

УДК: 573; 57:51-76

Распределение мононуклеотидных повторов в бактериальных хромосомах: А/Т-треки преобладают над G/C-треками

С. С. Киселев^а, В. М. Комаров, И. С. Масулис, О. Н. Озолин

Учреждение Российской академии наук Институт биофизики клетки РАН,
142290, Московская обл., г. Пушкино, ул. Институтская, д. 3

E-mail: ^а anthyllium@gmail.com

Получено 26 мая 2010 г.

В данной работе исследовано распределение мононуклеотидных треков разной длины в 342 хромосомах зубактерий и 69 хромосомах архей. Несмотря на то, что число анализируемых повторов проявляет зависимость от нуклеотидного состава, в 73 % случаев (301 хромосома, в том числе и в 90 хромосомах, обогащенных GC-парами) было обнаружено преобладание poly(dA)_n- и poly(dT)_n-треков над poly(dG)_n- и poly(dC)_n-треками. В природных ДНК число А/Т-треков чаще всего в 2 раза выше, чем в случайных нуклеотидных последовательностях с аналогичным АТ/GC-составом и длиной. Обсуждаются возможные причины появления подобной асимметрии в распределениях.

Ключевые слова: мононуклеотидные повторы, А/Т-треки, G/C-треки

Distribution of mononucleotide repeats in bacterial chromosomes: A/T-tracts dominate on G/C-tracts

S. S. Kisselev, V. M. Komarov, I. S. Masulis, O. N. Ozoline

Institute of Cell Biophysics RAS, Institutskaya street 3, Pushchino, Moscow region, 142290, Russia

Abstract. – This study analyzes the abundance of mononucleotide tracts of different length in 342 eubacterial and 69 archaeal chromosomes. Despite the fact that the amount of analyzed repeats depends on nucleotide content, the predominance of poly(dA)_n- and poly(dT)_n-tracts on poly(dG)_n- and poly(dC)_n-tracts was found in 301 chromosomes (73 % of events), including 90 GC-rich chromosomes. In natural DNAs amount of A/T-tracts usually appeared to be two-fold higher than in randomized nucleotide sequences with the same AT/GC-content and length. Possible reasons of this asymmetry in distribution are discussed.

Keywords: mononucleotide repeats, A/T-tracts, G/C-tracts

Citation: *Computer Research and Modeling*, 2010, vol. 2, no. 2, pp. 183–187 (Russian).

Введение

Различного типа повторяющиеся нуклеотидные последовательности присутствуют в геномах как эукариот, так и прокариот. Принято считать, что большинство из них появились в результате генетических перестроек и дупликаций, т. е. их совокупность в некоторой степени отражает эволюционный путь конкретного биологического вида. Значительная часть повторяющихся элементов в ДНК представляет собой короткие простые повторы [Katti, Ranjekar, Gupta, 2001; van Belkum, 1998], в том числе мононуклеотидные треки, детальному исследованию которых посвящено относительно мало работ [Coenye, Vandamme, 2005; Dechering, 1998]. К настоящему времени накопились данные, свидетельствующие о некоторой неравномерности в распространении poly(dA)_n-, poly(dT)_n-, poly(dG)_n- и poly(dC)_n-повторов у прокариот. Она проявляется в преимущественном распространении poly(dA/T)-треков по сравнению с poly(dG/C)-повторами [Coenye, Vandamme, 2005; Dechering, 1998; Gur-Arie et al, 2000]. Вместе с тем биологическая значимость таких элементов геномов остается практически неисследованной, не понятно, существует ли какая-то общая закономерность в распространенности треков различных типов.

Целью данной работы был анализ распространения мононуклеотидных повторов (poly(dA)_n, poly(dT)_n, poly(dG)_n, poly(dC)_n) в репрезентативной выборке полностью расшифрованных геномов прокариот.

В рамках этой цели были рассмотрены следующие задачи:

1. Определение частот встречаемости мононуклеотидных повторов в геномах архей и эубактерий.
2. Сравнение частот встречаемости гомополимерных треков в хромосомных ДНК с таковыми в случайных нуклеотидных последовательностях, аналогичных по своему АТ/ГС-составу и длине природным ДНК.

Модели и методы

В работе использованы нуклеотидные последовательности 411 бактериальных хромосом (342 эубактериальных и 69 архейных), взятые из базы данных GenBank (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>). Среди хромосом архей 42 были с GC-составом менее 50 % и 27 – с GC-составом более 50 %, среди хромосом эубактерий – 171 и 171 соответственно.

Для поиска мононуклеотидных треков использовали программу DNA Tool [Ozoline, Deev, Arkhipova, 1997]. Количественный анализ осуществляли при помощи Microsoft Excel 2003. При этом учитывались треки длиной от 5 нуклеотидов и выше. Такой выбор был обусловлен тем, что мононуклеотидные треки длиной менее 5 нуклеотидов обладают менее выраженными структурообразующими свойствами, что ограничивает возможность их участия в различных биологических процессах.

Число мононуклеотидных АТ- и Г/С-треков в конкретной хромосомной ДНК сравнивали с количеством таких же треков в случайной нуклеотидной последовательности, имеющей аналогичный природной ДНК АТ/ГС-состав и длину. В качестве критерия отличия частот встречаемости гомонуклеотидных треков в геномных ДНК от таковых в случайных нуклеотидных последовательностях использовали величину

$$R_i = v_i^o / v_i^e,$$

где v_i^o – наблюдаемое число гомоолигомерных треков в i -й хромосоме, а v_i^e – ожидаемое число таких же треков в случайной последовательности, аналогичной по своему АТ/ГС-составу и длине i -й хромосоме. Значения v_i^e определяли отдельно для каждого из четырех видов повторов по формуле, предложенной de Wachter [de Wachter, 1981]:

$$\tau_{1,n} = p_i^n (1 - p_i) \times [(L - n - 1)(1 - p_i) + 2],$$

где $\tau_{1,n}$ – ожидаемое число гомоолигомерных треков из n нуклеотидов в последовательности длиной L нуклеотидов, p_i – частота встречаемости каждого основания в последовательности.

Результаты

Как и ожидалось, во всех хромосомах с GC-составом менее 50 % число А/Т-треков превышало число G/C-треков (рис. 1а), а в 110 хромосомах со средним значением GC-состава 65 % число G/C-треков было больше, чем число А/Т-треков (рис. 1б). Это значит, что распределение мононуклеотидных повторов проявляет определенную зависимость от нуклеотидного состава. Тем не менее для 81 эубактериальной и 9 архейных хромосом с GC-составом более 50 % было обнаружено явное преобладание poly(dA)_n- и poly(dT)_n-повторов, а не poly(dG)_n- и poly(dC)_n-треков. Таким образом, большинство исследованных хромосом (301 из 411 исследованных, т. е. в 73 % случаев) характеризуются преимущественным содержанием poly(dA)_n- и poly(dT)_n-треков, что предполагает их особое биологическое значение.

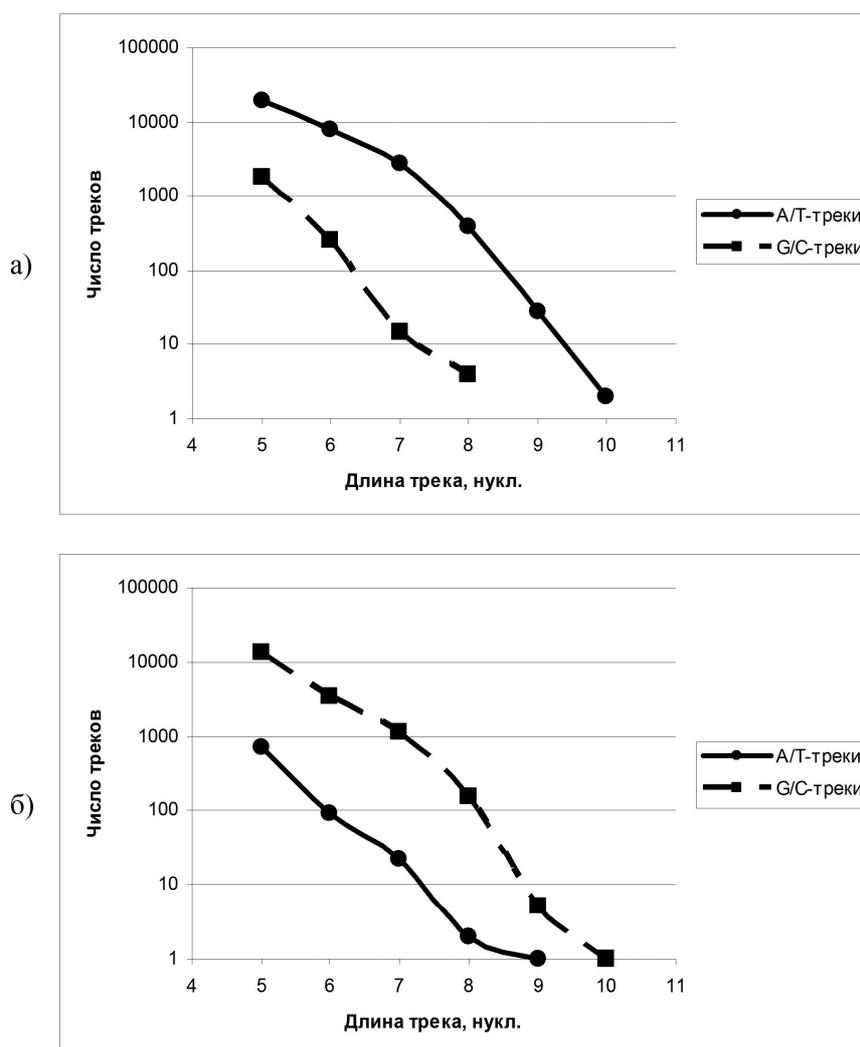


Рис. 1. Примеры распределения А/Т- и G/C-треков в конкретных хромосомных ДНК:
 а) *Bacillus subtilis* subsp. *subtilis* str. 168 (NC_000964) (длина – 4214630 п. н., GC-состав – 43,5 %);
 б) *Thermus thermophilus* HB27 (NC_005835) (длина – 1894877 п. н., GC-состав – 69,4 %)

Помимо этого, профиль распределения коэффициентов **R**, отражающих отклонение числа конкретных треков в геномных ДНК от их количества в случайных нуклеотидных последовательностях, отличался для А/Т-треков и G/C-треков (рис. 2). В подавляющем боль-

шинстве случаев $R_{(A+T)}$ был больше 1 (в 382 хромосомах, т. е. в 93 %), а $R_{(G+C)}$, наоборот, меньше 1 (в 294 хромосомах, т. е. в 72 %). Обе гистограммы выявляют наличие максимумов в профилях распределения R . Расположение этих максимумов свидетельствует о том, что число А/Т-треков в исследованных геномных ДНК чаще всего в 2 раза превышает их число в случайных нуклеотидных последовательностях, а G/C-треки, наоборот, встречаются в два раза реже.

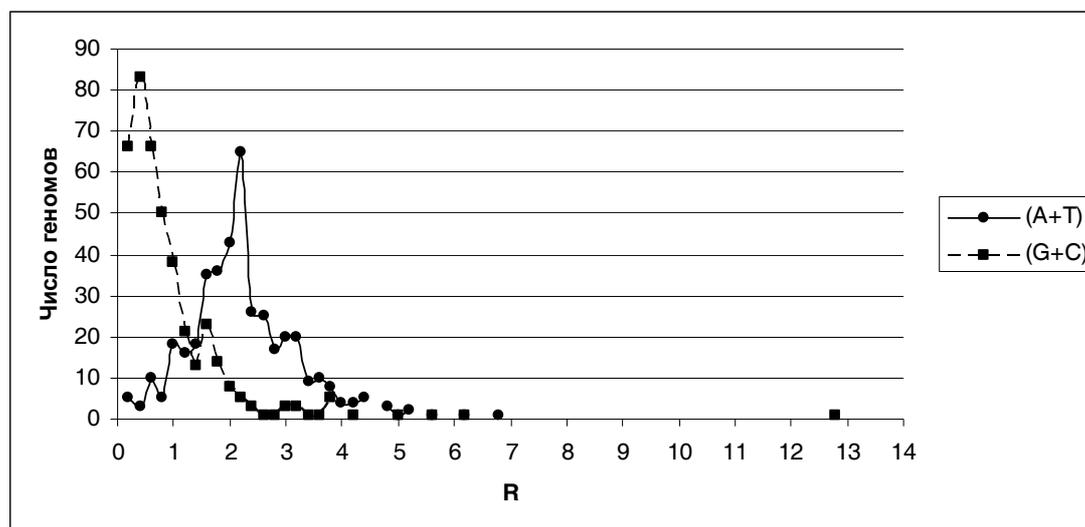


Рис. 2. Гистограмма распределения геномов по коэффициентам $R_{(A+T)}$ и $R_{(G+C)}$. Формула для вычисления R приведена в разделе «Модели и методы». Приведены данные для 411 хромосомных ДНК. Число геномов определено в последовательных интервалах $\Delta R = 0,2$

Вполне возможно, что обнаруженная асимметрия распространенности А/Т- и G/C-треков отражает некоторую внутреннюю особенность их структурной организации. В основе этого явления, например, может лежать исходно пониженный структурный полиморфизм водородного связывания уотсон-криковских АТ-пар по сравнению с GC-парами. Как показано в работах [Комаров, 1998; Kabanov, Komarov, 2002; Komarov, Polozov, Konoplev, 1992] GC-пары, в отличие от АТ-пар, характеризуются 4-х кратным вырождением водородного связывания оснований. Для них реализуются две структуры «пропеллероподобного» типа (правый и левый пропеллер) и две симметричные структуры «ступенькообразного» типа на одних и тех же водородных связях, в то время как АТ-пары оснований образуют только две «пропеллероподобные» структуры комплементарного Н-спаривания, а «ступенькообразные» структуры у них отсутствуют. Меньшая гетерогенность может способствовать большей устойчивости А/Т-треков, что может быть важным для их биологической функции.

Заключение

По результатам данного исследования можно сделать следующие выводы:

1. В хромосомных ДНК прокариот наблюдается общая тенденция преимущественного присутствия гомонуклеотидных А/Т-треков, которые преобладают над G/C-треками не только в хромосомах с GC-составом менее 50 %, но и в ряде хромосом с высоким содержанием GC-пар.
2. Природные ДНК отличаются от случайных нуклеотидных последовательностей частотой встречаемости гомонуклеотидных трекков, причем параллельно с накоплением А/Т-треков наблюдается уменьшение числа G/C-треков.

Список литературы

- Van Belkum A., Scherer S., van Alphen L., Verbrugh H.* Short-sequence DNA repeats in prokaryotic genomes // *Microbiol. Mol. Biol. Rev.* 1998. Vol. 62. P. 275–293.
- Coenye T., Vandamme P.* Characterization of mononucleotide repeats in sequenced prokaryotic genomes. // *DNA Res.* 2005. Vol. 12. P. 221–233.
- Dechering K. J., Cuelenaere K., Konings R. N., Leunissen J. A.* Distinct frequency-distributions of homopolymeric DNA tracts in different genomes // *Nucleic Acids Res.* 1998. Vol. 26. P. 4056–4062.
- Gur-Arie R., Cohen C. J., Eitan Y., Shelef L., Hallerman E. M., Kashi Y.* Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism // *Genome Res.* 2000. Vol. 10. P. 62–71.
- Kabanov A. V., Komarov V. M.* Polymorphism of hydrogen bonding in the short double helices of oligonucleotides: semiempirical quantum chemical study // *Int. J. Quant. Chem.* 2002. Vol. 88. P. 579–587.
- Katti M. V., Ranjekar P. K., Gupta V. S.* Differential distribution of simple sequence repeats in eukaryotic genome sequences // *Mol. Biol. Evol.* 2001. Vol. 18. P. 1161–1167.
- Komarov V. M., Polozov R. V., Konoplev G. G.* Non-planar structure of nitrous bases and non-coplanarity of Watson-Crick pairs // *J. Theor. Biol.* 1992. Vol. 155. P. 281–294.
- Ozoline O. N., Deev A. A., Arkhipova M. V.* Non-canonical sequence elements in the promoter structure. Cluster analysis of promoters recognized by *Escherichia coli* RNA polymerase // *Nucleic Acids Res.* 1997. V. 25. P. 4703–4709.
- De Wachter R.* The number of repeats expected in random nucleic acid sequences and found in genes // *J. Theor. Biol.* 1981. Vol. 91. P. 71–98.
- Комаров В. М.* Квантово-химическое полуэмпирическое исследование полиморфизма уотсон-криковского спаривания азотистых оснований // *Биофизика.* 1998. Т. 43. № 6. С. 967–974.