

УДК: 579:252

Промоторные островки в геноме *E. coli*: сравнительный анализ с АТ-богатыми последовательностями

**Глазунова О.А.¹, Киселев С.С.¹, Шавкунов К.С.^{1,2},
Быков А.А.^{1,3}, Панюков В.В.⁴, Озолин О.Н.^{1,2*}**

¹*Институт биофизики клетки, Российская академия наук, Пущино, Московская область, 142290, Россия*

²*Пушинский государственный естественно-научный институт, Пущино, Московская область, 142290, Россия*

³*Нижегородский государственный университет, биологический факультет, Нижний Новгород, 603950, Россия*

⁴*Институт математических проблем биологии, Российская академия наук, Пущино, Московская область, 142290, Россия*

Аннотация. Функциональные свойства *промоторных островков* генома кишечной палочки, т.е. участков с аномально высокой плотностью сигналов транскрипции, сопоставлены со свойствами геномных областей, аномально обогащённых АТ-парами. Установлено, что компьютерные выборки участков обоих типов по соответствующим критериям частично перекрываются, а их функциональные свойства во многом похожи. Однако *промоторные островки* отличаются от АТ-богатых последовательностей большей способностью к синтезу коротких олигонуклеотидов, которые могут быть целевыми продуктами этих необычных участков или побочными продуктами их супрессированного состояния. В *островках* чаще, чем в АТ-богатых участках и намного чаще, чем в местах расположения обычных промоторов встречаются инвертированные повторы. Так как структурные элементы такого типа обычно являются мишенями для взаимодействия с димерами и тетрамерами регуляторных белков, можно предположить, что инициация транскрипции с *островковых* промоторов контролируется регуляторными сетями клеток и, следовательно, их короткие РНК-продукты могут быть нужны для нормальной жизнедеятельности бактерий. Об этом, в частности, свидетельствует высокое содержание в клетках кишечной палочки олигонуклеотида, синтез которого с *островкового* промотора гена *yjgL* подтвержден экспериментально.

Ключевые слова: *промоторные островки, АТ-богатые участки генома, abortивный синтез, нетранслируемые РНК, горизонтальный перенос генов.*

ВВЕДЕНИЕ

Ранее в геноме кишечной палочки (*E. coli*) мы обнаружили необычные по функциональным свойствам участки [1–4]. Из-за аномально высокой плотности в них потенциальных мест инициации транскрипции мы назвали эти участки *промоторными островками* [1] и установили, что высокая способность *островков* формировать

* Ozoline@rambler.ru

транскрипционные комплексы совсем не соответствует их низкой транскрипционной активности [2, 3]. Гетерохроматин-подобное состояние *островков* оказалось оправданным их ассоциацией с генами, полученными *E. coli* в результате горизонтального переноса [2, 3]. Такой обмен генетической информацией широко распространён в природе, так как может предоставить эволюционные преимущества микроорганизму, случайно получившему полезный ген [5]. Но большинство чужих генов не нужны клеткам, и поэтому промотеры, в полной мере способные инициировать транскрипцию [4], в большинстве случаев подвергаются замалчиванию с помощью специальных ксеногенных ингибиторов [6, 7]. У *E. coli* таким ингибитором является гистон-подобный белок H-NS, который преимущественно связывается с участками ДНК, обогащёнными А/Т-парами [8–10], в том числе с *промоторными островками* [3]. Поэтому необычное соотношение длинных и коротких РНК-продуктов, синтезируемых с *островков*, вполне может быть простым следствием их гетерохроматин-подобного состояния. В таком случае аналогичное соотношение должно наблюдаться и на других АТ-богатых участках, т.е. других местах преимущественного связывания H-NS. Не исключено, однако, что короткие РНК-продукты *островков* играют какую-то особую биологическую роль. В таком случае их синтез должен контролироваться специальными регуляторными сетями и поэтому может отличаться у *островков*, обогащённых сигналами транскрипции, и в участках генома, отобранных по высокому содержанию А/Т-пар.

В этой работе мы выявили наиболее АТ-богатые последовательности, соответствующие по размеру минимальному *островку*, и сравнили с ними их функциональные свойства.

МЕТОДЫ И АЛГОРИТМЫ

Геном кишечной палочки и наборы геномных последовательностей

В работе был использован геном *E. coli* K12 MG1655, взятый из NCBI GenBank (NC_000913.2), и карта расположения генов, взятая из RegulonDB [11]. Сравнительный анализ был проведён для пяти типов геномных областей. Первый набор состоял из 78 *промоторных островков*, отобранных ранее [1, 2] по наличию ≥ 8 потенциальных точек инициации транскрипции в каждом скользящем окне размером 100 н.п. на протяжении ≥ 300 н.п. Размер последовательностей в этом наборе варьировал от 301 до 1102. Второй тип исследованных участков был представлен 95 АТ-богатыми последовательностями длиной 300 н.п. Он был сформирован с помощью программы ContentOf (см. ниже). Число А/Т-пар в этих последовательностях варьировало в диапазоне от 216 до 238, так что процентная вариация составила $72\div 79\%$, что намного выше процентного содержания А/Т-пар в хромосоме *E. coli* (49.2%). Три других типа геномных областей, содержащих множественные функциональные промотеры, одиночные промотеры и контрольные последовательности, были использованы нами ранее для сравнительного анализа структурно-функциональных свойств *промоторных островков* и обычных промоторов [2]. Каждый набор состоял из 78 последовательностей. Набор одиночных промоторов, содержащий минимальное число сигналов транскрипции, достаточное для адекватной экспрессии генов, включал последовательности стандартного размера, начиная с позиции -250 до позиции $+50$ относительно точки инициации транскрипции. В набор с множественными промоторами были включены последовательности, включающие от 3 до 9 экспериментально найденных точек инициации транскрипции. Их размер варьировал в диапазоне от 318 до 847 н.п. и охватывал всю область, заключённую между указанными выше границами всех индивидуальных промоторов. Координаты точек инициации транскрипции этих и одиночных промоторов были взяты из

RegulonDB [11]. Контрольные последовательности имели стандартную длину в 300 н.п. Они были выбраны из более обширной компиляции, использованной в работе [1], по минимальному значению показателя соответствия идеальному промотору, рассчитанному алгоритмом PlatProm.

Анализ генома *E. coli* программой ContentOf

Программа ContentOf входит в программный пакет aSHAPE [12]. По входным параметрам, длине участка FragSz и списку нуклеотидов Lst эта программа строит гистограмму распределения участков заданной длины по количеству в них целевых нуклеотидов (список Lst). Построение гистограммы выполняется просмотром каждого участка с шагом в одну нуклеотидную пару. Анализируя гистограмму, пользователь имеет возможность извлечь часть гистограммы, и сохранить её в виде файла для дальнейшей работы. Для отбора участков с повышенным содержанием А/Т-пар были использованы следующие параметры: FragSz = 300, Lst = {АТ}.

Оценка плотности распределения потенциальных промоторов

Для поиска сигналов инициации транскрипции, удовлетворяющих критерию структурно-модульного сходства с промоторами, использовали алгоритм PlatProm [1]. Он сканирует обе нити ДНК и вычисляет степень соответствия (*вес*) каждой позиции потенциальной точке инициации транскрипции. Рассчитанный *вес* считали значимым, если он превышал фоновый уровень на величину равную четырём стандартным отклонениям (StD), значения которых вычислялись по ранее сформированному набору контрольных последовательностей [1].

Поиск прямых и инвертированных повторов

Для обнаружения повторяющихся мотивов нуклеотидной последовательности в сравниваемых участках генома использовали программу Unipro UGENE (версия 1.14.2) [13]. Биологически значимыми в рамках решаемых задач считали прямые и инвертированные повторы длиной ≥ 6 н.п., разделённые спейсером от 5 до 10 н.п.

Анализ данных по распределению мест контакта РНКП и Н-NS в геномной ДНК

Для оценки способности РНК-полимеразы (РНКП) и белка Н-NS взаимодействовать с разными участками генома было использовано два набора данных [14, 15], полученных методом специфической иммунопреципитации ДНК-белковых комплексов (ChIP-on-chip). Этот метод позволяет обнаруживать места связывания любого целевого белка с бактериальной хромосомой непосредственно в живой клетке. Для этого используются специфические антитела, с помощью которых комплексы целевого белка с ДНК извлекаются из выделенного из клеток и фрагментированного хроматина (ChIP). Экстрагированная из комплексов ДНК затем гибридизуется с предназначенными для этой цели микроматрицами (chip). В случае РНКП эффективность связывания с геномом определяли как \log_2 отношения гибридизационных сигналов на микроматрице, полученных с ДНК после специфической иммунопреципитации комплекса, и сигналов, полученных с контрольной ДНК, не подвергнутой иммунопреципитации. Геномный участок считали взаимодействующим с РНКП, если среднее значение величин \log_2 всех проб внутри исследуемой области было положительным. Для оценки эффективности взаимодействия с Н-NS были использованы данные, уже обработанные авторами [15]. В этом случае геномные области считали взаимодействующими с Н-NS, если они перекрывались с сайтами связывания, по крайней мере, на 20 н.п.

Дифференциальный анализ экспрессии

РНК-продукты, синтезируемые на разных участках генома, анализировали так же, как описано ранее [2, 3]. В качестве исходных данных были использованы результаты до сих пор единственного эксперимента по полногеномному скринингу 5'-концевых последовательностей клеточных РНК с полностью доступными данными [16]. На вход программы RNAMatcher [2] подавался массив из прочтений (reads) длиной в 44 нуклеотида, находящихся на 5'-концах клеточных РНК. Программа искала позиции, соответствующие каждому из них в геноме и подсчитывала общее число одинаковых прочтений в массиве. При совпадении всех 44 нуклеотидов с геномной последовательностью образцу присваивалась соответствующая координата, а сам он извлекался из общего массива. После извлечения таких образцов искали прочтения с меньшим числом совпадений на 5'-конце. Образцы, совпадающие с геномом на 28–44 нуклеотида, считали представителями полноразмерных РНК. Образцы, совпадающие только по 9–20 5'-концевым нуклеотидам (при условии наличия адаптерной последовательности на 3'-конце) – считали коротким олигонуклеотидом.

Ассоциация исследованных участков генома с горизонтально перенесёнными генами

Взаимное расположение исследуемых областей и генов, предположительно полученных кишечной палочкой в результате горизонтального переноса, определяли с использованием данных, полученных пятью группами [17–21]. Сравниваемые области считали ассоциированными с чужими генами, если они располагались внутри длинных *геномных островков*, предсказанных GIST [20] или IslandViewer [21], а также перекрывались с регуляторными областями или кодирующими последовательностями чужих генов не менее чем на 100 н.п. [17–19].

Целевая регистрация коротких РНК, синтезируемых с *островковых* промоторов

Клетки *E. coli* K12 MG1655 растили на среде LB до ранней стационарной фазы. Внутриклеточное содержание коротких РНК определяли для клеток, выращенных при 37°C в 100 мл культуры в колбе объёмом 0.5 литра, снабжённой рассекателями (условия оптимального роста) или в условиях микроаэрации (две 50 мл пробирки были на 80% заполнены средой и плотно закрыты). Скорость перемешивания обеих культур была одинаковой. Через 8 часов после инокуляции среды ночной культурой клетки собирали центрифугированием. Фракцию коротких РНК получили с использованием набора miRvana™ (Ambion, США). Затем к 3'- и 5'-концам полученных РНК с помощью T4-РНК-лигазы были присоединены адаптеры с известной последовательностью (набор реактивов NEBNext^R Multiplex Small RNA Library Prep Set, New England Biolabs). Для обнаружения искомого продукта использовали две пары праймеров. В каждой паре один праймер соответствовал 5'-концу 5'-концевого или 3'-концевого адаптера, а второй – стыку между 3'-концом адаптера и искомой РНК. ПЦР проводили на программируемом термостате ДТ-322 (ДНК-технология, Россия). Ампликоны фракционировали в 2% агарозном геле, содержащем бромистый этидий и фотографировали в УФ-свете.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Промоторные островки только частично перекрываются с участками генома, наиболее обогащёнными А/Т-парами

Отбор наиболее АТ-богатых участков генома *E. coli* осуществляли среди фрагментов длиной 300 н.п. Число А/Т-пар в таких фрагментах варьировало в

диапазоне от 71 (23.7%) до 238 (79.3%) со средним значением 147.6. Для получения набора АТ-богатых фрагментов, по размеру сопоставимого с уже имеющимся наборами, каждый из которых состоял из 78 фрагментов, было использовано пороговое значение ≥ 216 А/Т-пар.

Таблица 1. Координаты левых границ АТ-богатых участков в геноме *E. coli* K12 MG1655 и количество в них А/Т-пар

83984_224	953799_219	1752597_222	2783004_216	3453432_221	4258373_216
156970_223	986363_216	1811105_217	2784026_220	3467822_216	4266437_216
237009_216	996808_227	1984520_216	2882221_226	3579989_218	4280712_217
330784_216	1102507_223	2031798_224	2901625_228	3581077_218	4324863_225
384000_228	1196754_220	2054647_216	2903460_219	3582489_218	4435575_221
389066_221*	1210337_220	2101896_218	2986249_216	3632404_216	4474835_229
400358_220	1211312_226	2104051_218	2988958_227	3648953_217	4539632_231
522097_224	1211712_217	2190227_230	2989846_217	3651672_216	4540652_217
527819_218	1214738_216	2342291_219	2993025_226	3767666_238	4554376_225
567808_219	1463065_230	2453650_216	2993691_216	3795016_225	4570120_216
576058_218	1528094_223	2467454_230	3117099_219	3797231_232	* синим цветом выделены А/Т-богатые участки, не менее, чем на 50% погружённые в кодирующие последовательности.
582518_233	1542767_218	2468158_232	3134360_217	3798815_218	
583280_216	1544015_217	2589013_220	3183150_229	3802238_224	
584855_217	1570098_219	2627556_218	3267089_223	3802871_230	
651121_217	1580563_218	2758162_223	3285166_225	3834657_216	
735388_216	1596164_218	2772310_221	3358908_227	4233532_217	
752055_222	1622551_218	2772814_222	3383267_216	4249396_220	

В результате было отобрано 95 последовательностей, которые оказались распределёнными по всему геному и, аналогично *промоторным островкам*, были обнаружены не только в межгенных участках, но и внутри генов (таблица 1). Пятьдесят восемь из них (61%) находились внутри или поблизости от *промоторных островков* (рис. 1), причём в двух случаях два АТ-богатых участка попали в один *островок*. Однако 37 (38.9%) обогащённых А/Т-парами последовательностей с *островками* не пересекались, а 22 *островка* (28.2%) не перекрывались с фрагментами АТ-богатого набора. Следовательно, похожесть *островковых* последовательностей на промоторные обусловлена не только их обогащённостью А/Т-парами.

Частота встречаемости инвертированных повторов в *островках* выше, чем в других наборах

Поиск промоторов с использованием алгоритма PlatProm базируется на наличии в них целого ряда структурно-функциональных модулей. Кроме высоко- и малоконсервативных мотивов, распознаваемых сменными σ -субъединицами РНКП или белками транскрипционного комплекса, к ним также относятся разного типа А/Т-треки, создающие анизотропные изгибы, гибкие звенья и термодинамически нестабильные площадки в ДНК. Именно поэтому одиночные и множественные промоторы имеют более высокое содержание А/Т-пар, чем последовательности, взятые из кодирующих участков генов (рис. 2,А). *Промоторные островки* по этому параметру выглядят как суперпромоторы и мало отличаются от АТ-богатых последовательностей (рис. 2,А).

Важными атрибутами обычных промоторов являются также прямые и инвертированные повторы [1, 22, 23]. Большая часть из них являются мишенями для взаимодействия с регуляторными белками. Поэтому, даже не располагая контекстами всех нуклеотидных последовательностей, распознаваемых факторами транскрипции, PlatProm использует их дискриминаторный потенциал. Контекст этих модулей не обязательно является АТ-богатым, поэтому их наличие или отсутствие даёт независимый ориентир для сравнения исследуемых наборов.

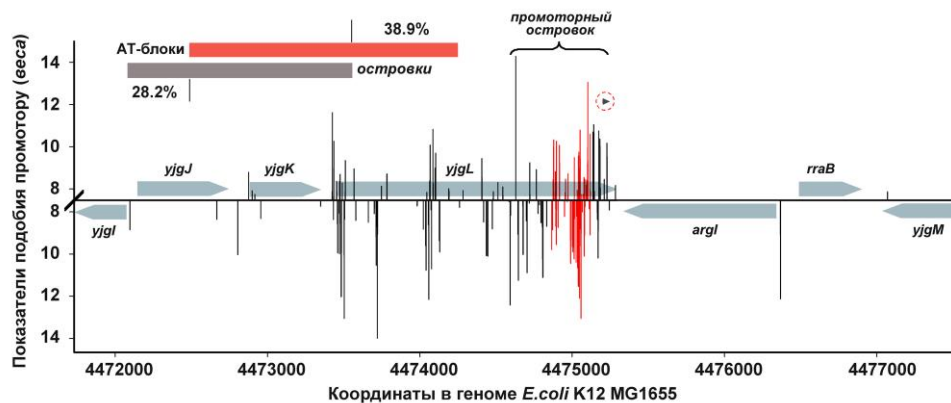


Рис. 1. Распределение потенциальных точек инициации транскрипции (столбики), предсказанных алгоритмом PlatProm. Высота столбиков выше и ниже оси X соответствует *весу*, рассчитанному для верхней и нижней нитей генома, соответственно. Расположение генов и направление их транскрипции показано серыми стрелками. Расположение *промоторного островка* показано фигурной скобкой. AT-богатый участок выделен красным цветом. Стрелкой в пунктирной окружности отмечено место синтеза и размер единственного РНК-продукта, зарегистрированного в области *островка* в работе [16]. На вставке слева схематически показана степень перекрытия наборов, содержащих *промоторные островки* и AT-богатые последовательности.

Поиск повторов проводился с использованием программы Unipro UGENE (версия 1.14.2) [13]. Его результаты показаны на рис. 2, В и С. Неожиданно оказалось, что по наличию прямых повторов наборы, содержащие одиночные и множественные промоторы, не отличались от контрольных последовательностей, которые были взяты из кодирующих участков генов. Возможно, это объясняется заложенным в них потенциалом кодирования регулярных структур белка. Даже простые треки из одинаковых аминокислот вполне могут быть закодированы триплетами, создающими прямые повторы в ДНК. Ещё более удивительным оказалось одинаковое содержание тандемных повторов в одиночных и множественных промоторах. Возможно, это объясняется их редким использованием для взаимодействия с регуляторами. Процент *островков*, содержащих прямые повторы, оказался больше, чем доля таких же AT-богатых последовательностей, но указанное выше равенство трёх других наборов не позволяет расценивать это, как значимый факт.

Характер распределения инвертированных повторов в анализируемых группах оказался совсем другим (рис. 2, С). Только 29% образцов контрольной группы имели такие модули. Процент содержащих их последовательностей в трёх промоторных группах возрастал с увеличением числа потенциальных промоторов. Это в точности соответствует ожидаемому, если инвертированные повторы считать мишенями для взаимодействия с зеркально-симметричными димерами регуляторных белков. Количество таких сайтов в группе множественных промоторов должно быть больше ещё и потому, что определённый вклад в их число в этом случае вносят консервативные элементы дивергентных промоторов. Тот факт, что *островки* сохранили эту тенденцию, косвенно указывает на то, что эволюционно они отбирались с учётом промоторной функции. В случае AT-богатых последовательностей эта тенденция нарушается: при большей плотности в них потенциальных промоторов (рис. 2, D и пример на рис. 1), процент последовательностей, имеющих инвертированные повторы, оказался ниже, чем у *островков*. Это, по-видимому, может свидетельствовать о разном характере эволюционного давления на последовательности этих двух типов и, следовательно, о возможном различии их функций.

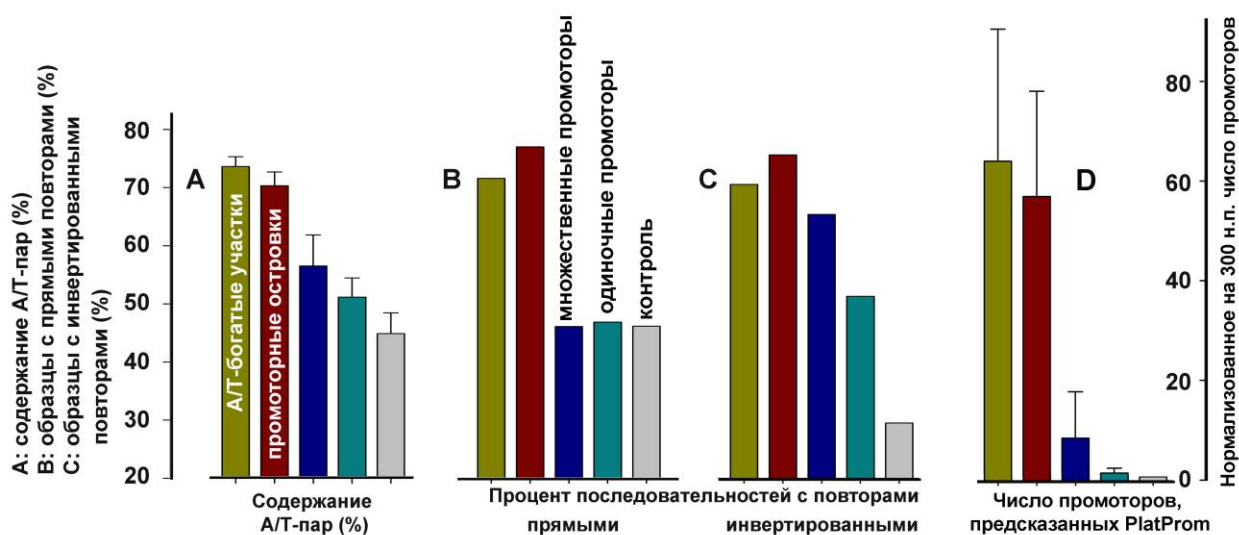


Рис. 2. Содержание А/Т-пар (А), прямых (В) и инвертированных (С) повторов, а также потенциальных промоторов (D) в сравниваемых наборах геномных последовательностей. Содержание А/Т-пар и промоторов, предсказанных PlatProm с достоверностью $p < 0.00004$ было посчитано для каждой последовательности, усреднено по каждому из пяти наборов и представлено в виде среднего по набору \pm StD. Для поиска повторяющихся мотивов была использована программа Unipro UGENE (версия 1.14.2) [13]. Значимыми считали повторы длиной ≥ 6 н.п.

Промоторные островки и АТ-богатые последовательности отличаются профилем синтезируемых РНК-продуктов

Ранее нами было обнаружено, что 75 из 78 *островков* (96.2%) находятся рядом с генами, приобретёнными кишечной палочкой в результате горизонтального переноса [2, 3]. Это значительно превышает процент одиночных и множественных промоторов с аналогичной локализацией (50.0% и 53.9%, соответственно). Доля таких участков среди АТ-богатых последовательностей практически не отличалась от *островковой* (92.6%), что согласуется с представлением о том, что чужеродные гены находятся в особом генетическом окружении, обогащённом А/Т-парами [20]. Возможно, именно эта особенность создаёт основу для их преимущественного связывания с гистон-подобным белком бактериального нуклеоида – Н-NS [6–9, 24, 25]. Поэтому не удивительно, что для выборки обогащённых А/Т-парами последовательностей мы обнаружили такую же высокую способность взаимодействия с этим белком (рис. 3,А). Поскольку эти участки перекрываются с *островками* (вставка рис. 1), часто находятся в регуляторных областях генов (таблица 1) и имеют высокую плотность промотор-подобных мест (рис. 2, D), нет ничего удивительного и в том, что они взаимодействуют с РНКП так же хорошо, как *островки* и лучше, чем последовательности других наборов (рис. 3, В). Вопрос заключался в том, отличаются ли они от *островков* по относительной способности инициировать синтез коротких и длинных РНК?

Для того чтобы ответить на этот вопрос, был выполнен дифференциальный анализ данных, опубликованных в работе [16]. Эти данные содержат информацию о 5'-концевых последовательностях длиной до 44 нуклеотидов, присутствующих в бактериальных клетках РНК. Сам анализ осуществлялся так же, как описано ранее [2, 3], только представителями полноразмерных продуктов считали все прочтения (reads), с минимальным соответствием геному в 28 нуклеотидов, а раньше это было сделано только для последовательностей длиной 44 н. Представителями коротких РНК-продуктов считали прочтения, совпадающие с геномом по 9–20 5'-концевым нуклеотидам. Используемые нами данные [16] свидетельствуют об инициации

синтеза РНК длиной ≥ 28 нуклеотидов только в 15 *островках* (19%) и 15 АТ-богатых последовательностях (16%). Это гораздо меньше доли одиночных и множественных промоторов, успешно начинающих синтез полноразмерных РНК (83% и 87%, соответственно). А когда число продуктов, синтезируемых в транскрибируемых участках, было нормировано на размеры анализируемых областей, то эти наборы стали отличаться ещё больше (рис. 3,С). Таким образом, используя бóльший объём данных (прочтения длиной 28–44 нуклеотида), мы подтвердили низкую способность *промоторных островков* вести синтез протяжённых РНК [2, 3] и обнаружили такое же свойство у АТ-богатых последовательностей.

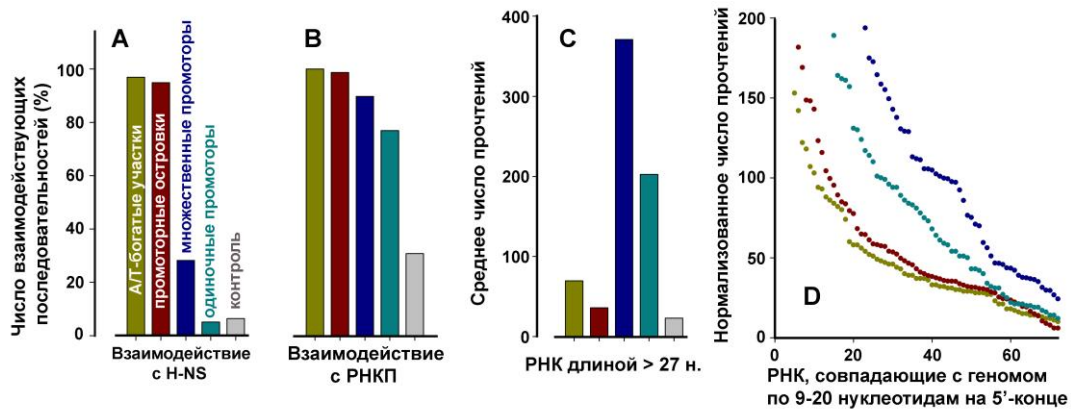


Рис. 3. Способность сравниваемых наборов геномных последовательностей взаимодействовать с H-NS (A) и РНКП (B), а также инициировать синтез протяжённых (C) и коротких РНК (D). Способность исследуемых наборов взаимодействовать с H-NS и РНКП оценивали по числу последовательностей, формирующих соответствующие комплексы (данные работ [14, 15]). Для оценки транскрипционной активности использовали результаты RNA-seq [16]. Прочтения разной длины в пределах указанных на панелях C и D диапазонов, суммировались для каждой геномной области. Для *островков* и множественных промоторов, которые имеют нестандартную длину, эта сумма нормировалась на 300 н.п. На панели C показано среднее число длинных прочтений в транскрибируемых последовательностях. Ранжированные графики панели D показывают нормализованное на длину фрагментов количество коротких прочтений. Их цветовые обозначения соответствуют остальным панелям.

Прочтения длиной 9–20 нуклеотидов были зарегистрированы практически для каждой из исследованных последовательностей. Больше всего их было обнаружено для регуляторных областей с множественными и одиночными промоторами. Число потенциально коротких РНК, стартующих внутри *островков* превышало количество аналогичных продуктов АТ-богатых последовательностей. То есть, имея похожую геномную локализацию (часто рядом с чужими генами) и очень похожие структурно-функциональные свойства, *промоторные островки* отличаются от АТ-богатых последовательностей несколько повышенной способностью к синтезу коротких РНК.

С промоторов *островка* гена *yjgL* идёт синтез коротких РНК

Дифференциальный анализ позволил нам оценить соотношение длинных и коротких РНК-продуктов в разных участках генома. Но полученные результаты ещё не позволяют утверждать, что в бактериальной клетке идёт синтез именно тех РНК, которые были найдены среди опубликованных прочтений. Во-первых, многие из них могут быть продуктами деградации. Во-вторых, 5'-концевой адаптер, использованный в работе [16], был присоединён к РНК с помощью T4 РНК-лигазы, но 3'-концевой адаптер имел 9 случайных нуклеотидов на 3'-конце. Он вводился в ДНК-копию в реакции обратной транскрипции, и случайные последовательности гарантировали получение копий всех РНК. Но такая стратегия не позволяла точно определить 3'-конец

продукта, так как 9-нуклеотидная случайная последовательность могла найти комплементарную мишень и внутри РНК. Используемый нами дифференциальный анализ снизил вероятность ошибки, так как для отбора потенциально коротких РНК вначале были отсеяны все прочтения с более длинным соответствием геному. Кроме того, для этого отбора брали только такие образцы, на 3'-конце которых была адаптерная последовательность. Тем не менее, неоднозначность остаётся и чтобы её устранить необходимо полногеномное прочтение коротких РНК после их предварительного лигирования адаптерами по обоим концам. Такое лигирование было выполнено нами для фракции коротких РНК из клеток, выращенных в разных условиях.



Рис. 4. Детекция РНК длиной 26 нуклеотидов, синтезируемой с *островкового* промотора гена *yjgL* (см. рис. 1) обнаруженной при анализе результатов, полученных методом RNA-seq [16]. Схематически показан протокол пробоподготовки коротких РНК фирмы NEB (А) и предложенная нами дальнейшая стратегия целевой амплификации (В). Последовательность искомой РНК: ААТССАСССГГААААТТТАСГТАТАГС. ПЦР-продукты были получены с использованием праймеров 5'-ААТГАТАСГГСГАССАСС-3' (1), 5'-СТСТТССГАТСТТГСТАТАСГТААА-3' (2), 5'-ТССГАСГАТСААТССАСС-3' (3), и 5'-СААГСАГАААГАСГГСАТАС-3' (4). Подчёркнуты участки, соответствующие искомой РНК. Продукты ПЦР анализировали в 2% агарозе. Проба без кДНК была использована в качестве негативного контроля. Гель был прокалиброван маркерами фирм Евроген (слева) и NEB (справа).

На рис. 4 схематически показана стратегия пробоподготовки. Для присоединения адаптеров мы использовали реактивы и протокол фирмы NEB (рис. 4,А). Для обнаружения РНК, предположительно синтезируемой с *островкового* промотора, была использована ПЦР с особыми парами праймеров, в каждой из которых один праймер соответствовал стыку между 3'-концом адаптера и искомым олигонуклеотидом (рис. 4,В). На рис. 4,С в качестве примера показан ПЦР-продукт, полученный для РНК *островка* гена *yjgL* (отмечена окружностью на рис. 1). При точном соответствии данным RNA-seq и использовании праймеров 1 и 2, его длина должна быть 95 н.п., а с праймерами 3 и 4 – 101 н.п., но для обоих продуктов возможны вариации по длине (рис. 4,В). Оказалось (рис. 4,С), что в условиях обычного роста в бактериальных клетках присутствует не только искомая РНК (26 нуклеотидов), но и более длинный продукт (~250 нуклеотидов). Так как он получен с парой праймеров 3 + 4, мы предполагаем, что его синтез начинается на том же промоторе, что и более короткой РНК. В любом случае, этот эксперимент свидетельствует о том, что синтез 26-ти нуклеотидной РНК идёт не с одного из сильных промоторов рядом расположенной АТ-богатой области, а с *островкового* промотора.

ЗАКЛЮЧЕНИЕ

Переход геномных исследований от изучения отдельных генов к анализу целых геномов позволил выявить ряд структурных и функциональных элементов, к которым, в частности, относятся *промоторные островки*. Само их появление рядом с горизонтально приобретёнными генами может потребовать значительной коррекции существующих представлений о механизмах ускоренной эволюции. Не менее важной является очень низкая способность *островков* начинать синтез полноразмерных РНК. По уже имеющимся данным [1–3] это может быть результатом гетерохроматизации *островков* гистон-подобным белком H-NS, чьей преимущественной мишенью являются последовательности, обогащённые А/Т-парами. В таком случае структурно-функциональные свойства *островков* должны быть похожи на свойства АТ-богатых последовательностей. Данное исследование было предпринято для проверки этого предположения, которое во многом подтвердилось. Однако оказалось, что *островки* чаще, чем АТ-богатые участки генома, и гораздо чаще, чем обычные промоторы, содержат инвертированные повторы. Такие структурные модули обычно являются мишенями для взаимодействия с димерами и тетрамерами регуляторных белков. Поэтому можно предположить, что инициация транскрипции с *островковых* промоторов в большей степени контролируется регуляторными сетями клеток, чем с АТ-богатых последовательностей.

Ранее было установлено, что *островки* способны синтезировать короткие РНК, но являются ли они целевыми продуктами этих участков, или побочными продуктами их супрессированного состояния – это вопрос, который пока остаётся открытым. Большая, по сравнению с АТ-богатыми участками, способность *островков* к синтезу таких продуктов (рис. 3,С), более частое присутствие в них потенциальных мишеней для взаимодействия с регуляторными белками (рис. 2,С) и подтверждённый синтез короткой РНК из гена *yjgL* (рис. 4,С) являются аргументами в пользу их биологического значения. *Промоторные островки* микробных геномов, например, могут играть роль своеобразных фабрик для производства особых коротких РНК. И хотя биологическая роль таких РНК ещё требует концептуального осознания, результаты сравнительного анализа свидетельствуют о разном характере эволюционного отбора в отношении *промоторных островков*, отобранных по высокой плотности и по равномерному распределению потенциальных сигналов транскрипции, и АТ-богатых последовательностей, отобранных по максимально высокому содержанию А/Т-пар в участке генома заданного размера.

Исследование выполнено за счёт грантов Российского научного фонда (проект № 14-14-00985) и Российского фонда фундаментальных исследований (проект № 12-07-00530).

СПИСОК ЛИТЕРАТУРЫ

1. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons from genome-scale promoter mapping. *Nucl. Acids Res.* 2009. V. 37. P. 4919–4931.
2. Panyukov V.V., Ozoline O.N. Promoters of *Escherichia coli* versus promoter islands: function and structure comparison. *PLoS ONE.* 2013. V. 8. Article No. e62601.
3. 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. *Mathematical Biology and Bioinformatics.* 2013. V. 8. P. t12–t26.
4. Shavkunov K.S., Tutukina M.N., Masulis I.S., Ozoline O.N. Promoter islands: the novel elements in bacterial genomes. *J. Biomol. Struct. Dyn.* 2011. V. 28. P. 1128–1129.
5. Syvanen M. Evolutionary implications of horizontal gene transfer. *Annu. Rev. Genet.* 2012. V. 46. P. 341–358.

6. 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 No. e81.
7. Navarre W.W., Porwollik S., Wang Y., McClelland M., Rosen H., Libby S.J., Fang F.C. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science.* 2006. V. 313. P. 236–238.
8. Dorman C.J. Horizontally acquired homologues of the nucleoid-associated protein H-NS: implications for gene regulation. *Mol. Microbiol.* 2010. V. 75. P. 264–267.
9. Dorman C.J. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat. Rev. Microbiol.* 2013. V. 11. P. 349–355.
10. Dorman C.J. H-NS-like nucleoid-associated proteins, mobile genetic elements and horizontal gene transfer in bacteria. *Plasmid.* 2014. V. 75. P. 1–11.
11. Salgado H., Peralta-Gil M., Gama-Castro S., Santos-Zavaleta A., Muñiz-Rascado L., García-Sotelo J.S., Weiss V., Solano-Lira H., Martínez-Flores I., Medina-Rivera A., Salgado-Ororio G., Alquicira-Hernández S., Alquicira-Hernández K., López-Fuentes A., Porrón-Sotelo L., Huerta A.M., Bonavides-Martínez C., Balderas-Martínez Y.I., Pannier L., Olvera M., Labastida A., Jiménez-Jacinto V., Vega-Alvarado L., Del Moral-Chávez V., Hernández-Alvarez A., Morett E., Collado-Vides J. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucl. Acids Res.* 2013. V. 41. P. D203–D213.
12. Панюков В.В., Назипова Н.Н., Озолин О.Н. Пакет программ aSHAPE для изучения пространственной конформации участков бактериального генома. *Математическая биология и биоинформатика.* 2011. Т. 6. С. 211–227.
13. Okonechnikov K., Golosova O., Fursov M., the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012. V. 28. P. 1166–1167.
14. 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.
15. Kahramanoglou C., Seshasayee A.S.N., 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*. *Nucl. Acids Res.* 2011. V. 39. P. 2073–2091.
16. Dornenburg J.E., DeVita A.M., Palumbo M.J., Wade J.T. Widespread antisense transcription in *Escherichia coli*. *mBio.* 2010. V. 1. Article No. e00024-10.
17. Lawrence J.G., Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA.* 1998. V. 95. P. 9413–9417.
18. 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.
19. 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 No. R4.
20. 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 No. e33759.
21. 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.
22. Ozoline O.N., Deev A.A. Predicting antisense RNAs in the genomes of *Escherichia coli* and *Salmonella typhimurium* using promoter-search algorithm PlatProm. *J. Bioinf. Comput. Biol.* 2006. V. 4. P. 443–454.
23. Kiselev S.S., Ozoline O.N. Structure-specific modules as indicators of promoter DNA in bacterial genomes. *Mathematical Biology and Bioinformatics.* 2011. V. 6. P. t1–t13.

24. Oshima T., Ishikawa S., Kurokawa K., Aiba H., Ogasawara N. *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res.* 2006. V. 13. P. 141–153.
25. Dorman C.J. H-NS, the genome sentinel. *Nat. Rev. Microbiol.* 2007. V. 5. P. 157–161.

Материал поступил в редакцию 09.12.2014, опубликован 22.12.2014.