

УДК 577.053

У АКТИНОБАКТЕРИЙ ЧИСЛО ДЛИННЫХ ШПИЛЕК В МЕЖГЕННЫХ ТРЕЙЛЕРНЫХ ОБЛАСТЯХ ВЕЛИКО ПО СРАВНЕНИЮ С ДРУГИМИ ОБЛАСТЯМИ ГЕНОМА

© 2007 г. Е. В. Любецкая, А. В. Селиверстов*, В. А. Любецкий

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

Поступила в редакцию 30.06.2006 г.

Принята к печати 31.08.2006 г.

Ключевые слова: вторичная структура ДНК, экспрессия генов.

ACTINOBACTERIA HAVE A LOT MORE LONG HAIRPINS IN TRAILER REGIONS THAN IN OTHER REGIONS, by E. V. Lyubetskaya, A. V. Seliverstov*, V. A. Lyubetsky (Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia; *e-mail: slvstv@iitp.ru).

Key words: DNA secondary structure, gene expression.

В геномах бактерий нередко чередуются гены, расположенные на комплементарных цепях ДНК и направленные навстречу друг другу (так называемые *конвергоны*). Если эти гены интенсивно транскрибируются, то естественно предположить, что в *области между ними* находится терминатор транскрипции, необходимый для прекращения транскрипции и снятия конформационного напряжения, возникающего в этом месте при образовании супервитков ДНК. Эту область далее будем называть *сходящейся трейлерной областью*. В геномах *Bacillus subtilis* и *Escherichia coli* роль такого терминатора играет так называемый *классический терминатор транскрипции*, т.е. относительно короткая (5–10 пар комплементарных нуклеотидов в спаренном участке – черенке) GC-богатая шпилька с небольшой петлей, за которой следует T-богатый участок. У *B. subtilis* и *E. coli* опероны обычно заканчиваются именно таким терминатором. Однако у многих актинобактерий, в том числе микобактерий, классические терминаторы относительно редки [1, 2].

Нами предложен новый алгоритм, который обеспечивает массовый и очень быстрый поиск шпилек большой мощности с заданными параметрами черенка и петли в любой нуклеотидной последовательности. Алгоритм применен для поиска шпилек в сходящихся трейлерных областях и в других областях генома актинобактерий. С помощью этого алгоритма в геномных областях перечисленных ниже типов проведен подсчет *длинных* (от 17 п.н. в черенке) шпилек, а также шпилек, которые могут образовать *крест* – опре-

деленную пару шпилек на комплементарных цепях ДНК [3]. Мы предполагаем, что такие *крестообразные структуры* формируют неканоническую вторичную структуру ДНК, которая служит для снятия конформационного напряжения и терминации транскрипции. Кроме того, в разного типа геномных областях актинобактерий алгоритм обнаружил существенно различные количества длинных шпилек. Особенно много таких шпилек найдено в сходящихся трейлерных областях. Сами шпильки не принадлежат к числу известных концевых терминаторов транскрипции или регуляторов инициации транскрипции, описанных ранее [4], отличаясь от них длиной и положением в геномной последовательности. В геноме актинобактерий найдено гораздо больше пар шпилек, образующих крестообразные структуры, чем таких же пар в геномных областях того же типа у *B. subtilis* и *E. coli*. У большинства актинобактерий шпильки не сопровождаются T-богатым участком и заметно длиннее классического терминатора транскрипции.

Известно, что у актинобактерий экспрессия генов биосинтеза аминокислот часто регулируется на уровне трансляции [5], в то время как у γ - и α -протеобактерий экспрессия этих генов предположительно регулируется на уровне транскрипции [6]. Поэтому в представленной работе рассмотрены области генома актинобактерий, потенциально способные содержать связанные с терминацией транскрипции вторичные структуры. В этой связи особый интерес представляют участки, расположенные после генов тРНК и генов с высоким уровнем экспрессии. Изучение этих генов полезно для предсказания границ оперонов и различие-

* Эл. почта: slvstv@iitp.ru

Частота шпилек с заданной длиной черенка (l) в различных областях генома актинобактерий, *B. subtilis* и *E. coli*

l	1	2а	2б	3	4
<i>Corynebacterium efficiens</i>					
25	0.23	0.38	1 (0)	0.00	0.00
23	0.23	0.76	2 (0)	0.00	0.00
20	0.68	6.06	16 (11)	0.00	0.00
17	2.60	13.64	37 (29)	0.17	0.31
15	4.98	21.59	57 (48)	0.43	1.85
10	20.61	45.83	121 (95)	31.79	8.31
<i>Mycobacterium tuberculosis</i>					
25	0.31	3.29	11 (11)	0.07	0.18
23	0.38	4.19	14 (14)	0.07	0.18
20	0.54	5.09	17 (17)	0.07	0.36
17	1.30	5.99	20 (20)	0.14	0.53
15	1.83	6.29	21 (21)	0.57	0.89
10	15.44	10.78	36 (35)	35.83	5.35
<i>Propionibacterium acnes</i>					
25	0.00	0.00	0	0.00	0.00
23	0.00	0.00	0	0.00	0.00
20	0.00	0.96	2 (2)	0.04	0.00
17	0.45	8.17	17 (16)	0.09	0.00
15	1.20	12.98	27 (23)	0.13	0.00
10	15.19	30.29	63 (57)	24.29	6.34
<i>Streptomyces coelicolor</i>					
25	0.08	0.95	8 (8)	0.00	0.00
23	0.15	2.25	19 (18)	0.01	0.00
20	0.54	3.08	26 (25)	0.02	0.17
17	1.73	9.60	81 (78)	0.09	0.41
15	3.61	14.45	122 (117)	0.35	1.00
10	31.07	37.32	315 (306)	23.09	14.36
<i>Bifidobacterium longum</i>					
25	0.45	1.51	3 (3)	0.05	0.00
23	0.45	2.02	4 (4)	0.05	0.00
20	1.05	5.56	11 (10)	0.05	0.00
17	1.79	13.64	27 (23)	0.43	0.45
15	4.19	21.21	42 (36)	0.65	1.80
10	28.25	47.47	94 (75)	28.42	11.26
<i>Leifsonia xyli</i>					
25	0.11	0.00	0	0.00	0.00
23	0.11	0.00	0	0.00	0.00
20	0.65	2.78	4 (4)	0.05	0.00
17	1.63	4.86	7 (7)	0.34	0.00
15	2.28	7.64	11 (11)	1.01	0.47
10	21.93	22.92	33 (33)	39.28	9.48

Окончание

<i>l</i>	1	2a	2б	3	4
<i>Escherichia coli</i>					
25	0.00	0.00	0	0	0.00
23	0.00	0.00	0	0	0.00
20	0.00	0.00	0	0	0.00
17	1.38	0.22	1 (1)	0.02	0.19
15	4.15	2.91	13 (8)	0.13	0.56
10	77.67	47.65	213 (159)	18.56	11.52
<i>Bacillus subtilis</i>					
25	0.00	0.00	0	0	0.00
23	0.00	0.00	0	0	0.00
20	0.13	0.51	2 (1)	0	0.00
17	0.89	3.55	14 (6)	0.02	0.20
15	2.74	13.96	55 (12)	0.11	0.20
10	25.86	67.77	267 (97)	18.71	13.64

Примечание. Приведено отношение (%) числа найденных нашим алгоритмом шпилек при фиксированной минимальной длине черенка *l* к числу всех областей данного типа в геноме. Столбцы 1–4 соответствуют следующим областям: 1 – все лидерные области; 2a и 2б – все сходящиеся трейлерные области; все кодирующие области; 4 – все расходящиеся области. В столбце 2б указано число шпилек, найденных во всех сходящихся трейлерных областях; в скобках приведено число шпилек без Т-богатого участка.

ния генов с высоким уровнем экспрессии и их паралоги с низким уровнем транскрипции [7].

Нуклеотидные последовательности геномов бактерий получены из базы данных GenBank, ftp.ncbi.nlm.nih.gov/genbank. Брали последовательности полностью секвенированного генома одного штамма от каждого представленного в базе данных вида актинобактерий – рассматривали только виды, у которых такой штамм имеется. Для сравнения использовали *E. coli* K12 и *B. subtilis* штамм 168. Длинные шпильки обнаружены после интенсивно транскрибируемых генов белков и тРНК. Ниже в таблице приведена часть наших данных.

Итак, в геноме рассматриваются участки четырех типов: все лидерные области, все сходящиеся трейлерные области (конвергоны), все кодирующие области, все расходящиеся лидерные области (дивергоны). На каждом участке отбирали не более одной шпильки с самым длинным черенком. При этом шпильки должны удовлетворять условиям: длина черенка не меньше порогового значения *l*, измеряемого числом пар нуклеотидов. Черенок может содержать односторонние выпячивания не более чем из 2 н., а длина петли черенка не должна превышать 15 н.

Для конвергонов белоккодирующих генов типичны наиболее длинные шпильки. Максимальные длины черенков шпилек в сходящихся трейлерных областях достигают 31 п.н. у *Mycobacterium tuberculosis*, *M. bovis* и *Streptomyces coelicolor*,

29 п.н. у *S. avermitilis* и 27 п.н. у *Corynebacterium diphtheriae*. В отличие от актинобактерий (таблица) у *B. subtilis* и особенно у *E. coli* шпильки с длиной черенка более 17 п.н. встречаются много реже. В соответствующем конвергоном столбце таблицы (в скобках) указана частота (доля) шпилек в этой области, после которых нет Т-богатого участка длиной не менее 7 п.н., содержащего не более двух исключений. Сам Т-богатый участок искали в интервале –5...+12 н. от последнего нуклеотида черенка шпильки.

Результаты проведенного анализа позволяют сделать следующие выводы. Если хотя бы один из двух направленных навстречу друг другу генов актинобактерий активно транскрибируется, то за ним часто следует потенциальная длинная крестообразная структура ДНК. Эта структура предположительно может выполнять две функции. Во-первых, снимать конформационные напряжения, возникающие при интенсивной транскрипции (с участием или без участия топоизомеразы). Во-вторых, изменение конформации ДНК может обеспечивать терминацию транскрипции.

Большинство найденных шпилек значительно длиннее обычных классических терминаторов транскрипции, и они имеют ряд особенностей. Например, за ними нет Т-богатых участков. Возникающая вторичная структура симметрична и может играть роль терминатора для РНК-полимеразы, движущейся по каждой из цепей ДНК. Косвенно в пользу такой гипотезы говорит боль-

шая частота указанных структур именно в сходящихся трейлерных областях по сравнению с областями другого типа.

Компьютерная реализация нашего алгоритма осуществлена Л.И. Рубановым. Авторы благодарят М.С. Гельфанда за обсуждение и ценные замечания, а также рецензента за содержательные замечания, способствовавшие улучшению текста. В создании программного обеспечения принимал участие М.А. Ширшин, которому авторы глубоко благодарны.

Работа частично поддержана программой ISTC (грант 2766).

СПИСОК ЛИТЕРАТУРЫ

1. Washio T., Sasayama J., Tomita M. 1998. Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res.* **26**, 5456–5463.
2. Unniraman S., Prakash R., Nagaraja V. 2002. Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res.* **30**, 675–684.
3. Кнорре Д.Г., Мызина С.Д. 2003. *Биологическая химия*. М.: Высшая школа, 107.
4. Wadkins R.M. 2000. Targeting DNA Secondary Structures. *Curr. Med. Chem.* **7**, 1–15.
5. Seliverstov A.V., Putzer H., Gelfand M.S., Lyubetsky V.A. 2005. Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol.* **5**, 54.
6. Vitreschak A.G., Lyubetskaya E.V., Shirshin M.A., Gelfand M.S., Lyubetsky V.A. 2004. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiol Lett.* **234**, 357–370.
7. Ishchukov I.M., Likhoshvai V.A., Matushkin Yu.G. 2004. A new algorithm for recognizing the operon structure of prokaryotes. *Proceedings of the fourth international conference on bioinformatics of genome regulation and structure*. Новосибирск: ред.-изд. отдел ИЦиГ, 73–76.