

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВСТРЕЧАЕМОСТИ НУКЛЕОТИДНЫХ W-ТРЕКОВ И S-ТРЕКОВ В ГЕНОМАХ УМЕРЕННО GC-БОГАТЫХ МИКРООРГАНИЗМОВ

Киселев С.С., Комаров В.М., Масулис И.С., Деев А.А.¹, 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: anthyllium@gmail.com

¹Учреждение Российской академии наук Институт теоретической и экспериментальной биофизики РАН, Россия, 142290, г. Пущино, ул. Институтская, 3

Смешанные нуклеотидные последовательности, содержащие dA и dT (W-треки), dG и dC (S-треки) широко распространены среди прокариот и эукариот. Существующие данные свидетельствуют о вариабельности частот их встречаемости в геномах микроорганизмов из разных таксономических групп.

Ранее нами было продемонстрировано, что мононуклеотидные poly(dA)_n- и poly(dT)_n-треки преобладают над poly(dG)_n- и poly(dC)_n-треками в геномах прокариот. Данная тенденция наблюдается более чем в 70% случаев, причём не только для AT-богатых микроорганизмов, но даже для бактерий с GC-составом более 50%.

В данной работе проведено исследование распространения W-треков и S-треков в геномах умеренно GC-богатых прокариот, содержащих от 50% до 60% процентов GC-пар. Нуклеотидные последовательности и аннотации полностью секвенированных геномов были взяты из банка данных NCBI GenBank (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>). Учитывали треки длиной от 5 нуклеотидов и выше. Критерием отличия частот встречаемости смешанных треков в геномных ДНК от случайных нуклеотидных последовательностей служила величина $R_i = v_i^o / v_i^e$, где v_i^o — наблюдаемое число треков в *i*-ой хромосоме, а v_i^e — ожидаемое число таких же треков в случайной последовательности, аналогичной по своему AT/GC-составу *i*-й хромосоме.

Также были изучены особенности распределения W-треков и S-треков внутри генов и межгенных областей. Для разбиения исходной последовательности геномной ДНК на кодирующие и не кодирующие участки использовали компьютерную программу DNA Tool.

Оказалось, что в межгенных участках количество W-треков часто превышает число S-треков, но при этом внутри генов наблюдается зависимость встречаемости треков от GC-состава.

Характер распространения смешанных треков в геномах умеренно GC-богатых микроорганизмов (без учёта генов и не кодирующих областей) демонстрирует более сильную зависимость от GC-состава по сравнению с встречаемостью мононуклеотидных poly(dA)_n- и poly(dT)_n-треков.

Обсуждаются возможные причины доминирования W-треков над S-треками в не кодирующих участках геномов.