

УДК 577.214.4

ТАКСОН-СПЕЦИФИЧНАЯ РЕГУЛЯЦИЯ SOS-ОТВЕТА У γ -ПРОТЕОБАКТЕРИЙ

© 2007 г. Л. В. Сычева^{1*}, Е. А. Пермина², М. С. Гельфанд^{1,3}

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

²ГосНИИгенетика, Москва, 113545

³Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Москва, 127994

Поступила в редакцию 07.11.2006 г.

Принята к печати 22.01.2007 г.

Система SOS-ответа – каскад последовательных реакций, индуцируемых повреждением клеточной ДНК. Гены, непосредственно участвующие в реализации этих реакций, регулируются белком LexA, который связывается со специфическими нуклеотидными последовательностями (сайтами) в их 5'-некодирующей области. Присутствие такой последовательности в регуляторной области гена служит критерием, по которому с помощью специальных методов можно выявить потенциально регулируемые SOS-ответом гены. В данной работе изучены гены, регуляция которых специфична для геномов определенной таксономической группы (Enterobacteriales, Pasteurellales, Vibrionales, Pseudomonadales и Alteromonadales). Среди них обнаружены гены, участие которых в SOS-ответе ранее не было установлено, но которые имеют консервативный сайт связывания с белком LexA в регуляторной области и функцию, связанную, возможно, с ответом клетки на повреждение ДНК. Среди этих генов: *mfd*, продукт которого в случае остановки транскрипции из-за повреждения ДНК способствует репарации матричной цепи; ген *VC0082*, кодирующий рекомбиназу; *VP2449*, отвечающий за устойчивость к воздействию ксенобиотиков. Описаны состав и эволюция LexA-регулона γ -протеобактерий.

Ключевые слова: SOS-ответ, LexA, γ -протеобактерии, регуляция транскрипции, сравнительная геномика.

TAXON-SPECIFIC REGULATION OF SOS-RESPONSE IN GAMMA-PROTEOBACTERIA, by L. V. Sycheva^{1*}, E. A. Permina², M. S. Gelfand^{1,3} (¹Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, 119992 Russia; *e-mail: lada.sychova@gmail.com; ²State Scientific Center "GosNII Genetika", Moscow, 113545 Russia; ³Kharkevich Institute for Information Transmission Problems Russian Academy of Sciences, Moscow, 127994 Russia). SOS-response system is a cascade of reactions induced by DNA damage in a cell. Genes participate in these reactions are regulated by the LexA protein binding to specific sequence in their upstream regions. The criterion for selection of genes putatively responsible for the SOS-response is the presence of such sequence. Genes with taxon-specific regulation in Enterobacteriales, Pasteurellales, Vibrionales, Pseudomonadales and Alteromonadales were analyzed using comparative genomic approaches. Some genes have conserved sites in regulatory region and suitable function, although their function in SOS-response has not been studied in experiment. The list of such genes includes *mfd*, which encodes a product repairing the mother chain in case of DNA damage-caused transcription stop; *VC0082*, which encodes a recombinase, and *VP2449*, responsible for xenobiotics resistance. Overall, this study characterized the content and evolution of the LexA regulon in gamma-proteobacteria are described here.

Key words: SOS-response, LexA, gamma-proteobacteria, regulation of transcription, comparative genomics.

Система SOS-ответа – каскад последовательных реакций, индуцируемых повреждением ДНК клетки [1]. Впервые возможность существования такой системы предположили Дефай (Defais M.) и соавт. [2], а затем эту гипотезу подтвердили и развили Рэдмен (Radman M.) и соавт. [3] SOS-ответ защищает клетку от повреждений ДНК, вызванных жестким химическим воздействием, радиоак-

тивным излучением, высоким давлением [4] или температурой. Известно несколько систем клетки, обеспечивающих сохранность генетического материала, например, рекомбинационная репарация, мутагенная репарация, эксцизионная репарация, репарация неспаренных нуклеотидов и репарация, осуществляемая ДНК-фотолиазой [5]. Система SOS-ответа хорошо изучена у *Escherichia coli*. Это касается как механизма регуляции, так и

* Эл. почта: lada.sychova@gmail.com

генов, продукты которых участвуют в восстановлении поврежденной ДНК [6].

В обычных условиях регулятор LexA репрессирует все гены SOS-ответа, связываясь со специфической последовательностью, называемой SOS-боксом (консенсус ТАСТГТАТАТАТАСAGTA) [7], в 5'-некодирующей области регулируемого гена. SOS-регулон не однороден по распределению сайтов связывания LexA в регулируемых им генах и, кроме этого, LexA обладает разным сродством к этим сайтам [8]. При повреждении ДНК белок RecA связывается с ее одноцепочечными участками (возникшими вследствие повреждения) и образует нуклеопротеиновые филаменты, переходя при этом в активированную форму – RecA*. В результате взаимодействия RecA* и LexA происходит протеолитическое расщепление белка LexA по связи Ala84–Gly85 и дерепрессия генов SOS-ответа.

После устранения последствий повреждения количество активированных молекул RecA снижается, что приводит к восстановлению пула нерасщепленных молекул LexA. Для функционирования механизма обратной связи клетке необходимо уменьшить количество молекул RecA, поэтому этот механизм включает LexA-зависимую репрессию гена *recA*.

В состав LexA-регулона входят гены, отвечающие за репарацию разрывов в дочерней и двойной цепочке ДНК и рекомбинацию (*recA*), гены полимеразы (*umuDC*, *dinP*), рекомбиназы (*recA*, *recN*), нуклеазы, гены, отвечающие за эксцизионную репарацию (*uvrAB*), а также гены, кодирующие хеликазу (*uvrD*) и ингибитор клеточного деления (*sulA*) [6]. Одним из первых в реакцию SOS-ответа вступает белок Ssb (single-stranded binding protein), который связывается с одноцепочечными участками ДНК [9]. Молекулы Ssb входят в состав полимеразного комплекса, поэтому связывание необходимо для последующей репликации. Гены *umuD* и *umuC* кодируют белки, составляющие ДНК-полимеразный комплекс (полимераза V). Иногда полимеразы V подвергается неортологичной замене. В частности, у *Caulobacter crescentus* оперон, состоящий из трех генов – *imuA*, *imuB* и *dnaE2*, кодирует каталитическую субъединицу полимеразы III и регулируется RecA [10].

Сложный фермент, образованный продуктами генов *umuD* и *umuC*, а также белками RecA и Ssb, работает быстрее, чем вегетативная полимеразы, но делает больше ошибок. Показано, что у *Pseudomonas putida* мутагенез, ассоциированный с полимеразой IV, не обусловлен запуском SOS-ответа, потому что транскрипция с промотора гена *dinB*, кодирующего полимеразу IV, не зависит от SOS-активности клетки. При SOS-ответе участок, в нуклеотидной последовательности которого возникла ошибка, вырезается нуклеазой – сложным ферментом, состоящим из продуктов

генов *uvrA* и *uvrB*. Продукты генов *recA* и *recN* и хеликазы UvrD осуществляют рекомбинационную репарацию ДНК.

Повреждение ДНК может привести к гибели клетки; решающим фактором в этом случае оказывается то, способна ли SOS-система восстановить генетический материал или нет. Механизм запуска SOS-ответа активен в течение всего жизненного цикла клетки и, в частности, во время деления, которое может остановиться в результате SOS-ответа до устранения повреждений в ДНК, что обеспечивается работой белка Sula.

Ортологи LexA встречаются в разных таксономических группах. Например, в геноме *Bacillus subtilis* регулятор SOS-системы – белок DinR, ортологичный белку LexA из *E. coli*. Сравнение DinR с его гомологами из протео- и грамположительных бактерий выявило консервативные участки внутри С-концевого домена, предположительно ответственного за автокаталитическое расщепление белка [1]. У γ -протеобактерий и грамположительных бактерий НТН-домен (helix–turn–helix) белка-регулятора консервативен, но этого недостаточно для сохранения мотива сигнала. DinR способен к автокаталитическому расщеплению подобно белку LexA *E. coli*, а ген *recA* репрессируется при связывании DinR с его промоторной областью [11]. Сайт связывания DinR в *B. subtilis* (Cheo-бокс) имеет палиндромную структуру, однако его консенсус (CGAACATATGTTC) не совпадает с консенсусом SOS-бокса у протеобактерий [12].

Изучение регулона LexA у α -протеобактерий с помощью экспериментальных методов и *in silico* [13] показало возможность участия в системе SOS-ответа нескольких новых генов, в том числе гена *parE*, кодирующего субъединицу В ДНК-топоизомеразы IV, а также гена *comM*, продукт которого – Mg²⁺-хелатаза, регулирует работу полимеразы во время SOS-ответа. В геномах подгрупп *Rhodobacter*, *Sinorhizobium*, *Agrobacterium*, *Caulobacter* и *Brucella* обнаружен консервативный tandemный повтор GTTC-N₇-GTTC в LexA-связывающей области [14].

Возможность предсказания сайтов связывания рассматривали ранее с использованием таких методов, как алгоритм конструирования консенсуса [15], алгоритм максимизации математического ожидания [16], анализ частоты встречаемости олигонуклеотидов [17], алгоритмы выборочного метода Гиббса [18] (табл. 1).

Для поиска регуляторных мотивов использовали также консенсусную последовательность [19]. При сравнении нуклеотидной последовательности, в которой ищут регуляторный сайт, с консенсусом используют индекс несоответствия HI (heterology index). Последовательности, индекс несоответствия которых меньше 15 (HI < 15), считают

Таблица 1. Методы предсказания сайтов связывания регуляторов экспрессии генов

Суммарное количество генов	Способ поиска	Нуклеотидная последовательность	Ссылка
69	Консенсусная последовательность	TACTG-(TA) ₅ -CAGTA	[19]
39	Консенсусная последовательность	CTG-N ₁₀ -CAG	[13]
19	Консенсусная последовательность, матрица позиционных весов	TACTGT-(AT) ₂ -ACAT-A/C-CAG-T/C-A	[17]
54	Матрица позиционных весов	TACTG-(TA) ₅ -CAGTA	Авторы данной работы

регуляторными и связывающими LexA с большой вероятностью.

Опубликован алгоритм поиска регуляторных мотивов из трех шагов [20]. Сначала в геномной ДНК проводится поиск совпадений с последовательностью регуляторного мотива, затем находки рекурсивно фильтруются и строится консенсусная матрица. В конечном итоге отбирают те мотивы, у которых индекс отклонения от консенсусной последовательности меньше восьми.

Завершающий шаг – программа автоматически собирает функциональные аннотации для найденных потенциальных членов регулона, используя сервер TBLASTN и базу данных GenBank.

Разработан также алгоритм поиска связанных между собой метаболических модулей, на уровне генома представляющих собой регулоны [21]. Для предсказания генов, входящих в регулон, и изучения бактериальных геномов используют методы теории информации, программирования и байесовой статистики. Результаты анализа новых бактериальных геномов вносят в базу данных RegulonDB (http://www.cifn.unam.mx/Computational_Genomics/regulondb/). Для поиска регуляторных сигналов используют консенсусную матрицу, по которой программа Patser строит весовую матрицу.

Ранее мы изучили (неопубликованные данные) ядро регулона у представителей таксона γ -протеобактерий. Предполагается, что внутри группы филогенетически близких организмов регулон консервативен. Это предположение лежит в основе изучения регулонов в геномах тех организмов, регуляция потенциальных членов регулона у которых еще не показана экспериментально.

Цель представленной работы состояла в изучении генов, регуляция которых специфична для геномов определенной таксономической группы, т.е. в полном описании состава LexA-регулона γ -протеобактерий. Эти данные могут быть использованы для изучения эволюции LexA-регулона.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

SOS-ответ у γ -протеобактерий изучали методом сравнительного анализа регуляции генов LexA-регулона [22]. В ходе работы сначала было подтверждено присутствие сайтов регуляции в 5'-некодирующей области тех генов, регуляция которых белком LexA установлена экспериментально. Затем среди генов, регулируемых LexA, выявили те, которые имеют сайты регуляции исключительно в геномах представителей определенной филогенетической группы (такие гены называются регулируемыми таксон-специфически). Критериями, позволяющими отнести ген к группе регулируемых таксон-специфически, помимо присутствия потенциального сайта регуляции в 5'-некодирующей области, были соотношения минимального количества геномов, в которых ген имеет регуляторный сайт (50%), и общего количества секвенированных геномов в изучаемой группе, а также функция этого гена.

Сравнительный анализ потенциальных регуляторных сайтов связывания проведен с помощью программы GenomeExplorer [23].

При поиске сигнала в нуклеотидной последовательности следует учитывать, в каких областях выше вероятность его найти. Регуляторную функцию выполняют, как правило, последовательности в 5'-некодирующей области, поэтому потенциальные сайты искали именно в этих областях с границами в позициях -200 и +50 относительно аннотированного начала гена.

Вес сайта вычисляли при помощи матрицы позиционных весов (PWM) (табл. 2), которую строили с использованием программы SignalX из пакета GenomeExplorer по формуле

$$W(b, k) = \log [N(b, k) + 0.5] -$$

$$- 0.25 \sum_{i=A, C, G, T} \log [N(i, k) + 0.5],$$

где $N(b, k)$ – количество появлений нуклеотида b в позиции k .

Вес предполагаемого сайта определяли как сумму весов нуклеотидов в соответствующих позициях:

$$Z(b_1 \dots b_k) = \sum_{k=1 \dots k} W(b_k, k),$$

где k – длина сайта. Основание логарифма выбирали так, чтобы на случайных последовательностях распределение Z имело математическое ожидание 0 и дисперсию 1. Вес Z позволяет оценить значимость индивидуального сайта.

Рассматривали сайты с весом не менее 3.75. При таком пороге и параметрах поиска в геноме *E. coli* нашли 92 потенциальных сайта.

При построении множественных выравниваний и филогенетических деревьев использовали программы ClustalX [24] и PhyIip [25]. Для изображения филогенетических деревьев использовали программу GeneMaster.

Рассмотрено 18 полных геномов γ -протеобактерий пяти подгрупп: **Enterobacteriales** (*E. coli* (ECC) [25, 26], *Salmonella typhi* (STY) [28], *S. typhimurium* (STM) [29], *Shigella flexneri* (SFX) [30], *Yersinia enterocolitica* (YEN) [31], *Y. pestis* (YPE) [26], *Photobacterium luminescens* (PLU) [32]), **Pasteurellales** (*Haemophilus ducreyi* (HDU) [33], *H. influenzae* (HIN) [34], *Pasteurella multocida* (PMU) [35]), **Vibrionales** (*Vibrio cholerae* (VCH) [26], *V. parahaemolyticus* (VPA) [36], *V. vulnificus* (VVU) [37], *Photobacterium profundum* (PPR) [38]), **Pseudomonadales** (*P. aeruginosa* (PAE) [26], *P. putida* (PPU) [38], *P. syringae* (PST) [38]) и **Alteromonadales** (*Shewanella oneidensis* (SON) [39]).

При поиске гомологов в базах данных GenBank и TCDB [40] использовали программу BLASTP с параметрами, устанавливаемыми по умолчанию [41], а также базу данных кластеров ортологических генов COG [42].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Ядро регулона

Ядром регулона считают гены, которые регулируются у филогенетически далеких организмов и заведомо регулируются внутри группы филогенетически близких организмов. Участок связывания регулятора у таких генов очень консервативен.

К наиболее консервативному ядру SOS-регулона относятся гены *lexA*, *recA* и *recN*. Эти гены имеют сайты связывания с белком LexA во всех исследованных геномах.

Enterobacteriales

В геномах *E. coli*, *S. typhi*, *S. typhimurium*, *S. flexneri*, *Y. enterocolitica*, *Y. pestis*, *P. luminescens* показана регуляция белком LexA таких эксперимен-

Таблица 2. Матрица позиционных весов для поиска сайтов связывания LexA

A	C	G	T
-0.01	0.03	-0.15	0.12
0.23	-0.05	-0.01	-0.17
-0.15	0.46	-0.15	-0.15
-0.15	-0.15	-0.15	0.46
-0.15	-0.15	0.46	-0.15
-0.13	-0.13	-0.05	0.30
0.28	-0.07	-0.14	-0.07
-0.22	0.07	-0.22	0.37
0.23	-0.31	0.10	-0.03
0.08	-0.07	-0.19	0.18
0.15	-0.14	-0.05	0.04
0.01	-0.01	-0.11	0.11
0.23	-0.05	-0.17	-0.01
0.05	0.03	-0.21	0.13
0.22	0.15	-0.31	-0.07
-0.15	0.46	-0.15	-0.15
0.46	-0.15	-0.15	-0.15
-0.15	-0.15	0.46	-0.15
-0.11	0.01	-0.11	0.21
0.12	-0.15	0.01	0.01

тально установленных членов SOS-регулона, как *lexA*, *recA*, *recN*, *uvrB*, *uvrD*, *dinI*, *ftsK*, *sulA*, *ssb* и *dinP* (табл. 3).

Известны также гены ядра регулона, регуляторный сайт перед которыми обнаружен не у всех представителей группы. Сайт регуляции перед геном *guvA* найден в шести геномах, у *Y. pestis* он отсутствует. В геномах *Y. enterocolitica*, *Y. pestis* и *P. luminescens* перед генами *uvrA* и *dinG* нет регуляторного сайта. У *S. typhi* нет и самого гена *dinG*.

Ген *umuD* отсутствует в геноме *Y. pestis* и *P. luminescens*. Регуляторный сайт перед геном *dinD* найден в двух случаях, а в геномах *S. typhi*, *S. typhimurium*, *Y. enterocolitica*, *Y. pestis* и *P. luminescens* не обнаружен сам ген.

Потенциально новый участник SOS-регулона, регуляция которого осуществляется таксон-специфически – ген *udjM* (рис. 1), сайт связывания LexA в его 5'-некодирующей области обнаружен у всех изученных организмов группы за исключением *P. luminescens*. Ген относится к группе ортологических генов, кодирующих связанную с мембраной металл-зависимую гидролазу (COG1988).

В шести случаях регуляторный мотив найден в гене *uscR*, кодирующем трансформирующий ДНК белок (рис. 2). Он принадлежит к группе ортологических генов, кодирующих регуляторы специфич-

Таблица 3. Ядро регулона и потенциальные таксон-специфичные гены в геномах представителей порядка *Enterobacteriales*

Ген*	Аннотация	ECC	STY	STM	SFX	YEN	YPE	PLU
<i>lexA</i>	Регулятор SOS-системы	+	+	+	+	+	+	+
<i>recA</i>	Корегулятор SOS-системы	+	+	+	+	+	+	+
<i>recN</i>	Репарация ДНК	+	+	+	+	+	+	+
<i>ruvA</i>	Хеликаза структуры Холлидея	+	+	+	+	+	-	+
<i>uvrA</i>	Эндонуклеаза ABC, субъединица А	+	+	+	+	-	-	-
<i>uvrB</i>	Эндонуклеаза ABC, субъединица В	+	+	+	+	+	+	+
<i>uvrD</i>	ДНК-хеликаза II	+	+	+	+	+	+	+
<i>umuD</i>	Регулятор полимеразного комплекса	+	+	+	+	+	0	0
<i>dinD</i>	Белок D, индуцируемый разрушением ДНК	+	0	0	+	0	0	0
<i>dinG</i>	АТР-зависимая хеликаза	+	0	+	+	-	-	-
<i>dinI</i>	Белок I, индуцируемый разрушением ДНК	+	+	+	+	+	+	+
<i>ftsK</i>	Белок клеточного деления	+	+	+	+	+	+	+
<i>sulA</i>	Ингибитор клеточного деления	+	+	+	+	+	+	+
<i>ssb</i>	Связывающий одноцепочечную ДНК белок	+	+	+	+	+	+	+
<i>dinP</i>	ДНК-полимераза IV	+	+	+	+	+	+	+
<i>yccR</i>	ДНК-трансформирующий белок	+	+	+	+	+	+	0
<i>ydjQ</i>	Нуклеазная субъединица экзонуклеазного комплекса	+	+	+	+	-	-	-
<i>otsB</i>	Регулятор осмотического давления	+	+	+	+	0	0	0
<i>sbmC</i>	Ингибитор гиразы	+	+	+	0	+	0	0
<i>yfiK</i>	Переносчик треонина	+	+	+	+	0	0	0
<i>ygjF</i>	G/U-мисметч-специфическая ДНК-гликозилаза	+	+	+	0	0	0	0
<i>ydjM</i>	Мембраносвязанная металл-зависимая гидролаза	+	+	+	+	+	+	-

Примечание. Здесь и далее: названия геномов расшифрованы в разделе “Экспериментальная часть”. “+” – перед геном есть потенциальный сайт регуляции; “-” перед геном нет потенциального сайта регуляции; 0 – ортологичный ген в геноме отсутствует.
* Названия генов взяты из генома *E. coli*.

ных для компетентных клеток генов (COG3070). Этого гена нет в геноме *P. luminescens*.

В четырех геномах LexA предположительно регулирует гены *otsB* (регулятор осмотического давления), *ydjQ* (нуклеазная субъединица экзонуклеазного комплекса, COG0322) (рис. 3), *sbmC* (ингибитор гиразы, COG3449), а также *yfiK* (транспортный белок, COG1280) (рис. 4).

В трех геномах – *E. coli*, *S. typhi* и *S. typhimurium* – регуляторный сайт присутствует в гене *ygjF*, ко-

дирующем G/U-мисметч-специфичную ДНК-гликозилазу.

Pasteurellales

В геномах *H. influenzae*, *H. ducreyi*, *P. multocida* белок LexA регулирует гены *lexA*, *recN*, *ruvA*, *recA* и *uvrD* (табл. 4).

Сайт связывания белка LexA, расположенный перед генами *uvrA* и *ssb*, которые входят в группу

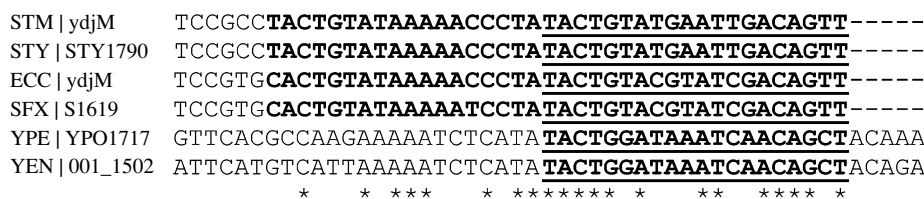


Рис. 1. Выравнивание 5'-некодирующих областей генов *ydjM*. Жирным выделен один предполагаемый сайт связывания LexA, подчеркнут – другой. Звездочками обозначены позиции, в которых у всех последовательностей в выравнивании находится один и тот же нуклеотид.

ECC yccR	TGTGAGT TACTGTATGAATGTACAGTA CATCCAGTGACGAC
SFX S1025	TGTGAGT TACTGTATGGATGTGCAGTA CATCCAGTGACAAC
STM yccR	TGTGAGT TACTGTATATTTCATACAGTAC -CCCTGTGGCGAT
STY STY1094	TGTGAGT TACTGTATATTTCATACAGTAC -TCCTGTGGCGAT
YEN 001_1314	TGTGGT TCCTGTATATATATACAGTA GTCACTGTT-TGTT
YPE YPO1437	TGTAGT TGCTGTATGTATATACAGTA GTCACTGTT-TAAT
	*** * ***** * ***** * **

Рис. 2. Выравнивание 5'-некодирующих областей генов *yccR*. Жирным выделен предполагаемый сайт связывания LexA. Звездочками обозначены позиции, в которых у всех последовательностей в выравнивании находится один и тот же нуклеотид.

экспериментально подтвержденных генов SOS-системы, оказался не очень устойчивым: перед *uvrA* он обнаружен в геномах *H. influenzae* и *P. multocida*, а перед *ssb* – в геномах *H. influenzae* и *H. ducreyi*.

Потенциальной таксон-специфичной регуляцией обладает ген *mfd* (потенциальный фактор транскрипционной репарации), сайт связывания LexA перед которым найден в двух геномах – *H. influenzae* и *P. multocida* – из трех рассмотренных.

Ближайший гомолог гена *mfd* из порядка Pasteurellales, функция которого изучена экспериментально – *mfd E. coli*. В случае повреждения ДНК матричной цепи во время транскрипции белок Mfd *E. coli* способствует ее репарации, взаимодействуя с белком UvrA [43]. В геноме *E. coli* перед геном *mfd* отсутствует потенциальный сайт регуляции. Несмотря на это, существование функциональной связи между Mfd и системой SOS-ответа делает вероятной регуляцию гена *mfd* белком LexA, по крайней мере, в некоторых геномах.

Vibrionales

В геномах *V. cholerae*, *V. parahaemolyticus*, *V. vulnificus*, *P. profundum* из генов ядра LexA-регулона *E. coli* наиболее консервативная регуляция у генов *lexA*, *recN*, *uvrA*, *uvrD*, *ruvA* и *dinP* (табл. 5).

В геноме всех рассмотренных организмов выявлен сигнал перед геном *VC0081*, который кодирует потенциальный слабоспецифичный экспортер токсичных веществ и принадлежит к COG0697

STY STY1804	A TACTGGATGAATAACCAGTT AA
STM STM1309	A TACTGGATGAATAACCAGTT AA
ECC ydjQ	A CACTGGATAGATAACCAGCA TT
SFX S1602	A CACTGGATAGATAACCAGCA TT
	* ***** *****

Рис. 3. Выравнивание 5'-некодирующих областей генов *ydjQ*. Жирным выделен предполагаемый сайт связывания LexA. Звездочками обозначены позиции, в которых у всех последовательностей в выравнивании находится один и тот же нуклеотид.

(рис. 5). У *E. coli* не найдено ортологов гена *VC0081 V. cholerae*, поэтому *VC0081* с читается потенциальным таксон-специфичным участником SOS-ответа.

Кроме того, во всех рассмотренных геномах потенциальный сигнал найден перед геном *VC0082*, продуктом которого является рекомбиназа (рис. 6). Этот ген ортологичен гену *yigN E. coli*, который регулируется LexA. Однако расстояния от потенциального регуляторного участка до самого гена у *VC0081* и *VC0082 (yigN)* примерно одинаковы, что, возможно, указывает на принадлежность этих двух генов к дивергенту. Таким образом, регулируется ли ген *VC0081* белком LexA в действительности, неясно.

Менее консервативен сигнал перед генами *recG*, *rpoD*, *mutH*, *uvrB*, *intI*. Потенциальные сайты регуляции перед этими генами обнаружены лишь в двух геномах – *V. parahaemolyticus* и *V. vulnificus*. Ортологи этих генов у *V. cholerae* и *P. profundum* сайтов связывания LexA не имеют.

Pseudomonadales, Alteromonadales

При проведении сравнительно-геномного анализа геномы представителей порядков Pseudomonadales и Alteromonadales объединили в одну группу ввиду их филогенетической близости. В геномах *P. aeruginosa*, *P. putida*, *P. syringae* и *S. oneidensis* обнаружен регуляторный сигнал перед генами *lexA*, *recA* и *recN*, участие в SOS-регулоне которых установлено экспериментально (табл. 6).

Потенциальный регуляторный сигнал перед геном *topB*, кодирующим ДНК-полимеразу III, най-

ECC yfiK	TGTCGGG TACTGTCTACCAAACAGAG GAGATA
SFX yfiK	TGTCGGG TACTGTCTACCAAACAGAG GAGATA
STY STY2838	TATCGGG TACTGTCTGCTAAAACAGAG GAGATG
STM yfiK	TATCGGG TACTGTCTGCTAAAACAGAG GAGATG
	* ***** * *****

Рис. 4. Выравнивание 5'-некодирующих областей генов *yfiK*. Жирным выделен предполагаемый сайт связывания LexA. Звездочками обозначены позиции, в которых у всех последовательностей в выравнивании находится один и тот же нуклеотид.

Таблица 4. Ядро регулона и потенциальные таксон-специфичные гены в геномах представителей порядка Pasteurellales

Ген*	Аннотация	HIN	HDU	PMU
<i>HI0749</i>	<i>lexA</i> , регулятор SOS-системы	+	+	+
<i>HI0600</i>	<i>recA</i> , корегулятор SOS-системы	+	+	+
<i>HI0070</i>	<i>recN</i> , репарация ДНК	+	+	+
<i>HI0313</i>	<i>ruvA</i> , хеликаза структуры Холлидея	+	+	+
<i>HI0250</i>	<i>ssb</i> , связывающий одноцепочечную ДНК белок	+	+	+
<i>HI1188</i>	<i>uvrD</i> , ДНК-хеликаза II	+	+	+
<i>HI0249</i>	<i>uvrA</i> , эндонуклеаза ABC, субъединица A	+	–	+
<i>HI1258</i>	<i>mfd</i> , транскрипционно-репарационный фактор	+	–	+

* Названия генов взяты из генома *H. influenzae*.**Таблица 5.** Ядро регулона и потенциальные таксон-специфичные гены в геномах представителей порядка Vibrionales

Ген*	Аннотация	VCH	VPA	VVU	PPR
<i>VC0092</i>	<i>lexA</i> , регулятор SOS-системы	+	+	+	+
<i>VC0543</i>	<i>recA</i> , корегулятор SOS-системы	–	+	+	+
<i>VC0852</i>	<i>recN</i> , репарация ДНК	+	+	+	+
<i>VC1846</i>	<i>ruvA</i> , хеликаза структуры Холлидея	+	+	+	+
<i>VC0190</i>	<i>uvrD</i> , ДНК-хеликаза II	+	+	+	+
<i>VC0394</i>	<i>uvrA</i> , эндонуклеаза ABC, субъединица A	+	+	+	+
<i>VC0394</i>	<i>uvrB</i> , эндонуклеаза ABC, субъединица B	–	+	+	–
<i>VC2043</i>	<i>topB</i> , ДНК-топоизомераза III	+	+	+	+
<i>VC2287</i>	<i>dinP</i> , ДНК-полимераза IV	+	+	+	+
<i>VCA0291</i>	<i>intI4</i> , сайт-специфичная рекомбиназа	+	+	+	0
<i>VC0081</i>	Потенциальная пермеаза	+	+	+	+
<i>VC0082</i>	<i>yigN</i> , рекомбиназа	+	+	+	+
<i>VC0668</i>	<i>mutH</i> , белок репарации мисметчей в ДНК	–	+	+	–
<i>VC2711</i>	<i>recG</i> , АТР-зависимая ДНК-хеликаза	+	+	–	–
<i>VC0517</i>	<i>rpoD</i> , σ -фактор РНК-полимеразы	–	+	+	–
<i>VC1878</i>	<i>msbA</i> , АТР-зависимый транспортный белок	–	+	0	+

* Названия генов взяты из генома *V. cholerae*.

ден лишь в трех рассмотренных геномах (*P. putida*, *P. syringae* и *S. oneidensis*). Консервативность предсказанного сигнала не очень высока, при этом одно плечо палиндрома сохраняется лучше другого (рис. 7).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В группе Enterobacteriales найдено шесть потенциально регулируемых LexA генов, сигналы перед которыми обнаружены только у организмов из данной группы: *uscR* (трансформирующий

```

VPA | VP0093   TAGACACTGGATAAAATGTCCAGTTTGTTGGATGAAAAATC---
VVU | VV10911 TACATACTGGATAAAATGTCCAGTTTGTTGCACAGAAAAATCCTC
VCH | VC0081   TCGGCACTGGATAAAATGTCCAGTTTGTTGCGCGTTTATCTC-
PPR | PPR0114 TGGATACTGGATAAAATGTCCAGCA-GTTATCCGTTAACGA---
                *          *****          ***          *

```

Рис. 5. Выравнивание 5'-некодирующих областей генов *VC0081*. Жирным выделен предполагаемый сайт связывания LexA. Звездочками обозначены позиции, в которых у всех последовательностей в выравнивании находится один и тот же нуклеотид.

ДНК белок), *ydjQ* (нуклеазная субъединица экзонуклеазного комплекса), *sbmC* (ингибитор гиразы), *ygiF* (G/U-мисметч-специфичная ДНК-гликозилаза), *otsB* (регулятор осмотического давления), *ydjM* (связанная с мембраной металл-зависимая гидролаза), *yfiK* (транспортный белок).

В группе Pasteurellales обнаружен один ген, потенциально регулируемый LexA и имеющий сигнал только в геномах этой группы – ген *mfd*, продукт которого участвует в репарации и взаимодействует с вовлеченным в SOS-ответ белком UvrA.

В группе Vibrionales найдены два гена с потенциальной таксон-специфичной регуляцией – *VC0081*, кодирующий слабоспецифичный переносчик токсичных веществ, и *VC0082*, кодирующий рекомбиназу.

В объединенной группе, в которую вошли представители порядков Pseudomonadales и Alteromonadales, LexA, по-видимому, таксон-специфически регулирует ген ДНК-полимеразы III – *topB*. Ранее участие ДНК-полимеразы III в репарации ДНК при SOS-ответе известно не было.

Таким образом, во всех рассмотренных порядках (Enterobacteriales, Pausterellales, Vibrionales, Pseudomonadales и Alteromonadales) предсказано существование потенциально регулируемых LexA генов, которые можно считать регулируемые таксон-специфически (они имеют сигнал только в геномах одной группы). Это свидетельствует о существовании различий в эволюционном развитии SOS-систем у разных филогенетических групп бактерий, обусловленных средой обитания и другими специфичными для каждой группы факторами.

Одним из фильтров при отборе генов, потенциально входящих в SOS-регулон, было соответствие функции гена-кандидата реализуемым при SOS-ответе реакциям клетки. Функции отфильтрованных таким образом генов связаны с синтезом ДНК, так как это необходимый для репарации повреждений ДНК процесс, транспортом ксенобиотиков и поддержанием гомеостаза.

В некоторых случаях сайт связывания LexA перед геном недостаточно устойчив, однако функция этого гена представлялась связанной с SOS-ответом. Так, идентифицирован ряд генов, потенци-

```
VVU | VV10910 GCAACAA--ACTGGACATTTATCCAGTATGTAAA
VPA | VP0094 CCAACAA--ACTGGACATTTATCCAGTGTCTAAA
VCH | VC0082 GCAACAA--ACTGGACATTTATCCAGTGCCGAAA
PPR | PPR0115 CGGATAAC TGCTGGACATTTATCCAGTATCCACT
          * * * * * * * * * * * * * * * * * * * * *
```

Рис. 6. Выравнивание 5'-некодирующих областей генов *VC0082*. Жирным выделен предполагаемый сайт связывания LexA. Звездочками обозначены позиции, в которых у всех последовательностей в выравнивании находится один и тот же нуклеотид.

альный регуляторный сайт перед которыми присутствует лишь в одном из рассмотренных геномов. Достоверность их регуляции не может быть установлена с помощью только методов биоинформатики, однако их возможная роль в SOS-ответе заслуживает экспериментальной проверки, и потому мы приводим список этих генов.

В геноме *H. ducreyi* регуляторный сигнал встретился перед геном *radA*, кодирующим белок RadA, который отвечает за репарацию ДНК, и перед геном интегразы/рекомбиназы *HD0897*.

В одном случае выявлен сигнал перед геном *VP2379 V. parahaemolyticus*, который принадлежит кластеру ортологичных генов (*COG*) *3141*, неохарактеризованный рецептор В-клеток). К этому же кластеру ортологичных генов относится ген *yebG E. coli*, возможность регуляции которого белком LexA мы предположили ранее (неопубликованные данные).

У *V. parahaemolyticus* предполагаемый сайт связывания LexA обнаружен перед геном, кодирующим слабоспецифичный экспортер токсичных для клетки веществ (*VP2449*). Согласно результатам поиска в базе данных кластеров ортологичных генов, этот ген принадлежит к семейству ортологичных генов *COG0534*, к которому также относится ген *dinF E. coli*, отвечающий за устойчивость к ксенобиотикам, и который, как доказано, регулируется белком LexA и участвует в SOS-ответе.

В геноме *V. vulnificus* сайты регуляции расположены перед генами, входящими в состав транспозона – генами транспозазы (*VV12451*, *VV12456*, *VV12476*, *VV12517*, *VV12529*, *VV12539*, *VV12548*) и

Таблица 6. Ядро регулона и потенциальные таксон-специфичные гены в геномах представителей порядка Pseudomonadales, Alteromonadales

Ген*	Аннотация	PAE	PPU	PST	SON
<i>lexA</i>	<i>lexA</i> , регулятор SOS-системы	+	+	+	+
<i>recA</i>	<i>recA</i> , корегулятор SOS-системы	+	+	+	+
<i>recN</i>	<i>recN</i> , репарация ДНК	+	+	+	+
<i>topB</i>	<i>topB</i> , ДНК-топоизомераза III	0	+	+	+

* Названия генов взяты из генома *P. aeruginosae*.


```

PPU | topB  ACATCCCCGCTGACAAAAAACCTGTACATCCATCCAG-ATAAACTTGCCTCTGCGC
PST | topB  ACAGCGCCAAAACGCCTTTATCTGTATATCCATACAG-ATAAACTTGCCTCTGCGC
SON | topB  CCTGCTTTGGGCTAAATTA TACTGTAATATATCCAGTATTTAGTGTAGATATGAT-
          * *                * * * * * * * * * * * * * *

```

Рис. 7. Выравнивание 5'-некодирующих областей генов *topB*. Жирным выделен предполагаемый сайт связывания LexA. Звездочками обозначены те позиции, в которых у всех последовательностей в выравнивании находится один и тот же нуклеотид.

интегразы (VVI2401). Для мобильных элементов регуляция SOS-системой имеет важное значение, так как целостность ДНК хозяйской клетки – главное условие их нормального функционирования. Таким образом, присутствие регуляторных мотивов перед генами, продукты которых отвечают за вырезание и вставку мобильного элемента, представляется вполне естественным.

У *P. putida* найден сигнал перед геном *PP4068*, кодирующим регулятор транскрипции из семейства Cro/C1. Этот регулятор гомологичен репрессору C1 фага λ из *E. coli* (рис. 4). Известно, что репрессор C1 *E. coli* регулируется SOS-ответом, поэтому вероятность того, что ген *PP4068 P. putida* регулируется LexA, достаточно высока.

Кроме того, у *P. putida* сайт связывания LexA обнаружен перед геном *grpE*, кодирующим белок теплового шока. Можно предположить, что продукт этого гена может участвовать в SOS-ответе, если повреждение ДНК вызвано воздействием экстремальной температуры.

Интересно, что у *P. putida* и *P. syringae* ген *lexA* дублирован. С помощью филогенетического анализа показано, что гены *lexA-2 P. putida* и *lexA-1 P. syringae* – это ортологи, как *lexA-2 P. syringae* и *lexA-1 P. putida*. У *P. aeruginosa* в геноме есть только одна копия *lexA*. Этот ген имеет более высокую степень сходства с парой *lexA-2 P. syringae* и *lexA-1 P. putida*, чем с паралогичной парой генов. Потенциальные сайты связывания LexA обнаружены у *lexA P. aeruginosa*, *lexA-2 P. syringae* и *lexA-1 P. putida*. Возможно, паралоги генов, кодирующих настоящие регуляторы, *lexA-2 P. putida* и *lexA-1 P. syringae*, утратили свою функцию и больше не кодируют репрессор SOS-ответа LexA.

Эта работа частично поддержана Медицинским институтом Ховарда Хьюза (55005610), INTAS (05-1000008-8028) и Российской академией наук (программа “Молекулярная и клеточная биология”).

СПИСОК ЛИТЕРАТУРЫ

- Walker C.G. 1996. *The SOS response of Escherichia coli. Escherichia coli and Salmonella, Cellular and Molecular Biology*. Washington, DC: ASM Press. 1.
- Defais M., Fanquet P., Radman M., Errera M. 1971. Ultraviolet reactivation and ultraviolet mutagenesis of lambda in different genetic systems. *Virology*. **43**, 495–503.
- Radman M. 1974. *Phenomenology of an inducible mutagenic DNA repair pathway in Escherichia coli: SOS repair hypothesis. Molecular and Environmental Aspects of Mutagenesis*. Charles C. Thomas, Springfield, III.
- Aertsen A., van Houdt R., Vanoirbeek K., Michiels C.W. 2004. An SOS Response Induced by High Pressure in *Escherichia coli*. *J. Bacteriol.* **186**, 6133–6141.
- Qiu X., Sundin G.W., Wu L., Zhou J., Tiedje J.M. 2005. Comparative analysis of differentially expressed genes in *Shewanella oneidensis* MR-1 following exposure to UVC, UVB and UVA radiation. *J. Bacteriol.* **187**, 3556–3564.
- Michel B. After 30 years of study, the bacterial SOS response still surprises us. 2005. *PLoS Biol.* **33**, 1174–1176.
- Little J.W., Mount D.W., Yanisch-Perron C.R. 1981. Purified *lexA* protein is a repressor of the *recA* and *lexA* genes. *Proc. Natl. Acad. Sci. USA*. **78**, 4199–4203.
- Nuyts S., van Mellaert L., Barbe S. et al. 2001. Insertion or Deletion of the Cheo Box Modifies Radiation Inducibility of *Clostridium* Promoters. *Appl. Environ. Microbiol.* **67**, 4464–4470.
- Brent R., Ptashne M. 1981. Mechanism of action of the *lexA* gene product. *Proc. Natl. Acad. Sci. USA*. **78**, 4204–4208.
- Galhardo R.S., Rocha R.P., Marques M.V., Menck C.F.M. 2005. An SOS-regulated operon involved in damage-inducible mutagenesis in *Caulobacter crescentus*. *Nucleic Acids Res.* **33**, 2603–2614.
- Winterling K.W., Levine A.S., Yasbin R.E., Woodgate R. 1997. Characterization of DinR, the *Bacillus subtilis* SOS Repressor. *J. Bacteriol.* **179**, 1698–1703.
- Winterling K.W., Chafin D., Hayes J.J., Sun J., Levine A.S., Yasbin R.E., Woodgate R. 1998. The *Bacillus subtilis* DinR Binding Site: Redefinition of the consensus sequence. *J. Bacteriol.* **180**, 2201–2211.
- Erill I., Jara M., Salvador N., Escribano M., Campoy S., Barbe J. 2004. Differences in LexA regulon structure among Proteobacteria through *in vivo* assisted comparative genomics. *Nucleic Acids Res.* **32**, 6617–6626.
- Mazon G., Erill I., Campoy S., Cortes P., Forano E., Barbe J. 2004. Reconstruction of the evolutionary history of the LexA-binding sequence. *Microbiology*. **150**, 3783–3795.
- Stormo G.D., Hartzell G.W. 1989. Identifying protein-binding sites from unaligned DNA fragments. *Proc. Natl. Acad. Sci. USA*. **86**, 1183–1187.
- Lawrence C.E., Reilly A.A. 1990. An EM algorithm for the identification and characterization of common sites in unaligned biopolymers sequence. *Proteins*. **7**, 41–51.
- van Helden J., André B., Collado-Vides J. 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* **281**, 827–842.

18. Lawrence C.E., Altschul S.F., Boguski M.S., Liu J.S., Neuwald A.F., Wootton J.C. 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*. **262**, 208–214.
19. de Henestrosa A. F., Ogi T., Aoyagi S., Chafin D., Hayes J.J., Ohmori H., Woodgate R. 2000. Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Molec. Microbiol.* **35**, 1560–1572.
20. Erill I., Escribano M., Campoy S., Barbe J. 2003. *In silico* analysis reveals substantial variability in the gene contents of the gamma proteobacteria LexA regulon. *Bioinformatics*. **19**, 2225–2236.
21. Resendis-Antonio O., Freyre-Gonzalez J.A., Menchaca-Mendez R., Gutierrez-Rios R.M., Martinez-Antonio A., Avila-Sanchez C., Collado-Vides J. 2005. Modular analysis of the transcriptional regulatory network of *E. coli*. *Genetics*. **21**, 16–20.
22. Гельфанд М.С., Миронов А.А. 1999. Компьютерный анализ регуляторных сигналов в полных бактериальных геномах. *Молекуляр. биология*. **33**, 772–778.
23. Миронов А.А., Винокурова Н.П., Гельфанд М.С. 2000. Программное обеспечение анализа бактериальных геномов. *Молекуляр. биология*. **34**, 253–262.
24. 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.
25. Felsenstein J. 1989. PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics*. **5**, 164–166.
26. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
27. Blattner F.R., Plunkett G., Bloch C.A. et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science*. **277**, 1453–1474.
28. <http://genome.wustl.edu/gsc/>
29. McClelland M., Sanderson K.E., Spieth J. et al. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*. **413**, 852–856.
30. Wei J., Goldberg M.B., Burland V. et al. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect. Immun.* **71**, 2775–2786.
31. <http://www.sanger.ac.uk/Projects/Microbes/>
32. Duchaud E., Rusniok C., Frangeul L. et al. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* **21**, 1307–1313.
33. <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=AE017143>
34. Cole S.T., Eiglmeier K., Parkhill J. et al. 2001. Massive gene decay in the leprosy bacillus. *Nature*. **6823**, 1007–1011.
35. 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.
36. Makino K., Oshima K., Kurokawa K. et al. 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet*. **361**, 743–749.
37. Kim Y.R., Lee S.E., Kim C.M. et al. 2003. Characterization and pathogenic significance of *Vibrio vulnificus*. Antigens preferentially expressed in septicemic patients. *Infect. Immun.* **71**, 5461–5471.
38. <http://www.tigr.org/cgi-bin/>
39. Heidelberg J., Paulsen I., Nelson K. et al. 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nat. Biotechnol.* **20**, 1118–1123.
40. <http://tcdb.ucsd.edu/>
41. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
42. Tatusov R.L., Koonin E.V., Lipman D.J. 1997. A genomic perspective on protein families. *Science*. **278**, 631–637.
43. Selby C.P., Sancar A. 1995. Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *J. Biol. Chem.* **270**, 4882–4889.