

УДК 577.1

ПУРИНОВЫЙ РЕГУЛОН ГАММА-ПРОТЕОБАКТЕРИЙ. ДЕТАЛЬНОЕ ОПИСАНИЕ

© 2002 г. Д. А. Равчеев, М. С. Гельфанд, А. А. Миронов, А. Б. Рахманинова

Государственный научно-исследовательский институт генетики и селекции промышленных микроорганизмов,
Москва 113545; факс: (095) 315-05-01; e-mail: masha@imb.ac.ru

Поступила в редакцию 09.11.2001 г.

Методами сравнительной геномики исследована структура пуринового регулона в семи геномах гамма-протеобактерий: *Escherichia coli*, *Salmonella typhi*, *Yersinia pestis*, *Haemophilus influenzae*, *Pasteurella multocida*, *Actinobacillus actinomycetemcomitans* и *Vibrio cholerae*. Полидромный сайт связывания пуринового репрессора, имеющий консенсус ACGCAAACGTTTGCCT, хорошо сохраняется перед генами ферментов, участвующих в синтезе инозинмонофосфата из фосфорибозилпирофосфата, в переносе одноуглеродных групп, а также перед генами ряда транспортных белков. Эти гены можно считать ядром пуринового регулона. Физиологически особенно важной представляется регуляция генов *purC* и *gsvTHP/fold*, так как PurR-сайт обнаружен перед неортологичными, но функционально заменимыми генами. С другой стороны, PurR-сайт плохо сохраняется перед ортологами ряда генов пуринового регулона *E. coli*: генами, вовлеченными в общий азотный обмен, биосинтез пиримидинов, синтез АМФ и ГМФ из ИМФ, а также перед геном самого пуринового репрессора. Предсказано, что к пуриновым регулонам исследованных бактерий могут относиться гены: *upp*, принимающий участие в синтезе пиримидинов; *uraA*, кодирующий транспортер урацила; *serA*, участвующий в биосинтезе серина; *fold*, ответственный за превращение N5,N10-металинтетрагидрофолата в N10-формилтетрагидрофолат; *rpiA*, задействованный в метаболизме рибозы; а также гены белков с неизвестной функцией: *yhhQ* и *ydiK*. Было показано, что PurR-сайт может иметь различное строение в разных геномах. Так, в геномах бактерий из групп Pasteurellaceae и Vibrionaceae наблюдается тенденция к уменьшению консервативности второй и пятнадцатой позиций сайта.

Пуриновый регулон включает гены, транскрипция которых регулируется пуриновым репрессором (PurR). Регулон хорошо изучен в *Escherichia coli* [1], предпринимались попытки к исследованию пуриновой регуляции в *Haemophilus influenzae* [2] и *Salmonella typhimurum* [3].

Ядро регулона в *E. coli* составляют гены, кодирующие ферменты синтеза пуриновых нуклеотидов *de novo*: *purF*, *purC*, *guaBA*, *purA*, *purHD*, *purMN*, *purB*, *purEK*, *purT*, *purL*, а также ген собственно репрессора *purR* [1]. В этом регулон входит также ряд генов, имеющих прямое или косвенное отношение к синтезу нуклеотидов: *prsA*, *purC*, *purD*, *codB*, *codA*, *gsvTHP*, *glyA*, *speAB*, *glnB*. Функциональная роль этих генов проиллюстрирована на рис. 1. Для всех перечисленных выше генов экспериментально показана пуриновая регуляция [1, 4–9].

Кроме того, в работе [2] было предсказано, что в данный регулон, могут входить также гены ряда транспортных белков, а именно нуклеозид-специфичного каналобразующего белка Tsx из *E. coli*, гипотетических транспортеров ксантина и урацила – YjcD, YieG, HI0125 (ортолога YjcD из *H. influenzae*) и гипотетических транспортеров пуринов YgfU и YieE из *E. coli*.

В настоящее время секвенировано около 2000 бактериальных геномов. Представляло интерес определить, насколько сохраняется описанная выше структура регулона в различных геномах. Задача настоящей работы состояла в определении генов, входящих в пуриновый регулон у гамма-протеобактерий.

Для решения поставленной задачи использовались методы сравнительной геномики [2]. Так, ген считался относящимся к пуриновому регулону, если одновременно выполнялись два условия: во-первых, перед опероном, содержащим этот ген, в последовательности ДНК можно найти возможный сайт связывания PurR, т.е. PurR-сигнал; во-вторых, PurR-сигнал сохраняется перед оперонами, содержащими ортологов данного гена по крайней мере в трех близко родственных видах из той же таксономической группы.

МАТЕРИАЛЫ И МЕТОДЫ

Исследованные геномы. Рассмотрено 16 геномов гамма-протеобактерий: полные геномы *E. coli* [10, 11], *Yersinia pestis* [10, 12], *Haemophilus influenzae* [10, 13], *Vibrio cholerae* [10], *Pseudomonas aeruginosa* [10], *Buchnera* sp. [10, 14], и *Xylella fastidiosa* [15] и предварительные последователь-

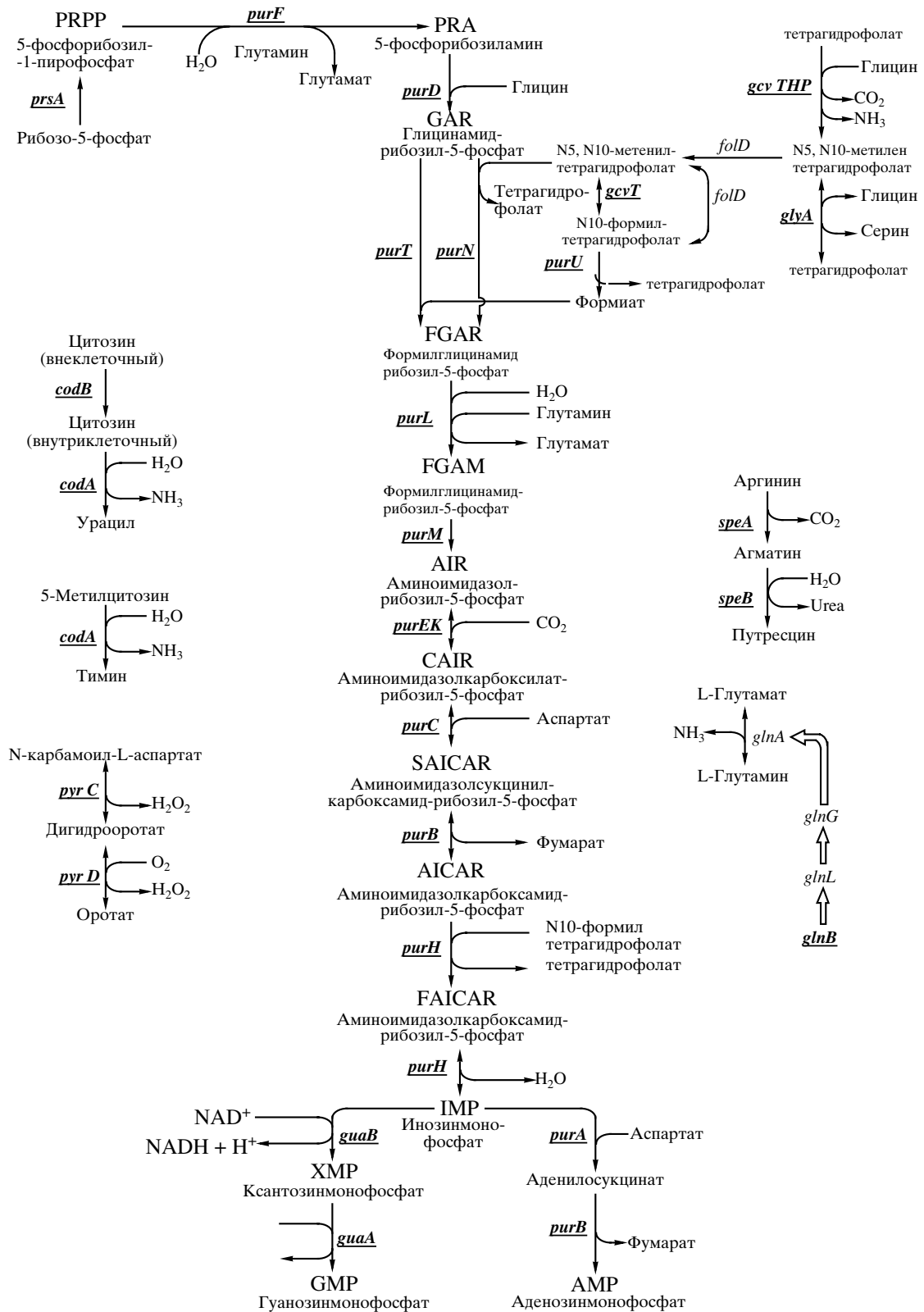


Рис. 1. Функциональная роль генов *E. coli*, входящих в пуриновый регулон. Приведены фрагменты карты метаболических путей *E. coli*. Для всех химических реакций указаны названия генов соответствующих ферментов. Названия генов, для которых экспериментально показана пуриновая регуляция, подчеркнуты.

ности геномов *Salmonella typhi*, *Klebsiella pneumoniae* [16], *Shewanella putrefaciens* [17], *Actinobacillus actinomycetemcomitans* [18], *Xylella almond* [19], *Pseudomonas fluorescens* [19], *Pseudomonas putida* и *Pseudomonas syringae* [20]. Близкие гомологи PurR обнаружены во всех геномах, за исключением *Buchnera* sp. Однако только семь из них можно считать ортологами PurR *E. coli*, остальные белки близко гомологичны не только PurR, но и RbsR и/или CytR *E. coli* (см. рис. 2,а).

Поэтому объектом настоящего исследования были выбраны гамма-протеобактерии, геном которых содержит явный ортолог PurR *E. coli*. Это геномы бактерий групп *Enterobacteriaceae* (*S. typhi*, *Y. pestis*), *Pasteurellaceae* (*H. influenzae*, *P. multocida*, *A. actinomycetemcomitans*) и *Vibrionaceae* (*V. cholerae*).

В геноме *K. pneumoniae* найден фрагмент ортолога PurR *E. coli*, содержащий только 37% последовательности PurR и находящийся на краю контига. На данном этапе работы фрагментарный геном *K. pneumoniae* был исключен из рассмотрения.

Поскольку геном *A. actinomycetemcomitans* также является фрагментарным, то в дальнейшем полученные в ходе его исследования данные не обсуждаются.

Процедура определения генов, входящих в PurR регулон. Экспериментально показано, что сигнал связывания PurR представляет собой палиндром из 16 нуклеотидов с консенсусом ACGCAAACGTTTGCGT [1].

Для поиска подобных сайтов в ДНК была использована матрица позиционных весов, составленная на основе экспериментально подтвержденных сайтов связывания PurR в геноме *E. coli* [2].

Ген считался входящим в PurR-регулон, если в геноме *E. coli* перед данным геном (или соответствующим опероном) в области $-300...+50$ пн относительно старта трансляции находится потенциальный PurR сайт с весом, превышающим пороговое значение 3.70;

перед ортологичными генами (или соответствующими оперонами) минимум в двух родственных геномах можно найти потенциальный PurR-сайт с весом выше 3.25, расположенный в области $-300...+100$ пн относительно старта трансляции.

Гены *prsA* и *purB* были рассмотрены как особые случаи, так как экспериментально показано, что PurR-сайт перед геном *prsA* находится на большом расстоянии от начала кодирующей последовательности (-357 пн в *E. coli*) [1, 4], а в случае гена *purB* сайт находится внутри кодирующей последовательности начиная с позиции $+185$ [1, 21, 22]. Поэтому в этих случаях область поиска сайта была расширена: до $-400...+100$ пн при поиске сайта для ортологов *prsA* и до $-300...+250$ пн при поиске сайта для ортологов *purB*.

В тех случаях, когда оперонная структура данного фрагмента ДНК неизвестна, гены относили к одному потенциальному оперону, если они имели одинаковое направление считывания, а расстояние между ними не превышало 100 пн. За начало такого потенциального оперона принимался ген, перед которым был найден PurR-сайт.

Для поиска PurR-сайтов, а также анализа ортологов использовался пакет программ Genome Explorer [23]. Для построения матриц позиционных весов использовалась программа SignalX [23].

Поиск гомологов в базе данных проводился с помощью программы BLAST [24].

Обозначения генов, данные в статье, соответствуют их ортологам в *E. coli*; для генов, не имеющих ортологов, приводятся идентификаторы из полных геномов или базы данных ERGO [25].

Выравнивание последовательностей белков и ДНК, а также построение филогенетических деревьев проводилось с помощью программы Clustal W [26].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Структура PurR-регулона в шести гамма-протеобактериях

Процедура поиска генов, входящих в PurR регулон (см. Материалы и методы), показала, что потенциальные PurR-сайты с порогом выше 3.70, можно найти перед 91 геном *E. coli*. Из них только в 38 случаях сайт сохранялся перед ортологичными генами минимум в двух родственных организмах. Соответственно, данные гены с наибольшей вероятностью входят в пуриновый регулон. Эти 38 генов можно разбить на семь групп, которые подробно описаны ниже.

Группа 1 – ген белка-репрессора purR. Область гена *purR* содержит в *E. coli* два экспериментально подтвержденных PurR-сайта, начинающихся с позиций -67 и $+21$ [1, 2, 27]. Такие же сайты были найдены перед *purR* в *S. typhi*, а в *Y. pestis* обнаружен только один сайт, находящийся перед кодирующей областью и имеющий меньший вес, чем в *E. coli* и *S. typhi* (таблица).

В геномах *H. influenzae*, *P. multocida*, *A. actinomycetemcomitans* и *V. cholerae* не удалось обнаружить значимый сайт перед *purR*. Можно думать, что в этих организмах нет регуляции транскрипции *purR* по механизму простой авторепрессии.

Группа 2 – гены, непосредственно вовлеченные в биосинтез пуринов. К этой группе относятся 15 генов *E. coli*: *purF*, *purC*, *guaB*, *guaA*, *purA*, *purH*, *purD*, *purM*, *purN*, *purB*, *purE*, *purK*, *purT*, *purL*, *prsA*. Известно, что транскрипция этих генов регулируется PurR [6, 21, 22, 28–30] и что *purHD*, *purEK*, *purMN* и *guaBA* являются полицистронными оперонами [1], а ген *purF* находится в

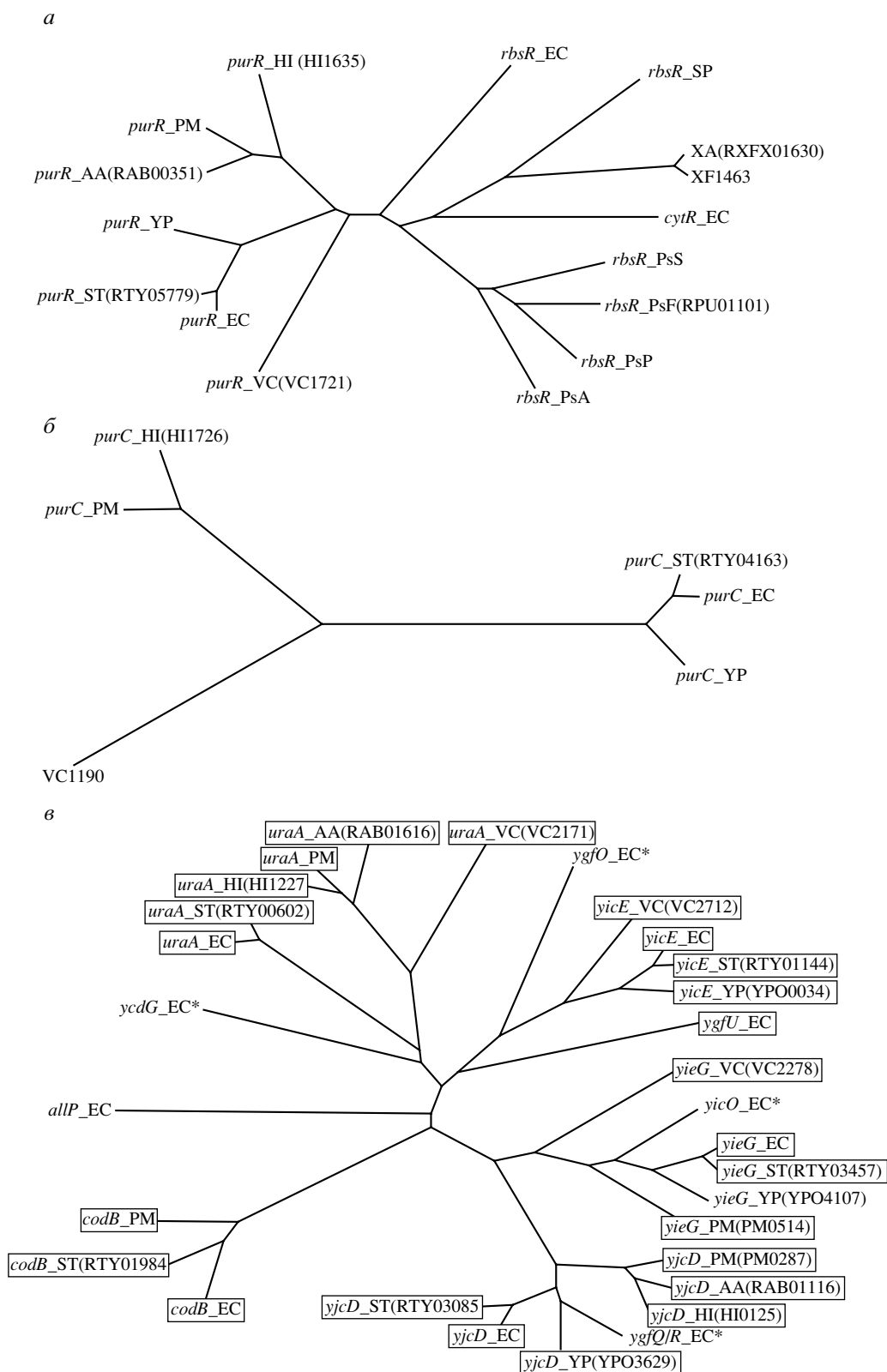


Рис. 2. Филогенетические деревья: *а* – для гомологов пуринового репрессора *E. coli* в гамма-протеобактериях; *б* – для белков, выполняющих функцию PurC в рассмотренных бактериях; *в* – для транспортных белков из рассмотренных бактерий; обведены гены, перед которыми обнаружен PurR-сайт; гены, находящиеся в потенциальных оперонах с PurR-сайтом, помечены звездочкой. Сокращения: EC – *E. coli*, ST – *S. typhi*, YP – *Y. pestis*, VC – *V. cholerae*, HI – *H. influenzae*, PM – *P. multocida*, AA – *A. actinomycetemcomitans*, PsA – *Pseudomonas aeruginosa*, PsP – *Ps. putida*, PsF – *Ps. fluorescens*, PsS – *Ps. syringae*, XA – *Xylella almond*, XF – *X. fastidiosa*.

Сайты перед генами, входящими в пуриновый регулон

Гены	<i>E. coli</i>	<i>S. typhi</i>	<i>Y. pestis</i>	<i>V. cholerae</i>	<i>H. influenzae</i>	<i>P. multocida</i>	<i>A. actinomyces-comitans</i>
Пуриновый репрессор							
<i>purR</i> ¹	4.76(-61) 4.07(27)	4.76(-59) 4.07(27)	3.61(-63) -	- -	- -	- -	- -
Синтез пуринов							
<i>prsA</i> ¹	4.72(-357)	4.72(-456)	-	4.45(-264)	3.60(-283)	3.91(-310)	-
<i>purC</i> ¹	4.67(-168)	4.67(-129)	-	4.86(-122)	4.46(-131)	4.73(-84)	0
<i>purA</i> ¹	4.17(-122) 3.77(-23)	4.17(-121) 3.77(-23)	- -	- -	- -	- -	3.59(-49)
<i>purT</i> ¹	5.08(-54)	5.08(-52)	4.67(-83)	4.84(-87)	0	-	0
<i>purL</i> ¹	4.94(-91)	4.94(-88)	4.94(-128)	4.81(-159)	4.73(-83)	5.08(-84)	0
<i>purB</i> ¹	4.51(185)	4.71(186)	4.32(186)	4.11(50)	3.26(183)	4.59(183)	4.59(201)
<i>purEK</i> ¹	4.61(-86)	4.61(-85)	4.61(-117)	4.86(-90)	4.46(-77)	4.60(-77)	0
<i>cvpA-purF</i> ¹	4.93(-71)	4.93(-70)	4.93(-75)	4.91(-59)	4.93(-84)	4.93(-68)	0
<i>purMN</i> ¹	4.55(-80)	4.55(-193)	4.55(-243)	4.93(-83)	4.55(-62)	4.55(-73)	3.63(24)
<i>PurHD</i> ¹	4.76(-122)	4.76(-124)	4.65(-118)	3.64(-188)	5.01(-65)	4.78(-74)	0
<i>guaBA</i> ¹	4.21(-68)	4.21(-66)	4.31(-75)	3.74(-95)	-	-	-
Синтез пиримидинов							
<i>codBA</i> ¹	4.68(-83)	4.85(-75)	0	0	0	4.86(-77)	0
<i>upp</i> ³	4.63(-233)	4.63(-348)	4.63(-195)	4.91(-86)	3.35(98)	3.49(-189)	-
<i>uraA</i> ³	*	*	0	3.45(-186)	*	*	3.68(-186)
<i>pyrC</i> ¹	4.54(-66)	4.54(-63)	4.32(-92)	3.49(-50)	0	0	0
<i>pyrD</i> ¹	4.51	4.51	4.45	-	-	-	-
Азотный обмен							
<i>speA</i> ¹	4.28(-133)	3.67(-132)	-	-	0	3.89(-54)	0
<i>speB</i> ¹	*	*	0	3.55(-66)	0	*	0
<i>glnB</i> ¹	4.06(-82)	-	-	-	-	3.51(-179)	-
Метаболизм глицина и одноуглеродных фрагментов							
<i>gcvTHP</i> ¹	4.31(-105)	4.48(-106)	3.60(-96)	0	0	0	0
<i>folD</i> ³	-	-	-	4.86(-123)	3.91(-44)	4.44(-40)	3.38(88)
<i>glyA</i> ¹	4.39(-133)	4.39(-124)	4.12(-321)	5.00(-72)	*	*	-
<i>serA</i> ³	3.72(-94)	3.40(-50)	4.08(-100)	4.76(-98)	*	*	*
<i>rpiA</i> ³	-	-	-	-	4.61(-55)	4.34(-103)	3.96(-36)
Трансмембранные белки							
<i>tcx</i> ²	4.57(-153)	4.57(-157)	0	0	0	0	0
<i>gltS</i> ³	4.48(-226)	4.46(-175)	4.58(-185)	-	-	0	-
<i>yieE</i> ²	4.46(-69) 4.15(29)	4.29(-57) 4.15(41)	4.46(-34)	4.86(-226) 4.53(-169)	0	0	0
<i>yieG</i> ²	4.21(-114)	4.21(-119)	-	4.53(-40)	0	4.21(-199)	0
<i>yjcD</i> ²	4.22(-273) 3.87(-131)	4.22(-263) 3.87(-120)	4.29(-207) 3.61(-74)	0	4.85(-85)	4.17(-82) 4.68(-94)	4.53(-79)
Белки с неизвестной функцией							
<i>yhhQ</i> ³	3.87(-160)	3.87(31)	3.45(-55)	3.91(-162)	-	3.54(-136)	-
<i>ydiK</i> ³	4.02(-181)	3.53(-181)	3.28(-89)	0	0	0	0
<i>ydiJ</i> ³	4.16(-223)	4.06(-98)	3.28(69)	-	-	-	-
Рибонуклеаза Т							
<i>rnt</i> ³	3.79(28)	3.79(28)	-	3.41(28)	3.47(37)	-	0

Примечание. * – ген имеет ортолога, находящегося в опероне, регулируемом PurR;

“-” – ген имеет нерегулируемого PurR ортолога;

0 – ортолога для данного гена не найдено.

¹ Пуриновая регуляция гена была известна ранее.

² Пуриновая регуляция гена предсказана в работе [2].

³ Пуриновая регуляция гена предсказана в данной работе.

одном опероне с геном белка CvpA, отвечающего за синтез колицина V [1].

Опероны *cvpA-purF*, *purEK*, *purHD* и *purMN* и ген *purL* сохраняются во всех рассмотренных геномах, за исключением генома *A. actinomycetemcomitans*, в котором из всех перечисленных генов только *purM* имеет ортолога (ID = 75%). Во всех случаях перед геном или опероном находится значимый PurR-сайт (см. таблицу). В *P. multocida* между *purH* и *purD* находится ген гипотетического мембранного белка DsaA (см. рис. 3,а). Интересно, что кластер *cvpA-purF* и PurR-сайт перед ним сохраняются во всех геномах, таким образом, в соответствии с принятыми критериями, этот оперон входит в пуриновый регулон. Однако остается неясным, какое отношение к метаболизму пуринов имеет мембранный белок CvpA, необходимый для синтеза бактериального токсина, колицина V.

Ген *purT* имеет ортологов со значимым PurR-сайтом в *S. typhi*, *Y. pestis* и *V. cholerae*, не имеет ортологов в *H. influenzae* и *A. actinomycetemcomitans*, а в *P. multocida* есть ортолог *purT* (ID = 66%), перед которым не обнаружен PurR-сайт (см. таблицу). Отсутствие в геноме гена *purT* и/или отсутствие PurR регуляции можно объяснить тем, что функция гена *purT* является практически дублирующей функцию *purN* (см. рис. 1).

В случае *purB* единственный значимый PurR-сайт обнаружен, как и следовало ожидать, внутри кодирующей области гена [21]. Этот сайт сохраняется во всех ортологах *purB* (см. таблицу). Маловероятно, что это связано с гомологией белковых последовательностей *purB* и его ортологов: сайт находится в области, где аминокислотные последовательности *purB* *E. coli*, *S. typhi*, *Y. pestis*, *H. influenzae*, *P. multocida* и *A. actinomycetemcomitans* мало похожи друг на друга, а в *V. cholerae* сайт находится вообще в другом месте кодирующей области.

Ген *purC* имеет ортологов только в *S. typhi* и *Y. pestis*. Дополнительный поиск по базе данных ERGO [25] показал, что в геномах *H. influenzae*, *P. multocida* и *V. cholerae* есть функциональные аналоги данного фермента, являющиеся далекими гомологами (менее 25% тождественных позиций) PurC из *E. coli* (см. рис. 2,б). По-видимому, данная ситуация является еще одним примером так называемых “неортологичных замен” [31]. Во всех случаях, за исключением *Y. pestis*, найден сайт с большим весом (см. таблицу). В геноме *Y. pestis* перед *purC* находится транспозаза, а PurR-сайт не обнаружен. Наличие регуляторного сигнала перед негомологичными, но функционально равноправными генами подтверждает принадлежность *purC* к пуриновому регулону. Вообще говоря, стандартный алгоритм, используемый в настоящей работе, не позволяет обнаружить подобные

ситуации, поэтому все случаи отсутствия ортологов исследуемого гена требуют дополнительной проверки.

Ген *purA* имеет ортологов во всех рассмотренных геномах, однако PurR-сайты перед этим геном обнаружены только в *E. coli*, *S. typhi* и *A. actinomycetemcomitans* (см. таблицу).

Гены *guaB* и *guaA* имеют ортологов во всех рассмотренных геномах. В *E. coli* эти гены образуют оперон *guaBA*, и этот кластер сохраняется в *S. typhi*, *Y. pestis*, *V. cholerae* и *H. influenzae*. В *P. multocida* между *guaB* и *guaA* встроены гипотетический ген, не имеющий явных гомологов. Оперонная структура нарушается в *A. actinomycetemcomitans*: гены расположены в разных местах генома. PurR-сайт перед опероном был обнаружен в *S. typhi*, *Y. pestis* и *V. cholerae* и не найден ни перед *guaB*, ни перед *guaA* в *H. influenzae*, *P. multocida* и *A. actinomycetemcomitans*. Интересно, что в *V. cholerae* сайт перед опероном *guaBA* слабый, но между *guaB* и *guaA* обнаружен сайт с большим весом (см. рис. 3,б).

В *E. coli* наиболее вероятный промотор, а также экспериментально подтвержденный PurR-сайт перед *prsA* находятся в кодирующей области предыдущего гена изопентенил-монофосфат киназы, *ispE*. Подобная ситуация наблюдается в остальных геномах, при этом позиция сайта относительно старта трансляции *prsA* коррелирует с расстоянием между *prsA* и *ispE* (см. рис. 3,в). Более того, в геномах *H. influenzae*, *A. actinomycetemcomitans* и *V. cholerae* расстояние между генами *prsA* и *ispE* очень мало: от 5 до 45 пн. В данной ситуации достоверное выделение PurR-сайта на фоне гена не представляется возможным, и вопрос о регуляции гена *prsA* остается открытым.

Сопоставляя описанные выше результаты со схемой на рис. 1, можно увидеть, что все гены пути синтеза ИМФ из рибозо-5-фосфата имеют консервативный PurR-сайт. Соответственно, можно предположить, что их транскрипция регулируется пуриновым репрессором у всех рассмотренных бактерий. В то же время наблюдается тенденция к потере такого сайта и соответственно типа регуляции у генов, регулирующих синтез АМФ и ГМФ из ИМФ, т.е. перед опероном *guaBA* и геном *purA*.

Группа 3 – гены, вовлеченные в биосинтез пиримидинов. Группа включает 5 генов: *codB*, *codA*, *purC*, *purD*, *upp*. Известно, что транскрипция генов *codB*, *codA*, *purC* и *purD* в *E. coli* находится под регуляцией PurR и что *codBA* является полицистронным опероном [1, 5, 32]. Считают также, что *upp* находится в одном опероне с геном урациловой пермеазы *uraA* и гипотетическим геном *b2496* [33].

Оперон *codBA* и PurR-сайт перед первым геном оперона – *codB* полностью сохраняются в *S. typhi* и *P. multocida*. Оперон утрачен в *H. influenzae* и *V. cholerae*. В *Y. pestis* и *A. actinomycetemcomitans*

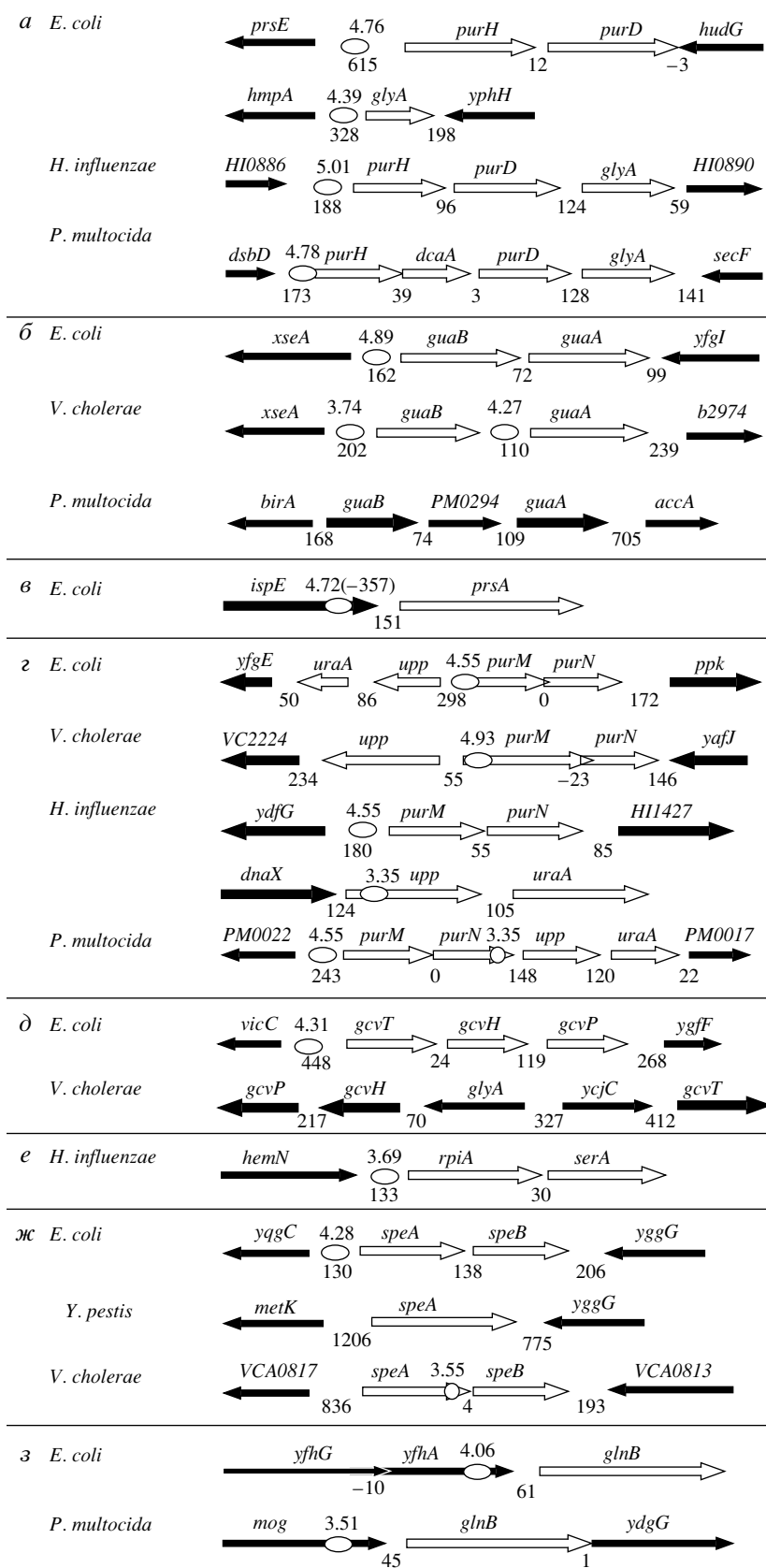


Рис. 3. Примеры оперонных структур, наблюдаемых в рассмотренных геномах (а-з). Пояснения в тексте.

<i>E. coli</i>				<i>Y. pestis</i>				<i>H. influenzae</i>				<i>P. multocida</i>				<i>V. cholerae</i>			
a	c	g	t	a	c	g	t	a	c	g	t	a	c	g	t	a	c	g	t
-0.27	-0.21	-0.02	-0.07	0.29	-0.21	-0.02	-0.06	0.29	-0.32	-0.11	0.14	0.23	-0.36	0.06	0.06	0.29	-0.21	-0.02	-0.06
0.01	0.14	-0.01	-0.14	0.01	0.18	-0.06	-0.25	0.19	0.07	-0.08	-0.18	0.06	0.08	-0.02	-0.12	0.09	0.12	0.06	-0.26
-0.04	-0.13	0.48	-0.31	-0.04	-0.32	0.39	-0.04	0.06	-0.24	-0.42	-0.24	0.03	-0.23	0.43	-0.23	-0.30	-0.11	0.42	-0.02
0.05	0.19	-0.19	-0.06	-0.03	0.20	-0.17	0.00	0.04	0.30	-0.33	-0.02	0.02	0.29	-0.34	0.02	0.07	0.27	-0.21	-0.13
0.38	-0.13	-0.13	-0.13	0.50	-0.23	-0.23	-0.04	0.51	-0.17	-0.17	-0.17	0.51	-0.17	-0.17	-0.17	0.55	-0.18	-0.18	-0.18
0.56	-0.06	-0.25	-0.25	0.56	-0.19	-0.19	-0.19	0.51	-0.17	-0.17	-0.17	0.51	-0.17	-0.17	-0.17	0.55	-0.18	-0.18	-0.18
0.32	-0.05	-0.43	0.05	0.27	-0.09	-0.37	0.20	0.33	-0.11	-0.11	-0.11	0.37	-0.27	-0.27	0.16	0.26	-0.07	-0.22	0.03
-0.20	0.60	-0.20	-0.20	-0.19	0.56	-0.19	-0.19	-0.17	-0.51	-0.17	-0.17	-0.21	0.46	-0.21	-0.03	-0.18	0.55	-0.18	-0.18
-0.20	-0.20	0.60	-0.20	-0.19	-0.19	0.56	-0.19	-0.17	-0.17	-0.51	-0.17	-0.03	-0.21	0.46	-0.21	-0.18	-0.18	0.55	-0.18
0.05	-0.43	-0.05	0.32	0.20	-0.37	-0.09	0.27	-0.11	-0.11	-0.11	0.33	0.16	-0.27	-0.27	0.37	0.03	-0.22	-0.07	0.26
-0.25	-0.25	-0.06	0.56	-0.19	-0.19	-0.19	0.56	-0.17	-0.17	-0.17	0.51	-0.17	-0.17	-0.17	0.51	-0.18	-0.18	-0.18	0.55
-0.13	-0.13	-0.13	0.38	-0.04	-0.23	-0.23	0.50	-0.17	-0.17	-0.17	0.51	-0.17	-0.17	-0.17	0.51	-0.18	-0.18	-0.18	0.55
-0.06	-0.19	0.19	0.05	0.00	-0.17	0.20	-0.03	-0.02	-0.33	0.30	0.04	0.02	-0.34	0.29	0.02	-0.13	-0.21	0.27	0.07
-0.31	0.48	-0.13	-0.04	-0.04	0.39	-0.32	-0.04	-0.24	-0.42	-0.24	0.06	-0.23	0.43	-0.23	0.03	-0.02	0.42	-0.11	-0.30
-0.14	-0.01	0.14	0.01	-0.25	-0.06	0.18	0.01	-0.18	-0.08	0.07	0.19	-0.12	-0.02	0.08	0.06	-0.26	0.06	0.12	0.09
-0.07	-0.02	-0.21	-0.27	-0.06	-0.02	-0.21	0.29	0.14	-0.11	-0.32	0.29	0.06	0.06	-0.36	0.23	-0.06	-0.02	-0.21	0.29

Рис. 4. Матрицы позиционных весов для PurR-сайтов в различных геномах.

не найдены ортологи гена *codB*, ортологи *codA* есть, но перед ними не обнаружены значимые PurR-сайты. Можно предположить, что функциональное значение имеет PruR-регуляция транскрипции только гена *codB*, кодирующего цитозинный транспортер (см. ниже).

Ген *purC* сохраняет PurR-сайт в геномах *S. typhi*, *Y. pestis* и *V. cholerae*, а в *H. influenzae*, *P. multocida* и *A. actinomycetemcomitans* ортологи *purC* не обнаружены. Ген *purD* имеет ортологов во всех рассмотренных бактериях, но значимый PurR-сайт перед этим геном найден только в *E. coli*, *S. typhi* и *Y. pestis*. Таким образом, *purC* и *purD*, гены основного пути биосинтеза пиримидинов на основе глутамина, регулируются пуриновым репрессором только в энтеробактериях, такая регуляция возможна также для *purC* в *V. cholerae*, а в геномах бактерий группы *Pasteurellaceae* существование данного метаболического пути представляется сомнительным.

Опероны *urp-uraA* и *purMN* образуют дивергон в геномах *E. coli* и *S. typhi* (см. рис. 3,з). PurR-сайт перед опероном *purMN* описан выше (см. также таблицу). Несмотря на большое расстояние между этим сайтом и стартом трансляции гена *urp*, консервативное взаиморасположение оперонов позволяет предположить одновременную и совместную регуляцию транскрипции обоих оперонов пуриновым репрессором. В пользу этого предположения говорят результаты подробного рассмотрения соответствующих локусов в других геномах. Дивергентное расположение гена *urp* и *purMN* в PurR-сайт между ними сохраняются в геномах *E. coli*, *S. typhi*, *Y. pestis* и *V. cholerae*. В геноме *H. influenzae* опероны *urp-uraA* и *purMN* расположены в разных местах, значимых PurR-сайт найден в самом начале гена *urp* (см. таблицу), при этом такой сайт не обнаружен в других ортологах *urp*. В геноме *P. multocida* возможно существова-

ние единого оператора *purMN-urp-uraA*. Обнаружен сильный PurR-сайт перед этим опероном, а также дополнительный значимый сайт перед *urp* в самом конце гена *purN*. В других ортологах *purN* такого сайта не найдено. Во всех геномах, содержащих *uraA*, PurR-сайт располагается либо перед опероном *urp-uraA*, либо непосредственно перед геном *uraA*, как в случае *V. cholerae* и *A. actinomycetemcomitans* (см. рис. 3,з). Следовательно, можно думать, что ген *uraA* входит в пуриновый регулон во всех исследованных бактериях, а ген *urp* – в *Enterobacteriaceae*, *V. cholerae* и, с большой степенью вероятности, в *Pasteurellaceae*.

Группа 4 – гены, участвующие в метаболизме глицина и переносе одноуглеродных фрагментов. Группа включает 5 генов *E. coli*: *gcvT*, *gcvH*, *gcvP*, *glyA* и *serA*. Известно, что транскрипция генов *gcvT*, *gcvH*, *gcvP*, *glyA* находится под регуляцией PurR и что *gcvTHP* является полицистронным опероном в *E. coli* [1].

Оперон *gcvTHP* кодирует белки системы, ответственной за использование глицина в качестве источника одноуглеродных фрагментов, переносимых коферментными формами тетрагидрофолата, а также участвующей во взаимопревращениях метенил- и формилтетрагидрофолатов (см. рис. 1). Последнюю функцию может выполнять также ген *fold* *E. coli*.

Оперон *gcvTHP* и PurR-сайт перед ним сохраняются в *S. typhi* и *Y. pestis*. Данный оперон имеет другую структуру в *V. cholerae* (см. рис. 3,д), сайт при этом не сохраняется. В *H. influenzae*, *P. multocida* и *A. actinomycetemcomitans* данный оперон полностью утрачен. Поэтому можно думать, что в *Pasteurellaceae* глицин не является непосредственным источником атомов углерода, необходимых для синтеза пуринов.

С другой стороны, в *E. coli* перед *fold* найден слабый PurR-сайт с весом ниже порогового значе-

ния, находящийся в кодирующей области (см. таблицу). Этот слабый сайт сохраняется в *S. typhi* и *Y. pestis*, что скорее всего связано с консервативностью самих кодирующих областей. И, напротив, в геномах *V. cholerae*, *H. influenzae*, *P. multocida* обнаружен сильный PurR-сайт, лежащий перед ортологами *folD* (см. таблицу).

Так как в геномах, где *gcvT* утрачен (*H. influenzae*, *P. multocida*) или не регулируется PurR (*V. cholerae*), потенциальную пуриновую регуляцию приобретает ген *folD*, можно предположить, что важным консервативным членом пуринового регулона является ген, ответственный за образование N10-формилтетрагидрофолата.

Ген *glyA* осуществляет взаимопревращения глицина и серина, что позволяет использовать серин в качестве источника одноуглеродных фрагментов (см. рис. 1). Данный ген имеет ортологов со значимым PurR-сайтом в геномах *S. typhi*, *Y. pestis* и *V. cholerae* (см. таблицу). В *H. influenzae* и *P. multocida* ген *glyA* располагается непосредственно за опероном *purHD*, и в данном случае нельзя исключать возможности существования единого оперона *purHD-glyA* (см. рис. 3,а). И только в *A. actinomycetemcomitans* перед геном *glyA* значимого PurR-сайта не найдено.

Продукт гена *cerA*, фермент фосфоглицератдегидрогеназа, участвует в начальных стадиях биосинтеза серина. Этот ген имеет ортологов во всех рассмотренных геномах. Значимый PurR-сайт перед геном *serA* обнаружен в геномах *E. coli*, *S. typhi*, *Y. pestis* и *V. cholerae* (см. таблицу). В геномах *H. influenzae*, *P. multocida* и *A. actinomycetemcomitans* перед *serA* на расстоянии 18–30 пн располагается ген *rpiA*, что допускает существование оперона *rpiA-serA* (см. рис. 3,е). Ген *rpiA* кодирует фермент, синтезирующий рибозо-5-фосфат, необходимый для синтеза пуринов (см. рис. 1). Во всех этих геномах перед *rpiA-serA* обнаружен сильный PurR-сайт (см. таблицу). Таким образом, можно думать, что пуриновый репрессор регулирует транскрипцию *serA* в *Enterobacteriaceae* и *V. cholerae* и *rpiA-serA* в *Pasteurellaceae*.

Группа 5 – гены, участвующие в азотном обмене. Группа включает 3 гена *E. coli*: *speA*, *speB* и *glnB*. Известно, что транскрипция всех этих генов находится под регуляцией PurR в *E. coli* [4] и что *speAB* является полицистронным опероном в *E. coli* [1].

PurR-сайт перед опероном *speAB* найден в *E. coli*, *S. typhi* и *P. multocida*. В *Y. pestis* не обнаружен ортолог *speA*, а перед ортологом *speB* нет значимого PurR-сайта (см. таблицу). В *V. cholerae* найден оперон, ортологичный *speAB*, однако значимый PurR-сайт обнаружен только в конце гена *speA* перед геном *speB* (см. рис. 3,ж), что в принципе допускает регуляцию транскрипции гена *speB* пуриновым репрессором. В *H. influenzae* и *A. acti-*

nomycetemcomitans оперон *speAB* утрачен полностью.

Перед кодирующим регуляторный белок геном *glnB* в геноме *E. coli* найден сильный PurR-сайт, лежащий, однако, в конце кодирующей области предыдущего гена (см. рис. 3,з). Ортологи *glnB* есть во всех рассмотренных геномах, но значимый PurR-сайт найден только в геноме *P. multocida* и, так же как и *E. coli*, этот сайт расположен в кодирующей области предыдущего гена (см. рис. 3,з). Скорее всего, ген *glnB* не является консервативным членом пуринового регулона, и наличие сайтов перед ним в геномах *E. coli* и *P. multocida* следует рассматривать как исключение.

Группа 6 – гены транспортных белков. В настоящее время экспериментально показан пуриновая регуляция только одного гена транспортного белка – *codB* (см. выше). С другой стороны, в работе [2] было предсказано, что в пуриновый регулон входят гены порина *tsx*, а также транспортных белков *yieG*, *yjcD* и *yicE*. Кроме того, в настоящей работе получены данные в пользу того, что ген урацилового транспортера *uraA* также входит в пуриновый регулон (см. выше).

Для всех описанных выше транспортеров, а также их паралофов было построено филогенетическое дерево (см. рис. 2,в). Как и следовало ожидать, дерево имеет три ветви, соответствующие трем семействам транспортных белков: NCS1, NCS2 и YICO/YIEG/YJCD, по классификации, принятой в базе данных SWISS-PROT [34].

Так как гены транспортных белков дублированы в большинстве геномов, ортологичность была установлена по сохранению регуляции или же по оперонной структуре.

Семейство NCS1 в рассмотренных геномах представлено транспортером цитозина – *CodB* и его ортологами, а также гипотетическим транспортером аллантаина AIP из *E. coli*, не имеющим ортологов в остальных геномах. Как обсуждалось выше, ген *codB* относится к пуриновому регулону, в то время как перед *allP* отсутствует значимый PurR-сайт (см. таблицу).

Семейство NCS2 представлено *UraA*, его ортологами, *YicE* и его ортологами, а также гипотетическими белками *YgfU* и *YgfO*. Пуриновая регуляция транскрипции *yicE* в геноме *E. coli* была предсказана ранее [2]. В настоящей работе показано, что ортологи гена *yicE* есть в *S. typhi*, *Y. pestis* и *V. cholerae*. Перед всеми этими генами были найдены значимые PurR-сайты (см. таблицу). Значимый PurR-сайт найден и перед *ygfU* *E. coli*, однако его нельзя подтвердить, так как в рассмотренных геномах нет ортологов *ygfU*. Интересно, что в геномах *E. coli*, *S. typhi* и *Y. pestis* ген *yicE* образует дивергон с геном Na⁺-зависимого транспортера глутамата *gltS*. Наличие PurR-сайта между этими генами допускает регуляцию транскрип-

ции *gltS* в *Enterobacteriaceae*. С другой стороны, в геномах, где *gltS* располагается отдельно, перед ним отсутствует PurR-сайт.

В рассмотренных геномах семейство гипотетических транспортеров YICO/YIEG/YJCD представлено генами *yjcD*, *yieG*, *yicO* и *ygfQ/R* *E. coli*, а также ортологами *yjcD* и *yieG* из других геномов. Как видно из рис. 2, в PurR-сайт есть перед всеми ортологами *yjcD*. PurR-сайт перед *yieG* сохраняется в *S. typhi*, *V. cholerae* и *P. multocida*, но пропадает в *Y. pestis*. Ген *yicO*, по-видимому, является одиночным геном или первым геном в опероне. Значимый PurR-сайт перед этим геном не найден. Ген *ygfQ/R* вместе с геном *egfO* из семейства NCS2 входит в генный кластер, имеющий отношение к окислительному метаболизму гуанина. Хотя перед кластером имеется значимый PurR-сайт, обсуждать регуляцию данного кластера не представляется корректным, так как он не имеет реальных ортологов в остальных рассмотренных геномах. В состав подобных кластеров, не имеющих ортологов в других геномах, входят также гены *ycdG* и *ygfU* из семейства NCS2.

Ген *tsx* кодирует нуклеозидспецифичный канал внешней мембраны *E. coli*. Среди изученных геномов только в геноме *S. typhi* есть ортолог *tsx*, перед ним найден PurR-сайт (см. таблицу).

Таким образом, обсуждаемые в этом разделе данные подтверждают ранее сделанное предсказание о том, что гены *yieG*, *yjcD* и *yicE* входят в пуриновый регулон [2].

Группа 7 – гены белков с неизвестной функцией. В настоящей работе значимый PurR-сайт обнаружен перед тремя генами с неизвестной функцией – *yhhQ*, *ydiJ* и *ydiK*.

Во всех рассмотренных геномах найдены ортологи *yhhQ* а в *S. typhi*, *Y. pestis*, *V. cholerae* и *P. multocida* перед геном присутствует значимый PurR-сайт (см. таблицу).

В геномах *E. coli*, *S. typhi* и *Y. pestis* обнаружен значимый PurR-сайт между генами *ydiJ* и *ydiK*, образующими дивергон. В геномах *V. cholerae*, *H. influenzae*, *P. multocida* и *A. actinomycetemcomitans* ортологов *ydiK* нет, а перед геном *ydiJ* нет значимого PurR-сайта (см. таблицу). Тем самым есть основание полагать, что PurR-сайт между генами *ydiJ* и *ydiK* в *Enterobacteriaceae* относится к гену *ydiK*.

Таким образом, не исключено, что транскрипция генов с неизвестной функцией – *yhhQ* и *ydiK* регулируется пуриновым репрессором.

В геноме *E. coli* были обнаружены также PurR-сайты с весом более 3.7 в EBG-системе перед геном бета-галактозидазы – *ebgA*, между генами дивергона *ygeV–ygeW*, продукт последнего похож на орнитинкарбамоилтрансферазу, перед геном гипотетического транскрипционного регулятора *ydjF*, перед геном гипотетического ассоциирован-

ного с мембраной белка *ydeK*, перед геном гипотетического трансмембранного белка *yjeO* и перед генами с неизвестной функцией *ydaU*, *yhaB* и *yhiE*. Все эти гены не имеют ортологов ни в одном из рассмотренных геномов, и поэтому их принадлежность к пуриновому регулону нельзя ни подтвердить, ни опровергнуть в рамках данного исследования.

В соответствии с использованной процедурой (см. Материалы и методы), к пуриновому регулону следовало бы отнести и ген рибонуклеазы *T-rnt*, в котором PurR-сайт расположен в самом начале кодирующей области (данные см. в таблице). Однако в этом случае консервативность сайта в разных геномах, по-видимому, связана с консервативностью белок-кодирующей области.

СРАВНЕНИЕ СТРУКТУРЫ PurR-САЙТА В ПЯТИ ГЕНОМАХ ГАММА-ПРОТЕОБАКТЕРИЙ

Как видно из рис. 2, филогенетическое дерево для гомологов пуринового репрессора содержит три ветви, содержащие строгих ортологов белка из *E. coli*, соответствующих PruR энтеробактерий, PurR группы *Pasteurellaceae*, и пуриновому репрессору из *V. cholerae*. Поэтому представляло интерес проверить, насколько изменяется пуриновый сигнал в различных геномах.

Для этого были построены матрицы позиционных весов для PurR-сайтов из геномов *E. coli*, *Y. pestis*, *H. influenzae*, *P. multocida* и *V. cholerae* (см. рис. 4). В качестве обучающих выборок были взяты PurR-сайты перед всеми генами, принадлежащими к предсказанным пуриновым регулонам.

Как видно из рис. 4, структура палиндромного PurR-сайта сохраняется без существенных изменений в *Y. pestis*, в то время как в геномах *H. influenzae*, *P. multocida* и *V. cholerae* наблюдается тенденция к заменам во второй и пятнадцатой позициях сайта. Поскольку эти заметы нуклеотидов происходят одновременно в двух симметричных позициях, вероятность того, что они являются следствием случайных флуктуаций, мала.

Остальные позиции сайта являются довольно консервативными. Таким образом, структура PurR-сигнала может изменяться, что необходимо учитывать при изучении пуриновой регуляции в различных геномах.

Работа была частично поддержана грантами INTAS (№ 99-1476) и ННМИ (№ 55000309).

СПИСОК ЛИТЕРАТУРЫ

1. *Neidhardt F.C.* Escherichia coli and Salmonella. Cellular and Molecular Biology. Washington: ASM Press, 1996. V. 1. Ch. 34. P. 561–579.

2. Mironov A.A., Koonin E.V., Roytberg M.A., Gelfand M.S. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes // Nucl. Acids Res. 1999. V. 27. № 14. P. 2981–2989.
3. Meng L.M., Nygaard P. Identification of hypoxanthine and guanine as the co-repressors for the purine regulon genes of *Escherichia coli* // Mol. Microbiol. 1990. V. 4. № 12. P. 2187–2192.
4. He B., Choi K.Y., Zalkin H. Regulation of *Escherichia coli* *glnB*, *prsA*, and *speA* by the purine repressor // J. Bacteriol. 1993. V. 175. № 11. P. 3598–3606.
5. Choi K.Y., Zalkin H. Regulation of *Escherichia coli* *purC* by the purine regulon repressor protein // J. Bacteriol. 1990. V. 172. № 6. P. 3201–3207.
6. Rolfes R.J., Zalkin H. *Escherichia coli* gene *purR* encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence, and interaction with the *purF* operator // J. Biol. Chem. 1988. V. 263. № 26. P. 19649–19652.
7. Steiert J.G., Rolfes R.J., Zalkin H., Stauffer G.V. Regulation of the *Escherichia coli* *glyA* gene by the *purR* gene product // J. Bacteriol. 1990. V. 172. № 7. P. 3799–3803.
8. Steiert J.G., Kubu C., Stauffer G.V. The PurR binding site in the *glyA* promoter region of *Escherichia coli* // REMS Microbiol. Lett. 1992. V. 78. № 2–3. P. 299–304.
9. Lorenz E., Stauffer G.V. RNA polymerase, PurR and MetR interactions at the *glyA* promoter of *Escherichia coli* // Microbiology. 1996. V. 142. № 7. P. 1819–1824.
10. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
11. Blattner F.R., Plunkett G., Bloch C.A. et al. The complete genome sequence of *Escherichia coli* K-12 // Science. 1997. V. 277. № 5331. P. 1453–74.
12. Parkhill J., Wren B.W., Thomson N.R. et al. Genome sequence of *Yersinia pestis*, the causative agent of plague // Nature. 2001. V. 413. № 6855. P. 523–527.
13. Cole S.T., Eiglmeier K., Parkhill J. et al. Massive gene decay in the leprosy bacillus // Nature. 2001. V. 409. № 6823. P. 1007–1011.
14. Shigenobu S., Watanabe H., Hattori M. et al. Genome sequence of the endocellular symbiont of aphids *Buchnera* sp. APS // Nature. 2000. V. 407. № 6800. P. 81–86.
15. <http://aeg.lbi.ic.unicamp.br/xf/>
16. <http://genome.wustl.edu/gsc/>
17. <http://www.sanger.ac.uk/Projects/Microbes/>
18. <http://www.genome.ou.edu/act.html>
19. http://spider.jgi-psf.org/JGI_microbial/html/
20. <http://www.tigr.org/cgi-bin/>
21. He B., Zalkin H. Repression of *Escherichia coli* *purB* is by a transcriptional roadblock mechanism // J. Bacteriol. 1992. V. 174. № 22. P. 7121–7127.
22. He B., Smith J.M., Zalkin H. *Escherichia coli* *purB* gene: cloning, nucleotide sequence, and regulation by *purR* // J. Bacteriol. 1992. V. 174. № 1. P. 130–136.
23. Миронов А.А., Винокурова Н.П., Гельфанд М.С. Программное обеспечение анализа бактериальных геномов // Молекуляр. биология. 2000. Т. 34. № 2. С. 253–262.
24. <http://www.ncbi.nlm.nih.gov/BLAST/>
25. <http://www.ig:8082/WIT2/>
26. Thompson J.D., Higgins D.G., Gibson T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice // Nucl. Acids Res. 1994. V. 22. № 22. P. 4673–4680.
27. Rolfes R.J., Zalkin H. Autoregulation of *Escherichia coli* *purR* requires two control sites downstream of the promoter // J. Bacteriol. 1990. V. 172. № 10. P. 5758–5766.
28. He B., Zalkin H. Regulation of *Escherichia coli* *purA* by purine repressor, one component of a dual control mechanism // J. Bacteriol. 1994. V. 176. № 4. P. 1009–1013.
29. Meng L.M., Kilstrup M., Nygaard P. Autoregulation of PurR repressor synthesis and involvement of *purR* in the regulation of *purB*, *purC*, *purL*, *purMN* and *guaBA* expression in *Escherichia coli* // Eur. J. Biochem. 1990. V. 187. № 2. P. 373–379.
30. He B., Shiau A., Choi K.Y. et al. Genes of the *Escherichia coli* pur regulon are negatively controlled by a repressor-operator interaction // J. Bacteriol. 1990. V. 172. № 8. P. 4555–4562.
31. Galperin M.Y., Koonin E.V. Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operon disruption // In Silico Biol. 1998. V. 1. № 1. P. 55–67.
32. Wilson H.R., Turnbough C.L., Jr. Role of the purine repressor in the regulation of pyrimidine gene expression in *Escherichia coli* K-12 // J. Bacteriol. 1990. V. 172. № 6. P. 3208–3213.
33. Cheng Y., Dylla S.M., Turnbough C.L., Jr. A long T. A tract in the upp initially transcribed regions is required for regulation of upp expression by UTP2-dependent reiterative transcription in *Escherichia coli* // J. Bacteriol. 2001. V. 183. № 1. P. 221–228.
34. <http://www.expasy.ch/sprot/>