

УДК 577.1

РЕГУЛЯЦИЯ ДЫХАНИЯ У ЭНТЕРОБАКТЕРИЙ: СОПОСТАВЛЕНИЕ ДАННЫХ ПО АНАЛИЗУ ЭКСПРЕССИИ НА БИОЧИПАХ И СРАВНИТЕЛЬНО-ГЕНОМНОГО АНАЛИЗА

© 2007 г. М. О. Цыганова¹, М. С. Гельфанд^{1,2,3}, Д. А. Равчев^{1,2*}

¹Факультет биоинженерии и биоинформатики Московского государственного университета им. М.В. Ломоносова, Москва, 119992

²Институт проблем передачи информации Российской академии наук, Москва, 127994

³Государственный научно-исследовательский институт генетики и селекции промышленных микроорганизмов “ГосНИИгенетика”, Москва, 113545

Поступила в редакцию 27.04.2006 г.

Принята в печать 19.06.2006 г.

Проведен сравнительный анализ данных о составе Fnr- и ArcA-модулонов, полученных экспериментально с помощью биочипов, и данных о составе Fnr- и ArcA-регулонов, полученных с помощью сравнительно-геномного подхода. Показано, что регуляторный каскад с участием белков Fnr и ArcA вносит существенный вклад в расширение Fnr-модулона по отношению к Fnr-регулону за счет оперонов ArcA-модулона. Обнаружено 26 оперонов с консервативными сайтами связывания Fnr и 16 оперонов с консервативными сайтами связывания ArcA. На основании данных о составе Fnr- и ArcA-регулонов выделена группа генов, так называемое “ядро” регулонов, с высококонсервативными сайтами связывания Fnr и ArcA и высокими значениями весов для этих сайтов.

Ключевые слова: аэробное и анаэробное дыхание, Fnr, ArcA, регулон, модулон, сравнительная геномика.

REGULATION OF RESPIRATION IN ENTEROBACTERIALES: THE COMPARISON OF MICROARRAY AND COMPARATIVE GENOMIC DATA, by M. O. Tsiganova¹, M. S. Gelfand^{1,2,3}, D. A. Ravchev^{1,2*} (¹Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, 119992 Russia; *e-mail: ravchev@iitp.ru; ²Institute for Information Transmission Problems Russian Academy of Sciences, Moscow, 127994 Russia; ³State Scientific Center “GosNII Genetica”, Moscow, 113545 Russia). Microarrays are widely used for gene expression profiling. In the case of prokaryotes such arrays usually provide data about composition of modulons, groups of genes whose expression is influenced by a single regulatory system or external stimulus. Unlike modulons, regulons include only genes directly controlled by regulatory systems. Here we compared the structures of the Fnr and ArcA modulons and regulons. The data about modulon composition were taken from published microarray assays, whereas regulons were characterized using comparative genomic approaches. The Fnr and ArcA regulons were shown to contain 26 and 16 operons, respectively. Ten operons had high-score and highly conserved site for both Fnr and ArcA. These genes are the “core of regulons”. Remarkably, all “core genes” encode enzymes involved in aerobic respiration and central metabolism. The Fnr-ArcA regulatory cascade plays an important role in expansion of the Fnr modulon.

Key words: Enterobacteriales, aerobic and anaerobic respiration, Fnr, ArcA, regulon, modulon, comparative genomics.

В настоящее время для изучения экспрессии генов часто используют биочипы, которые позволяют одновременно получить и обработать информацию о работе большого числа генов. Несмотря на очевидные преимущества, этот метод имеет также и ряд существенных недостатков, связанных в основном со сложностью технологии, высоким уровнем шума и ошибок эксперимента [1]. Метод биочипов предоставляет данные

о составе модулона. Модулон – это группа генов, экспрессия которых изменяется при мутациях в гене белка-регулятора или при изменении условий среды. В случае регуляции в состав модулона входят гены, экспрессия которых прямо или опосредованно регулируется некоторой регуляторной системой.

Помимо экспериментальных технологий для изучения регуляции транскрипции в прокариотических организмах в настоящее время широко ис-

* Эл. почта: ravchev@iitp.ru

пользуются методы биоинформатики. Методы биоинформатики позволяют проводить поиск потенциальных регуляторных сайтов и выделять регулоны: группы генов, экспрессия которых непосредственно регулируется определенной регуляторной системой. Сравнительно-геномный метод, заключающийся в проверке консервативности потенциальных сайтов связывания регуляторов, значительно повышает качество предсказаний. Такой подход предполагает, что ген прямо регулируется изучаемым регулятором, если перед ортологичными ему генами в родственных геномах сохраняются потенциальные сайты связывания этого регулятора. Непрерывное увеличение числа расшифрованных геномов значительно расширяет возможности этого подхода как количественные, так и качественные [2]. Из анализа индивидуальных регуляторных систем следует, что сравнительный подход является мощным методом функционального описания бактериальных геномов [3–7].

Интуитивно понятно, что в состав модулонов входит больше генов, чем в состав соответствующих регулонов. Так как экспрессия большинства генов контролируется более чем одним регулятором, которые формируют сложные регуляторные каскады, то, по-видимому, одна из причин расширения модулона по отношению к регулону – влияние этих каскадов. Существуют и другие причины расширения модулонов. Так, в простейшей ситуации, если регулятор I транскрипции, регулирует метаболический путь синтеза вещества, которое является эффектором для регулятора II транскрипции, то регулон регулятора II будет входить в модулон I.

Для оценки вклада регуляторных каскадов в расширение модулона были выбраны факторы транскрипции Fnr и ArcA, контролирующие дыхание энтеробактерии *Escherichia coli* K12. Белки Fnr и ArcA могут служить как активаторами, так и репрессорами, в зависимости от расположения сайтов их связывания относительно промоторов регулируемого оперона. Белок Fnr – это глобальный регулятор, который переключает метаболизм с аэробного на анаэробный путь [8]. Регулятор транскрипции ArcA входит в состав двухкомпонентной системы ArcA-ArcB, где ArcB – трансмембранный сенсорный белок, чувствительный к состоянию мембранных хинонов [9, 10]. По всей видимости, система ArcA-ArcB отвечает за переключение метаболизма с дыхания на брожение. В системе регуляции дыхания *E. coli* белок ArcA занимает подчиненное положение по отношению к Fnr: известно, что транскрипция гена *arcA* регулируется Fnr [11], т.е. эти белки образуют регуляторный каскад.

Целью данной работы было выделение Fnr- и ArcA-регулонов и оценка вклада каскада Fnr-ArcA-

сА-в расширение модулона Fnr по отношению к Fnr-регулону.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Геномы. Для выявления Fnr- и ArcA-регулонов с помощью сравнительно-геномных методов использовали геномы бактерий порядка *Enterobacteriales*, включающего *E. coli* и близкородственные организмы. Полные нуклеотидные последовательности геномов *E. coli* K12 (EC) [18], *Salmonella typhi* Ty2 (STY) [19], *Salmonella typhimurium* LT2 (STM) [20], *Yersinia pestis* KIM (YP) [21], *Pectobacterium carotovorum* subsp. *atroseptica* SCRI1043 (PCA) [22], *Photobacterium luminescens* subsp. *laumondii* TTO1 (PHL) [23] получены из базы данных GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>). Неполные последовательности геномов *Y. enterocolitica* (YE), *Serratia marcescens* Db11 (SM) и *P. asymbiotica* ATCC43949 (PHA) взяты с интернет-сайта Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/>), *P. chrysanthemii* штамм 3937 (PCH) – с интернет-сайта The Institute For Genomic Research (<http://www.tigr.org>). Во всех геномах найдены ортологи генов белков Fnr и ArcA.

Выделение модулонов. Экспериментальные данные о составе модулона взяты из работ [12, 13], в которых влияние белков-регуляторов Fnr [12] и ArcA [13] на уровень экспрессии большого числа генов *E. coli* изучали с использованием биочипов. В этих работах [12, 13] определено также изменение уровня экспрессии генов. По влиянию на экспрессию генов эффекты каждого фактора разделили на три группы:

- активация – при увеличении уровня экспрессии не менее чем в 2 раза;
- отсутствие эффекта – уровень экспрессии изменяется в любую сторону менее чем в 2 раза;
- репрессия – при уменьшении уровня экспрессии не менее чем в 2 раза.

Тем самым гены, уровень экспрессии которых менялся в 2 раза и более, были отнесены к модулону. Гены, уровень экспрессии которых изменялся менее чем в 2 раза, в состав модулона не включались. Подобный критерий принято использовать при изучении регуляции экспрессии генов [14].

На следующем этапе работы на основе этой информации были выделены опероны, данные об изменении экспрессии которых можно считать противоречивыми. Данные считали противоречивыми, если один регулятор оказывал противоположный эффект (в соответствии с приведенными критериями) на разные гены одного оперона, или мутации в гене регулятора не влияли на один из генов оперона, тогда как экспрессия других генов того же оперона изменялась более чем в 2 раза.

Определение состава регулона. Принадлежность гена к регулону определяли с использованием метода проверки соответствия. В его основе лежит сравнительный анализ геномных последовательностей родственных организмов. Ген считали входящим в регулон, если сайт связывания перед ним сохранялся более чем в трех родственных геномах. Однако, если сайт был обнаружен только перед генами *E. coli* K12, *S. typhimurium* LT2 и *S. typhi* Ty1, то ген не включали в регулон, поскольку степень родства между этими организмами очень высока, и сохранение сайтов может быть лишь следствием высокой степени сходства некодирующих областей.

В нашей работе мы несколько модифицировали критерий, по которому определяли принадлежность (или непринадлежность) гена к регулону. Ген считали членом регулона, если он входил в состав модулона и одновременно с этим имел в регуляторной области консервативный сайт связывания белка-регулятора. Т.е. описанный подход основан на совмещении экспериментальных данных и результатов сравнительной геномики. Гены, не входящие в состав модулонов, не рассматривали.

Поиск потенциальных сайтов связывания проводили с помощью метода матриц позиционных весов нуклеотидов [15]. С этой целью использовали матрицы для поиска сайтов связывания белков Fng [16] и ArcA [17], описанные ранее. Поиск сайтов проводили в потенциальных регуляторных областях генов – в пределах $-400...+100$ п.н. относительно старта трансляции. Исключение было сделано для оперона *cydAB*, который, как показано экспериментально, содержит два сайта связывания ArcA на расстоянии -637 и -719 п.н. от старта трансляции. Поэтому для этого оперона поиск потенциальных ArcA-сайтов проводили в пределах от $-750...+100$ п.н. относительно старта трансляции. Сайты считали значимыми, если их вес превышал пороговое значение: 3.50 для Fng-сайтов и 4.25 для ArcA-сайтов. Потенциальный Fng-сайт обнаружен в регуляторной области 15% генов *E. coli*, а ArcA-сайт – в регуляторной области 7% генов.

Программное обеспечение. Для поиска ортологических генов и потенциальных регуляторных сайтов использовали пакет программ “Genome Explorer” [24].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Определение состава Fng- и ArcA-модулонов E. coli

Fng- и ArcA-модулоны *E. coli* выделены по результатам двух экспериментальных исследований [12, 13]. В этих работах представлены данные об изменении в два и более раз уровня экспрессии

174 оперонов; из них 151 оперон отнесен к Fng-модулону, и 135 – к ArcA-модулону (табл. 1).

Определение состава Fng- и ArcA-регулонов E. coli

В каждом опероне, входящем в Fng- или ArcA-модулон, проведен поиск сайтов связывания Fng и ArcA (см. Экспериментальную часть). В результате потенциальный сайт связывания Fng был обнаружен перед 38 оперонами (табл. 2), а сайт связывания ArcA – перед 23 оперонами (табл. 3).

Проверка консервативности найденных потенциальных сайтов связывания Fng и ArcA с помощью метода проверки соответствия показала, что в 26 случаях из 38 для Fng и в 16 случаях из 23 для ArcA сайт сохраняется перед ортологичными генами как минимум в трех родственных геномах (табл. 2, табл. 3).

Эти опероны с высокой вероятностью входят в соответствующие регулоны, поскольку сайты связывания перед ними консервативны. В 15 случаях сайт сохраняется перед ортологичными генами в восьми, девяти или 10 родственных геномах, в 17 случаях сайт сохраняется в пяти и более родственных геномах, и только в 10 – в четырех или трех родственных геномах (табл. 2, табл. 3).

Так как консервативность потенциальных сайтов связывания проверяли на оперонах, входящих в Fng- и ArcA-модулоны по данным анализа на биочипах, то все опероны, перед которыми найден такой сайт, можно отнести к Fng- или ArcA-регулону.

Члены модулона, которые не имеют в регуляторной области консервативного сайта, по формальному критерию не могут быть отнесены к регулону. Таких оперонов в случае Fng – семь (табл. 2), а в случае ArcA – пять (табл. 3). Замечено, что эти опероны не относятся к оперонам центрального метаболизма, а функции некоторых из них вовсе не известны (табл. 4). Кроме того, практически все сайты перед этими генами имеют низкий вес – менее 3.75 для Fng и менее 4.75 для ArcA. Отсутствуют и какие-либо независимые экспериментальные данные, указывающие на влияние белков Fng и ArcA на экспрессию этих оперонов, поэтому они были исключены из состава соответствующих регулонов.

Если гены оперона имеют ортологов только в *E. coli*, *S. typhimurium* и *S. typhi*, мы не можем говорить о принадлежности этих генов к регулонам (см. Экспериментальную часть). Сайты связывания Fng обнаружены перед пятью такими оперонами (табл. 2), ArcA-сайты – перед двумя (табл. 3).

В состав Fng-регулона входят 26 оперонов (табл. 2). Ранее выявили регуляцию белком Fng восьми из них: *narXL* [25], *nirBDC-cysG* [26], *pdhR-aceEF-lpdA* [27], *dcuB-fumB* [28], *ndh* [29] и *cydAB*

Таблица 1. Сопоставление Fnr- и ArcA- модулонов и регулонов

Оперон	Эффект Fnr	Fnr-регулон	Эффект ArcA	ArcA-регулон
<i>xylE, mhpABCDFE, ygfK, fdhF, yhiUV, caiTABCDE, yjiZ, yjeH, araBAD, yhcL, fliLMNOPQR, nadC, yjeT, eutK, aat, nikABCDE, yidKJ</i>	+		Нет данных	
<i>yhjRQONML, ybeRS, cmtBA-yggP-yggFDC, yqiGK, sbm-ygfDGH, yjeMNO, proVWX, ycbPONME, emrKY, parC, yfbTS, gcvTH, yceL-grxB, dfp-dut-ttk, aroKB-damX-dam-rpe-gph-trpS, dadAX</i>	+		0	
<i>thiMD-yegX, rpoN-yhbH-ptsN-yhbJ</i>	–		0	
<i>CusCFBA</i>	+		#	
<i>tgt-yajC-secDF, purMN</i>	–	*	0	
<i>NarXL</i>	+	*	0	
<i>thrS-infC-rpmI-rplT, rpsT</i>	–	*	–	
<i>metBL, yjiHG-iada-yjiE, nanATEK-yhcH, upp-uraA, ibpB-yidE, glgCAP, ybhK, nirBDC-cysC, rhaBAD, pstSCAB-phoU, aroP</i>	+	*	+	
<i>cydAB, nuoABCDEFGHIJKLMN</i>	–	*	#	*
<i>dcuB-fumB</i>	+	*	+	*
<i>pdhR-aceEF-lpdA¹, cyoABCDE, ndh, atpIBEFHAGDC, icdA, mdh, gltA</i>	–	*	–	*
<i>potFGHI, gapA, yadF, lpdA¹</i>	–		–	*
<i>yigI-rarD-yigGF, acrEF</i>	+		+	*
<i>hycBCDEFGHI, lysU, yhiW, ycdM, ydbA₂, ftsYEX, yjcRQP, yjcP, ybdL, umuDC, kbl-tdh, xdhA-ygeT, pyrD, mobABC, yhjW, recC, narU, ygiF-glnE-hldE, map-glnD-dapD-b0165-yaeI, yjcS, b0540-tra5₂, tynA, mrcA, yheEFG-hofFGH-yheHIJK-pshM-hofD-bfr, yadQR, speC, yhjJ, yabM, yhhTU, dinG, yhiH-yhhJ, sbcDC, ycfT, yaaU, yaaJA, yi4I, yi21₅-yi22₅-yqiGHI, frdABCD, yafZ, aroM, yghQ, nadA-pnuC, yhjCD, degQS, yhdRP, yihR, cdh, menFD-yfbB-menBCE-pmrD, yhcRQPO, yjihH, yeiTA, yjdHG, yicONM, yhfZYXWVUTS, sbp, ilvY, yiaY, fhuACDB, yigN, fliFGHIJK</i>	+		+	
<i>gatYZABCD₂, ybjX, yfiA, yceD-rpmF, trpLEDCBA, gpmA, ptsI-crr, rpsL-rpsG-fusA-tufA, rplM-rpsI, rpsP-yfiA-trmD-rplS, yajG-ampG, oppABCDEF, rpsA, ompA, argT-hisJQMP, manXYZ, ptsG, yggTUV, rpmBG, nrdAB, gapC₁-gapC₂, artPIQM, ydcF</i>	–		–	
<i>wecDE-wzxE-yifM₁, xylFGHR, ptr-recBD, tyrB, yhjX, cvpA-purF, yraPONM, yjfHIJKLMC, yphGFEDCB, yhgF, dnaX-ybaB-recB, hisS-yfgM-der-yfgJ-xseA, yjiON, idnDOTR-yjgR, yjhATS</i>	0		+	
<i>rbsDACBKR, ycdC</i>	0		–	
<i>b1172-ymgD, yhjE, aceBAK</i>	Нет данных		+	
<i>YjcR</i>	Нет данных		–	

Примечание. “+” – Регулятор активирует экспрессию гена, “–” – регулятор репрессирует экспрессию оперона, “0” – регулятор не влияет на экспрессию оперона, “#” – данные противоречивы, * – ген отнесен к регулону.

¹ Оперон может транскрибироваться с двух промоторов, расположенных перед генами *pdhR* и *lpdA*.

[30], *cyoABCDE* [30] и *nuoABCDEFGHIJKLMN* [31]. Обнаружено также влияние мутации в гене *fnr* на уровень экспрессии оперона *icdA*, однако поиск сайтов связывания Fnr перед этим опероном не проводили [32]. Предсказанные в нашей работе консервативные сайты связывания Fnr перед геном *icdA* позволяют утверждать, что транскрипция этого гена прямо регулируется Fnr, т.е. ген *icdA* должен быть отнесен к Fnr-регулону. Ранее с

использованием методов биоинформатики была предсказана регуляция оперона *upp-uraA* белком Fnr [16]. В настоящей работе мы предсказали 16 новых оперонов, потенциальных членов Fnr-регулона (табл. 2).

По функциям 26 оперонов рассматриваемого регулона Fnr можно разделить на три группы (табл. 4):

Таблица 2. Потенциальные сайты связывания Fnr

Оперон	Геном									
	EC	STM	STY	YE	YP	SM	PHL	PHA	PCA	PCH
Гены, включенные в регулон										
<i>tgt-yajC-secD-secF*</i>	4.12 (-31) 3.83 (-187)	3.83 (-322)	3.73 (-244)	–	4.27 (41)	3.56 (-319)	4.14 (-49)	3.56 (-392)	–	–
<i>purMN*</i>	4.46 (-244)	4.28 (-325)	4.28 (-322)	4.09 (-263)	4.09 (-311)	–	5.03 (-138)	5.08 (-125)	4.01 (-310)	4.09 (-303)
<i>narXL</i>	5.15 (-278) 4.20 (-240)	5.05 (-282)	5.05 (-281)	–	3.53 (-279)	5.00 (-291)	4.03 (46)	4.03 (41)	4.70 (-203)	5.00 (-352) 4.41 (-89)
<i>thrS-infC-rpml-rplT*</i>	3.67 (-69)	–	–	3.59 (-258) 3.98 (-73)	3.57 (-319) 3.81 (-73) 3.57 (-77)	–	3.73 (-278) 3.79 (-65)	3.79 (-64)	3.49 (21)	–
<i>rpsT*</i>	3.85 (1)	3.85 (1)	3.85 (1)	0	3.85 (-1)	0	4.62 (-265)	0	0	0
<i>yjiHG-iadA-yjiE*</i>	3.60 (8)	4.41 (-69)	4.41 (-281)	–	0	0	0	0	3.49 (-4)	0
<i>nanATEK-yhcH*</i>	4.36 (47)	3.70 (47)	3.70 (47)	–	3.87 (-98)	0	0	0	0	0
<i>Upp-uraA</i>	4.46 (-67)	4.28 (-102)	4.28 (-104)	4.09 (-99)	4.09 (-66)	–	5.03 (-102)	5.08 (-30)	4.01 (-167)	4.09 (-96)
<i>ibpB-yidE*</i>	3.98 (-173)	–	3.40 (-179)	3.89 (-286) 4.25 (-191)	4.25 (-306)	0	0	0	3.62 (-144)	3.96 (-103)
<i>glgCAP*</i>	3.93 (-169)	3.58 (-169)	3.58 (-159)	–	3.76 (-169)	3.53 (-234)	0	0	–	–
<i>ybhK*</i>	4.42 (-146)	4.32 (-145)	4.32 (-147)	3.67 (-183) 3.99 (-115)	3.99 (-191)	–	3.59 (-266)	3.67 (-100)	3.54 (-118)	–
<i>nirBDC-cysG</i>	5.15 (-73)	5.15 (-73)	5.15 (-71)	5.15 (-72) 3.70 (-321) 3.67 (-1) 3.66 (-343)	5.15 (-70)	3.64 (-315)	0	0	5.00 (-72)	3.48 (-175)
<i>rhaBAD*</i>	3.64 (-397)	3.64 (-397) 3.52 (-323)	3.64 (-399) 3.52 (-325)	0	3.58 (-128)	0	0	0	–	–
<i>pstSCAB-phoU*</i>	4.22 (-270)	3.56 (-111)	3.56 (-113)	3.45 (-128)	3.72 (-278)	–	4.04 (-34)	3.84 (-105)	3.53 (-50)	–
<i>metBL*</i>	3.79 (-73)	–	–	4.03 (-213)	3.61 (-16)	–	4.50 (-65)	0	3.69 (50)	3.92 (-171)
<i>aroP*</i>	3.80 (-91)	–	–	–	–	–	3.62 (-78)	–	3.67 (0)	3.61 (86)
<i>pdhR-aceEF-lpdA</i>	4.43 (-116)	4.43 (-117)	4.43 (-117)	3.84 (-397) 4.11 (-148)	4.04 (-64)	4.35 (-118)	3.50 (-380)	3.82 (-354)	3.87 (-114)	3.40 (-118)
<i>gltA*</i>	3.67 (-172)	3.69 (-212)	3.69 (-214)	3.45 (-366)	0	–	3.75 (-109)	–	–	–
<i>nuoABCDEFGHIJKLMN</i>	3.59 (-432)	4.34 (-198)	4.34 (-198)	3.50 (-382)	–	0	3.69 (-392)	–	0	0

Таблица 2. Окончание

Оперон	Геном									
	EC	STM	STY	YE	YP	SM	PHL	PHA	PCA	PCH
<i>cyoABCDE</i>	3.37 (-155)	–	–	–	3.50 (-222)	3.67 (86)	3.80 (-337)	–	–	3.77 (83)
<i>dcuB-fumB*</i>	3.73 (-282)	–	–	0	0	4.11 (-79)	0	0	3.64 (-277) 4.31 (-241) 3.54 (-98)	4.31 (-107)
<i>ndh</i>	5.03 (-151)	3.58 (-200) 4.65 (-151) 3.94 (-71)	3.58 (-198) 4.65 (-149) 3.94 (-69)	–	4.90 (-149)	4.64 (-109)	4.18 (-149)	4.52 (-151)	4.90 (-336) 3.52 (62)	5.05 (-340) 3.60 (75)
<i>atpIBEFHAGDC*</i>	3.92 (-110)	3.92 (-121)	3.92 (-123)	3.93 (-224)	3.51 (-307)	3.65 (80)	3.64 (-224)	3.53 (-225)	3.70 (83)	3.78 (74)
<i>icdA*</i>	3.60 (-147)	–	–	–	–	3.51 (-239) 4.41 (-4)	4.50 (-248) 3.60 (-2)	3.92 (-250) 3.60 (-4)	3.60 (-1)	3.56 (-317)
<i>mdh*</i>	3.93 (-258)	3.93 (-236)	3.93 (-236)	3.79 (-255)	–	–	3.87 (-85)	3.45 (-46)	–	–
<i>cydAB</i>	4.78 (-346) 4.57 (-294) 3.98 (-249)	5.08 (-348) 4.14 (-296) 3.81 (-251)	5.08 (-350) 4.14 (-298) 3.81 (-253)	3.90 (-376) 4.84 (-312) 4.31 (-261)	4.84 (-313) 4.48 (-262)	4.08 (-356)	–	–	4.84 (-399) 3.55 (-347)	3.87 (-397) 4.84 (-374)
Гены, не включенные в регулон										
<i>nikABCDE</i>	3.73 (-87)	0	0	0	0	0	0	0	0	–
<i>kbl-tdh</i>	4.37 (-122)	–	–	3.95 (-315)	0	0	0	0	0	0
<i>ybjX</i>	3.79 (-257)	–	–	3.53 (-333)	0	0	0	0	0	0
<i>ybdL</i>	3.54 (-343)	–	–	–	0	–	0	0	–	3.54 (-333)
<i>ftsYEX</i>	3.65 (-399)	–	–	–	–	–	–	0	–	–
<i>yjcRQP</i>	4.02 (-387)	0	0	0	0	3.52 (56)	0	0	0	0
<i>aceEF-lpdA</i>	4.29 (-132)	–	–	–	–	0	–	–	–	–
Гены, принадлежность которых к регулону на основе формального критерия определить невозможно										
<i>yidKJ</i>	3.86 (-239)	0	0	0	0	0	0	0	0	0
<i>cusCFBA</i>	3.55 (-184)	0	0	0	0	0	0	0	0	0
<i>yjcP</i>	4.42 (-146)	0	0	0	0	0	0	0	0	0
<i>potFGHI</i>	3.51 (-398)	3.79 (-99)	3.79 (-97)	–	–	–	0	0	–	–
<i>umuDC</i>	4.20 (-247) 4.02 (-1)	–	–	0	0	0	0	0	0	0

Примечание. Условные обозначения для геномов см. в Экспериментальной части. Указаны веса сайтов, в скобках приведено положение сайта относительно старта трансляции первого гена оперона. “–” – Сайт перед опероном отсутствует, “0” – ортологов данных генов не обнаружено. Жирным выделены гены “ядра”.

* Предсказанные в данной работе новые члены Fng –регулона.

Таблица 3. Потенциальные сайты связывания ArgA

Оперон	Геном									
	EC	STM	STY	YE	YP	SM	PHL	PHA	PCA	PCH
Гены, включенные в регулон										
<i>pdhR-aceEF-lpdA</i> ^{*,1}	4.42 (-230) 5.06 (-232)	4.85 (-85)	4.85 (-85)	4.85 (-116)	4.85 (-32)	4.85 (-86)	4.85 (-348)	4.85 (-322)	4.85 (-82)	4.85 (-86)
<i>lpdA1</i>	4.57 (-223) 4.85 (-84)	4.39 (-108)	4.39 (-105)	4.68 (-248)	4.68 (-138)	–	–	–	4.39 (-139)	4.39 (-120)
<i>cyoABCDE</i>	4.92 (-260) 4.95 (-249)	4.97 (-261) 4.69 (-250)	4.97 (-263) 4.69 (-252)	4.46 (-302)	4.71 (-301)	4.53 (-319) 4.82 (-286)	4.77 (-185)	4.83 (-186)	–	–
<i>dcuB-fumB</i> [*]	4.28 (-395) 4.53 (-274) 4.32 (-256) 4.95 (-390)	–	–	4.34 (-285)	–	4.11 (-79)	0	0	3.64 (-277) 4.31 (-241) 3.54 (-98)	4.31 (-107)
<i>ndh</i>	5.15 (-45) 4.65 (-34)	5.21 (-45) 4.65 (-34) 4.58 (-66)	5.21 (-43) 4.65 (-32) 4.58 (-64)	–	4.94 (-31)	4.71 (-388)	4.75 (-31)	5.13 (-33)	4.27 (-218)	4.27 (-221)
<i>atpIBEFHAGDC</i> [*]	4.59 (-339)	–	–	–	–	4.28 (-88)	4.60 (-290) 4.28 (-163)	–	–	–
<i>icdA</i>	4.91 (-111)	4.61 (-111)	4.61 (-109)	4.35 (-40)	4.73 (11)	4.73 (-41)	4.91 (-46)	4.91 (-48)	4.53 (-38)	4.53 (-38)
<i>mdh</i>	5.19 (-263)	5.19 (-241)	5.19 (-241)	–	–	–	4.71 (-285)	4.41 (-375)	–	5.05 (-318)
<i>cydAB</i>	5.94 (-586)	4.91 (-408)	4.91 (-406)	5.32 (-382)	4.60 (-343)	5.30 (-477)	4.92 (-537)	4.85 (-259)	4.47 (-617)	4.70 (-442)
<i>nuoABCDEFGHIJKLMN</i>	4.46 (-163)	4.46 (-163)	4.46 (-163)	4.46 (-171)	4.46 (-172)	0	5.70 (-540)	5.50 (-541)	0	0
<i>gltA</i>	4.34 (-393) 5.15 (-654)	4.24 (-129)	4.24 (-131)	4.40 (-365) 4.34 (-191)	4.83 (-250)	–	–	–	–	–

Таблица 4. Функции генов, перед которыми обнаружены потенциальные сайты связывания Fnr и/или ArcA

Оперон	Функция	Fnr-регу- лон	ArcA-ре- гулон	Ссылка
<i>pdhR-aceEF-lpdA</i>	Пируватдегидрогеназный комплекс	+	+	[46, 47]
<i>cyoABCDE</i>	Цитохром- <i>bo</i> -оксидаза	+	+	[39]
<i>dcuB-fumB</i>	DcuB – Транспортёр C4-дикарбоновых кислот FumB – Фумаратгидратаза	+	+	[40, 41]
<i>Ndh</i>	NADH-дегидрогеназа II	+	+	[37]
<i>atpIBEFHAGDC</i>	АТФ-синтаза	+	+	[45]
<i>icdA</i>	NADPH-зависимая изоцитратдегидрогеназа	+	+	[43]
<i>Mdh</i>	Малатдегидрогеназа	+	+	[42]
<i>gltA</i>	Цитратсинтаза	+	+	[44]
<i>cydAB</i>	Цитохром- <i>bd</i> -оксидаза	+	+	[39]
<i>nuoABCDEFGHIJKLMN</i>	NADH-дегидрогеназа I	+	+	[38]
<i>yigI-rarD-yigG-yigF</i>	Функция неизвестна	–	+	[12]
<i>gapA</i>	Глицеральдегид-3-фосфат-дегидрогеназа	–	+	[49]
<i>acrEF</i>	Гены устойчивости к акрифлавину	–	+	[50]
<i>yadF</i>	Ангидраза	–	+	[51]
<i>aroP</i>	Транспортёр аминокислот	+	–	[52]
<i>gatYZABCDR</i>	Утилизация галактитола	–	–	[53]
<i>lysU</i>	Лизил-тРНК-синтаза	–	–	[54]
<i>ycdM</i>	Предполагаемая монооксигеназа	–	–	[12]
<i>potFGHI</i>	Система транспорта путресцина	+	–	[55]
<i>hycBCDEFGHI</i>	Формиат-водород-лиаза	–	–	[56]
<i>ydbA</i>	Функция неизвестна	–	0	[12]
<i>yhiW</i>	Предполагаемый регулятор транскрипции	–	0	[57]
<i>tgt-yajC-secD-secF</i>	Tgt – тРНК-транскриптаза YajC – функция неизвестна SecD, SecF – система секреции белков	+	–	[58, 59]
<i>purMN</i>	PurM – фосфорибозиламиноимидазолсинтаза PurN – фосфорибозилглицинамид-формилтрансфераза	+	–	[60]
<i>narXL</i>	Двухкомпонентная система регуляции нитрат-нитрит-ного дыхания	+	–	[61]
<i>thrS-infC-rpmI-rplT</i>	ThrS – треонил-тРНК-синтаза InfC – фактор инициации трансляции IF-3 RpmI – рибосомный белок L35 RplT – рибосомный белок L20	+	–	[62–65]
<i>rpsT</i>	Рибосомный белок S20	+	–	[66]
<i>iadA</i>	Изоаспартилдипептидаза	+	–	[67]
<i>nanATEK-yhcH</i>	Гены катаболизма сиаловой кислоты	+	–	[68]
<i>upp-uraA</i>	Upp – урацил-фосфорибозилтрансфераза UraA – транспортёр урацила	+	–	[69]

Таблица 4. Окончание

Оперон	Функция	Fnr-регулон	ArcA-регулон	Ссылка
<i>ibpB-yidE</i>	IbpB – белок теплового шока YidE – предполагаемый транспортный белок	+	–	[70, 71]
<i>glgCAP</i>	Гены биосинтеза гликогена	+	–	[72–74]
<i>ybhK</i>	Функция неизвестна	+	–	[75]
<i>nirBDC-cysG</i>	NirBC – нитритредуктаза NirD – транспортер нитрита CysG – сирогемсинтаза	+	–	[76]
<i>rhaBAD</i>	Гены метаболизма рамнозы	+	–	[77]
<i>pstSCAB-phoU</i>	Система транспорта фосфатов	+	–	[78]
<i>metBL</i>	Гены биосинтеза метионина	+	–	[79]
<i>nikABCDE</i>	Гены транспорта никеля	–	–	[80]
<i>kbl-tdh</i>	Kbl – 2-амино-3-кетобутират-СоА-лигаза Tdh – L-треонин-3-дегидрогеназа	–	–	[81]
<i>ybjX</i>	Функция неизвестна	–	–	[75]
<i>ybdL</i>	Функция неизвестна	–	–	[75]
<i>ftsYEX</i>	Гены регуляции деления клетки	–	–	[82]
<i>yjcRQP</i>	Функция неизвестна	–	–	[75]
<i>yidKJ</i>	Функция неизвестна	0	–	[71]
<i>yjcP</i>	Функция неизвестна	0	–	[75]
<i>umuDC</i>	ДНК-полимераза V	0	–	[83]
<i>cusCFBA</i>	Система экспорта меди	–	–	[84]

Примечание. “+” – Оперон относится к регулону, “–” – оперон не относится к регулону, “0” – определение принадлежности оперона к регулону невозможно. Жирным выделены гены “ядра”.

1. Опероны, участвующие в процессах дыхания и центрального метаболизма: *pdhR-aceEF-lpdA*, *ndh*, *nuoABCDEFGHIJKLMN*, *atpIBEFHAGDC*, *cyoABCDE*, *mdh*, *icdA*, *gltA*, *cydAB*, *dcuB-fumB*, *narXL*, *glgCAP*, *nirBDC-cysG*, *pstSCAB-phoU*;

2. Опероны, не участвующие в процессе дыхания: *tgt-yajC-secDF*, *purMN*, *thrS-infC-rpmI-rplT*, *rpsT*, *yhjHG-iadA-yjiE*, *nanATEK-yhcH*, *upp-uraA*, *ibpB-yidE*, *rhaBAD*, *metBL*, *aroP*;

3. Ген, функция которого неизвестна или экспериментально не определена: *ybhK*.

В состав ArcA-регулона входят 16 оперонов. Восемь из них, *cyoABCDE* [30], *cydAB* [30], *icdA* [32], *lpdA* [33], *nuoABCDEFGHIJKLMN* [35], *gltA* [34], *mdh* [35] и *ndh* [36], как показано экспериментально, регулируются белком ArcA. Следовательно, мы предсказали восемь новых оперонов – потенциальных членов регулона ArcA (табл. 3).

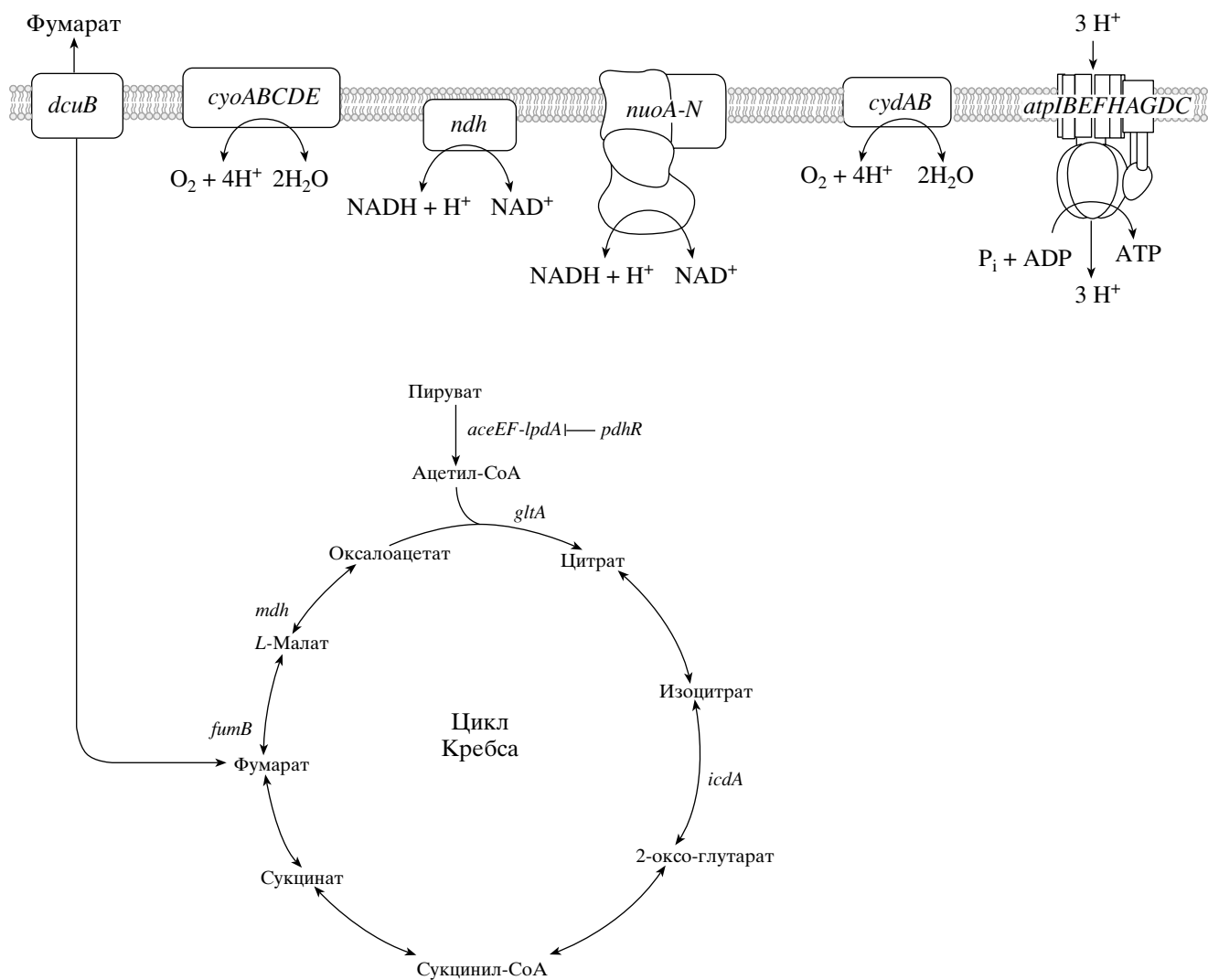
Рассматриваемые опероны также можно разделить на три функциональные группы (табл. 4):

1. Опероны, участвующие в процессах дыхания и центрального метаболизма: *pdhR-aceEF-lpdA*, *ndh*, *nuoABCDEFGHIJKLMN*, *atpIBEFHAGDC*, *cyoABCDE*, *mdh*, *icdA*, *gltA*, *cydAB*, *dcuB-fumB*, *gapA*, *lpdA*, *potFGHI*;

2. Опероны, не участвующие в процессе дыхания: *acrEF*;

3. Опероны, функция которых неизвестна или не определена экспериментально: *yigI-rarD-yigGF*, *yadF*.

Среди оперонов Fnr- и ArcA-регулонов можно выделить группу из 10 оперонов, которые входят одновременно в оба регулона (табл. 1), формируя так называемое “ядро”. Опероны, составляющие эту группу, характеризуются рядом особенностей. Сайты перед этими оперонами имеют наибольшие из возможных весов. Так, сайт связывания Fnr перед геном *ndh* имеет вес 5.03, а во многих других случаях веса сайтов связывания Fnr превышают 3.75. Что касается сайтов связывания ArcA, то только в опероне *atpIBEFHAGDC* вес ArcA



Функции генов “ядра” (пояснения в тексте).

сА-сайта ниже 4.75, все же остальные опероны содержат в своих регуляторных областях сайты связывания ArgA с весом, значительно превышающим эту величину и достигающим, в случае *mdh*, 5.19.

Практически все сайты связывания перед рассматриваемыми оперонами высококонсервативны. Так, Fnr-сайты перед оперонами *pdhR-aceEF-lpdA*, *ndh*, *atpIBEFHAGDC* и сайт связывания ArgA перед *pdhR-aceEF-lpdA*, *icdA*, *cydAB* сохраняются в 10 родственных геномах, а ArgA-сайт перед опероном *ndh* сохраняется в девяти родственных геномах.

Кроме того, многочисленные экспериментальные данные указывают на существование или возможность существования предсказанных нами сайтов перед некоторыми оперонами “ядра”. Ранее обнаружили регуляцию белком Fnr оперонов *pdhR-aceEF-lpdA* [27], *dcuB-fumB* [28], *ndh* [29], *cyd-*

AB [30], *cyoABCDE* [30] и *nuoABCDEFGHIJKLMN* [31] и модуляцию оперона *icdA* белком Fnr [32]. Белок ArgA регулирует опероны *lpdA* [27], *cyoABCDE* [30], *cydAB* [30], *icdA* [32], *gltA* [34], *mdh* [35], *nuoABCDEFGHIJKLMN* [35] и *ndh* [36].

Гены “ядра” кодируют, как правило, белки, участвующие в процессах дыхания и центрального метаболизма (рисунок). По выполняемым функциям эти гены разделены на следующие группы:

1. Гены NADH-дегидрогеназ: *ndh* [37] и *nuoABCDEFGHIJKLMN* [38];
2. Гены цитохром-*bo*- и *bd*-редуктаз, соответственно *cyoABCDE* и *cydAB* [39].
3. Гены метаболизма C4-дикарбоновых кислот: *dcuB-fumB* [40, 41]. Продукты этих генов – ферменты и белки-транспортеры, которые поставляют субстраты для фумаратного дыхания.

Таблица 5. Соотношение Fnr- и ArcA-регулонов и модулонов

Модулон/оперон	Группа, №						
	1	2	3	4	5	6	7
Fnr-модулон	+	+	+	+	+	+	–
Fnr-регулон	–	+	+	+	–	–	–
ArcA-регулон	–	–	–	+	+	–	–
ArcA-модулон	–	–	+	+	+	+	+
Число оперонов (генов)	36 (104)	3 (9)	13 (33)	10 (40)	6 (14)	83 (204)	21 (66)

Примечание. “+” – Группа оперонов входит в модулон или регулон; “–” – группа оперонов не входит в модулон или регулон.

Кроме того, белок FumB это один из ферментов цикла Кребса.

4. Гены ферментов цикла Кребса: *mdh* [42], *icdA* [43], *gltA* [44].

5. Гены субъединиц АТФ-синтазы: *atpIBEF-HAGDC* [45].

6. Гены, отвечающие за образование ацетилкофермента А из пирувата: *pdhR-aceEF-lpdA* [46, 47].

Соотношение сопряженных модулонов и регулонов. Вклад регуляторных каскадов в расширение модулонов

Fnr – глобальный регулятор дыхания у *E. coli*, который прямо или опосредованно влияет на экспрессию не только генов дыхательных ферментов, но и генов, кодирующих регуляторы более низкого уровня, например ArcA. Таким образом, ArcA представляет собой регулятор второго уровня. Ожидалось, что Fnr-модулон помимо генов Fnr-регулона будет включать гены ArcA-модулона. Для определения роли регуляторного каскада Fnr-ArcA в расширении Fnr-модулона по сравнению с Fnr-регулоном мы сопоставили полученные нами данные о составе регулона с данными, полученными с помощью биочипов, и составили схему, визуализирующую соотношение Fnr-регулона, Fnr-модулона, ArcA-регулона и ArcA-модулона (табл. 5). Как уже сказано, к Fnr-регулону нами были отнесены 26 оперонов. Большая часть, а именно 125 оперонов Fnr-модулона, не находится под прямой Fnr-регуляцией. В тоже время, 89 из этих оперонов входят в ArcA-модулон. Эти соотношения иллюстрируют идею расширения модулона регулятора более высокого уровня за счет модуляции регулятором более низкого уровня. Ни в Fnr-модулон, ни в один из двух исследованных регулонов не входит 21 оперон из ArcA-модулона. Скорее всего, это обусловлено высоким уровнем “шума” в результатах, полученных с помощью биочипов, или же связано с влиянием других регуляторов, экспрессия генов которых контролируется Fnr или ArcA.

Следует заметить, что вне Fnr-модулона не выявлены гены, относящиеся к ArcA-регулону. В то же время, опероны *narXL*, *queA-tgt-yajC-secD-secF*, *purMN* входят в Fnr-регулон и не входят в состав ArcA-модулона.

Нельзя исключить и возможность расширения модулонов за счет неизвестных регуляторов. Так, в геноме *E. coli* предсказано примерно 150 генов белков-регуляторов транскрипции, точная функция которых неизвестна [48].

Например, экспрессия оперона *narXL*, кодирующего систему нитрат-нитритного дыхания, регулируется белком Fnr. Ожидается, что этот регуляторный каскад также будет вносить вклад в расширение Fnr-модулона как регулятор второго уровня.

Противоречивые экспериментальные данные

В процессе определения членов модулонов по сформулированному в разделе “Экспериментальная часть” критерию были выделены опероны, данные об изменении экспрессии которых противоречивы. В эту группу вошли четыре оперона (табл. 6).

В пределах этой группы можно выделить две подгруппы оперонов:

– данные о влиянии регулятора Fnr противоречивы – три оперона (*cusCFBA*, *rpsJ-rplCDWB-rpsS-roIV-rpsC-rplP-rpmC-rpsQ*, *rplNXE-rpsNH-rplFR-rpsE-rpmD-rplO-prlA-rpmJ-rpsMKD-rpoA-rplQ*);

– данные о влиянии регулятора ArcA противоречивы – три оперона (*nucABCDEFGHIJKLMN*, *rpsJ-rplCDWB-rpsS-roIV-rpsC-rplP-rpmC-rpsQ*, *rplNXE-rpsNH-rplFR-rpsE-rpmD-rplO-prlA-rpmJ-rpsMKD-rpoA-rplQ*).

При этом перед опероном *cusCFBA* найден консервативный Fnr-сайт, а перед опероном *nucABCDEFGHIJKLMN* – консервативные Fnr- и ArcA-сайты. Перед опероном *nucABCDEFGHIJKLMN* экспериментально выявлен сайт связывания белка ArcA [31], что согласуется с нашими результатами.

Таблица 6. Противоречивые данные о влиянии регулятора на уровень экспрессии генов

Ген	Эффект Fnr [12]	Эффект ArcA [13]
Оперон <i>cusCFBA</i>		
<i>cusA</i>	+	+
<i>cusB</i>	+	+
<i>cusC</i>	0	Не изучен
Оперон <i>nuoABCDEFGHIJKLMN</i>		
<i>nuoE</i>	–	–
<i>nuoB</i>	Не изучен	–
<i>nuoF</i>	Не изучен	0
<i>nuoG</i>	Не изучен	0
Оперон <i>rplCDWB-rpsS-roIV-rpsC-rplP-rpmC-rpsQ</i>		
<i>rpsJ</i>	–	–
<i>rplB</i>	0	0
<i>rplC</i>	0	0
<i>rplD</i>	0	0
<i>rpmC</i>	+	+
Оперон <i>rplNXE-rpsNH-rplFR-rpsE-rpmD-rplO-prlA-rpmJ-rpsMKD-rpoA-rplQ</i>		
<i>prlA</i>	–	–
<i>rplX</i>	Не изучен	–
<i>rplO</i>	0	0
<i>rpsE</i>	0	0
<i>rplQ</i>	0	0
<i>rpsH</i>	0	0
<i>rplE</i>	0	0
<i>rplF</i>	Не изучен	0

Примечание. “–” – Репрессия, “0” – нет эффекта, “+” – активация.

Несмотря на противоречивость данных, полученных с помощью микрочипов, мы включили оперон *nuoABCDEFGHIJKLMN* в состав модулонов Fnr и ArcA и оперон *cusCFBA* – в Fnr-модулон, руководствуясь опубликованными данными в первом случае и нашим предсказанием – во втором.

Кроме того, по данным [12] оперон *cydAB* не относится к ArcA-модулону, хотя экспериментально установлено присутствие перед ним сайта связывания ArcA [35]. Поэтому оперон *cydAB* также включен в ArcA-модулон.

Противоречивыми являются данные о группе из 21 оперона, которые входят в ArcA-модулон, но не входят в Fnr-модулон. Указанные противоречия свидетельствуют о достаточно высоком уровне шума в рассмотренных исследованиях.

Таким образом, данные, полученные с использованием биочипов, можно рассматривать лишь как промежуточный этап изучения регуляторной системы. Однако совмещение их с данными, полученными методами сравнительной геномики, позволяет получить более точную информацию

как о составе, так и об эволюции регуляторных систем.

Работа была частично поддержана Медицинским институтом Ховарда Хьюза (55000309), Фондом поддержки российской науки и Программой “Молекулярная и клеточная биология” Российской академии наук.

СПИСОК ЛИТЕРАТУРЫ

1. Hatfield G.W., Hung S., Baldi P. 2003. Differential of DNA microarray gene expression data. *Mol. Microbiol.* **47**, 871–877.
2. Gelfand M.S., Koonin E.V., Mironov A.A. 2000. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res.* **28**, 695–705.
3. Rodionov D.A., Dubchak I., Arkin A., Alm E., Gelfand M.S. 2004. Reconstruction of regulatory and metabolic pathways in metal-reducing delta-proteobacteria. *Genome Biol.* **5**, R90.

4. Panina E.M., Mironov A.A., Gelfand M.S. 2001. Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res.* **29**, 5195–5206.
5. Panina E.M., Mironov A.A., Gelfand M.S. 2003. Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proc. Natl. Acad. Sci. USA.* **100**, 9912–9917.
6. 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.
7. Равчев Д.А., Рахманинова А.Б., Миронов А.А., Гельфанд М.С. 2005. Регуляция нитрат-нитритного дыхания гамма-протеобактерий: исследование методами сравнительной геномики. *Молекуляр. биология.* **39**, 832–846.
8. Melville S.B., Gunsalus R.P. 1996. Isolation of an oxygen-sensitive Fnr protein of *Escherichia coli*: interaction at activator and repressor sites of Fnr-controlled genes. *Proc. Natl. Acad. Sci. USA.* **93**, 1226–1231.
9. Malpica R., Franco B., Rodriguez C., Kwon O., Georgellis D. 2004. Identification of a quinone-sensitive redox switch in the ArcB sensor kinase. *Proc. Natl. Acad. Sci. USA.* **101**, 13318–13323.
10. Georgellis D., Kwon O., De Wulf P., Lin E.C. 1998. Signal decay through a reverse phosphorelay in the Arc two-component signal transduction system. *J. Biol. Chem.* **273**, 32864–32869.
11. Compan I., Touati D. 1994. Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA. *Mol. Microbiol.* **11**, 955–964.
12. Salmon K., Hung S.P., Mekjian K., Baldi P., Hatfield G.W. 2003. Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and Fnr. *J. Biol. Chem.* **278**, 29837–29855.
13. Salmon K.A., Hung S.P., Steffan N.R., Krupp R., Baldi P., Hatfield G.W., Gunsalus R.P. 2005. Global gene expression profiling in *Escherichia coli* K12: The effects of oxygen availability and ArcA. *J. Biol. Chem.* **280**, 15084–15096.
14. *Escherichia coli* and *Salmonella*. *Cellular and Molecular Biology*. Ed. Neighart F.C. 1996. Washington: ASM Press.
15. Гельфанд М.С., Миронов А.А. 1999. Компьютерный анализ регуляторных сигналов в полных бактериальных геномах. *Молекуляр. биология.* **33**, 772–778.
16. Герасимова А.В., Родионов Д.А., Миронов А.А., Гельфанд М.С. 2001. Компьютерный анализ регуляторных сигналов в бактериальных геномах. Участки связывания Fnr. *Молекуляр. биология.* **35**, 1001–1009.
17. Favorov A.V., Gelfand M.S., Gerasimova A.V., Ravchev D.A., Mironov A.A., Makeev V.J. 2005. A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length. *Bioinformatics.* **21**, 2240–2245.
18. Blattner F.R., Plunkett G. III, Bloch C.A. et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science.* **277**, 1453–1474.
19. 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.
20. McClelland M., Sanderson K.E., Spieth J. et al. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature.* **413**, 852–856.
21. Deng W., Burland V., Plunkett G. et al. 2002. Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol.* **184**, 4601–4611.
22. Bell K.S., Sebaihia M., Pritchard L. et al. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. USA.* **101**, 11105–11110.
23. Duchaud E., Rusniok C., Frangeul L. et al. 2003. The genome sequence of the entomopathogenic bacterium *Photobacterium luminescens*. *Nature Biotechnol.* **21**, 1307–1313.
24. Миронов А.А., Винокурова Н.П., Гельфанд М.С. 2000. Програмное обеспечение анализа бактериальных геномов. *Молекуляр. биология.* **34**, 253–262.
25. Takahashi K., Hattori T., Nakanishi T., Nohno T., Fujita N., Ishihama A., Taniguchi S. 1994. Repression of *in vitro* transcription of the *Escherichia coli* *fnr* and *narX* genes by FNR protein. *FEBS Lett.* **340**, 59–64.
26. Jayaraman P.S., Gaston K.L., Cole J.A., Busby S.J. 1988. The *nirB* promoter of *Escherichia coli*: location of nucleotide sequences essential for regulation by oxygen, the Fnr protein and nitrite. *Mol. Microbiol.* **2**, 527–530.
27. Quail M.A., Haydon D.J., Guest J.R. 1994. The *pdhR-aceEF-lpd* operon of *Escherichia coli* expresses the pyruvate dehydrogenase complex. *Mol. Microbiol.* **12**, 95–104.
28. 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.
29. Green J., Guest J.R. 1994. Regulation of transcription at the *ndh* promoter of *Escherichia coli* by Fnr and novel factors. *Mol. Microbiol.* **12**, 433–444.
30. Cotter P.A., Gunsalus R.P. 1992. Contribution of the *fnr* and *arcA* gene products in coordinate regulation of cytochrome o and d oxidase (*cyoABCDE* and *cydAB*) genes in *Escherichia coli*. *FEMS Microbiol. Lett.* **70**, 31–36.
31. 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 regulators. *Mol. Microbiol.* **16**, 521–534.
32. Chao G.C., Shen J., Tseng C.P., Park S.J., Gunsalus R.P. 1997. Aerobic regulation of isocitrate dehydrogenase gene (*icd*) expression in *Escherichia coli* by the *arcA* and *fnr* gene products. *J. Bacteriol.* **179**, 4299–4304.
33. Cunningham L., Georgellis D., Green J., Guest J.R. 1998. Co-regulation of lipoamide dehydrogenase and 2-oxoglutarate dehydrogenase synthesis in *Escherichia coli*: characterisation of an ArcA binding site in the *lpd* promoter. *FEMS Microbiol. Lett.* **169**, 403–408.
34. Park S.J., McCabe J., Turna J., Gunsalus R.P. 1994. Regulation of the citrate synthase (*gltA*) gene of *Escherichia coli* in response to anaerobiosis and carbon supply: role of the *arcA* gene product. *J. Bacteriol.* **176**, 5086–5092.

35. Lynch A.S., Lin E.C. 1996. Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*: characterization of DNA binding at target promoters. *J. Bacteriol.* **178**, 6238–6249.
36. Liu X., Wulf P. 2004. Probing the ArcA-P modulon of *Escherichia coli* by whole genome transcriptional analysis and sequence recognition profiling. *J. Biol. Chem.* **279**, 12588–12597.
37. Calhoun M.W., Gennis R.B. 1993. Demonstration of separate genetic loci encoding distinct membrane-bound respiratory NADH dehydrogenases in *Escherichia coli*. *J. Bacteriol.* **175**, 3013–3019.
38. Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. 1993. The gene locus of the proton-translocating NADH: ubiquinone oxidoreductase in *Escherichia coli*. Organization of the 14 genes and relationship between the derived proteins and subunits of mitochondrial complex I. *J. Mol. Biol.* **233**, 109–122.
39. Chepuri V., Lemieux L., Au D.C.T., Gennis R.B. 1990. The sequence of the *cyo* operon indicates substantial structural similarities between the cytochrome *o* ubiquinol oxidase of *Escherichia coli* and the aa3-type family of cytochrome *c* oxidases. *J. Biol. Chem.* **265**, 11185–11192.
40. Bell P.J., Andrews S.C., Sivak M.N., Guest J.R. 1989. Nucleotide sequence of the Fnr-regulated fumarase gene (*fumB*) of *Escherichia coli* K-12. *J. Bacteriol.* **171**, 3494–3503.
41. Six S., Andrews S.C., Roberts R.E., Uden G., Guest J.R. 1993. Construction and properties of *Escherichia coli* mutants defective in two genes encoding homologous membrane proteins with putative roles in anaerobic C4-dicarboxylic acid transport. *Biochem. Soc. Trans.* **21**, 342S.
42. McAlister-Henn L., Blaber M., Bradshaw R.A., Nisco S.J. 1987. Complete nucleotide sequence of the *Escherichia coli* gene encoding malate dehydrogenase. *Nucleic Acids Res.* **15**, 4993.
43. Apostolakos D., Menter P.A., Rampsch B.J., Reeves H.C., Birge E.A. 1982. Genetic map position of the cistron coding for isocitrate dehydrogenase in *Escherichia coli* K-12. *Current Microbiol.* **7**, 45–47.
44. Ner S.S., Bhayana V., Bell A.W., Giles I.G., Duckworth H.W., Bloxham D.P. 1983. Complete sequence of the *gl-tA* gene encoding citrate synthase in *Escherichia coli*. *Biochemistry.* **22**, 5243–5249.
45. Kanazawa H., Mabuchi K., Futai M. 1982. Nucleotide sequence of the promoter region of the gene cluster for proton-translocating ATPase from *Escherichia coli* and identification of the active promoter. *Biochem. Biophys. Res. Commun.* **107**, 568–575.
46. Haydon D.J., Quail M.A., Guest J.R. 1993. A mutation causing constitutive synthesis of the pyruvate dehydrogenase complex in *Escherichia coli* is located within the *pdhR* gene. *FEBS Lett.* **336**, 43–47.
47. Stephens P.E., Darlison M.G., Lewis H.M., Guest J.R. 1983. The pyruvate dehydrogenase complex of *Escherichia coli* K12. Nucleotide sequence encoding the pyruvate dehydrogenase component. *Eur. J. Biochem.* **133**, 155–162.
48. Perez-Rueda E., Collado-Vides J. 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28**, 1838–1847.
49. Branlant G., Branlant C. 1985. Nucleotide sequence of the *Escherichia coli* *gap* gene. Different evolutionary behavior of the NAD⁺-binding domain and of the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* **150**, 61–66.
50. Klein J.R., Henrich B., Plapp R. 1991. Molecular analysis and nucleotide sequence of the *envCD* operon of *Escherichia coli*. *Mol. Gen. Genet.* **230**, 230–240.
51. Merlin C., Masters M., McAteer S., Coulson A. 2003. Why is carbonic anhydrase essential to *Escherichia coli*? *J. Bacteriol.* **185**, 6415–6424.
52. Honore N., Cole S.T. 1990. Nucleotide sequence of the *aroP* gene encoding the general aromatic amino acid transport protein of *Escherichia coli* K-12: homology with yeast transport proteins. *Nucleic Acids Res.* **18**, 653–653.
53. Nobelmann B., Lengeler J.W. 1995. Sequence of the *gat* operon for galactitol utilization from a wild-type strain EC3132 of *Escherichia coli*. *Biochim. Biophys. Acta.* **1262**, 69–72.
54. Leveque F., Plateau P., Dessen P., Blanquet S. 1990. Homology of *lysS* and *lysU*, the two *Escherichia coli* genes encoding distinct lysyl-tRNA synthetase species. *Nucleic Acids Res.* **18**, 305–312.
55. Pistocchi R., Kashiwagi K., Miyamoto S., Nukui E., Sadakata Y., Kobayashi H., Igarashi K. 1993. Characteristics of the operon for a putrescine transport system that maps at 19 minutes on the *Escherichia coli* chromosome. *J. Biol. Chem.* **268**, 146–152.
56. Boehm R., Sauter M., Boeck A. 1997. Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenlyase components. *Science.* **277**, 1453–1474.
57. Jiang Y., Yang F., Zhang X., Yang J., Chen L., Yan Y., Nie H., Xiong Z., Wang J., Dong J., Xue Y., Xu X., Zhu Y., Chen S., Jin Q. 2005. The complete sequence and analysis of the large virulence plasmid pSS of *Shigella sonnei*. *Plasmid.* **54**, 149–159.
58. Reuter K., Slany R., Ullrich F., Kersten H. 1991. Structure and organization of *Escherichia coli* genes involved in biosynthesis of the deazaguanine derivative queuine, a nutrient factor for eukaryotes. *J. Bacteriol.* **173**, 2256–2264.
59. Gardel C., Johnson K., Jacq A., Beckwith J. 1990. The *secD* locus of *E. coli* codes for two membrane proteins required for protein export. *EMBO J.* **9**, 3209–3216.
60. Smith J.M., Daum H.A. 1986. Nucleotide sequence of the *purM* gene encoding 5'-phosphoribosyl-5-aminoimidazole synthetase of *Escherichia coli* K12. *J. Biol. Chem.* **261**, 10632–10636.
61. Nohno T., Noji S., Taniguchi S., Saito T. 1989. The *narX* and *narL* genes encoding the nitrate-sensing regulators of *Escherichia coli* are homologous to a family of prokaryotic two-component regulatory genes. *Nucleic Acids Res.* **17**, 2947–2957.
62. Mayaux J.-F., Fayat G., Fromant M., Springer M., Grunberg-Manago M., Blanquet S. 1983. Structural and transcriptional evidence for related *thrS* and *infC* expression. *Proc. Natl. Acad. Sci. USA.* **80**, 6152–6156.
63. Sacerdot C., Fayat G., Dessen P., Springer M., Plumbridge J.A., Grunberg-Manago M., Blanquet S. 1982. Sequence of a 1.26-kb DNA fragment containing the

- structural gene for *E. coli* initiation factor IF3: presence of an AUU initiator codon. *EMBO J.* **1**, 311–315.
64. Wada A., Sako T. 1987. Primary structures of and genes for new ribosomal proteins A and B in *Escherichia coli*. *J. Biochem.* **101**, 817–820.
65. Fayat G., Mayaux J.-F., Sacerdot C., Fromant M., Springer M., Grunberg-Manago M., Blanquet S. 1983. *Escherichia coli* phenylalanyl-tRNA synthetase operon region. Evidence for an attenuation mechanism. Identification of the gene for the ribosomal protein L20. *J. Mol. Biol.* **171**, 239–261.
66. Mackie G.A. 1981. Nucleotide sequence of the gene for ribosomal protein S20 and its flanking regions. *J. Biol. Chem.* **256**, 8177–8182.
67. Burland V.D., Plunkett G., Sofia H.J., Daniels D.L., Blattner F.R. 1995. Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. *Nucleic Acids Res.* **23**, 2105–2119.
68. Kalivoda K.A., Steenbergen S.M., Vimr E.R., Plumbridge J. 2003. Regulation of sialic acid catabolism by the DNA binding protein NanR in *Escherichia coli*. *J. Bacteriol.* **185**, 4806–4815.
69. Anderson P.S., Smith J.M., Mygind B. 1992. Characterization of the *upp* gene encoding uracil phosphoribosyltransferase of *Escherichia coli* K12. *Eur. J. Biochem.* **204**, 51–56.
70. Allen S.P., Polazzi J.O., Gierse J., Easton A.M. 1992. Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. *J. Bacteriol.* **174**, 6938–6947.
71. Burland V.D., Plunkett G., Daniels D.L., Blattner F.R. 1993. DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: organizational symmetry around the origin of replication. *Genomics.* **16**, 551–561.
72. Baecker P.A., Furlong C.E., Preiss J. 1983. Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADP-glucose synthetase as deduced from the nucleotide sequence of the *glgC* gene. *J. Biol. Chem.* **258**, 5084–5088.
73. Kumar A., Larsen C.E., Preiss J. 1986. Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADP-glucose:alpha-1,4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the *glgA* gene. *J. Biol. Chem.* **261**, 16256–16259.
74. Choi Y.-L., Kawamukai M., Utsumi R., Sakai H., Komano T. 1989. Molecular cloning and sequencing of the glycogen phosphorylase gene from *Escherichia coli*. *FEBS Lett.* **243**, 193–198.
75. Oshima T., Aiba H., Baba T. et al. 1996. A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7–28.0 min region on the linkage map. *DNA Res.* **3**, 137–155.
76. Peakman T., Crouzet J., Mayaux J.F., Busby S.J.W., Mohan S., Harborne N., Wootton J., Nicolson R., Cole J.A. 1990. Nucleotide sequence, organisation and structural analysis of the products of genes in the *nirB-cysG* region of the *Escherichia coli* K-12 chromosome. *Eur. J. Biochem.* **191**, 315–323.
77. Moralejo P., Egan S.M., Hidalgo E.F., Aguilar J. 1993. Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. *J. Bacteriol.* **175**, 5585–5594.
78. Amemura M., Makino K., Shinagawa H., Kobayashi A., Nakata A. 1985. Nucleotide sequence of the genes involved in phosphate transport and regulation of the phosphate regulon in *Escherichia coli*. *J. Mol. Biol.* **184**, 241–250.
79. Duchange N., Zakin M.M., Ferrara P., Saint-Girons I., Park I., Tran S.V., Py M.-C., Cohen G.N. 1983. Structure of the *metJBLF* cluster in *Escherichia coli* K12. Sequence of the *metB* structural gene and of the 5'- and 3'-flanking regions of the *metBL* operon. *J. Biol. Chem.* **258**, 14868–14871.
80. Navarro C., Wu L.-F., Mandrand-Berthelot M.-A. 1993. The *nik* operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. *Mol. Microbiol.* **9**, 1181–1191.
81. Aronson B.D., Ravnika P.D., Somerville R.L. 1988. Nucleotide sequence of the 2-amino-3-ketobutyrate coenzyme A ligase (*kbl*) gene of *E. coli*. *Nucleic Acids Res.* **16**, 3586–3586.
82. Gill D.R., Hatfull G.F., Salmond G.P.C. 1986. A new cell division operon in *Escherichia coli*. *Mol. Gen. Genet.* **205**, 134–145.
83. Kitagawa Y., Akaboshi E., Shinagawa H., Horii T., Oga-wa H., Kato T. 1985. Structural analysis of the *umu* operon required for inducible mutagenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* **82**, 4336–4340.
84. Yamamoto K., Ishihama A. 2005. Transcriptional response of *Escherichia coli* to external copper. *Mol. Microbiol.* **56**, 215–227.