

УДК 577.1

## РЕГУЛЯЦИЯ НИТРАТ-НИТРИТНОГО ДЫХАНИЯ ГАММА-ПРОТЕОБАКТЕРИЙ: ИССЛЕДОВАНИЕ МЕТОДАМИ СРАВНИТЕЛЬНОЙ ГЕНОМИКИ

© 2005 г. Д. А. Равчев<sup>1,2\*</sup>, А. Б. Рахманинова<sup>1,2</sup>, А. А. Миронов<sup>1,2,3</sup>, М. С. Гельфанд<sup>1,2,3</sup>

<sup>1</sup>Московский государственный университет им. М.В. Ломоносова, Москва, 119992

<sup>2</sup>Институт проблем передачи информации Российской академии наук, Москва, 127994

<sup>3</sup>Государственный научный центр Российской Федерации “ГосНИИгенетика”, Москва, 117545

Поступила в редакцию 09.03.2005 г.

Нитрат и нитрит – наиболее предпочтительные акцепторы электронов в отсутствие молекулярного кислорода. В организме гамма-протеобактерии *Escherichia coli* регуляция нитрат-нитритного дыхания осуществляется двумя гомологичными факторами транскрипции NarL и NarP. Данная регуляторная система интенсивно изучается в течение более чем двадцати лет, однако многие ключевые вопросы, в том числе и структура сайта связывания белка NarL, до сих пор остаются нерешенными. Применение методов сравнительной геномики позволило установить, что в большинстве организмов из группы гамма-протеобактерий регуляция осуществляется одиночным белком NarP. Изучение NarP-регулона в десяти геномах показало, что его структура в разных геномах может весьма различаться, однако механизм регуляции генов восстановления нитрата и нитрита является крайне консервативным. Обнаружена корреляция между эволюцией системы нитрат-нитритного дыхания и изменениями в структуре регуляторной системы. Обнаружение NarP-сайтов перед геном глобального регулятора FNR и генами аэробного метаболизма, *cydAB*, *mdh* и *sucAB*, указывает на то, что роль белка NarP в регуляции дыхания изменяется в процессе эволюции. Обнаружено 35 новых членов обобщенного регулона, установлена авторегуляция оперона *narQP* в организмах сем. Vibrionaceae.

**Ключевые слова:** сравнительная геномика, гамма-протеобактерии, нитрат-нитритное дыхание, NarP, NarL, NarX, NarQ.

COMPARATIVE GENOMICS ANALYSIS OF NITRATE AND NITRITE RESPIRATION IN GAMMA PROTEOBACTERIA, by D. A. Ravcheev<sup>1,2\*</sup>, A. B. Rakhmaninova<sup>1,2</sup>, A. A. Mironov<sup>1,2,3</sup>, M. S. Gelfand<sup>1,2,3</sup> (<sup>1</sup>Moscow State University, Department of Bioengineering and Bioinformatics, Moscow 119992, Russia, \*e-mail: ravcheev@iitp.ru; <sup>2</sup>Institute For Information Transmission Problems Russian Academy of Sciences, Moscow, 127994, Russia; <sup>3</sup>State Scientific Center “GosNII Genetika”, Moscow 117545, Russia). Nitrate and nitrite are preferred respiration oxidants during anaerobic conditions. In *Escherichia coli* such nitrate- and nitrite respiration is controlled by homologous transcriptional factors NarL and NarP. Although this system was intensively studied during the last two decades, the exact mechanisms of regulation and the structure of the NarL binding signals remained elusive. By the use of comparative genomics approach it was determined that most of the gammaproteobacteria contained only NarP protein. Regulog analysis revealed that whole structure of NarP regulons varied in different genomes and only regulation of nitrate and nitrite reduction system seemed to be highly conservative. Correlation between changes in the respiration system and the presence of the single regulatory system was shown. Conservative NarP binding sites upstream of *fnr* gene and genes for aerobic metabolism point to alteration in NarP role in respiration control during evolution. Thirty five new regulog members were determined and autoregulation of *narQP* operon in Vibrionaceae genomes was predicted.

**Key words:** comparative genomics, gamma-proteobacteria, nitrate and nitrite respiration, NarP, NarL, NarQ, NarX.

Многие прокариотические организмы, например, *Escherichia coli*, могут использовать различные субстраты для дыхания, что позволяет им существовать в широком диапазоне условий. В отсутствие

кислорода клетки “предпочитают” получать энергию за счет нитрат-нитритного дыхания [1]; этот тип дыхания регулируется в *E. coli* с помощью удвоенной двухкомпонентной системы, включающей гомологичные сенсорные белки NarQ и NarX и гомологичные факторы транскрипции NarL и NarP (Nar – Nitrate reductase regulator). Полагают,

Принятые сокращения: ЦТК – цикл трикарбоновых кислот; ТМОА – триметиллоксид азота.

\* Эл. почта: ravcheev@iitp.ru

что такая регуляторная система необходима для настройки системы дыхания в присутствии двух альтернативных акцепторов электронов: нитрата и нитрита [2]. Оба упомянутых выше фактора транскрипции могут быть как репрессорами, так и активаторами – в зависимости от расположения сайтов их связывания относительно промотора регулируемого оперона [1].

Для данной регуляторной системы характерно частичное разделение функций между гомологичными белками (рис. 1). Так, NarX взаимодействует только с белком NarL, при этом по-разному реагируя на нитрат и нитрит, тогда как NarQ способен к взаимодействию с обоими регуляторами и одинаково реагирует на оба субстрата [3]. С другой стороны, белок NarP не связывается с сайтами связывания NarP, тогда как NarL может взаимодействовать и с сайтами связывания NarP, и со своими собственными [4] (рис. 1). Таким образом, в случае *E. coli* функционирует единый NarL-NarP-регулон, который разделен нами на следующие группы генов.

1. Гены, обеспечивающие восстановление нитрата и нитрита: *narGHJI*, *narK* [5], *napFDABCDEF-ccmABCDEFGH* [6], *nrfABCDEFG*, *nirBDC-cysG* [7]; экспрессия этих генов регулируется в зависимости от соотношения нитрата и нитрита в среде.

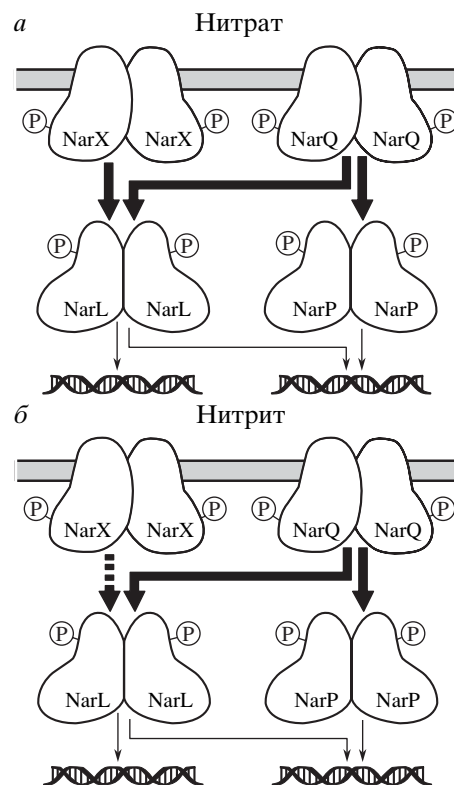
2. Гены дегидрогеназ доноров электронов: *fdnGHI* [8], *hyaABCDE*, *hybOABCDEFG* [9], *nuoABCDEFGHIJKLMN* [10]. Поскольку продукты данных генов необходимы для формирования полных дыхательных цепей, их экспрессия активируется белками NarL и NarP.

3. Гены редуктаз акцепторов электронов: *dmsABC* [11], *torCAD* [12], *frdABCD* [5]; они кодируют редуктазы акцепторов электронов, использующихся в анаэробных условиях в отсутствие нитрата и нитрита. В присутствии последних экспрессия данных генов репрессируется.

4. Гены метаболизма четырехуглеродных дикарбоновых кислот: *dcuB-fumB* [13]. Продукты этих генов – ферменты и белки-транспортёры, которые поставляют субстраты для фумаратного дыхания, гораздо менее эффективного, чем нитрат-нитритное. Поэтому эти гены репрессируются факторами NarL и NarP.

5. Гены ферментов брожения: *focA-pflB* [14], *adhE* [15]. Поскольку брожение является менее выгодным способом получения энергии, чем дыхание, эти гены также репрессируются в присутствии нитрата и нитрита.

Несмотря на то, что регуляцию нитрат-нитритного дыхания активно исследуют в течение последних двадцати лет, на многие важные вопросы до сих пор нет ответа. Целью данной работы стало изучение регуляции нитрат-нитритного дыхания в группе гамма-протеобактерий. сосре-



**Рис. 1.** Система дифференцированного ответа на нитрат и нитрит в *E. coli*: а – в присутствии нитрата; б – в присутствии нитрита. Участки связывания регуляторных белков с ДНК показаны тонкими стрелками; взаимодействие между сенсорными и регуляторными белками – жирными. Сплошные стрелки – фосфорилирование регуляторных белков, пунктирные – дефосфорилирование.

доточив свое внимание на организмах, в которых имеется только один регулятор NarP, мы исследовали структуру NarP-регулонов в геномах десяти гамма-протеобактерий и описали обобщенный для исследованных геномов NarP-регулон.

## УСЛОВИЯ ЭКСПЕРИМЕНТА

**Процедура определения генов, входящих в обобщенный регулон.** Для поиска потенциальных сайтов связывания NarP (далее – “потенциальные сайты”) использовали метод матриц позиционных весов нуклеотидов [16], метод построения которой описан в разделе “Результаты”.

Для определения принадлежности гена к регулону применяли метод проверки соответствия. Ген рассматривается как член регулона, если его регуляторной области обнаружен потенциальный сайт связывания, который сохраняется перед ортологичными генами в родственных геномах [16]. В данной работе такие сайты искали в области –400...+100 п.н. относительно старта трансляции гена. В случае, когда оперонная структура

данного фрагмента последовательности неизвестна, полагали, что гены принадлежат к одному оперону, если они имеют одинаковое направление считывания, а расстояние между ними не превышает 100 п.н.

В работе применяли новый метод выявления членов обобщенного регулона, основанный на попарном сравнении геномов организмов, относящихся к одной таксономической группе. В каждом из геномов производили поиск потенциальных NarP-сайтов. Далее, если сайты перед ортологичными генами сохраняются, то вычленили потенциальные члены обобщенного регулона. Если сайт перед геном сохраняется, как минимум, в трех геномах организмов из одной группы, то ген рассматривался как член обобщенного регулона. В случае таких генов проверяли также, имеются ли такие сайты перед генами-ортологами в организмах из других групп. Этот подход использовали при изучении геномов бактерий семейств Pasteurellaceae и Vibrionaceae.

Поскольку в данной работе исследовали только геномы двух организмов из группы Enterobacteriaceae – *Y. pestis* и *Y. enterocolitica*, то описанный выше подход не мог быть использован. В данном случае ген рассматривали как возможный член обобщенного регулона, если потенциальные сайты перед ним обнаруживались в обоих геномах. В таком случае проверяли наличие сайтов перед его ортологами в других геномах, и, если таковые удавалось обнаружить, ген также рассматривали как член обобщенного регулона.

Некоторые ферменты, задействованные в системе дыхания, существуют в нескольких изоформах, которые являются продуктом паралогичных генов. Регуляция таких паралогов рассматривается отдельно (см. раздел “Обсуждение результатов”).

**Программное обеспечение.** Для поиска потенциальных сайтов связывания NarP, а также анализа ортологов использовали пакет программ Gene-Explorer [17], а для построения матриц позиционных весов – программу SignalX [17]. Поиск гомологов в базах данных проводили с помощью программы BLAST [18]. Для множественного выравнивания аминокислотных и нуклеотидных последовательностей использовали программу ClustalX [19]. Филогенетические деревья строили по методу максимального правдоподобия (Maximum Likelihood Method) [20] с помощью программы proml из программного пакета PHYLIP 3.63 (<http://evolution.genetics.washington.edu/>). Для построения диаграмм Лого, отображающих структуру регуляторного сигнала, использовали программу WebLogo [21] (<http://weblogo.berkeley.edu/>).

**Объекты исследования.** Изучали 13 геномов различных организмов, относящихся к группе гамма-протеобактерий. Полные последовательности геномов *Escherichia coli* K-12 (EC) [22], *Sal-*

*monella typhi* Ty2 (ST) [23], *Erwinia carotovora* s. *atroseptica* (EO) [24], *Yersinia pestis* KIM (YP) [25], *Haemophilus ducreyi* 35000HP (HD), *Haemophilus influenzae* Rd (HI) [26], *Pasteurella multocida* Pm70 (PM) [27], *Vibrio cholerae* O1 (VC) [28], *Vibrio parahaemolyticus* RIMD 2210633 (VP) [29] и *Vibrio vulnificus* CMCP6 (VV) [30] взяты из базы данных GenBank [31]. Полная последовательность генома *Yersinia enterocolitica* (YE) взята с интернет-сайта (<http://www.sanger.ac.uk/>). Предварительная последовательность генома *Actinobacillus actinomycetemcomitans* HK1651 (AA) также взята с интернет-сайта (<http://www.genome.ou.edu/>). Предварительная последовательность генома *Vibrio fischeri* ES114 (VF) взята из базы данных GOLD [32].

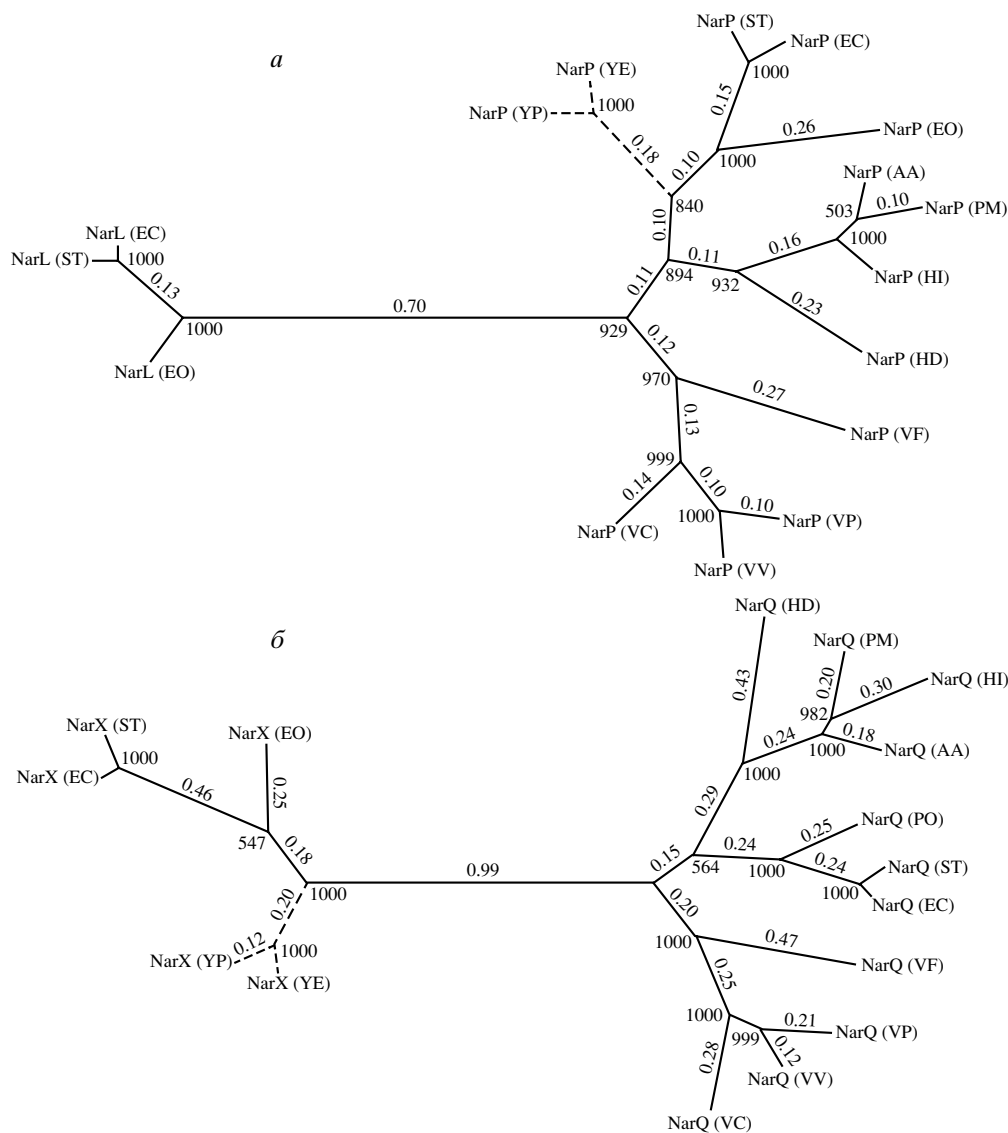
## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

### Эволюция регуляторной системы

На первом этапе исследования производили поиск ортологичных белков регуляторной системы *E. coli* в геномах организмов трех бактериальных семейств: Enterobacteriaceae (*S. typhi*, *E. carotovora*, *Y. pestis*, *Y. enterocolitica*), Pasteurellaceae (*P. multocida*, *A. actinomycetemcomitans*, *H. ducreyi*, *H. influenzae*) и Vibrionaceae (*V. cholerae*, *V. parahaemolyticus*, *V. vulnificus*, *V. fischeri*). Удвоенная регуляторная система обнаружена лишь в геномах организмов *S. typhi* и *E. carotovora*, являющихся ближайшими родственниками *E. coli*. На филогенетических деревьях для белков NarL/NarP (рис. 2а) и NarQ/NarX (рис. 2б) белки из различных организмов образуют два кластера, разделенных длинной веткой. По-видимому, дупликация двухкомпонентной системы произошла до расхождения эволюционных ветвей, соответствующих семействам Enterobacteriaceae, Pasteurellaceae и Vibrionaceae. В тех геномах, где обнаруживается лишь одиночная система, ген *narP* существует совместно с геном *narQ*. Исключение составляют геномы *Y. pestis* и *Y. enterocolitica*, в которых обнаружена нестандартная пара NarP-NarX (рис. 2а, б). Поэтому аминокислотные последовательности белков NarX из этих бактерий рассмотрены более подробно.

Ранее показано, что дифференцированный ответ на нитрат и нитрит осуществляется за счет периплазматического домена NarX [33] (рис. 3а). Тем не менее, очевидных различий между последовательностями периплазматических модулей белков NarX и NarQ выявить не удалось [34].

С другой стороны, особенности строения центрального модуля (рис. 3а) позволяют достаточно точно разграничить NarX и NarQ. Центральный модуль имеет аминокислотную последовательность, гомологии с которой не найдено нигде, кроме белков NarX и NarQ. Функция данного домена остается неизвестной. Важной особеннос-



**Рис. 2.** Филогенетические деревья для белков NarL/NarP (а) и NarX/NarQ (б). Деревья построили по методу максимального правдоподобия (Maximum Likelihood Method). Условные обозначения для геномов – см. “Условия эксперимента”. Ветви, соответствующие белкам из *Y. pestis* и *Y. enterocolitica*, показаны пунктиром. Цифры на ветвях обозначают ожидаемую долю аминокислотных замен; цифры в узлах – статистическую значимость данного узла (percentage of bootstrap replications; максимум – 1000).

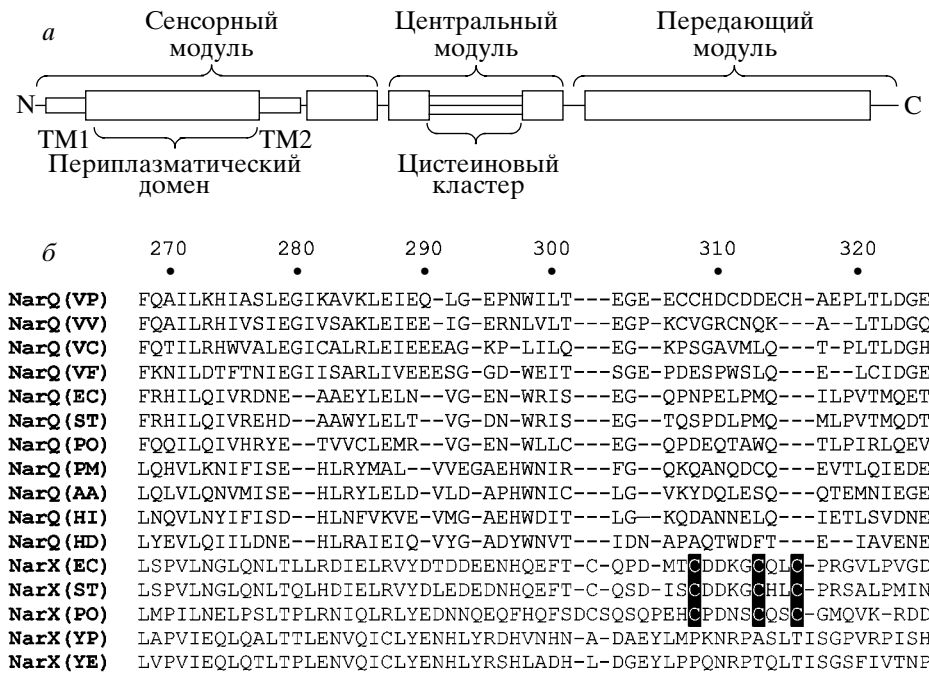
тью модуля NarX *E. coli* является кластер цистеинов в позициях 308, 313 и 316 NarX. Эти цистеиновые остатки сохраняются в последовательностях NarX из геномов других протеобактерий, и замена любого из них приводит к резкому снижению активности белка [34].

На рис. 3 представлено множественное выравнивание белков NarX и NarQ всех организмов, рассмотренных в данной работе (рис. 3б). Как видно, в последовательностях белков из *Y. pestis* и *Y. enterocolitica* отсутствуют характерные для NarX цистеиновые остатки. Таким образом, хотя сенсорные белки из бактерий рода *Yersinia* близки по последовательности к NarX, они, по всей вероятности, вы-

полняют функцию NarQ, т.е. реакция в ответ на нитрат и нитрит одинакова.

#### Построение матрицы для поиска потенциальных сайтов связывания NarP

Поскольку данные о структуре сайтов связывания NarL весьма противоречивы [35, 36], а наличие удвоенной регуляторной системы крайне затрудняет изучение регуляции, мы сосредоточили наше внимание на изучении NarP-зависимой регуляции в десяти геномах, в которых присутствует ген *narP*, но отсутствует *narL*.



**Рис. 3.** а – Структура сенсорных белков NarX и NarQ [34]; TM – трансмембранные сегменты. б – Фрагмент выравнивания аминокислотных последовательностей центрального модуля белков NarX и NarQ. Сверху указаны позиции аминокислот для NarX *E. coli*. Консервативные цистеиновые остатки зачернены. Условные обозначения для геномов – см. “Условия эксперимента”.

Нами установлено, что ген *narP* всегда обнаруживается в геномах, имеющих гены для периплазматических нитрат- и нитрит-редуктаз (соответственно, *nar* и *nrf*) и гены экспорта гема в периплазму (*ccm*), в которых имеются различные оперонные перестройки (табл. 1). Исключение составляют геномы *Y. pestis*, *Y. enterocolitica* и *V. cholerae*, в которых *nrf*-генов нет. В случае *E. coli* показано, что регуляция экспрессии всех этих генов осуществляется белком NarP [6, 7]. Сайт связывания NarP

представляет собой палиндром с консенсусом TACYUMTNNAKRRGTA [4], и сайты такого вида обнаруживаются в различных геномах перед оперонами, содержащими гены *nar*, *nrf* и *ccm* (табл. 2). Нами составлена обучающая выборка и построены матрица для поиска сайтов связывания NarP и диаграмма Лого (рис. 4).

В геномах проводили поиск сайтов с весом выше порогового значения 3.50. Однако в геноме *A. actinomycetemcomitans* перед генами – потенци-

**Таблица 1.** Оперонная структура для генов *nar*, *ccm* и *nrf*

Геномы	Опероны		
	<i>nar</i>	<i>ccm</i>	<i>nrf</i>
<i>Y. pestis</i> , <i>Y. enterocolitica</i>	<i>narFDABC</i>	<i>ccmABCDEFGH</i>	Отсутствует
<i>P. multocida</i>	<i>narFDAGHBC</i>	<i>ccmABCDEFGH</i>	<i>nrfABCDEFGH</i>
<i>A. actinomycetemcomitans</i>	<i>narDAGHBC</i>	<i>ccmABCDEFGH</i>	<i>nrfABCDEXFGH</i>
<i>H. influenzae</i>	<i>narFDAGHBC</i>	<i>ccmABCDEFGH</i>	<i>nrfABCD nrfEXFGH</i>
<i>H. ducreyi</i>	<i>narFDAGHBC</i>	<i>ccmABCDEF ccmGH</i>	<i>nrfABCD nrfEXFGH</i>
<i>V. vulnificus</i> , <i>V. parahaemolyticus</i>	<i>narFDABC</i>	<i>ccmABCDEFGH</i>	<i>nrfA↔nrfBCDEXF</i> <i>nrfG</i>
<i>V. cholerae</i>	<i>narFDABC</i>	<i>ccmABCDEFGH</i>	Отсутствует
<i>V. fischeri</i>	<i>narFDABC</i>	<i>ccmABCDEFGH</i>	<i>nrfA↔nrfBCDEXF</i> <i>nrfG</i>

Примечание: “↔” – гены образуют дивергон.

**Таблица 2.** Потенциальные сайты связывания NarP, вошедшие в обучающую выборку (см. пояснения в тексте). Указано положение сайта относительно начала рамки считывания

Геном	Оперон	Сайт	Вес	Положение
<i>Y. pestis</i>	<i>napFDABC</i>	T a A C T C T a a A G A G T a A	4.48	-184
	<i>ccmABCDEFGH</i>	T A C C t c T A T A a g G G T A	4.67	-113
<i>Y. enterocolitica</i>	<i>napFDABC</i>	T a A C T C T a a A G A G T a A	4.48	-182
	<i>ccmABCDEFGH</i>	T A C C t c T A T A a g G G T A	4.67	-113
<i>P. multocida</i>		T A c T a C g T A a G g A t T A	3.83	-79
	<i>ccmABCDEFGH</i>	c A C C T C T a a A G A G G T a	4.80	-99
	<i>nrfABCDEFXG</i>	T A a C T c T a a A t A G g T A	4.78	-289
<i>A. actinomycetemcomitans</i>		T A C t T a T t t A g A g G T A	4.71	-206
	<i>nrfABCDEFXG</i>	T A C t a A T a a A T a g G T A	4.32	-343
<i>H. influenzae</i>	<i>ccmABCDEFGH</i>	T A c t T C T a a A G A g t T A	4.79	-85
<i>H. ducreyi</i>	<i>napFDAGHBC</i>	T a c C T a a A T g g A G t g A	3.62	-81
	<i>ccmABCDEFGH</i>	T A C c T C T a a A G A a G T A	4.94	-86
<i>V. vulnificus</i>		T A C t a T a a a A a g G T A	3.82	-117
	<i>napFDABC</i>	T A C C C C c T A a G G G G T A	4.71	-126
	<i>ccmABCDEFGH</i>	T c a C T C T A T A G A G g t A	4.49	-147
	<i>nrfBCDEFX</i>	T A C C C C T a a A G G G G T A	4.91	-304
<i>V. parahaemolyticus</i>	<i>nrfA</i>	T A C C C C T t t A G G G G T A	4.91	-193
	<i>napFDABC</i>	T A C C T C c T A a G A G G T A	4.97	-124
	<i>napGH</i>	T A C C T C t T A t G A G G T A	4.88	-82
	<i>nrfBCDEFX</i>	T A C C c C T a a A G t G G T A	4.88	-300
		T a c c T C t T A g G A a t a A	4.21	-196
<i>V. cholerae</i>	<i>nrfA</i>	T A C C a C T t t A G g G G T A	4.88	-190
		T t a t T C c T A a G A g g t A	4.21	-294
	<i>napF</i>	T A C C T C c T A a G A G G T A	4.97	-79
		c A c C t C T t t A G g G c T a	4.16	-153
<i>V. fischeri</i>	<i>ccmABCDEFGH</i>	T A A C t a c a a a g g G T T A	4.23	-143
	<i>napF</i>	T A a C C a t T A t g G G g T A	4.23	-85
	<i>ccmA</i>	T A C C t a c a a a g t G G T A	4.51	-190
	<i>nrfA</i>	T A a C a C T T A A G a G g T A	4.93	-234
		T A C c t c T T A A t t a G T A	4.62	-144
		t A C C t c T t t A a t G G T g	4.33	-218
	<i>nrfB</i>	T A c C t C T T A A G t G t T A	4.93	-253
	T A C t a a T T A A g a g G T A	4.62	-343	
	c A C C a t T a a A g a G G T a	4.33	-269	

альными членами регулона – зачастую находили лишь сайты с весом ниже принятого порогового значения. Поэтому для данного генома этой бактерии нами установлено пороговое значение 3.25. При этих условиях потенциальные сайты связывания NarP обнаруживаются в каждом геноме перед приблизительно четырьмя сотнями генов. Ясно, что отдельные предсказания таких сайтов недопустимы. Однако неоднократно было показано, что после применения процедуры проверки соот-

ветствия (см. “Условия эксперимента”) могут остаться лишь единичные ложные показания [37].

#### Структура обобщенного NarP-регулона в исследованных геномах

В соответствии с процедурой поиска членов обобщенного регулона, описанной в разделе “Условия эксперимента”, к нему можно отнести 77 генов, организованных, как минимум, в 29 оперо-



**Рис. 4.** Диаграмма Лого для сайтов связывания NarP. По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

нов. В исследованных геномах обобщенный NarP-регулон включает в себя почти все гены, входящие в объединенный NarL-NarP-регулон генома *E. coli* (табл. 3). Исключение составляют лишь некоторые гены, для которых в исследованных геномах ортологов не было найдено. Это гены из оперонов *narGHJI*, *narK*, *hya*, *hyb*, *fumB* и *nuo*. Ортологи генов оперона найдены лишь в геномах *Y. pestis* и *Y. enterocolitica*, однако перед ними не обнаружено значимых NarP-сайтов. Таким образом, регуляция дыхания в ответ на нитрат и нитрит в этих организмах осуществляется одиночной двухкомпонентной системой.

Несколько оперонов включено в обобщенный NarP-регулон на основании формального критерия, описанного в разделе “УСЛОВИЯ ЭКСПЕРИМЕНТА”, однако при анализе множественного выравнивания регуляторных областей этих

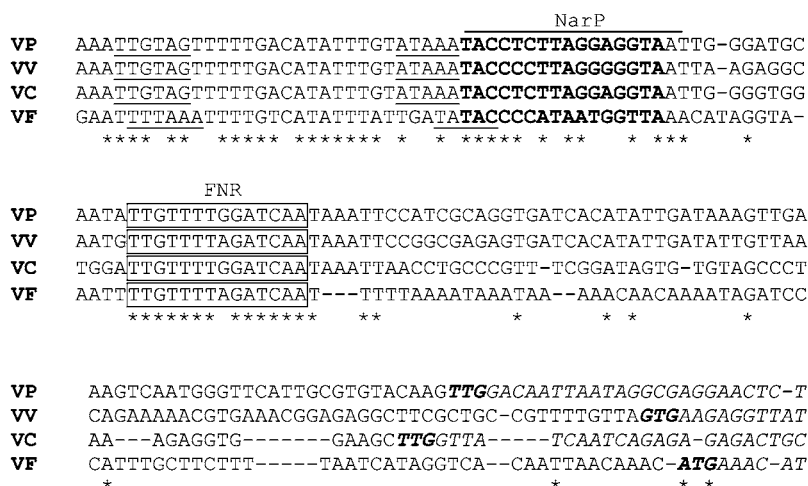
оперонов показано, что потенциальные сайты неконсервативны, и, следовательно, данные предсказания не могут рассматриваться как вполне достоверные. Таковыми оперонами являются *ldhA*, *hemR*, *gcvA* и *sucAB*. В то же время, экспериментальная проверка полученных результатов компьютерного анализа представляется нам вполне целесообразной.

За счет обнаружения новых членов обобщенного регулона к пяти функциональным группам генов, перечисленным в разделе “Введение”, добавлено еще две. Каждая из семи функциональных групп более подробно описана далее.

**Гены регуляторных белков.** Наиболее неожиданным оказался тот факт, что потенциальные NarP-сайты обнаруживаются перед генами белков, участвующих в регуляции дыхания. В геномах *P. multocida*, *A. actinomycetemcomitans*, *H. influenzae*, *V. cholerae* и *V. fischeri* такие сайты найдены перед геном *fnr*, продукт которого представляет собой глобальный регулятор дыхания, осуществляющий переключение между аэробным и анаэробным метаболизмом.

Во всех геномах группы Vibrionaceae потенциальные NarP-сайты располагаются перед опероном *narQP* в весьма консервативном участке регуляторной области после потенциальных промоторов. Другой консервативный участок содержит потенциальные сайты связывания белка FNR (рис. 5).

**Гены восстановления нитрата и нитрита.** Ядро обобщенного NarP-регулона составляют гены, непосредственно участвующие в восстановлении нитрата и нитрита. Это гены периплазматической нитратредуктазы (*nap*), периплазматической (*nrf*) и цитоплазматической (*nir*) нитритредук-



**Рис. 5.** Множественное выравнивание регуляторных областей оперона *narQP* геномов бактерий сем. Vibrionaceae. Условные обозначения для геномов – см. “Условия эксперимента”. Кодрующие области показаны курсивом, стартовые кодоны – жирным курсивом. Потенциальные сайты связывания NarP показаны жирным прямым шрифтом, FNR – выделены рамкой. Потенциальные –35 и –10 последовательности промоторов подчеркнуты. Консервативные позиции обозначены звездочками.

Таблица 3. Потенциальные сайты перед генами, входящими в обобщенный NarP-регулон

Оперон	Ссылка	Геном													
		Enterobacteriaceae				Pasteurellaceae				Vibrionaceae					
		YP	YE	PM	AA	HI	HD	VV	VP	VC	VF				
<i>fnr</i>	*	-	-	3.72 (-23)	3.41 (-30)	3.58 (-71)	-	-	-	4.73 (-188)	-	-	4.84 (-60)	3.90 (-176)	
<i>narQP</i>	*	-	-	-	-	-	-	-	-	4.99 (-139)	-	-	4.99 (-94)	4.25 (-117)	
<i>napFDAGHBC</i>	[6]	4.48 (-184)	4.48 (-182)	3.24 (-251)	3.47 (-135)	3.92 (-87)	3.62 (-81)	4.73 (-126) <sup>1)</sup>	4.99 (-124) <sup>1)</sup>	4.90 (-82) <sup>2)</sup>	4.17 (-153) <sup>1)</sup>	4.99 (-79) <sup>1)</sup>	4.25 (-85) <sup>1)</sup>		
<i>сstABCDEF</i>	[6]	4.67 (-113)	4.67 (-113)	4.80 (-99)	3.57 (-99)	4.79 (-85)	4.95 (-86)	4.48 (-147)	3.53 (-145)	4.92 (-304) <sup>3)</sup>	4.23 (-143)	4.89 (-300) <sup>3)</sup>	4.52 (-190)		
<i>nrfABCDEF</i>	[7]	0	3.83 (-179)	4.97 (-289)	4.33 (-343)	3.52 (-71)	3.32 (-182)	3.87 (-202) <sup>3)</sup>	4.22 (-196) <sup>3)</sup>	4.47 (-17)	0	4.47 (-17)	4.94 (-253) <sup>3)</sup>		
<i>nirBDC-cysG</i>	[7]	4.75 (-99)	4.75 (-101)	0	0	0	0	4.00 (-105)	4.00 (-107)	4.47 (-17)	0	4.47 (-17)	4.25 (-113)		
Гены восстановления альтернативных акцепторов электронов															
<i>torCAD</i>	[12]	0	0	4.14 (-268)	0	0	0	0	0	-	-	-	-	3.77 (-31)	
<i>torYZ</i>	*	0	4.35 (-179)	0	3.64 (-63)	4.13 (-41)	3.57 (-201)	0	4.29 (-61)	4.56 (-45)	4.46 (-67)	4.42 (-87)	4.42 (-87)		
<i>dmsABC</i>	[11]	3.90 (-178)	0	3.92 (-38)	3.68 (-44)	4.28 (-104)	0	0	0	0	0	0	0		
<i>frdABCD</i>	[5]	-	-	4.36 (-141)	4.20 (-270)	4.02 (-150)	-	3.54 (-85)	3.54 (-87)	4.36 (-134)	3.84 (-318)	4.07 (-118)	3.84 (-318)		
<i>cydAB</i>	*	4.02 (-372)	-	-	3.44 (-107)	3.60 (-133)	4.05 (-152)	4.59 (-389)	4.04 (-359)	3.52 (-101)	3.84 (-78)	3.52 (-101)	4.07 (-118)		
Гены дегидрогеназ доноров электронов															
<i>fdnGHI</i>	[8]	0	0	3.68 (-234)	3.51 (-111)	-	0	0	0	0	0	0	0	0	
<i>fdoGHI</i>	*	3.64 (-177)	3.64 (-167)	0	0	0	0	0	0	0	0	0	0	0	
<i>fdhD</i>	*	3.64 (-10)	3.64 (-16)	3.68 (-24)	3.44 (-10)	-	-	0	-	-	0	-	0	0	
<i>nqrABCDEF</i>	*	-	-	4.43 (-376)	4.36 (-351)	-	4.27 (-308)	-	-	-	-	-	-	3.69 (-251)	
				4.13 (-351)											



Таблица 3. Окончание

Оперон	Ссылка	Геном												
		Enterobacteriaceae					Pasteurellaceae					Vibrionaceae		
		YP	YE	PM	AA	HI	HD	VV	VP	VC	VF			
<i>dcuA</i>	*	-	-	4.06 (-105)	3.45 (-19)	3.79 (-76)	-	-	3.65 (-22)	-	-	3.71 (-34)		
<i>dcuB</i>	[13]	-	-	4.56 (-22)	4.64 (-56)	0	4.34 (-97)	-	-	-	-	3.60 (-67)		
<i>aspA</i>	*	-	-	3.65 (-105)	3.85 (-392)	-	3.63 (-66)	4.89 (-224)	4.26 (-226)	0	-	4.64 (-141)		
<i>fumC</i>	*	-	-	3.65 (-64)	4.51 (-20)	3.62 (-62)	-	0	-	-	-	0		
Гены ферментов, участвующих в брожении														
<i>pflB</i>	[14]	3.52 (-366)	3.70 (-22)	3.71 (-17)	3.71 (-17)	-	3.62 (-204)	4.11 (-23)	4.11 (-23)	-	-	4.22 (-270)		
<i>yfiD</i>	*	-	0	-	0	-	3.98 (48)	-	-	-	3.65 (-220)	0		
<i>adhE</i>	[15]	-	-	3.76 (-77)	4.13 (-50)	0	0	4.24 (-207)	3.98 (-191)	4.03 (-191)	-	3.56 (-234)		
<i>pgk</i>	*	-	-	3.57 (-68)	-	3.50 (-71)	3.60 (-22)	-	-	-	-	-		
<i>ldhA</i>	*	-	-	0	-	-	0	-	3.52 (-18)	3.89 (-153)	-	3.55 (-240)		
Гены ферментов цикла трикарбоновых кислот														
<i>mdh</i>	*	-	-	4.25 (-21)	3.29 (-18)	4.33 (-20)	3.91 (-26)	-	-	-	-	-		
<i>sucAB</i>	*	-	-	3.51 (-43)	-	3.94 (-21)	3.66 (-78)	-	-	-	-	-		
Гены синтеза молибдоптеринового кофактора														
<i>moaABCDE</i>	*	-	-	-	4.82 (-296)	3.59 (-323)	3.57 (-239)	3.89 (-255)	3.77 (-217)	4.19 (-233)	-	4.13 (-257)		
<i>gcvA</i>	*	3.75 (-364)	-	3.57 (-141)	3.48 (-165)	3.57 (-130)	-	-	-	-	-	-		
Гены метаболизма серы														
<i>cysJH</i>	*	-	-	0	0	0	0	4.22 (-99)	4.22 (-27)	-	-	3.84 (-78)		

Примечание. Условные обозначения для геномов: см. "Условия эксперимента". В графе "Ссылка" указаны номера ссылок на статьи, в которых показана принадлежность данных генов к *NarL-NarP*-регулону *E. coli*.  
 "\*" – Экспериментальные данные о принадлежности сайта относительно начала рамки считывания.  
 "-" – Сайт перед опероном отсутствует, "0" – ортолог данных генов не обнаружено. Оперонные перестройки (см. табл. 1) обозначены индексами: <sup>1)</sup> сайт перед опероном *parFDABC*; <sup>2)</sup> сайт перед опероном *parGH*; <sup>3)</sup> сайт между оперонами *nrfA* и *nrfBCDEF*, позиция указана относительно начала гена *nrfB*.

таз, а также гены, ответственные за экспорт гема в периплазму и формирование NarP-комплекса (*ссм*). Потенциальные сайты перед этими оперонами обнаружены во всех случаях.

*Гены дегидрогеназ доноров электронов.* Ранее было показано, что только оперон *fdnGHI* формиатдегидрогеназы регулируется факторами NarL и NarP [38]. Этот оперон гомологичен оперону *fdn* и выполняет ту же функцию [38]. Сайты перед опероном *fdo* обнаружены в геномах *Y. pestis* и *Y. enterocolitica*. Ген *fdhD* имеет потенциальный сайт в регуляторной области четырех генов. Точная функция этого гена неизвестна, но предполагается, что продукт его экспрессии участвует в работе формиатдегидрогеназ [39].

Новым членом обобщенного регулона является также оперон *nqrABCDEF*, кодирующий NADH-дегидрогеназу, экспортирующую ионы натрия [40].

*Гены восстановления альтернативных акцепторов электронов.* Потенциальные сайты связывания NarP обнаружены перед оперонами, кодирующими дегидрогеназы альтернативных акцепторов электронов. К данной группе относятся два новых члена регулона: опероны *torYZ* и *cydAB*. Первый из них кодирует вторую редуктазу триметилоксида азота, и гены *torY* и *torZ* являются паралогами соответственно генов *torC* и *torA* [41]. Второй, оперон *cydAB*, кодирует комплекс цитохром-*d*-оксидазы. Этот фермент катализирует перенос электронов с убихинола-8 на молекулярный кислород и участвует, таким образом, в аэробном дыхании [42]. Оперон является членом NarP-регулонов во многих геномах (табл. 3).

*Гены метаболизма четырехуглеродных дикарбоновых кислот.* NarL-зависимая регуляция оперона *dcuB-fumB* в *E. coli* экспериментально показана ранее [13]. В ряде исследуемых геномов потенциальные сайты располагаются перед геном *dcuB*, тогда как ортологов гена *fumB* не обнаружено ни в одном из геномов. Членом обобщенного регулона является и паралог *dcuB*, ген *dcuA*, также кодирующий белок-транспортер дикарбоновых кислот [13]. Кроме того, потенциальные сайты найдены перед двумя паралогами *fumB-fumC* и *aspA*. Ген *fumC*, так же как и *fumB*, кодирует фермент фумаразу [43], тогда как продуктом *aspA* является другой фермент метаболизма четырехуглеродных дикарбоновых кислот – аспартат-аммоний-лиаза [44].

*Гены ферментов, участвующих в брожении.* Ранее показано, что в объединенный NarL-NarP-регулон *E. coli* входят гены алкогольдегидрогеназы *adhE* [15] и пируват-формиат-лиазы *pflB* [14]. В геноме *E. coli* ген *pflB* может транскрибироваться как совместно с геном транспортера формиата *focA*, образуя транскрипт *focA-pflB*, так и со своего собственного промотора. Сайты связывания NarL располагаются перед опероном *focA-pflB*, перед

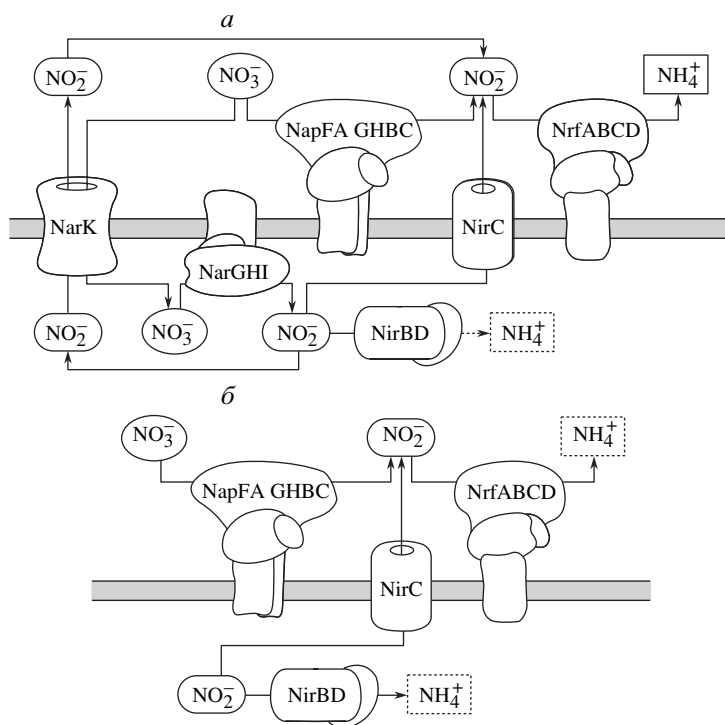
внутренним же промотором их не обнаружено [14]. Поскольку в геномах *Y. pestis* и *Y. enterocolitica* имеется потенциальный оперон *focA-pflB*, то весьма вероятно, что транскрипция данных генов осуществляется так же, как и в *E. coli*. На это указывает и тот факт, что в этих геномах обнаружены потенциальные NarP-сайты перед геном *focA* (данные не приводятся), однако такие же сайты обнаружены и перед геном *pflB*. В других исследуемых геномах данные гены располагаются отдельно, и перед одиночным геном *focA* не удается найти сайтов с хорошей гомологией. Отсюда можно предположить, что регуляция именно гена *pflB* является ключевой. По-видимому, еще один член регулона – паралог *pflB* – ген *yfiD* также кодирует пируват-формиатлиазу [45].

Наконец, консервативность сайтов в различных геномах позволяет отнести ген *ldhA*, кодирующий лактатдегидрогеназу [46], и ген *pgk*, кодирующий фосфоглицераткиназу [47], к обобщенному NarP-регулону. При этом сайты перед геном *pgk* в семействе Pasteurellaceae при множественном выравнивании регуляторных областей располагаются на одном и том же месте и перекрываются с потенциальными промоторами, тогда как потенциальные сайты перед *ldhA* в Vibrionaceae располагаются в разных местах (данные не приводятся), что ставит под сомнение регуляцию данного гена с помощью NarP.

*Гены ферментов цикла трикарбоновых кислот.* Два оперона, кодирующих ферменты цикла трикарбоновых кислот, по всей видимости, также входят в обобщенный NarP-регулон. Продуктом гена *mdh* является малатдегидрогеназа [48], тогда как оперон *sucAB* кодирует две субъединицы фермента 2-оксоглутаратдегидрогеназы [49]. Сайты перед двумя этими оперонами обнаружены только в геномах бактерий из семейства Pasteurellaceae (табл. 3), причем сайты перед геном *mdh* располагаются в одном месте, а сайты перед *sucAB* разбросаны (данные не приводятся).

*Гены синтеза молибдоптеринового кофактора.* В семи исследованных геномах потенциальные сайты связывания NarP обнаружены перед опероном *toaABCDE*. При этом сайты в четырех геномах представителей Vibrionaceae расположены в одном месте, тогда как сайты в геномах трех видов сем. Pasteurellaceae разбросаны и менее достоверны (данные не приводятся). Продукты генов оперона *toaABCDE* ответственны за ранние стадии синтеза молибдоптеринового кофактора [50].

В ряде геномов NarP-сайты локализуются перед геном *gcvA*, но их последовательности мало консервативны (данные не приводятся). Белок GcvA представляет собой транскрипционный фактор, активатор оперона *gcvTHP*. Продукты генов этого оперона участвуют в расщеплении глицина, который является источником переносимых фо-



**Рис. 6.** Система восстановления нитрата и нитрита в *E. coli* (а) и в организмах, геномы которых исследованы в данной работе (б).

латами одноуглеродных фрагментов [51]. Включение таких фрагментов, переносимых тетрагидрофолатами, – необходимая стадия биосинтеза молибдоптеринового кофактора [50].

**Гены метаболизма серы.** Обнаруженная нами регуляция оперона *cysJIN* специфична для организмов семейства *Vibrionaceae*. Белки, кодируемые данным опероном, участвуют в восстановлении 3'-фосфоаденилсульфата до сульфида [52].

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

### Структура системы нитрат-нитритного дыхания в исследованных геномах

В клетках *E. coli* сформировалась сложная система восстановления нитрата и нитрита (рис. 6а). Восстановление нитрата может происходить в периплазме за счет комплекса  $\text{NarABCD}$ , а в цитоплазме – комплекса  $\text{NarGHI}$ . Цитоплазматическое восстановление нитрата более выгодно энергетически в силу особенностей организации дыхательной цепи, однако при этом в цитоплазме накапливается токсичный для клетки нитрит. Поэтому здесь присутствует нитритредуктазный комплекс  $\text{NirBD}$ , выполняющий защитную функцию. Кроме того, нитрит выводится в периплазму с помощью транспортеров  $\text{NarK}$  и  $\text{NirC}$  [1]. В периплазме восстановление нитрита осуществляется с помощью сопряженного с мембраной  $\text{Nrf}$ -комплекса.

Более безопасный периплазматический путь используется, если концентрация нитрата невелика. При более высоких концентрациях он репрессируется, и активируется цитоплазматический путь [1]. Такая система восстановления нитрата и нитрита требует точной регуляции, что и привело в эволюции к удвоению регуляторной системы.

В геномах, изучавшихся в данной работе, есть гены только для периплазматической нитратредуктазы, т.е. здесь реализован более безопасный путь восстановления нитрата (рис. 6б). В этом случае достаточно более простой единичной системы. В качестве сенсорного белка в данной системе действует белок  $\text{NarQ}$ , одинаково реагирующий на нитрат и нитрит. В геномах бактерий рода *Yersinia* обнаружены ортологи гена *narX*. Однако некоторые изменения в аминокислотной последовательности, в частности, отсутствие в ней консервативных цистеинов, свидетельствует о том, что данный белок, по-видимому, выполняет функции белка  $\text{NarQ}$ , и его ответ на нитрат и нитрит не дифференцирован.

Однако в геномах *Y. pestis* и *Y. enterocolitica* нет генов для периплазматической нитритредуктазы, а в геноме *V. cholerae* нет генов ни для одной из нитритредуктаз. Поэтому организация системы восстановления нитрата и нитрита в этих трех организмах требует отдельного изучения, что выходит за рамки данного исследования.

*Авторегуляция и регуляторные каскады*

Как упоминалось ранее, во всех геномах группы *Vibrionaceae* перед опероном *narQP* имеются потенциальные NarP-сайты, что свидетельствует об авторегуляции данного оперона. В геноме *E. coli* автоматически регулируется оперон *narXL*, а гены *narQ* и *narP* не регулируются ни NarL, ни NarP [53]. По-видимому, центральная роль в регуляции нитрат-нитритного дыхания у *E. coli* принадлежит системе NarX-NarL. В организмах же, чьи геномы исследованы в данной работе, регуляция осуществляется одиночной системой NarQ-NarP, которая принимает на себя регуляторные функции, выполняемые двумя системами в *E. coli*.

Неожиданным оказалось то, что перед геном *fnr* в пяти геномах обнаруживаются потенциальные сайты (табл. 3). Продукт данного гена представляет собой глобальный регулятор дыхания, который осуществляет переключение между аэробным и анаэробным метаболизмом [1].

Следует заметить, что в геноме *E. coli* экспрессия оперона *narXL* угнетается белком FNR [53]. Таким образом, в организмах, исследованных в данной работе, вероятно, роль NarP в регуляторных каскадах отличается от роли этого белка у *E. coli*.

*Гены аэробного метаболизма, регулируемые NarP*

Как указывалось в разделе “Результаты”, в исследованных геномах обобщенный NarP-регулон включает в себя, помимо всего прочего, ряд генов, участвующих в аэробном метаболизме (*cydAB*, *mdh* и *sucAB*), тогда как у *E. coli* подобной ситуации не наблюдается. Оперон *cydAB* кодирует субъединицы аэробной редуктазы, и в *E. coli* его экспрессия репрессируется регулятором FNR в ответ на анаэробные условия [42]. Тем не менее, в восьми геномах перед данным опероном найдены хорошо выраженные сайты связывания NarP.

Кроме того, в данной работе к обобщенному NarP-регулону мы отнесли два оперона, продукты которых представляют собой ферменты цикла трикарбоновых кислот. В *E. coli* экспрессия генов цикла регулируется факторами транскрипции ArcA и FNR [54, 55]. Однако в геномах бактерий семейства *Pasteurellaceae* перед данными генами обнаружены потенциальные сайты связывания NarP.

Перечисленные факты, в совокупности с фактом обнаружения консервативных NarP-сайтов перед геном *fnr*, указывают на то, что в тех организмах, геномы которых исследованы в данной работе, роль белка NarP в регуляции дыхания более существенна, чем в *E. coli*.

*Гены синтеза молибдоптеринового кофактора, транспорта гема и метаболизма серы*

Известно, что присутствие молибдоптеринового кофактора, гемов и железосерных кластеров является характерной чертой компонентов дыхательных цепей [1]. Молибдоптериновый кофактор входит в состав ряда ферментов дыхательных цепей (редуктаз и дегидрогеназ). Так, данный кофактор необходим для работы комплексов, кодируемых оперонами *nrf*, *nap*, *torCAD*, *torYZ*, *dms*, *fdn* и *fdo* [1]. Поскольку для всех этих оперонов нами была предсказана NarP-зависимая регуляция, принадлежность генов синтеза кофактора к исследуемому регулону представляется вполне логичной.

В геномах *H. ducreyi*, *H. influenzae* и *P. multocida* потенциальные сайты связывания имеются и перед геном *hemR*, предположительно кодирующим транспортер гемина [26], однако сайты перед этим опероном крайне дивергированы (данные не приводятся), и поэтому принадлежность данного гена к обобщенному NarP-регулону сомнительна.

Не совсем ясно, какова роль потенциальных NarP-сайтов перед опероном *susJHI*, ответственным за восстановление соединений серы. Как упоминалось ранее, продукты данных генов участвуют в восстановлении 3'-фосфоаденилсульфата до сульфида [52]. Не исключено, что восстановленная сера необходима для перезарядки железосерных кластеров в составе компонентов дыхательных цепей.

*Регуляция дублицированных генов*

У *E. coli* известно достаточно большое число дубликаций генов, участвующих в процессах дыхания [1]. В данной работе нами рассмотрено четыре случая таких дубликаций для оперонов *fdnGHI/fdoGHI*, *torCAD/torYZ*, *pflB/yfiD* и *dcuA/dcuB*. Все перечисленные пары характеризуются следующей особенностью: продукты оперонов из одной и каждой пары выполняют одинаковую функцию, в то время как различия в регуляции экспрессии этих оперонов весьма существенны [1, 8, 13, 45].

Однако в ходе эволюции в механизмах регуляции могли произойти значительные изменения. Поэтому регуляция оперонов из одной пары рассматривалась нами совместно: в случае, если сайт сохраняется перед любым геном из пары, то к обобщенному регулону причислялись оба члена пары. Например, сайты перед опероном *fdoGHI* обнаружены в геномах *Y. pestis* и *Y. enterocolitica*, а сайты перед *fdnGHI* – в геномах *P. multocida* и *A. actinomycetemcomitans*. В остальных же геномах потенциальных сайтов перед этими оперонами либо нет, либо оба оперона отсутствуют. В общей сложности, сайты найдены в четырех геномах, на

основании чего оба оперона включены в NarP-регулон.

Опероны *torCAD* и *yfiD* отнесены к членам регулона, исходя из тех же соображений (см. табл. 3).

#### Регуляция генов NADH-дегидрогеназ

В геномах организмов из группы Pasteurellaceae потенциальные NarP-сайты имеются перед опероном *nqrABCDE*, который кодирует NADH-дегидрогеназу, экспортирующую ионы натрия. Данный фермент впервые обнаружен в организме морской бактерии *Vibrio alginolyticus* [40]. В этом случае экспорт ионов Na<sup>+</sup> против градиента концентрации происходит за счет окисления NADH; при этом осуществляется перенос электронов на убихиноны, т.е. Nqr-комплекс представляет собой компонент дыхательной цепи. Ортологи генов *nqr*-оперона обнаружены во всех исследованных геномах, однако NarP-зависимая регуляция данного оперона специфична для бактерий семейства Pasteurellaceae.

У бактерий *E. coli* имеется протон-экспортирующая NADH-дегидрогеназа, кодируемая опероном *nuo*, транскрипция которого активируется белками NarL и FNR [10]. Оперон *nuo* найден в геномах *Y. pestis* и *Y. enterocolitica*, но потенциальных NarP-сайтов перед ним нет. Поскольку гены оперонов *nuo* и *nqr* негомологичны, в данном случае имеет место негомологичная замена с частичным изменением функции.

#### Таксон-специфичная регуляция

В данной работе исследованы десять геномов гамма-протеобактерий, в которых NarP-зависимую регуляцию ранее никогда не изучали. Более того, по сравнению с *E. coli*, все эти геномы изучены недостаточно. Поэтому мы использовали новую методику выявления членов обобщенного регулона.

Ранее изучение обобщенного регулона началось с поиска потенциально сайтов связывания в хорошо изученном геноме, а затем, если сайты в геномах близкородственных организмов были консервативны, можно было делать вывод о принадлежности гена к обобщенному регулону [56]. Однако в данном случае подобный подход оказался невозможным, поскольку у *E. coli* нитрат-нитритное дыхание регулируется сразу двумя факторами транскрипции – NarL и NarP, причем достоверной информации о структуре сайтов связывания NarL практически нет. Поэтому мы были вынуждены прибегнуть к методике, основанной на попарном сравнении геномов организмов, относящихся к одной таксономической группе. Данная методика рассмотрена в разделе “Условия эксперимента”.

С помощью новой методики удалось выявить несколько новых членов обобщенного регулона, NarP-зависимая регуляция которых специфична для семейства. Так, регуляция оперонов *fumC*, *pgk*, *mdh* и *sucAB* оказалась специфичной для организмов семейства Pasteurellaceae, регуляция *ldhA* и *cysJHI* и авторегуляция оперона *narQP* – специфичной для организмов семейства Vibrionaceae. Регуляция другой группы генов не является в строгом смысле специфичной для семейства. Однако эти гены удалось идентифицировать как новые члены обобщенного регулона, поскольку первоначально они были выявлены как члены регулонов в одной таксономической группе, а затем сайты перед этими генами были обнаружены в геномах организмов из других групп. Таковыми являются гены *fnr*, *aspA* и *gcvA*.

Работа получила финансовую поддержку Медицинского института Ховарда Хьюза (55000309), Фонда поддержки российской науки, Программ “Молекулярная и клеточная биология” и “Происхождение и эволюция биосферы” Российской академии наук.

#### СПИСОК ЛИТЕРАТУРЫ

1. Gennis R.B., Stewart V. 1996 Respiration. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology*. Ed. Neidhart F.C. Washington: ASM Press, pp. 217–286.
2. Stewart V., Rabin R.S. 1995 Dual sensors and dual response regulators interact to control nitrate- and nitrite-responsive gene expression in *Escherichia coli*. In: *Two-component signal transduction*. Eds Hoch J.A., Silhavy T.J. Washington: ASM Press, pp. 233–252.
3. Rabin R.S., Stewart V. 1993. Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate- and nitrite-regulated gene expression in *Escherichia coli* K-12. *J. Bacteriol.* **175**, 3259–3268.
4. Darwin A.J., Tyson K.L., Busby S.J., Stewart V. 1997. Differential regulation by the homologous response regulators NarL and NarP of *Escherichia coli* K-12 depends on DNA binding site arrangement. *Mol. Microbiol.* **25**, 583–595.
5. Li J., Kustu S., Stewart V. 1994. *In vitro* interaction of nitrate-responsive regulatory protein NarL with DNA target sequences in the *fdnG*, *narG*, *narK* and *frdA* operon control regions of *Escherichia coli* K-12. *J. Mol. Biol.* **241**, 150–165.
6. Darwin A.J., Ziegelhoffer E.C., Kiley P.J., Stewart V. 1998. Fnr, NarP, and NarL regulation of *Escherichia coli* K-12 *napF* (periplasmic nitrate reductase) operon transcription *in vitro*. *J. Bacteriol.* **180**, 4192–4108.
7. Wang H., Gunsalus R.P. 2000. The *nrfA* and *nirB* nitrite reductase operons in *Escherichia coli* are expressed differently in response to nitrate than to nitrite. *J. Bacteriol.* **182**, 5813–5822.
8. Wang H., Gunsalus R.P. 2003. Coordinate regulation of the *Escherichia coli* formate dehydrogenase *fdnGHI* and

- fdhF* genes in response to nitrate, nitrite, and formate: roles for NarL and NarP. *J. Bacteriol.* **185**, 5076–5085.
9. Richard D.J., Sawers G., Sargent F., McWalter L., Boxer D.H. 1999. Transcriptional regulation in response to oxygen and nitrate of the operons encoding the [NiFe] hydrogenases 1 and 2 of *Escherichia coli*. *Microbiology.* **145**, 2903–2912.
  10. Bongaerts J., Zoske S., Weidner U., Uden G. 1995. Transcriptional regulation of the proton translocating NADH dehydrogenase genes (*nuoA-N*) of *Escherichia coli* by electron acceptors, electron donors and gene regulator. *Mol. Microbiol.* **16**, 521–534.
  11. Bearson S.M., Albrecht J.A., Gunsalus R.P. 2002. Oxygen and nitrate-dependent regulation of *dmsABC* operon expression in *Escherichia coli*: sites for Fnr and NarL protein interactions. *BMC Microbiol.* **2**, 13.
  12. Iuchi S., Lin E.C. 1987. The *narL* gene product activates the nitrate reductase operon and represses the fumarate reductase and trimethylamine N-oxide reductase operons in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* **84**, 3901–3905.
  13. Golby P., Kelly D.J., Guest J.R., Andrews S.C. 1998. Transcriptional regulation and organization of the *dcuA* and *dcuB* genes, encoding homologous anaerobic C4-dicarboxylate transporters in *Escherichia coli*. *J. Bacteriol.* **180**, 6586–6596.
  14. Kaiser M., Sawers G. 1995. Nitrate repression of the *Escherichia coli pfl* operon is mediated by the dual sensors NarQ and NarX and the dual regulators NarL and NarP. *J. Bacteriol.* **177**, 3647–3655.
  15. Chen Y.M., Lin E.C. 1991. Regulation of the *adhE* gene, which encodes ethanol dehydrogenase in *Escherichia coli*. *J. Bacteriol.* **173**, 8009–8013.
  16. Mironov A.A., Koonin E.V., Roytberg M.A., Gelfand M.S. 1999. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res.* **27**, 2981–2989.
  17. Миронов А.А., Винокурова Н.П., Гельфанд М.С. 2000. Програмное обеспечение анализа бактериальных геномов. *Молекуляр. биология.* **34**, 253–262.
  18. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z.M.W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
  19. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882.
  20. Felsenstein J. 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266**, 418–427.
  21. Schneider T.D., Stephens R.M. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100.
  22. Blattner F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science.* **277**, 1453–1474.
  23. Deng W., Liou S.R., Plunkett G., Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R. 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J. Bacteriol.* **185**, 2330–2337.
  24. Bell K.S., Sebahia M., Pritchard L., Holden M.T., Hyman L.J., Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J., Mungall K., Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J., Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H., Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S., Salmond G.P., Birch P.R., Parkhill J., Toth I.K. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. USA.* **101**, 11105–11110.
  25. Deng W., Burland V., Plunkett G., Boutin A., Mayhew G.F., Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C., Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V., Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R., Perry R.D. 2002. Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol.* **184**, 4601–4611.
  26. Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G.G., FitzHugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science.* **269**, 496–512.
  27. May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V. 2001. Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc. Natl. Acad. Sci. USA.* **98**, 3460–3465.
  28. Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature.* **406**, 477–483.
  29. Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K., Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S., Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T. 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet.* **361**, 743–749.
  30. Kim Y.R., Lee S.E., Kim C.M., Kim S.Y., Shin E.K., Shin D.H., Chung S.S., Choy H.E., Progulsk-Fox A., Hillman J.D., Handfield M., Rhee J.H. 2003. Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect. Immun.* **71**, 5461–5471.

31. Benson D.A., Boguski M.S., Lipman D.J., Ostell J., Ouellette B.F., Rapp B.A., Wheeler D.L. 1999. GenBank. *Nucleic Acids Res.* **27**, 12–17.
32. Bernal A., Ear U., Kyrpides N. 2001. Genomes OnLine Database (GOLD): a monitor of genome projects worldwide. *Nucleic Acids Res.* **29**, 126–127.
33. Stewart V., Chen L.L., Wu H.C. 2003. Response to culture aeration mediated by the nitrate and nitrite sensor NarQ of *Escherichia coli* K-12. *Mol. Microbiol.* **50**, 1391–1399.
34. Stewart V. 2003. Biochemical Society Special Lecture. Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria. *Biochem. Soc. Trans.* **31**, 1–10.
35. Baikalov I., Schroder I., Kaczor-Grzeskowiak M., Grzeskowiak K., Gunsalus R.P., Dickerson R.E. 1996. Structure of the *Escherichia coli* response regulator NarL. *Biochemistry.* **35**, 11053–11061.
36. Maris A.E., Sawaya M.R., Kaczor-Grzeskowiak M., Jarvis M.R., Bearson S.M., Kopka M.L., Schroder I., Gunsalus R.P., Dickerson R.E. 2002. Dimerization allows DNA target site recognition by the NarL response regulator. *Nature Struct. Biol.* **9**, 771–778.
37. Gelfand M.S., Laikova O.N. 2003. Prolegomena to the evolution of transcriptional regulation in bacterial genomes. In: *Frontiers in computational genomics*. Eds Galperin M.Y., Koonin E.V. Wymondham, U.K.: Caister Acad. Press, pp. 195–216.
38. Darwin A.J., Li J., Stewart V. 1996. Analysis of nitrate regulatory protein NarL-binding sites in the *fdnG* and *narG* operon control regions of *Escherichia coli* K-12. *Mol. Microbiol.* **20**, 621–632.
39. Plunkett G., Burland V., Daniels D.L., Blattner F.R. 1993. Analysis of the *Escherichia coli* genome. III. DNA sequence of the region from 87.2 to 89.2 minutes. *Nucleic Acids Res.* **21**, 3391–3398.
40. Tokuda H., Nakamura T., Unemoto T. 1981. Potassium ion is required for the generation of pH-dependent membrane potential and delta pH by the marine bacterium *Vibrio alginolyticus*. *Biochemistry.* **20**, 4198–4203.
41. Gon S., Patte J.C., Mejean V., Iobbi-Nivol C. 2000. The *torYZ* (*yecK-bisZ*) operon encodes a third respiratory trimethylamine N-oxide reductase in *Escherichia coli*. *J. Bacteriol.* **182**, 5779–5786.
42. Cotter P.A., Chepuri V., Gennis R.B., Gunsalus R.P. 1990. Cytochrome *o* (*cyoABCDE*) and *d* (*cydAB*) oxidase gene expression in *Escherichia coli* is regulated by oxygen, pH, and the *fnr* gene product. *J. Bacteriol.* **172**, 6333–6338.
43. Tseng C.P., Yu C.C., Lin H.H., Chang C.Y., Kuo J.T. 2001. Oxygen- and growth rate-dependent regulation of *Escherichia coli* fumarase (FumA, FumB, and FumC) activity. *J. Bacteriol.* **183**, 461–467.
44. Spiro S., Guest J.R. 1991. Adaptive responses to oxygen limitation in *Escherichia coli*. *Trends Biochem. Sci.* **16**, 310–314.
45. Wyborn N.R., Messenger S.L., Henderson R.A., Sowers G., Roberts R.E., Attwood M.M., Green J. 2002. Expression of the *Escherichia coli* *yfiD* gene responds to intracellular pH and reduces the accumulation of acidic metabolic end products. *Microbiology.* **148**, 1015–1026.
46. Jiang G.R., Nikolova S., Clark D.P. 2001. Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology.* **147**, 2437–2446.
47. Alefounder P.R., Perham R.N. 1989. Identification, molecular cloning and sequence analysis of a gene cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-phosphoglycerate kinase and a putative second glyceraldehyde 3-phosphate dehydrogenase of *Escherichia coli*. *Mol. Microbiol.* **3**, 723–732.
48. Park S.J., Cotter P.A., Gunsalus R.P. 1995. Regulation of malate dehydrogenase (*mdh*) gene expression in *Escherichia coli* in response to oxygen, carbon, and heme availability. *J. Bacteriol.* **177**, 6652–6656.
49. Park S.J., Chao G., Gunsalus R. 1997. Aerobic regulation of the *sucABCD* genes of *Escherichia coli*, which encode alpha-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase: roles of ArcA, Fnr, and the upstream *sdhCDAB* promoter. *J. Bacteriol.* **178**, 4138–4142.
50. Rajagopalan K.V. 1996. Biosynthesis of the molybdenum cofactor. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology*. Ed. Neidhart F.C. Washington: ASM Press, pp. 674–679.
51. Green J.M., Nichols B.P., Matthews R.G. 1996. Folate Biosynthesis, Reduction, and Polyglutamylation. In: *Escherichia coli and Salmonella. Cell. Mol. Biol.* Ed. Neidhart F.C. Washington: ASM Press, pp. 665–673.
52. Kredich N.M. 1996. Biosynthesis of Cysteine. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology*. Ed. Neidhart F.C. Washington: ASM Press, pp. 514–527.
53. Darwin A.J., Stewart V. 1995. Expression of the *narX*, *narL*, *narP*, and *narQ* genes of *Escherichia coli* K-12: regulation of the regulators. *J. Bacteriol.* **177**, 3865–3869.
54. Lynch A.S., Lin C.C. 1996. Responses to Molecular Oxygen. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology*. Ed. Neidhart F.C. Washington: ASM Press, pp. 1526–1537.
55. Cronan J.E., Laporte D. 1996. Tricarboxylic Acid Cycle and Glyoxylate Bypass. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology*. Ed. Neidhart F.C. Washington: ASM Press, pp. 206–216.
56. Равчеев Д.А., Гельфанд М.С., Миронов А.А., Рахманинова А.Б. 2002. Пуриновый регулон гамма-протеобактерий. Детальное описание. *Генетика.* **38**, 1203–1214.