

УДК: 577.218:579.253.43

Интеграция чужеродного генетического материала провоцирует локальный мутагенез в геноме бактерии-реципиента

**Глазунова О.А.¹, Шавкунов К.С.¹, Тутукина М.Н.¹,
Панюков В.В.², Озолин О.Н.*¹**

¹*Институт биофизики клетки РАН, Пущино, Московская область, Россия*

²*Институт математических проблем биологии РАН – филиал Института прикладной математики им. М.В. Келдыша РАН, Пущино, Московская область, Россия*

Аннотация. Обмен генетической информацией является одним из основных факторов бактериальной эволюции, но механизмы интеграции горизонтально перенесённых генов в регуляторные сети нового хозяина остаются практически не исследованными. Данная работа посвящена изучению этого адаптационного процесса, в том числе, участия в нём транскрипционного аппарата бактериальных клеток. Для этого два чужих гена в геноме *Escherichia coli* K12 MG1655 были заменены на новые копии из геномов предполагаемых доноров и начат хронический рост мутантных и контрольных культур. Через 2000 (для гена *sfmA*) и 4000 (для гена *udhZ*) генераций было проведено популяционное секвенирование ампликонов модифицированных и исходных участков генома, с последующим анализом обнаруженных в них спонтанных мутаций. Установлено, что в модифицированных участках чаще, чем в немодифицированных, возникают замены G/C-пар на A/T-пары, а замены A/T → G/C, наоборот, наблюдаются реже. Это значит, что в ответ на введение чужеродного генетического материала в геноме возникает адаптивная реакция, направленная на обогащение модифицированного участка A/T-парами. В процессе длительной эволюции это может привести как к «замалчиванию» нежелательного гена с участием специфического супрессора чужеродных генов H-NS, так и к созданию приемлемого для адекватной экспрессии промотора в составе промоторного островка.

Ключевые слова: *бактериальная эволюция, горизонтальный перенос генов, однонуклеотидный полиморфизм, промоторные островки.*

ВВЕДЕНИЕ

Особенностью бактериальной эволюции является способность прокариот приобретать и ассимилировать гены других организмов [1, 2]. Такой обмен генетической информацией называют горизонтальным переносом, в отличие от «вертикальной» передачи генетической информации по наследству дочерним клеткам. Чаще всего горизонтальный перенос происходит в результате конъюгации, совмещённой с репликацией, и трансдукции, но даже простое поглощение клеткой фрагмента свободной молекулы ДНК (трансформация) может завершиться его встраиванием в геном [3–5]. Во время горизонтального переноса передаются случайные фрагменты ДНК, содержащие кодирующие и некодирующие нуклеотидные

*ozoline@rambler.ru

последовательности, куски генов или целые группы генов. Длина переносимого фрагмента может меняться от нескольких нуклеотидных пар до нескольких тысяч пар [6]. Из-за постоянно идущего процесса приобретения и утраты генов, сравнение любой пары геномов одного вида часто обнаруживает очень большие вариации [7].

Случайно приобретённые гены можно отличить от хозяйских по особенностям нуклеотидного состава, частоте использования кодонов и филогенетически. Такой анализ полногеномных данных показал, что значительная часть функционально важных генов получены современными бактериями именно благодаря ассимиляции чужих генов [8, 9], а их приобретение в некоторых случаях привело к происхождению новых видов [10].

Механизм ассимиляции чужих генов в геноме нового хозяина пока малопонятен. Однако известно, что области генома, содержащие чужие гены, обогащены промотор-подобными последовательностями [11], а многие из них содержат необычные участки, названные *промоторными островками* [12–16]. Такие последовательности, длиной от 300 до 1100 н.п., были обнаружены в процессе поиска потенциальных промоторов в геноме *E. coli* [12], но привлекли особое внимание необычностью своих функциональных свойств. Так, используя данные ChIP-on-chip, полученные в работах [17, 18], было установлено, что островки хорошо взаимодействуют с РНК-полимеразой, а экспериментальные данные подтвердили их способность формировать транскрипционно компетентные комплексы [12]. Однако синтез РНК с островковых промоторов, судя по данным работы [19], обычно ограничивается короткими (9–14 н.) продуктами [13, 16]. То есть, транскрипционный аппарат удерживается на островках в состоянии abortивного синтеза, наименее затратного для ресурсов клетки. Основной вклад в эту супрессию вносит, по-видимому, гистон-подобный белок H-NS, так как по данным работы [20] он покрывает до 90 % их общей длины [13], а делеция гена *hns* активировала синтез РНК со всех исследованных островков [14]. Являясь специфическим ингибитором горизонтально перенесённых генов [21, 22], H-NS, следовательно, может использовать А/Т-богатые последовательности островков для репрессии синтеза ненужных клетке продуктов.

По оценкам разных авторов, количество генов, приобретенных *E. coli* в результате горизонтального переноса, составляет от 10 до 30 % [11, 23–25], и нижняя граница этого диапазона соответствует количеству островков, обнаруженных в геноме кишечной палочки. Причём 75 из 78 островков, выявленных PlatProm [12], и 370 из 434 островков со смешанной σ -специфичностью, обнаруженных унифицированным алгоритмом поиска промоторов PlatPromU [16], оказались расположенными рядом с генами, полученными *E. coli* в результате горизонтального переноса. Так как в геномах потенциальных доноров эти гены имеют обычные промоторные области, было высказано предположение, что формирование промотор-подобных сайтов является адаптивной реакцией генома на внедрение чужеродного генетического материала. С одной стороны, образовавшийся островок может «выключить» экспрессию ненужных клетке генов, благодаря способности эффективно связывать репрессор H-NS, а с другой, он же может создать промотор для интеграции нового гена в регуляторные сети нового хозяина.

Для того, чтобы подтвердить такое виртуозное сочетание биологических функций необходимо доказать, что перенос «ненужного» клетке гена провоцирует эволюционный процесс, направленный на обогащение самого гена и его генетического окружения А/Т-парами, которые являются необходимыми компонентами для эффективного связывания как белка-супрессора H-NS, так и РНК-полимеразы. Данная работа посвящена проверке этой гипотезы. В качестве модельных объектов выбраны два гена, полученные *E. coli*, по-видимому, из генома сальмонелл. Они ассоциированы с промоторными островками и транскрипционно практически не активны. Поэтому их можно считать «ненужными» для кишечной палочки. Эти гены были удалены из

генома, а вместо них вставлены такие же гены сальмонелл, содержащие нормальные промоторы. В мае 2015 года мы начали хронический рост исходного и мутантных штаммов, ежедневно пересевая культуры. Через 2000 генераций для одного рекомбинантного штамма и 4000 для другого, нуклеотидные последовательности вставок были подвергнуты популяционному секвенированию, и в данной работе приводится первый анализ полученных данных.

МАТЕРИАЛЫ И МЕТОДЫ

Использованные в работе геномы

В работе были использованы геномы и генные карты *Escherichia coli* K12 MG1655 (*E. coli*) и *Salmonella enterica* Typhimurium 14028S, взятые из NCBI GenBank (NC_000913.3 и NC_016856.1, соответственно).

Получение рекомбинантных штаммов *E. coli*

Фрагменты для рекомбинации были получены с использованием ДНК *S. enterica* Typhimurium 14028S (любезно предоставлена К. Нойхаузом, Технический университет Мюнхена). Для наработки фрагмента, ассоциированного с геном *sfmA*, использовали праймеры F1 и R1, а для фрагмента, ассоциированного с геном *udhZ*, – F2 и R2 (таблица 1). Параметры амплификации: денатурация при 95 °C – 3 мин; 30 циклов: 95 °C – 30 сек, 55/54 °C – 30 сек (для *sfmA* и *udhZ*, соответственно), 72 °C – 1.5 мин; досинтез 72 °C – 2 мин. Полимеразную цепную реакцию (ПЦР) проводили на амплификаторе Biometra T-1 Thermocycler (Biometra, Германия) в соответствии с протоколом, рекомендуемым производителем реактивов (Евроген и Альфа-фермент, Россия). Рекомбинацию осуществляли по методу gene doctoring, предложенному в работе [26]. Ген устойчивости к канамицину, введенный в процессе рекомбинации, не удаляли, чтобы сохранить возможность отбора рекомбинантных штаммов в случае контаминации культур посторонней микрофлорой. Последовательности вставок были проверены прямым секвенированием и не содержали мутаций.

Таблица 1. Праймеры, использованные в работе

№	Название	Последовательность праймеров*	Применение
1	F1	5'-GCCCCGAATTCGCGCCATGATTAACCAAAAAGA-3'	Синтез гена <i>sfmA</i> из генома <i>S. enterica</i> .
2	R1	5'-GCCTAAGCTTATCATCTGCTATGTTCTCTA-3'	
3	F2	5'-TAGAGAATTCTGTTATATTCTGTTG-3'	Синтез гена <i>udhZ</i> из генома <i>S. enterica</i> .
4	R2	5'-CCCTAAGCTTCGTAACGGC-3'	
5	F3	5'-CTGTCAGGTTTCGTTTCAG-3'	Синтез нативного и рекомбинантного <i>sfmA</i>
6	R3	5'-CATCATAGTTAATACGCCTTGTG-3'	
7	F4	5'-TGTGTGCTTGTGTCTCAGC-3'	Синтез рекомбинантной области <i>udhZ</i>
8	R4	5'-CAGTCATAGCCGAATAGCCT-3'	
9	F5	5'-TCGTTGAGTTCTACAGTCCA-3'	Синтез нативной области <i>udhZ</i>
10	R5	5'-СТААТАСГСТГСТГАСТСТГ-3'.	
11	F6	5'-AAAAGAATTCAACTTGTTAGACAATAAAT TGTTATCAAAGATGACTTCTGTTATATTCTGTTG-3'	Наработка фрагмента для гомологичной рекомбинации фрагмента с геном <i>udhZ</i> (рис 1)
12	R6	5'-СТААСТСГАСГТГТАТГАГААААГАТТА АСГССГАТАААСААТАТССТТАГТТСС-3'	
13	<i>udhZ</i> _forw	5'-AAAAGAGGATGAACTGTACCG3'	Экспрессионный анализ
14	<i>udhZ</i> _rev	5'-ACGCCTGCGATAACAATAT-3'	
15	<i>sfmA</i> _forw	5'-TGGCGGTACAATCACTT-3'	
16	<i>sfmA</i> _rev	5'-CAAGAATCTCAATCCCCGAC-3'	

*Двойным подчеркиванием в праймерах указаны сайты рестрикции.

Среды и условия роста

Условия для долгосрочного роста выбирали по данным экспрессионного анализа базы данных GenExpDB [27]. Согласно им, оба гена плохо транскрибируются практически во всех условиях роста, но экспрессия *udhZ* ещё и снижается в условиях анаэробного роста на минимальной среде М9 [28]. Поэтому клетки с заменой *sfmA* и ассоциированного с ним островка на ортолог из генома сальмонеллы (*E. coli-sfmA_rec*) и культуру немодифицированных клеток кишечной палочки (*E. coli-sfmA_control*) растили в обычных условиях для культивирования кишечной палочки (15 мл 0.5-кратной среды Лурия-Бертани в колбах объёмом 50 мл), а клетки с заменой гена *udhZ*, ассоциированного с двумя островками (*E. coli-ydhZ_rec*), и соответствующую контрольную культуру (*E. coli-ydhZ_control*), растили в 25 мл среды М9 + 0.5 % глюкозы в колбах объёмом 25 мл (умеренно анаэробный рост). Клетки мутантных культур выращивали в присутствии 50 мкг/мл канамицина. Скорость роста бактерий определяли по оптической плотности на спектрофотометре Jasco UVIDEС-4 (Jasco Spectroscopic, Япония) при длине волны 600 нм.

Анализ уровня экспрессии

Суммарную фракцию клеточных РНК выделяли из 25 мл хронически растущих культур с использованием реагента TRIzol (Амбион, США) и затем обрабатывали ДНКазой I (New England Biolabs, США) в соответствии с протоколами производителей. Праймеры, использованные для экспрессионного анализа, перечислены в таблице 1. кДНК получали в 25 мкл реакционной смеси, содержащей 2 мкг РНК и 80 U обратной транскриптазы RevertAid (Thermo Scientific, Литва) в соответствии с протоколом производителя. Два микролитра этой смеси использовали для ПЦР, которую проводили в объёме 25 мкл с 5 мкл qPCRmix-HS SYBR (Evrogen, Россия) на термоциклере DPlite 4 (ДНК технологии, Россия). Программа ПЦР включала первоначальную денатурацию при 95 °С (3 мин), а затем 35 циклов: плавление при 95 °С – 30 сек, отжиг при 54 °С – 40 сек, синтез при 72 °С – 30 сек. Флюоресценцию SYBR Green I измеряли в конце каждого цикла в течение 15 сек. Образцы без обратной транскриптазы были использованы в качестве отрицательного контроля. В качестве референсного продукта была использована бессмысловая РНК гена *usaA* [14]. Транскрипционную активность рекомбинантных генов оценивали по сравнению с ассимилированными ортологами *E. coli* методом $2^{-\Delta\Delta C}$, где C – определяемый прибором номер порогового цикла на динамической кривой синтеза ампликонов.

Подготовка библиотек и секвенирование

Бактериальные культуры, прошедшие 2000 или 4000 генераций, растили до стационарной фазы роста при 37 °С и собирали центрифугированием. ДНК выделяли с использованием набора реактивов Wizard Genomic DNA Purification Kit (Promega, США), согласно протоколу производителя. Для амплификации анализируемых геномных областей использовали пары праймеров F3-R3 (*E. coli-sfmA_rec* и *sfmA_control*), F4-R4 (*E. coli-ydhZ_rec*) и F5-R5 (*E. coli-ydhZ_control*). В ПЦР смесь (Евроген, Россия) вносили примерно 100 нг геномной ДНК. Реакцию проводили в соответствии со следующим протоколом: денатурация при 95 °С – 3 мин; 7 циклов: 95 °С – 30 сек, 54 °С – 30 сек, 72 °С – 5 мин; 23 цикла: 95 °С – 30 сек, 54 °С – 30 сек, 72 °С – 3 мин; досинтез 72 °С – 2 мин.

Очищенные с помощью QIAquick Gel Extraction Kit (QIAGEN, Германия), ампликоны двух контрольных штаммов, размером 1170 и 975 н.п. для *E. coli-sfmA_control* и *ydhZ_control*, соответственно, смешивали по 200 нг каждого и фрагментировали, используя Ion Xpress Plus Fragment Library Kit (Life technologies, США) на сегменты длиной 200–300 н.п. Ампликоны двух рекомбинантных штаммов

*E. coli-sfmA*_rec и *udhZ*_rec имели разные размеры (2683 и 907 н.п., соответственно). Поэтому их фрагментировали до смешивания, используя разные режимы. Полученные фрагменты очищали, используя Agencourt AMPure XP Reagent (Life Sciences, США), присоединяли к ним адаптеры, используя Ion Fragment Library Kit, и снова очищали Agencourt AMPure XP Reagent. Качество фрагментации оценивали на приборе E-Gel SizeSelect Agarose Gel (Thermo Fisher Scientific, США) и отбирали фрагменты длиной около 330 н.п. После очистки из геля отобранные фрагменты амплифицировали на приборе DTLite 4 (ДНК-Технология, Россия) в режиме: предварительная денатурация – 95 °С – 5 минут; 7 циклов: 95 °С – 15 секунд, 58 °С – 15 секунд, 70 °С – 1 минута. Количественную нормализацию библиотек осуществляли с использованием Ion Library Equalizer Kit (Thermo Fisher Scientific, США), после чего концентрация ДНК в пробах, измеренная на спектрофотометре Qubit® 3.0 Fluorometer (Invitrogen, США), составляла ~100 пМ.

Эмульсионную ПЦР и последующее обогащение образца сферами, содержащими ДНК, проводили на системе Ion OneTouch 2 (Life Technologies, США). Секвенирование проводили на чипах Ion 314™ Chip v2 на секвенаторе Ion Torrent PGM (Life Technologies, США) согласно протоколу производителя.

Анализ полученных данных

Полученные прочтения (reads), варьирующие по длине от 25 до 280/350 нуклеотидов (н.), фильтровали по качеству на web-сервере Galaxy (инструмент Filter by Quality) [29] с пороговым уровнем ≥ 20 , что соответствует 99 % вероятности того, что все нуклеотиды в каждом прочтении идентифицированы верно. Исходный формат данных предварительно конвертировали на сервере Galaxy с помощью Fastq Groomer.

Для дальнейшего анализа отфильтрованные прочтения нарезали на фрагменты длиной 25 н. Их картировали на соответствующие участки генома *E. coli* K12 MG1655 (контроль) или рекомбинантных штаммов с помощью программы Matcher [30], которая находится в свободном доступе по адресу <http://www.mathcell.ru/DnaRnaTools/Matcher.zip>. Остальные фрагменты анализировали на наличие прочтений с заменами (одна замена в одном прочтении) с помощью программы MisMatcher (доступна по запросу).

РЕЗУЛЬТАТЫ

Выбор генетических локусов для рекомбинации и поиск потенциального донора

Чтобы не ошибиться с выбором модельного объекта, мы заменили уже имеющиеся в хромосоме *E. coli* чужие гены, рядом с которыми есть островки, на их ортологи, взятые из генома потенциального донора. Для выбора модельных генетических локусов были проанализированы все чужие гены, расположенные рядом с 370 *смешанными промоторными островками* [16]. Ортологи искали программой BLAST (NCBI). В 68 случаях оказалось, что прилегающие к островку гены уникальны и не имеют гомологов в других геномах. Для 120 островков был найден ортолог только для одного из двух соседних генов.

Перенос такого гена, даже вместе с его регуляторной областью, не позволяет полностью сохранить генетическое окружение. Тем не менее, из этой группы был выбран ген *udhZ*, кодирующий фермент ассимиляции железа, потому что рядом с ним находятся два островка (рис. 1). В 107 случаях островок вместе с его окружением был найден в геномах нескольких или многих видов бактерий. Среди них мог быть первоначальный донор, но для его идентификации нужен сложный филогенетический анализ, который пока не проводился. В 13 случаях ассоциированные с островками гены, помимо *E. coli*, были найдены в геномах одного рода, но имели другое расположение на хромосоме, что также ограничивало возможность их использования

для модельного эксперимента. В трех случаях островки находились между конвергентными генами. Их биологическая роль мало понятна, поэтому такие участки тоже не рассматривались. Остальные 59 случаев были проанализированы более детально.

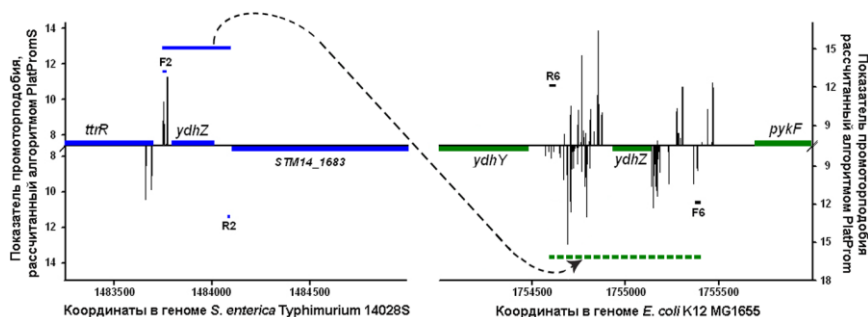


Рис. 1. Генетическое окружение гена *ydhZ* в геноме *S. enterica* Typhimurium 14028S (слева) и *E. coli* K12 MG1655 (справа). Цветными прямоугольниками показано расположение генов в двух геномах. Столбики соответствуют потенциальным точкам старта транскрипции, предсказанным алгоритмами, оптимизированными для поиска промоторов в соответствующих геномах. F2, R2, F6 и R6 – праймеры, использованные для рекомбинации (таблица 1).

Среди возможных доноров оказались *Salmonella enterica*, *Salmonella bongori*, *Citrobacter koseri*, *Citrobacter rodentium*, *Klebsiella oxytoca*, *Enterobacter cloacae*, *Dickeya dadantii*, *Cronobacter sakazakii*, *Yersinia enterocolitica*, *Edwardsiella tarda* и *Morganella morganii*. В каждом случае был получен профиль распределения предсказанных *in silico* точек старта в геноме потенциального донора. Для сканирования геномной ДНК сальмонелл был использован алгоритм PlatPromS, адаптированный к контексту их промоторов [31], а результаты сканирования целевой области сравнивали с профилем распределения стартовых точек транскрипции у *E. coli*, полученным алгоритмом PlatProm. В остальных случаях, из-за отсутствия обучающих промоторных компиляций, геномы потенциальных доноров и *E. coli* сканировали унифицированным алгоритмом PlatPromU, а пороговые значения для обеспечения одинакового уровня достоверности определяли, как описано в [32]. Оказалось, что в 12 геномах возможных доноров, целевые гены тоже имеют островки, что может указывать на возможность их горизонтального переноса из других геномов. Выбор модельного объекта среди остальных 47 кандидатов был осуществлён исходя из следующих соображений: доступность донора, близкая по размеру межгенная область, наличие информации о функциональной принадлежности соседних генов и характера зависимости их экспрессии от условий роста. В результате, в качестве второго модельного островка была взята область, расположенная между «родным» *folD* (кодирует фермент с дегидрогеназной и циклогидролазной активностями) и «чужим» – *sfmA* (белок аппарата адгезии) генами. Его генетическое окружение показано на рисунке 7 в статье [13].

В базе NCBI гомология с *ydhZ* с высокой степенью подобия (87 %) была найдена только в геномах сальмонелл, а гомология с *sfmA* (покрытие 83–86 %, идентичность 66–68 %) была найдена у близкородственных *E. coli* шигелл, в трех геномах *Enterobacter cloacae* и во всех геномах эволюционно более удалённой от *E. coli* *Salmonella enterica*. Сальмонеллы, следовательно, могли быть донором обоих генов для *E. coli*. Оказалось, что ортологи *sfmA* (*fimA*) и *ydhZ* у всех сальмонелл имеют обычные промоторы, а эти же гены разных штаммов *E. coli* – промоторные островки, которые, следовательно, могли быть продуктами адаптивной эволюции. Поэтому для рекомбинации были выбраны именно эти два гена. Их замена в геноме *E. coli* на ортологи из генома *S. enterica* практически не повлияла на рост бактериальных культур (рис. 2).

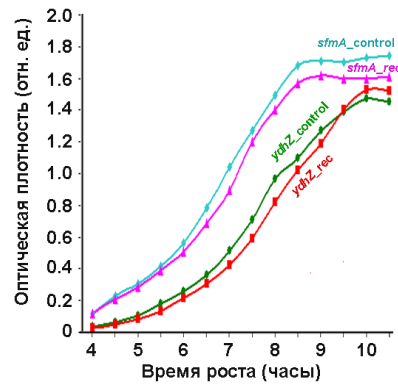


Рис. 2. Динамика роста исходных (*E. coli-sfmA_control* и *E. coli-ydhZ_control*) и мутантных (*E. coli-sfmA_rec* и *E. coli-ydhZ_rec*) культур сразу после рекомбинации. В процессе последующего роста взаимное расположение кривых существенным образом не изменялось. Отставание в росте контрольного и мутантного штаммов с заменой гена *ydhZ* от двух других культур объясняется менее комфортными условиями культивирования – недостаток кислорода и минеральная среда M9, вместо LB.

Экспрессия рекомбинантных генов изменяется в ходе хронического роста

Уровень экспрессии перенесённых генов оценивался в течение первых семи месяцев роста (рис. 3). Сразу после получения мутантов, РНК выделяли из клеток одного клона, выращенных до стационарной фазы. Впоследствии, для выделения РНК использовали растущие культуры, содержащие гетерогенную популяцию клеток, накопивших спонтанные мутации как в опытных, так и в контрольных образцах. Транскрипционную активность рекомбинантных штаммов сравнивали с уровнем экспрессии генов в параллельно растущих контрольных образцах.

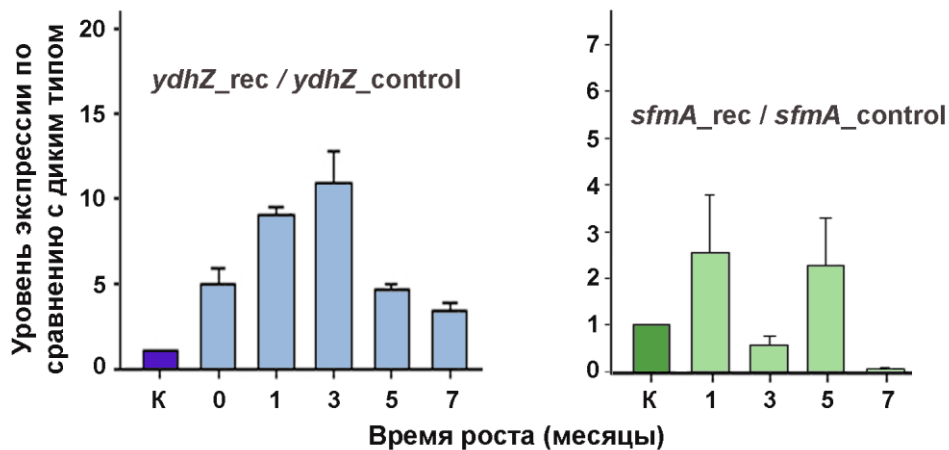


Рис. 3. Зависимость внутриклеточного содержания РНК-продуктов перенесённых генов от времени культивации. Количество продуктов, зарегистрированных в мутантных культурах, нормировались на количество продуктов, зарегистрированных в контрольных образцах (схематически показано столбиками К). Измерения, сделанные в нулевой точке для мутанта *E. coli-sfmA_rec* дали очень большие разбросы, поэтому соответствующие результаты не приводятся.

Транскрипционная активность перенесённого гена *ydhZ* оказалась выше его ортолога, удалённого из генома *E. coli* (столбики 0 и К на левой панели рис. 3). Это означает, что промотор гена *ydhZ* из генома сальмонеллы легко распознаётся транскрипционным аппаратом *E. coli*. Не исключено, что синтезируемая РНК является

плохим субстратом для рестриктаз *E. coli* и поэтому её больше в рекомбинантных клетках. Однако в течение трёх месяцев внутриклеточное содержание чужой РНК устойчиво и достоверно росло. Это значит, что в популяции рекомбинантных клеток чаще возникают такие спонтанные мутации, которые на порядок увеличивают продукцию чужого гена по сравнению с уже ассимилированным ортологом. Последующий рост сопровождался снижением экспрессии, что может означать успешное подавление клетками синтеза ненужного гена. Результаты, полученные для гена *sfmA*, оказались менее однозначными (правая панель на рис. 3), но и для них прослеживается аналогичная тенденция. В любом случае, обе рекомбинации вызвали адаптивный ответ генома реципиента.

Характер мутации в рекомбинированных областях отличается от спонтанных замен в исходных последовательностях

Библиотеки ампликонов рекомбинантных штаммов *E. coli_ydhZ_rec* (праймеры F4 и R4 в таблице 1) и *E. coli_sfmA_rec* (праймеры F3 и R3) секвенировали совместно на приборе Ion Torrent PGM с использованием набора Ion 314™ Chip v2. В результате было получено 406119 прочтений, а при совместном секвенировании контрольных областей – 560699 прочтений. После фильтрации по качеству с использованием порогового уровня $Q \geq 20$, обеспечивающего 99 % достоверность для всех нуклеотидов в каждом прочтении, и нарезания на фрагменты по 25 нуклеотидов, мы получили 174963 и 376222 прочтений для двух мутантных и двух контрольных областей, соответственно. Большая часть из них идеально картировалась на исследуемые области, но около одного процента содержали от одной до трёх точечных замен (таблица 2). Фрагменты с одной точечной заменой были использованы для дальнейшего анализа.

Таблица 1. Число идеально картируемых прочтений и прочтений с 1–3 ошибками для каждой культуры

Штамм	Размер Ампликона (н.п.)	Число поколений	Число совпадающих прочтений	Число 25 нуклеотидных прочтений с заменами			
				одна	две	три	%
<i>ydhZ_control</i>	975	4000	73787	559	136	132	1.1
<i>ydhZ_mutant</i>	907		36136	286	72	79	1.2
<i>sfmA_control</i>	1170	2000	272198	1983	385	363	0.99
<i>sfmA_mutant</i>	2683		146309	883	292	300	1.0

На рисунке 4 показано процентное содержание замен разного типа, выявленных с помощью программы MisMatcher. Одинаковые замены в одной позиции, зарегистрированные в разных прочтениях, учитывались как одна мутация. Мутации картировали по обеим нитям исследуемых областей, и те из них, которые приводили к появлению одной и той же комплементарной пары, тоже считались как одна замена, но разные мутации в одной позиции учитывались независимо.

Как и ожидалось, во всех культурах чаще всего возникают транзиции, то есть замены типа пурин-пурин и пиримидин-пиримидин в каждой нити ДНК. Однако в обеих мутантных культурах число спонтанных мутаций, приводящих к превращению G/C-пары в A/T-пару, оказалось больше, чем в контрольных клетках, а мутаций A/T→G/C, наоборот, меньше. Эта разница с контрольными культурами является ещё одним свидетельством адаптивного ответа бактериальных клеток на перенос генетического материала и соответствует предположению о том, что такая интеграция повышает частоту спонтанного появления A/T-пар вблизи модифицированного участка генома.

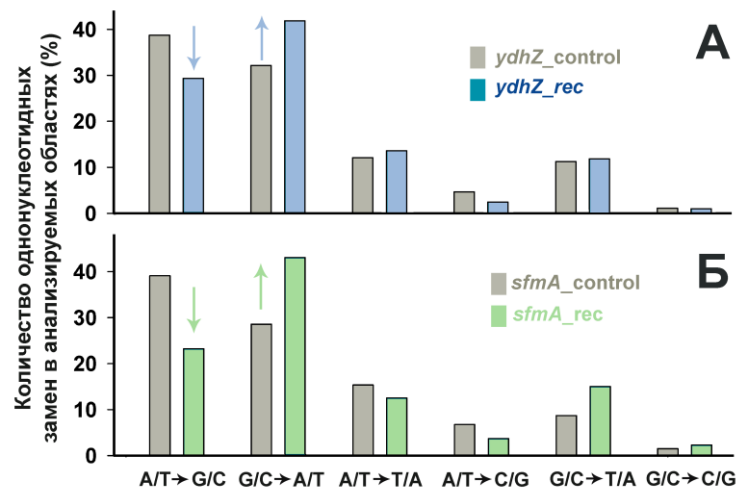


Рис. 4. Частота спонтанных мутаций в ампликонах, полученных из рекомбинантных участков геномов мутантных штаммов и из соответствующих областей параллельно с ними растущих клеток *E. coli* MG 1655. Показаны результаты популяционного секвенирования, т.е. прочтения ампликонов, полученных из отобранной смеси клеток с разным набором спонтанных мутаций. Данные приведены в процентах от всех спонтанных мутаций, зарегистрированных для каждой из анализируемых областей в прочтениях с одной заменой.

ОБСУЖДЕНИЕ

В работе приведены первые результаты долгосрочного эволюционного эксперимента, начатого полтора года назад и направленного на выяснение механизмов интеграции в геном чужеродных генов. Для этого в хромосоме *E. coli* были заменены два ранее ассимилированных гена вместе с находящимися рядом с ними промоторными островками на соответствующие ортологи *Salmonella enterica* Typhimurium 14028S, транскрибируемые с обычных промоторов (см. рис. 1, а также рис. 7 в статье [13]). Так как у кишечной палочки оба гена имеют низкий уровень экспрессии, были подобраны такие условия роста, которые не требуют их транскрипционной активности. Адаптивный ответ клеток на перенос чужеродного генетического материала был зарегистрирован по очень небольшому отличию в кривых роста (рис. 2), по изменению транскрипционной активности (рис. 3) и по особому характеру спонтанных мутаций в рекомбинантных областях (рис. 4). Все эти изменения проявляются только на популяционном уровне, так как ни одна из мутаций пока не распространилась на все клетки популяции.

В похожем, но более масштабном долгосрочном эксперименте, идущем без всякого эволюционного давления уже 29 лет в лаборатории Р. Ленского в Университете Мичигана, за 40000 генераций 12 параллельно растущих культур накопили в геноме *E. coli* всего 627 фиксированных замен [33, 34]. При этом в каждой популяции появлялись сотни тысяч точечных замен, но большинство из них оказались нейтральными и поэтому не закрепились в геномах. То есть, вероятность случайно получить хотя бы одну фиксированную мутацию в исследуемых нами рекомбинантных областях размером 1–2000 н.п. за 2–4000 генераций была очень мала ($\sim 10^{-5}$). Однако особый характер мутагенеза в них позволяет надеяться на ускоренную адаптивную эволюцию.

Замена G/C → A/T может появиться в результате дезаминирования метилированных цитидинов, и в бактериальных клетках имеются необходимые для этого ферменты. Кроме этого, для мутанта с замененным геном *sfmA* немного повышенной оказалась частота трансверсий G/C → T/A. Это может быть опосредовано окислением гуанинов до 8-оксогуанина, образующего неканоническую комплементарную пару с аденином. Во время следующего раунда репликации аденин становится матрицей для включения

тимина. Косвенно это подтверждается тем, что в анаэробно растущей культуре клеток с замкнутой геномной *udhZ* такого эффекта зарегистрировано не было.

В любом случае, полученные нами данные подтвердили предположение о том, что в ответ на введение чужеродного генетического материала, в геноме возникает адаптивная реакция, направленная на обогащение модифицированного участка А/Т-парами. В процессе длительной эволюции это может привести как к инактивации нежелательного гена с участием специфического супрессора чужеродных генов H-NS, так и к созданию приемлемого для адекватной экспрессии промотора в составе промоторного островка.

Экспрессионный анализ, а также разработка программ Matcher и Mismatcher были выполнены при финансовой поддержке Российского научного фонда (грант 14-14-00985, <http://rscf.ru/node/8>, руководитель ОНО, исполнители: ОАГ, МНТ, КСШ). Поиск модельных генов и получение мутантов выполнены при финансовой поддержке Российского фонда фундаментальных исследований (грант 16-04-01570, руководитель ОНО, исполнитель ОАГ). Авторы выражают благодарность В.В. Врублевской и О.С. Моренкову за получение анти-Dps антител.

СПИСОК ЛИТЕРАТУРЫ

1. Ochman H., Davalos L.M. The nature and dynamics of bacterial genomes. *Science*. 2006. V. 311. P. 1730–1733. doi: [10.1126/science.1119966](https://doi.org/10.1126/science.1119966).
2. Cohan F.M., Koeppel A.F. The origins of ecological diversity in prokaryotes. *Curr. Biol.* 2008. V. 18. P. R1024–R1034. doi: [10.1016/j.cub.2008.09.014](https://doi.org/10.1016/j.cub.2008.09.014).
3. Wybouw N., Pauchet Y., Heckel D.G., Van Leeuwen T. Horizontal gene transfer contributes to the evolution of *Arthropod Herbivory*. *Genome Biol. Evol.* 2016. V. 8. P. 1785–1801. doi: [10.1093/gbe/evw119](https://doi.org/10.1093/gbe/evw119).
4. Levin B.R. Frequency-dependent selection in bacterial populations. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 1988. V. 319. P. 459–472. doi: [10.1098/rstb.1988.0059](https://doi.org/10.1098/rstb.1988.0059).
5. Domingues S., Harms K., Fricke W.F., Johnsen P.J., da Silva G.J., Nielsen K.M. Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS Pathog.* 2012. V. 8. Article № e1002837. doi: [10.1371/journal.ppat.1002837](https://doi.org/10.1371/journal.ppat.1002837).
6. Ochman H., Lawrence J.G., Groisman E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000. V. 405. P. 299–304. doi: [10.1038/35012500](https://doi.org/10.1038/35012500).
7. Studier F.W., Daegelen P., Lenski R.E., Maslov S., Kim J.F. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *J. Mol. Biol.* 2009. V. 394. P. 653–680. doi: [10.1016/j.jmb.2009.09.021](https://doi.org/10.1016/j.jmb.2009.09.021).
8. Nakamura Y., Itoh T., Matsuda H., Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* 2004. V. 36. P. 760–766. doi: [10.1038/ng1381](https://doi.org/10.1038/ng1381).
9. McDaniel L.D., Young E., Delaney J., Ruhnau F., Ritchie K.B., Paul J.H. High frequency of horizontal gene transfer in the oceans. *Science*. 2010. V. 330. P. 50. doi: [10.1126/science.1192243](https://doi.org/10.1126/science.1192243).
10. Gogarten J.P., Doolittle W.F., Lawrence J.G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 2002. V. 19. P. 2226–2238. doi: [10.1093/oxfordjournals.molbev.a004046](https://doi.org/10.1093/oxfordjournals.molbev.a004046).
11. Huang Q., Cheng X., Cheung M.K., Kiselev S.S., Ozoline O.N., Kwan H.S. High density transcriptional initiation signals underline genomic islands in bacteria. *PLoS ONE*. 2012. V. 7. Article № e33759. doi: [10.1371/journal.pone.0033759](https://doi.org/10.1371/journal.pone.0033759).

12. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons in genome-scale promoter mapping. *Nucl. Acids Res.* 2009. V. 37. P. 4919–4931. doi: [10.1093/nar/gkp490](https://doi.org/10.1093/nar/gkp490).
13. Panyukov V.V., Ozoline O.N. Promoters of *Escherichia coli* versus promoter islands: function and structure comparison. *PLoS ONE*. 2013. V. 8. Article № e62601. doi: [10.1371/journal.pone.0062601](https://doi.org/10.1371/journal.pone.0062601).
14. Purtov Y.A., Glazunova O.A., Antipov S.S., Pokusaeva V.O., Fesenko E.E., Preobrazhenskaya E.V., Shavkunov K.S., Tutukina M.N., Lukyanov V.I., Ozoline O.N. Promoter Islands as a platform for interaction with nucleoid proteins and transcription factors. *J. Bioinform. Comput. Biol.* 2014. V. 12. P. 1441006. doi: [/10.1142/S0219720014410066](https://doi.org/10.1142/S0219720014410066).
15. Glazunova O.A., Kiselev S.S., Shavkunov K.S., Bykov A.A., Panyukov V.V., Ozoline O.N. Promoter islands in the genome of *E. coli*: comparative analysis against AT-rich sequences. *Math. Biol. Bioinform.* 2015. V. 10. P. t29–t38. doi: [10.17537/2015.10.t29](https://doi.org/10.17537/2015.10.t29).
16. Panyukov V.V., Kiselev S.S., Shavkunov K.S., Masulis I.S., Ozoline O.N. Mixed promoter islands as genomic regions with specific structural and functional properties. *Math. Biol. Bioinform.* 2013. V. 8. P. t12–t26. doi: [10.17537/2013.8.t12](https://doi.org/10.17537/2013.8.t12).
17. Reppas N.B., Wade J.T., Church G.M., Struhl K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell.* 2006. V. 24. P. 747–757. doi: [10.1016/j.molcel.2006.10.030](https://doi.org/10.1016/j.molcel.2006.10.030)
18. Herring C.D., Raffaele M., Allen T.E., Kanin E.I., Landick R., Ansari A.Z., Palsson B.O. Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J. Bacteriol.* 2005. V. 178. P. 6166–6174. doi: [10.1128/JB.187.17.6166-6174.2005](https://doi.org/10.1128/JB.187.17.6166-6174.2005).
19. Dornenburg J.E., DeVita A.M., Palumbo M.J., Wade J.T. Widespread antisense transcription in *Escherichia coli*. *mBio*. 2010. V. 1. Article № e00024-10. doi: [10.1128/JB.187.17.6166-6174.2005](https://doi.org/10.1128/JB.187.17.6166-6174.2005).
20. Kahramanoglou C., Seshasayee A.S., Prieto A.I., Ibberson D., Schmidt S., Zimmermann J., Benes V., Fraser G.M., Luscombe N.M. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.* 2011. V. 39. P. 2073–91. doi: [10.1093/nar/gkq934](https://doi.org/10.1093/nar/gkq934).
21. Lucchini S., Rowley G., Goldberg M.D., Hurd D., Harrison M., Hinton J.C. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* 2006. V. 2. Article № e81. doi: [10.1371/journal.ppat.0020081](https://doi.org/10.1371/journal.ppat.0020081).
22. Dorman C.J. H-NS, the genome sentinel. *Nat. Rev. Microbiol.* 2007. V. 5. P. 157–161. doi: [10.1038/nrmicro1598](https://doi.org/10.1038/nrmicro1598).
23. Langille M.G., Brinkman F.S. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics.* 2009. V. 25. P. 664–665. doi: [10.1093/bioinformatics/btp030](https://doi.org/10.1093/bioinformatics/btp030).
24. Lawrence J.G., Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA.* 1998. V. 95. P. 9413–9417. doi: [10.1073/pnas.95.16.9413](https://doi.org/10.1073/pnas.95.16.9413).
25. Price M.N., Dehal P.S., Arkin A.P. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.* 2008. V. 9. Article № R4. doi: [10.1186/gb-2008-9-1-r4](https://doi.org/10.1186/gb-2008-9-1-r4).
26. Lee D.J., Bingle L.E., Heurlier K., Pallen M.J., Penn C.W., Busby S.J., Hobman J.L. Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. *BMC Microbiol.* 2009. V. 9. P. 252. doi: [10.1186/1471-2180-9-252](https://doi.org/10.1186/1471-2180-9-252).
27. *E. coli* Gene Expression Database (GenExpDB). URL: <http://genexpdb.ou.edu/main/> (accessed November 26, 2016).
28. Sangurdekar D.P., Srienc F., Khodursky A.B. A classification based framework for quantitative description of large-scale microarray data. *Genome Biol.* 2006. V. 7. Article No. R32. doi: [10.1186/gb-2006-7-4-r32](https://doi.org/10.1186/gb-2006-7-4-r32).

29. Afgan E., Baker D., van den Beek M., Blankenberg D., Bouvier D., Cech M., Chilton J., Clements D., Coraor N., Eberhard C., Gruning B., Guerler A., Hillman-Jackson J., Von Kuster G., Rasche E., Soranzo N., Turaga N., Taylor J., Nekrutenko A., Goecks J. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucl. Acids Research*. 2016. V. 44. P. W3–W10. doi: [10.1093/nar/gkw343](https://doi.org/10.1093/nar/gkw343).
30. *Matcher*. URL: <http://www.mathcell.ru/DnaRnaTools/Matcher.zip> (accessed November 2016).
31. Ozoline O.N., Deev A.A. Predicting antisense RNAs in the genomes of *Escherichia coli* and *Salmonella typhimurium* using promoter-search algorithm PlatProm. *J. Bioinform. Comput. Biol.* 2006. V. 4. P. 443–454. doi: [10.1142/S0219720006001916](https://doi.org/10.1142/S0219720006001916).
32. Kiselev S.S., Ozoline O.N. Structure-specific modules as indicators of promoter DNA in bacterial genomes. *Math. Biol. Bioinform.* 2011. V. 6. P. t1–t13. doi: [10.17537/2011.6.t1](https://doi.org/10.17537/2011.6.t1).
33. Barrick J.E., Yu D.S., Yoon S.H., Jeong H., Oh T.K., Schneider D., Lenski R.E., Kim J.F. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*. 2009. V. 461. P. 1243–1247. doi: [10.1038/nature08480](https://doi.org/10.1038/nature08480).
34. Barrick J.E., Lenski R.E. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harb. Symp. Quant. Biol.* 2009. V. 74. P. 119–29. doi: [10.1101/sqb.2009.74.018](https://doi.org/10.1101/sqb.2009.74.018).

Рукопись поступила в редакцию 01.12.2016.

Дата опубликования 12.12.2016.