

**РАСПРЕДЕЛЕНИЕ МОНОНУКЛЕОТИДНЫХ ПОВТОРОВ В
БАКТЕРИАЛЬНЫХ ХРОМОСОМАХ:
А/Т-ТРЕКИ ПРЕОБЛАДАЮТ НАД G/C-ТРЕКАМИ**

Киселев С.С., Комаров В.М., Масулис И.С., Озолинь О.Н.

Учреждение Российской академии наук Институт биофизики клетки РАН, Россия,
142290, г. Пущино, ул. Институтская, 3, тел. 8(4967)73-94-04,
e-mail: anthyllum@gmail.com

Установление особенностей распределения повторяющихся нуклеотидных последовательностей в геномах различных групп организмов является важнейшей задачей сравнительной геномики. Известно, что подобные последовательности широко распространены как среди про-, так и эукариот. Существующие данные свидетельствуют о наличии некоторой неравномерности в распространении различных мономерных повторов, вариативности частот их встречаемости в геномах микроорганизмов разных таксономических групп. Однако биологическая значимость этих элементов геномов остается практически не исследованной, не ясно, существуют ли какие-то общие закономерности в распределении трексов различных типов.

В работе исследовалось распространение мономерных повторов в 411 бактериальных хромосомах (342 эубактериальных и 69 архейных) из банка данных NCBI GenBank. Была проанализирована встречаемость гомополимерных трексов (poly(A)_n , poly(T)_n , poly(G)_n и poly(C)_n) длиной от 5 нуклеотидов и выше. При этом в качестве критерия отличия частот встречаемости подобных трексов в геномных ДНК от случайных нуклеотидных последовательностей использовалась величина $R_i = v_i^o / v_i^e$, где v_i^o — наблюдаемое число гомоолигомерных трексов в i -ой хромосоме, а v_i^e — ожидаемое число таких же трексов в случайной последовательности, аналогичной по своему АТ/ГС-составу i -й хромосоме. Значения v_i^e определяли отдельно для каждого из четырех видов повторов по формуле, предложенной De Wachter:

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

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

Оказалось, что в большинстве хромосом (в 301 из 411 исследованных, т.е. в 73,2% случаев) poly(A)_n - и poly(T)_n -повторы значительно преобладали над poly(G)_n - и poly(C)_n -повторами. Подобное преобладание наблюдалось и в ряде случаев, когда хромосомы обладали GC-составом более 50% (81 эубактериальная и 9 архейных хромосом).

Обсуждаются возможные биофизические причины появления подобной асимметрии в распределениях.