Gammaproteobacteria	Enterobacteriaceae	53.70	–
35	<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. HN06	NC_017027.1	Gammaproteobacteria	Pasteurellaceae	40.22	–
36	<i>Pectobacterium atrosepticum</i> SCRI1043	NC_004547.2	Gammaproteobacteria	Enterobacteriaceae	51.00	–
37	<i>Pediococcus claussenii</i> ATCC BAA-344	NC_016605.1	Bacilli	Lactobacillaceae	37.00	–
38	<b><i>Propionibacterium acnes</i> 266</b>	NC_017534.1	Actinobacteria	Propionibacteriaceae	60.00	–
39	<i>Proteus mirabilis</i> HI4320	NC_010554.1	Gammaproteobacteria	Enterobacteriaceae	38.88	–
40	<i>Pseudovibrio sp.</i> FO BEG1	NC_016642.1	Alphaproteobacteria	Rhodobacteraceae	52.41	–
41	<i>Ralstonia solanacearum</i> FQY_4	NC_020799.1	Betaproteobacteria	Burkholderiaceae	66.79	–
42	<b><i>Rhizobium etli</i> CFN 42</b>	NC_007761.1	Alphaproteobacteria	Rhizobiaceae	61.05	–
43	<i>Rhodobacter capsulatus</i> SB 1003	NC_014034.1	Alphaproteobacteria	Rhodobacteraceae	66.60	–
44	<i>Rhodopirellula baltica</i> SH 1	NC_005027.1	Planctomycetes	Planctomycetaceae	55.40	–
45	<i>Rhodospseudomonas palustris</i> CGA009	NC_005296.1	Alphaproteobacteria	Bradyrhizobiaceae	64.99	–
46	<b><i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i> str. LT2</b>	NC_003197.1	Gammaproteobacteria	Enterobacteriaceae	52.22	–
47	<i>Serratia plymuthica</i> 4Rx13	NC_021591.1	Gammaproteobacteria	Enterobacteriaceae	56.09	–
48	<b><i>Shigella flexneri</i> 2a str. 301</b>	NC_004337.2	Gammaproteobacteria	Enterobacteriaceae	50.67	–
49	<i>Shimwellia blattae</i> DSM 4481 = NBRC 105725	NC_017910.1	Gammaproteobacteria	Enterobacteriaceae	56.60	–
50	<i>Sinorhizobium meliloti</i> 1021	NC_003078.1 (plasmid pSymB)	Alphaproteobacteria	Rhizobiaceae	62.16	–
51	<i>Spirochaeta thermophila</i> DSM 6192	NC_014484.1	Spirochaetes	Spirochaetaceae	61.90	–
52	<i>Streptococcus pyogenes</i> NZ131	NC_011375.1	Bacilli	Streptococcaceae	38.60	–
53	<i>Streptomyces sp.</i> PAMC26508	NC_021055.1	Actinobacteria	Streptomycetaceae	71.06	–
54	<i>Taylorella equigenitalis</i> 14/56	NC_021036.1	Betaproteobacteria	Alcaligenaceae	37.50	–
55	<i>Thermofilum pendens</i> Hrk 5	NC_008698.1	Thermoprotei	Thermofilaceae	57.68	–
56	<i>Thermovirga lienii</i> DSM 17291	NC_016148.1	Synergistetes	Synergistaceae	47.10	–
57	<i>Vibrio cholerae</i> O1 str. 2010EL-1786	NC_016445.1 (chromosome 1)	Gammaproteobacteria	Vibrionaceae	47.49	–
58	<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	NC_003919.1	Gammaproteobacteria	Xanthomonadaceae	64.74	–
59	<i>Xenorhabdus bovienii</i> SS-2004	NC_013892.1	Gammaproteobacteria	Enterobacteriaceae	45.00	–
60	<b><i>Yersinia pseudotuberculosis</i> IP 32953</b>	NC_006155.1	Gammaproteobacteria	Enterobacteriaceae	47.19	–

\*Жирным шрифтом выделены микроорганизмы с исследованным профилем распределения потенциальных сигналов транскрипции.

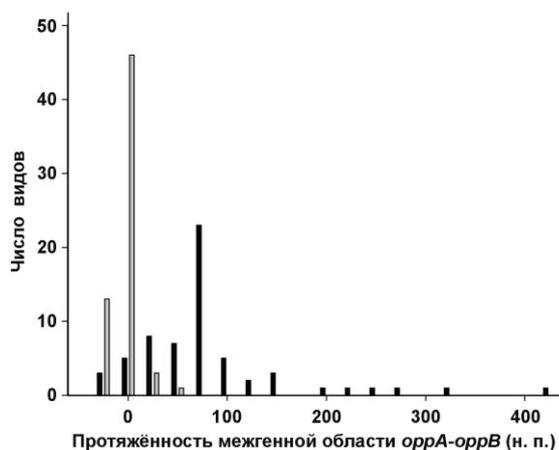
## РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Расстояние между генами *orrA* и *orrB* может быть достаточно для размещения функционального промотора

Для 60 видов микроорганизмов было проанализировано распределение межгенных расстояний для пар *orrA*→*orrB* и *orrB*→*orrC*. Анализировали только такие опероны,

порядок следования генов в которых соответствует рис. 1. Оказалось, что между генами *orrA* и *orrB* превалирует расстояние 75–80 нуклеотидных пар (н. п.), хотя в ряде случаев длина данного участка составляет более 100 н. п. (тёмные столбики на рис. 2). Гены *orrB* и *orrC* обычно расположены практически вплотную друг к другу или даже перекрываются, что соответствует отрицательным расстояниям на рис. 2 (светлые столбики).

Наличие протяженного интервала между *orrA* и *orrB* допускает физическую возможность формирования транскрипционного комплекса без затрагивания их кодирующих последовательностей. Так, РНК-полимераза при образовании иницирующего комплекса защищает от действия гидролизующих агентов участок ДНК длиной ~80 н. п., ограниченный, как правило, позициями  $-65 \div +20$  относительно стартовой точки транскрипции. При этом граница контакта с промотором в удалённой от старта (позиция +1) нетранскрибируемой области может варьировать от позиции  $-40$  до позиции  $-80$  [27]. Алгоритм PlatProm, адаптированный для поиска основных промоторов, распознаваемых  $\sigma^{70}$ -РНК-полимеразой, в геноме кишечной палочки не находит потенциальных промоторов между генами *orrA* и *orrB* [20].



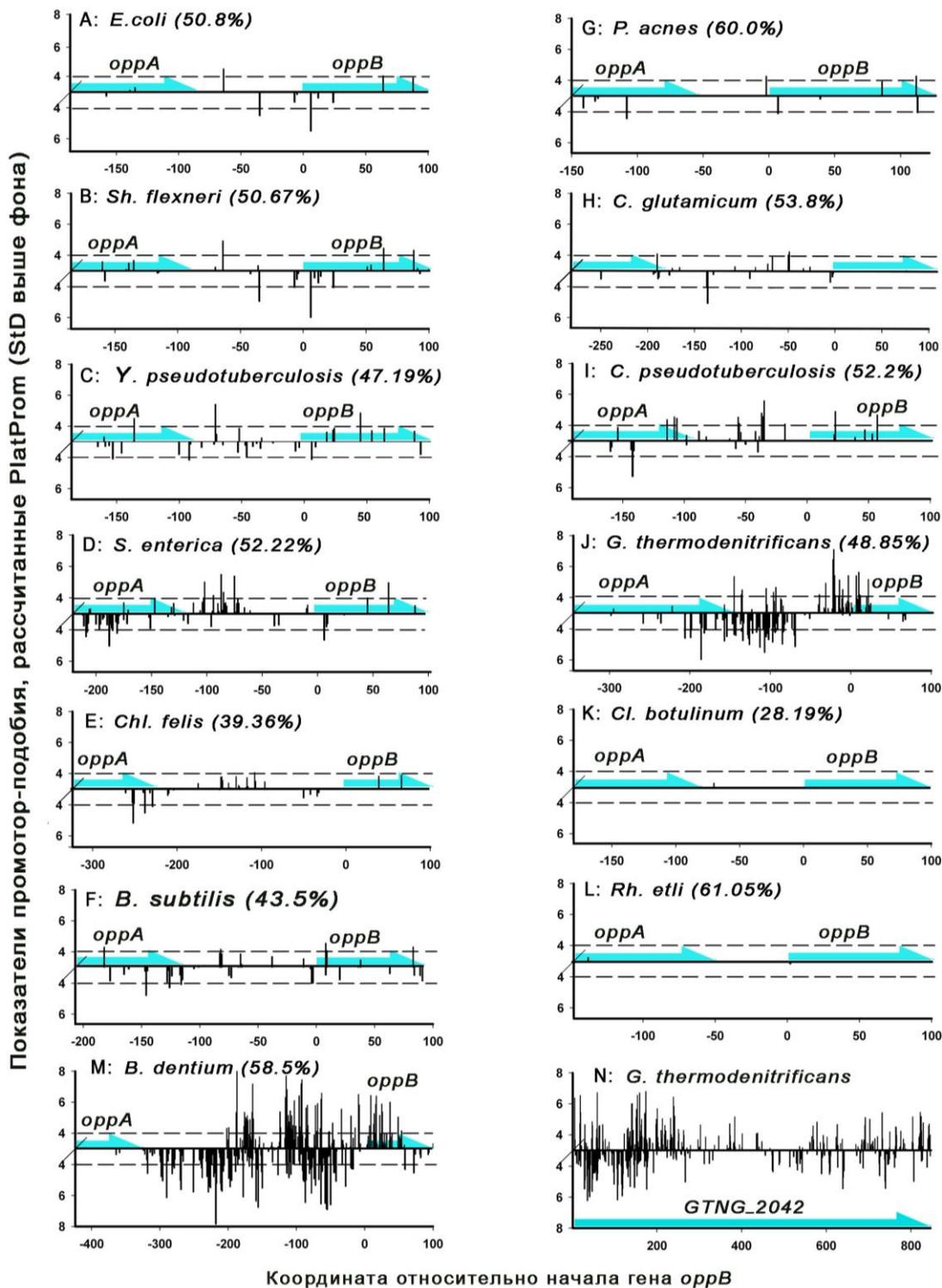
**Рис. 2.** Распределение длин межгенных участков *orrA-orrB* (тёмные столбики) и *orrB-orrC* (светлые столбики) для микроорганизмов, перечисленных в таблице 1.

Однако унифицированный алгоритм PlatPromU, учитывающий характеристики, инвариантные для промоторов разных  $\sigma$ -факторов, но игнорирующий контекст узнаваемых этими факторами консервативных элементов [23], предсказал стартовую точку для транскрипции *orrB* в позиции  $-64$  и два промотора для синтеза антисмысловых к ней РНК в позициях  $-35$  и  $+6$  (рис. 3А). То есть, в геноме *E. coli* предсказывается промотор, способный вносить дополнительный вклад в экспрессию генов *orrBCDF* и промоторы для антисмысловых РНК с потенциальной регуляторной функцией.

### Распределение потенциальных стартов транскрипции между генами *orrA* и *orrB* вариabelно даже у близкородственных видов

Ранее было установлено, что консервативность структурно-функциональной организации транскрипционного аппарата позволяет использовать PlatPromU для поиска промоторов и в других бактериальных геномах [23] при условии адекватного определения фонового уровня и величины стандартного отклонения (StD) для показателей промотор-подобия [23]. Для определения этих параметров были просканированы полные геномы 13 бактерий, у которых гены *orrA* и *orrB* разделены некодирующим участком 70–350 н. п. (выделены жирным шрифтом в таблице 1). У близкородственных *E. coli* энтеробактерий *Sh. flexneri* (рис. 3,В) и *Y. pseudotuberculosis*

(рис. 3,С), так же как у *S. enterica* (рис. 3,Д), дивергировавшей от них более 100 миллионов лет назад, в межгенной области *oppA-oppB* обнаружен похожий паттерн распределения сигналов инициации транскрипции.



**Рис. 3. А-М:** Локализация потенциальных стартов транскрипции, предсказанных алгоритмом PlatPromU в межгенной области *oppA-oppB* разных геномов (А-М) и вдоль всей кодирующей последовательности гена *GTNG\_2042* *G. thermodenitrificans* (N). Значения показателей промотор-подобия выражены в величинах стандартных отклонений, превышающих фоновый уровень. Условная координата «0» по оси абсцисс присвоена иницирующему кодону гена *oppB* (А-М) или *GTNG\_2042* (N). Пунктирными линиями на А-М показан уровень показателя промотор-подобия, обеспечивающий меньше одного ложноположительного сигнала на 12500 н. п. ( $p < 0.00004$ ).

Промоторы для встречного синтеза смысловой и антисмысловой РНК имеются также между генами *oppA-oppB* у *Chl. felis* (рис. 3,Е), *B. subtilis* (рис. 3,Ф) и *P. acnes* (рис. 3,Г), которые относятся к разным семействам. Но у *C. glutamicum* (рис. 3,И), *C. pseudotuberculosis* (рис. 3,И) и *G. thermodenitrificans* (рис. 3,Ж) смысловая транскрипция в направлении *oppB* не перекрывается с синтезом потенциального антисмыслового продукта, а у *Cl. botulinum* (рис. 3,К) и *Rh. etli* (рис. 3,Л) есть потенциальные промоторы только для одного направления. Наибольший интерес представляют *B. dentium* (рис. 3,М), у которого на протяжённом участке предсказываются множественные старты транскрипции. Можно было бы предположить, что мультипромоторные области в первую очередь обнаружатся у микроорганизмов с низким содержанием G/C-пар в геноме, так как компьютерный поиск промоторов в значительной степени ориентируется на АТ-богатые последовательности, необходимые для локального расплетания двойной спирали и активации промотора. Однако *B. dentium* следует скорее отнести к видам с повышенным содержанием G/C-пар (58.5%).

Обогащённые потенциальными промоторами области были выявлены ранее в 78 генетических локусах *E. coli* и были названы *промоторными островками* [20]. Оказалось, что 75 *промоторных островков* находятся рядом или входят в кодирующие области генов, полученных кишечной палочкой в процессе горизонтального переноса [28]. Наличие ассоциированных с ними множественных промотор-подобных мест рассматривается как следствие вынужденной ассимиляции новых генов и встраивания их в сложившиеся у данного вида регуляторные сети. Предполагается, что обладая парадоксально низким уровнем транскрипционной активности, *промоторные островки* сдерживают продукцию мРНК чужеродных генов, предоставляя при этом набор потенциальных промоторов для оптимизации уровня экспрессии ксеногенного генетического материала [28]. Поэтому было высказано предположение о возможности попадания гена *oppB* в геном *B. dentium* посредством горизонтального переноса.

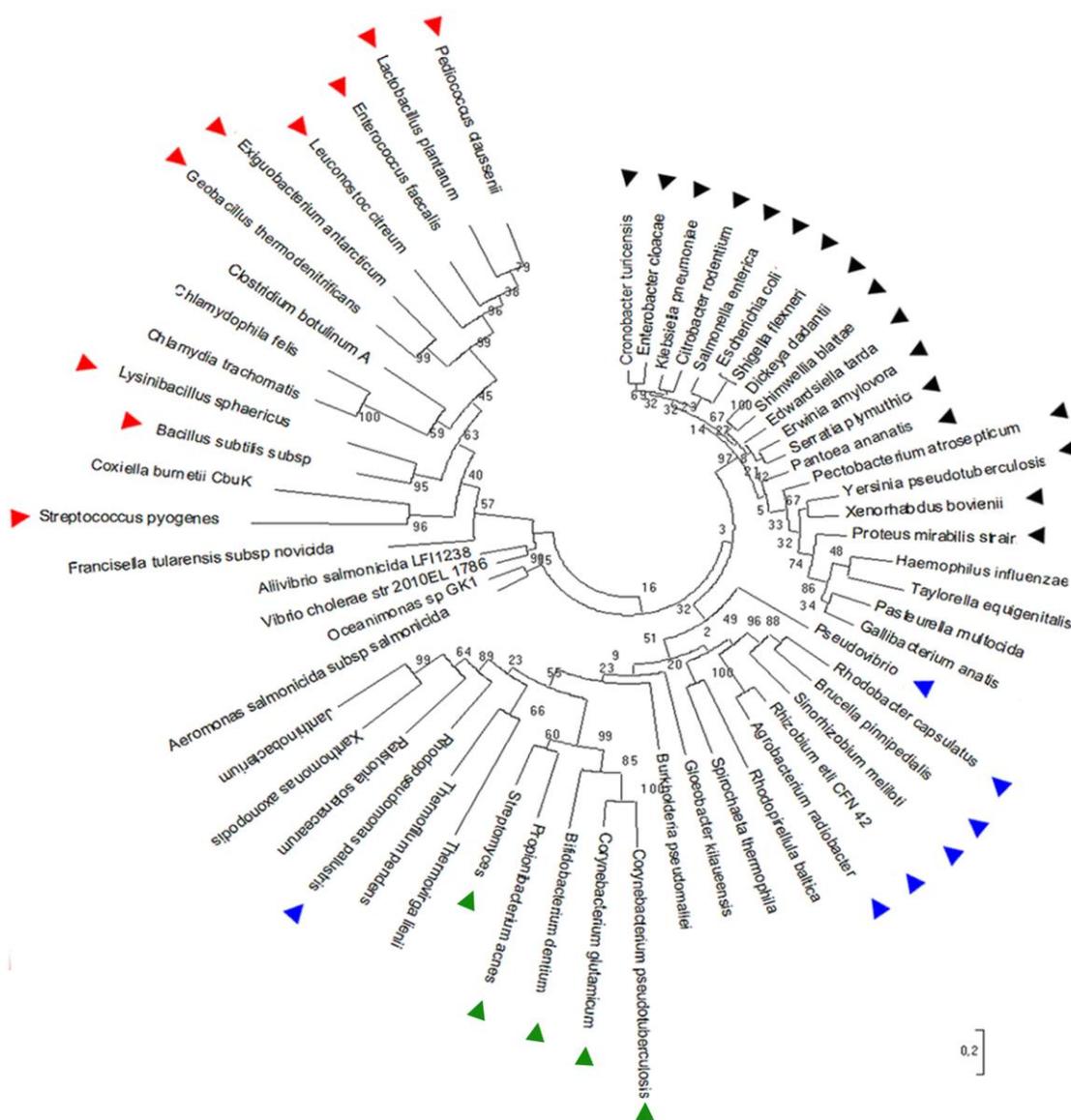
Различия в распределении предсказанных сигналов транскрипции пред геном *oppB* у филогенетически удалённых микроорганизмов представляется вполне естественным. Формирование регуляторных элементов генома и эволюционное становление генов детерминировано факторами отбора, имеющими разную природу и направленность. Изменчивость белок-кодирующих последовательностей подчинена задаче стабилизации оптимальной структуры, соответствующей характеру процессов, в которые данный белок вовлечен. Для регуляторных последовательностей определяющим является достижение баланса в уровне экспрессии различных продуктов, что в большей степени зависит от условий обитания организма и необходимости динамично реагировать на их изменение [29]. Даже у микроорганизмов, относящихся к филогенетически замкнутой группе эпсилон-протеобактерий, *Campylobacter jejuni* и *Helicobacter pylori*, наблюдается сильное разнообразие как во взаимном расположении генов-ортологов, так и в позиционировании экспериментально зарегистрированных стартовых точек транскрипции, ассоциированных с этими генами [30]. Поэтому не удивительно, что сходство в распределении предсказанных сигналов транскрипции прослеживается только у близкородственных видов (рис. 3,А-Д). Но принципиальная разница в их расположении в межгеновой области *oppA-oppB* *B. subtilis* (рис. 3,Ф) и *G. thermodenitrificans* (рис. 3,Ж), принадлежащих к одному семейству, так же как наличие *промоторного островка* перед *oppB* у *B. dentium* свидетельствуют о значительных вариациях в паттерне регуляторных сигналов.

### **Филогенетический анализ свидетельствует против горизонтального переноса гена *oppB* в геном *B. dentium***

Учитывая предпочтительное расположение *промоторных островков* рядом с чужими генами, нами была проверена возможность приобретения *B. dentium* гена *oppB*

в результате горизонтального переноса. Для реконструкции эволюционных связей по этому гену между 60 микроорганизмами (таблица 1) был проведён филогенетический анализ (рис. 4). В качестве стандартного эволюционного маркера был использован ген 16S рНК (рис. 5).

Энтеробактерии (чёрные символы на рис. 4 и 5) оказались сгруппированными в сходные по топологии кластеры как по 16S рНК, так и по последовательности гена *oppV*, хотя порядок ближайших соседей на последних узлах ветвления нередко не совпадал. Структура кластеров, выявляемых по филогении 16S рНК (рис. 5), хорошо воспроизводится для *oppV* у альфа-протеобактерий, за исключением *Rhodopseudomonas palustris* и *Pseudovibrio sp.* (синие символы) и представителей порядков Bacillales и Lactobacillales (красные символы), хотя в кладе для *oppV* попали четыре представителя других таксонов (рис. 4). Актинобактерии (зелёные символы), сохраняя мини-кластеризацию между видами на уровне класса, для 16S рНК и *oppV* оказались в составе разных макро-кластеров.



**Рис. 4.** Филогенетическое дерево, построенное в MEGA6 [25] с использованием метода наибольшего правдоподобия на основе анализа нуклеотидных последовательностей генов *oppV*. Значения бутстрап-поддержки в процентах от 2000 реплик указаны рядом с узлами ветвления. Длина масштабной линейки соответствует частоте нуклеотидных замен на позицию.

Наиболее очевидное несоответствие филогенетических связей, определяемых по последовательностям двух генов, наблюдается для *Spirochaeta thermophila* и *Rhodopseudomonas palustris* (рис. 4 и 5). Попадая в разряд одних из наиболее древних организмов по ранжированию, основанному на 16S рНК, *S. thermophila* обладает геном *oppB*, гомологичным *oppB* альфа-протеобактерий, а *oppB* *R. palustris* оказался в смешанной группе. В целом эволюция кодирующей последовательности *oppB* обнаруживает выраженную зависимость от ранней сегрегации микроорганизмов, выявляемую по сопоставлению с филогенетическими связями для 16S рНК. Организмы, для которых существенно меняется топология узлов на филогенетических деревьях, построенных для нуклеотидных последовательностей 16S рНК и *oppB*, являются перспективными кандидатами для изучения направленности изменений *oppB*.

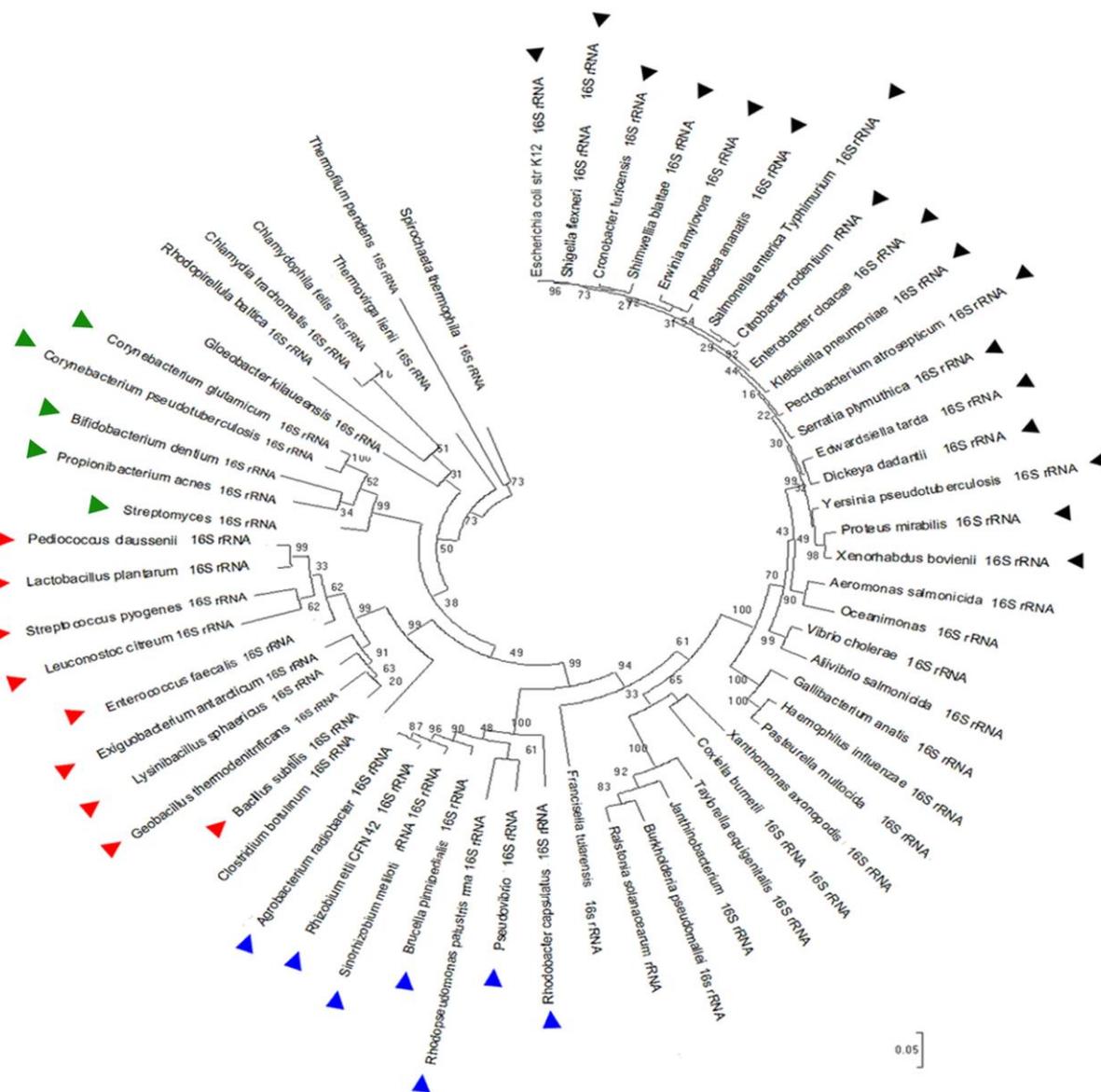
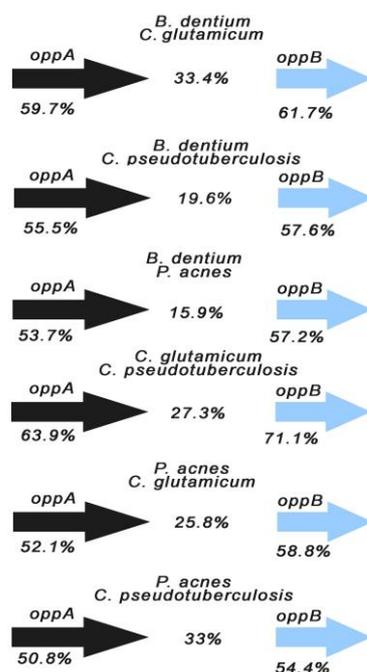


Рис. 5. Филогенетическое дерево, построенное для генов 16S рНК так, как описано в подписи к рис. 4.

Если бы промоторный островок перед *oppB* у *B. dentium* появился в ответ на интеграцию чужого гена, состав группы его ближайших соседей на филогенетических деревьях, построенных для 16S рНК и *oppB*, должен был бы отличаться, а ортологи в геномах из группы *oppB* должны были бы транскрибироваться с обычных промоторов.

Однако как раз для этой бактерии состав клады оказался устойчив (*C. glutamicum*, *C. pseudotuberculosis*, *P. acnes*, *Streptomyces sp.*), хотя по 16S рНК *B. dentium* оказался ближе к *P. acnes*, а по *oppB* – к ортологам из коринебактерий. При этом распределение потенциальных стартов транскрипции, предсказанных с помощью алгоритма PlatPromU, как и ожидалось, не выявляет промоторных островков у родственных организмов (рис. 3,G-I), хотя размер межгенной области *oppA-oppB* для представителей этой группы сильно варьирует.

Попарное выравнивание нуклеотидных последовательностей генов *oppA* и *oppB* с их ортологами из общей клады указывает на то, что *oppB* более консервативен, чем *oppA*, в том числе в паре *B. dentium* – *P. acnes* (три верхних схемы на рис. 6). Такая же закономерность наблюдается для трёх остальных пар. Это свидетельствует о совместной эволюции генов *oppA* и *oppB* у этих бактерий и против предположения о горизонтальном переносе *oppB* в геном *B. dentium*, то есть указывает на то, что наличие промоторного островка перед ним не является прямым следствием адаптивных процессов, обусловленных необходимостью ассимилировать чужой ген.



**Рис. 6.** Результаты попарного выравнивания нуклеотидных последовательностей кодирующих областей генов *oppA* и *oppB* и межгенных участков. На каждой схеме указаны пары микроорганизмов и процент идентичности пар оснований в сравниваемых последовательностях. Для глобального выравнивания был использован алгоритм Нидлмана и Вунша [31, 32].

Степень подобия межгенных областей оказалась гораздо ниже, чем кодирующих последовательностей (рис. 6). Наименьшее соответствие наблюдалось в паре *B. dentium* – *P. acnes* (15.9%). В значительной степени это обусловлено разницей в длине сравниваемых последовательностей (рис. 3). Так, при использовании алгоритма Смита-Ватермана [33, 34], осуществляющего локальное, а не глобальное выравнивание, степень идентичности в межгенных областях ожидаемо оказалась выше для всех пар, варьируя в диапазоне от 35% (*B. dentium* – *C. glutamicum*) до 51% (*P. acnes* – *C. pseudotuberculosis*), а для пары *B. dentium* – *P. acnes* был найден участок локального подобия с 46.7% идентичностью. В целом это свидетельствует о масштабных генетических перестройках в промоторной области *oppB* у *B. dentium*, которые нельзя объяснить локальными изменениями нуклеотидной последовательности.



относительной эффективности экспрессии генов *oppA* и *oppBCDF*, что предоставляет прекрасную модельную систему не только для изучения закономерностей экспрессии последовательных генов оперона, но и функционального потенциала *островка*.

## ЗАКЛЮЧЕНИЕ

Генетическая организация бактериального оперона *oppABCDF*, содержащего полный набор генов, необходимых для сборки функционального транспортера и предоставляющего возможность для их синхронной экспрессии, оказалась достаточно консервативной по признаку взаимного расположения генов *oppA* и *oppB*, а также *oppB* и *oppC*. Было обнаружено, что у большинства микроорганизмов не исключена возможность размещения дополнительного(ых) промотора(ов) для экспрессии внутренних генов *opp*-оперонов в межгенной области *oppA-oppB*. Интересным и неожиданным результатом стало обнаружение обогащенной потенциальными промоторами области перед геном *oppB* у *B. dentium*. Становление гена *oppB* трансмембранного компонента ABC-транспортера у организмов, эволюционно близких к *B. dentium* (*C. glutamicum*, *C. pseudotuberculosis* и *P. acnes*) не коррелировало с аналогичной изменчивостью регуляторной области. Попытка реконструировать историю появления *промоторного островка* у *B. dentium* привела к обнаружению в нём участка, гомологичного 5'-концевой кодирующей последовательности гена олигопептидпермеазы (*GNTG\_2042*) *G. thermodenitrificans*, белковый продукт которого может быть функциональным аналогом *OppB*. Важно, что короткий фрагмент ДНК с аномально высоким содержанием сигналов транскрипции (рис. 3, N и рис. 7) не был элиминирован эволюционным отбором, а трансформировался в протяжённый *промоторный островок* (рис. 3, M). Это может свидетельствовать о том, что экспрессия собственных генов бактерий может подвергаться такой же адаптации, сопряжённой с мультипликацией сигналов транскрипции, как и при ассимиляции чужеродного генетического материала.

Исследование выполнено за счёт гранта Российского научного фонда (проект №14-14-00985).

## СПИСОК ЛИТЕРАТУРЫ

1. Ames G.F.-L., Mimura C., Shyamala V. Bacterial periplasmic permeases belong to a family of transport proteins operating from *Escherichia coli* to human traffic ATPases. *FEMS Microbiol. Rev.* 1990. V. 75. P. 429–446. doi: [10.1111/j.1574-6968.1990.tb04110.x](https://doi.org/10.1111/j.1574-6968.1990.tb04110.x)
2. Davidson A.L., Dassa E., Orelle C., Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* 2008. V. 72. P. 317–364. doi: [10.1128/MMBR.00031-07](https://doi.org/10.1128/MMBR.00031-07).
3. Gilson E., Higgins C.F., Hofnung M., Ames G.F.-L., Nikaido H. Extensive homology between membrane-associated components of histidine and maltose transport systems of *Salmonella typhimurium* and *Escherichia coli*. *J. Biol. Chem.* 1982. V. 257. P. 9915–9918. URL: <http://www.jbc.org/content/257/17/9915.full.pdf> (дата обращения: 07.06.2015).
4. Navarro C., Wu L.F., Mandrand-Berthelot M.A. The *nik* operon of *Escherichia coli* encodes a periplasmic binding protein-dependent transport system for nickel. *Mol. Microbiol.* 1993. V. 9. P. 1181–1191. doi: [10.1111/j.1365-2958.1993.tb01247.x](https://doi.org/10.1111/j.1365-2958.1993.tb01247.x).
5. Koster W. ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B-12. *Res. Microbiol.* 2001. V. 152. P. 291–301. doi: [10.1016/S0923-2508\(01\)01200-1](https://doi.org/10.1016/S0923-2508(01)01200-1).
6. Henderson D.P., Payne S.M. *Vibrio cholerae* iron transport system: roles of heme and siderophore iron transport in virulence and identification of a gene associated with

- multiple iron transport systems. *Infect. Immun.* 1994. V. 62. P. 5120–5125. URL: <http://iai.asm.org/content/62/11/5120.full.pdf> (дата обращения: 07.06.2015).
7. Rodriguez G.M., Smith I. Identification of an ABC transporter required for iron acquisition and virulence in *Mycobacterium tuberculosis*. *J. Bacteriol.* 2006. V. 188. P. 424–430. doi: [10.1128/JB.188.2.424-430.2006](https://doi.org/10.1128/JB.188.2.424-430.2006).
  8. Saier M.H., Reddy V.S., Tamang D.G., Vastermark A. The transporter classification database. *Nucl. Acids Res.* 2014. V. 42. P. D251–D258. doi: [10.1093/nar/gkt1097](https://doi.org/10.1093/nar/gkt1097).
  9. *Transporter Classification Database*. URL: <http://www.tcdb.org> (дата обращения: 07.06.2015).
  10. Hiles I.D., Gallagher M.P., Jamieson D.J., Higgins C.F. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. *J. Mol. Biol.* 1987. V. 195. P. 125–142. doi: [10.1016/0022-2836\(87\)90332-9](https://doi.org/10.1016/0022-2836(87)90332-9).
  11. Hogarth B.G., Higgins C.F. Genetic organization of the oligopeptide permease (*opp*) locus of *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.* 1983. V. 153. P. 1548–1551. URL: <http://jb.asm.org/content/153/3/1548.full.pdf> (дата обращения: 07.06.2015).
  12. Lee E.M., Ahn S.H., Park J.H., Lee J.H., Ahn S.C., Kong I.S. Identification of oligopeptide permease (*opp*) gene cluster in *Vibrio fluvialis* and characterization of biofilm production by *oppA* knockout mutation. *FEMS Microbiol Lett.* 2004. V. 240. P. 21–30. doi: [10.1016/j.femsle.2004.09.007](https://doi.org/10.1016/j.femsle.2004.09.007).
  13. Kashiwagi K., Tshako M.H., Sakata K., Saisho T., Igarashi A., da Costa S.O., Igarashi K. Relationship between spontaneous aminoglycoside resistance in *Escherichia coli* and a decrease in oligopeptide binding protein. *J. Bacteriol.* 1998. V. 180. P. 5484–5488. URL: <http://jb.asm.org/content/180/20/5484.full.pdf> (дата обращения: 07.06.2015).
  14. Yu D., Pi B., Yu M., Wang Y., Ruan Z., Feng Y., Yu Y. Diversity and evolution of oligopeptide permease systems in staphylococcal species. *Genomics.* 2014. V. 104. P. 8–13. doi: [10.1016/j.ygeno.2014.04.003](https://doi.org/10.1016/j.ygeno.2014.04.003).
  15. Berntsson R.P., Smits S.H., Schmitt L., Slotboom D.J., Poolman B. A structural classification of substrate-binding proteins. *FEBS Lett.* 2010. V. 584. P. 2606–2617. doi: [10.1016/j.febslet.2010.04.043](https://doi.org/10.1016/j.febslet.2010.04.043).
  16. Medrano M.S., Ding Y., Wang X.G., Lu P., Coburn J., Hu L.T. Regulators of expression of the oligopeptide permease A proteins of *Borrelia burgdorferi*. *J. Bacteriol.* 2007. V. 189. P. 2653–2659. doi: [10.1128/JB.01760-06](https://doi.org/10.1128/JB.01760-06).
  17. Pletzer D., Lafon C., Braun Y., Kohler T., Page M.G., Mourez M., Weingart H. High-throughput screening of dipeptide utilization mediated by the ABC transporter DppBCDF and its substrate-binding proteins DppA1–A5 in *Pseudomonas aeruginosa*. *PLoS ONE.* 2014. V. 9. Article № e111311. doi: [10.1371/journal.pone.0111311](https://doi.org/10.1371/journal.pone.0111311).
  18. *RegulonDB*. URL: <http://regulondb.ccg.unam.mx> (дата обращения: 07.06.2015).
  19. Salgado H., Peralta-Gil M., Gama-Castro S., Santos-Zavaleta A., Muniz-Rascado L., Garcia-Sotelo J.S., Weiss V., Solano-Lira H., Martinez-Flores I., Medina-Rivera A., Salgado-Osorio G., Alquicira-Hernandez S., Alquicira-Hernandez K., Lopez-Fuentes A., Porrón-Sotelo L., Huerta A.M., Bonavides-Martinez C., Balderas-Martinez Y.I., Pannier L., Olvera M., Labastida A., Jimenez-Jacinto V., Vega-Alvarado L., Del Moral-Chavez V., Hernandez-Alvarez A., Morett E., Collado-Vides J. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucl. Acids Res.* 2013. V. 41. P. D203–D213. doi: [10.1093/nar/gks1201](https://doi.org/10.1093/nar/gks1201).
  20. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons from genome-scale promoter mapping. *Nucl. Acids Res.* 2009. V. 37. P. 4919–4931. doi: [10.1093/nar/gkp490](https://doi.org/10.1093/nar/gkp490).
  21. Панюков В.В., Киселев С.С., Шавкунов К.С., Масулис И.С., Озолин О.Н. Мультиспецифичные промоторные островки как участки генома с необычными

- структурными и функциональными свойствами. *Математическая биология и биоинформатика*. 2013. Т. 8. С. 432–448. doi: [10.17537/2013.8.432](https://doi.org/10.17537/2013.8.432).
22. *GenBank Catalog of Bacterial Genomes*. URL: <ftp://ftp.ncbi.nih.gov/genomes/Bacteria/> (дата обращения: 07.06.2015).
  23. Киселев С.С., Озолин О.Н. Структурообразующие модули как индикаторы промоторной ДНК в бактериальных геномах. *Математическая биология и биоинформатика*. 2011. Т. 6. С. 39–52. doi: [10.17537/2011.6.39](https://doi.org/10.17537/2011.6.39).
  24. Woese C.R., Fox G.E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA*. 1977. V. 74. P. 5088–5090. doi: [10.1073/pnas.74.11.5088](https://doi.org/10.1073/pnas.74.11.5088).
  25. Tamura K., Stecher G., Peterson D., Filipiński A., Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 2013. V. 30. P. 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197).
  26. Tamura K., Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 1993. V. 10. P. 512–526. URL: <http://mbe.oxfordjournals.org/content/10/3/512.full.pdf> (дата обращения: 07.06.2015).
  27. Ozoline O.N., Tsyganov M.A. Structure of open promoter complexes with *Escherichia coli* RNA polymerase as revealed by the DNaseI footprinting technique: compilation analysis. *Nucl. Acids Res.* 1995. V. 23. P. 4533–4541. doi: [10.1093/nar/23.22.4533](https://doi.org/10.1093/nar/23.22.4533).
  28. Panyukov V.V., Ozoline O.N. Promoters of *Escherichia coli* versus promoter islands: function and structure comparison. *PLoS ONE*. 2013. V. 8. Article № e62601. doi: [10.1371/journal.pone.0062601](https://doi.org/10.1371/journal.pone.0062601).
  29. Wang L., Wang F.F., Qian W. Evolutionary rewiring and reprogramming of bacterial transcription regulation. *J. Genet. Genomics*. 2011. V. 38. P. 279–288. doi: [10.1016/j.jgg.2011.06.001](https://doi.org/10.1016/j.jgg.2011.06.001).
  30. Porcelli I., Reuter M., Pearson B.M., Wilhelm T., van Vliet A.H. Parallel evolution of genome structure and transcriptional landscape in the Epsilonproteobacteria. *BMC Genomics*. 2013. V. 14. Article № 616. doi: [10.1186/1471-2164-14-616](https://doi.org/10.1186/1471-2164-14-616).
  31. *EMBOSS Needle*. URL: [http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html) (дата обращения: 07.06.2015).
  32. Needleman S.B., Wunsch C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 1970. V. 48. P. 443–453. doi: [10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4).
  33. *EMBOSS Water*. URL: [http://www.ebi.ac.uk/Tools/psa/emboss\\_water/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html) (дата обращения: 07.06.2015).
  34. Smith T.F., Waterman M.S. Identification of common molecular subsequences. *J. Mol. Biol.* 1981. V. 147. P. 195–197. doi: [10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
  35. *Microbial Nucleotide BLAST*. URL: [http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastSearch&BLAST\\_SPEC=MicrobialGenomes](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=MicrobialGenomes) (дата обращения: 07.06.2015).
  36. *EMBOSS Matcher*. URL: [http://www.ebi.ac.uk/Tools/psa/emboss\\_matcher/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_matcher/nucleotide.html) (дата обращения: 07.06.2015).

Материал поступил в редакцию 26.06.2015, опубликован 14.07.2015.