

Секция «Биоинженерия и биоинформатика»

Изучение распределения сайтов рестрикции в геномах прокариот

Русинов Иван Сергеевич

Студент

*Московский государственный университет имени М.В. Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия*

E-mail: is_rusinov@mail.ru

Система рестрикции-модификации (Р-М) - это ферментативная система прокариотической клетки, функцией которой является борьба с чужеродной ДНК, в частности с бактериофагами. Сайты узнавания компонентов систем Р-М называются сайтами рестрикции. Одним из механизмов, позволяющих бактериофагам успешно избегать рестрикции, является уменьшение количества сайтов рестрикции в ДНК [2]. Основываясь на статистической недопредставленности палиндромных последовательностей в нескольких прокариотических геномах, М.С. Гельфанд и Е.В. Кунин выдвинули предположение, что и прокариоты избегают сайтов узнавания закодированных в геноме систем Р-М [1].

В данной работе было изучено распределение 277 сайтов рестрикции и 86 других палиндромов в 2764 последовательностях хромосом и плазмид полных прокариотических геномов, для которых в базе данных REBASE была доступна информация о закодированных в них экспериментально подтвержденных или предсказанных системах Р-М. Недопредставленность сайтов оценивалась методом, предложенным С. Карлиным и соавторами [2].

Было показано, что распределение встречаемости сайтов в прокариотических геномах сдвинуто в область недопредставленности относительно распределения для эукариотических вирусов, которые не подвержены действию систем Р-М (см. рис. 1). Далее мы выбрали случаи закодированных в геноме эндонуклеаз рестрикции с известным сайтом узнавания и показали, что 50,9% таких сайтов недопредставлено, а перепредставлено только 4,1% сайтов, и 93,5% таких случаев - сайт CCGG в представителях рода *Helicobacter*.

Мы показали, что недопредставленных сайтов значительно больше, чем можно объяснить закодированными в геноме системами Р-М: в среднем в геноме недопредставлено 15,5 потенциальных сайтов рестрикции, при этом в среднем на один геном приходится около 4 закодированных в нем систем Р-М. Возможным объяснением может служить эволюционный след от систем Р-М, присутствовавших у предка организма, но потерянных к настоящему моменту. Для проверки этого предположения мы сравнили профили встречаемости потенциальных сайтов рестрикции близких родственников. Оказалось, что высоким сходством обладают профили для организмов, принадлежащих одному роду. Этому правилу не подчиняются плазмиды, профили для которых могут существенно отличаться даже от хромосомных профилей того же штамма. Подобное поведение плазмид может быть объяснено их высокой мобильностью. Таким образом, почти четырехкратный избыток недопредставленных сайтов рестрикции по сравнению с известными системами Р-М действительно можно объяснить эволюционным следом утраченных систем Р-М.

Литература

1. Gelfand M.S., Koonin E.V. Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes // Nucleic Acids Research. 1997; 25(12): 2430-2439.
2. Karlin S., Burge C., Campbell A.M. Statistical analyses of counts and distributions of restriction sites in DNA sequences // Nucleic Acids Research. 1992; 20(6): 1363-1370.

Слова благодарности

Хочу выразить благодарность моим научным руководителям: Ершовой А.С. и Алексе-евскому А.В.

Иллюстрации

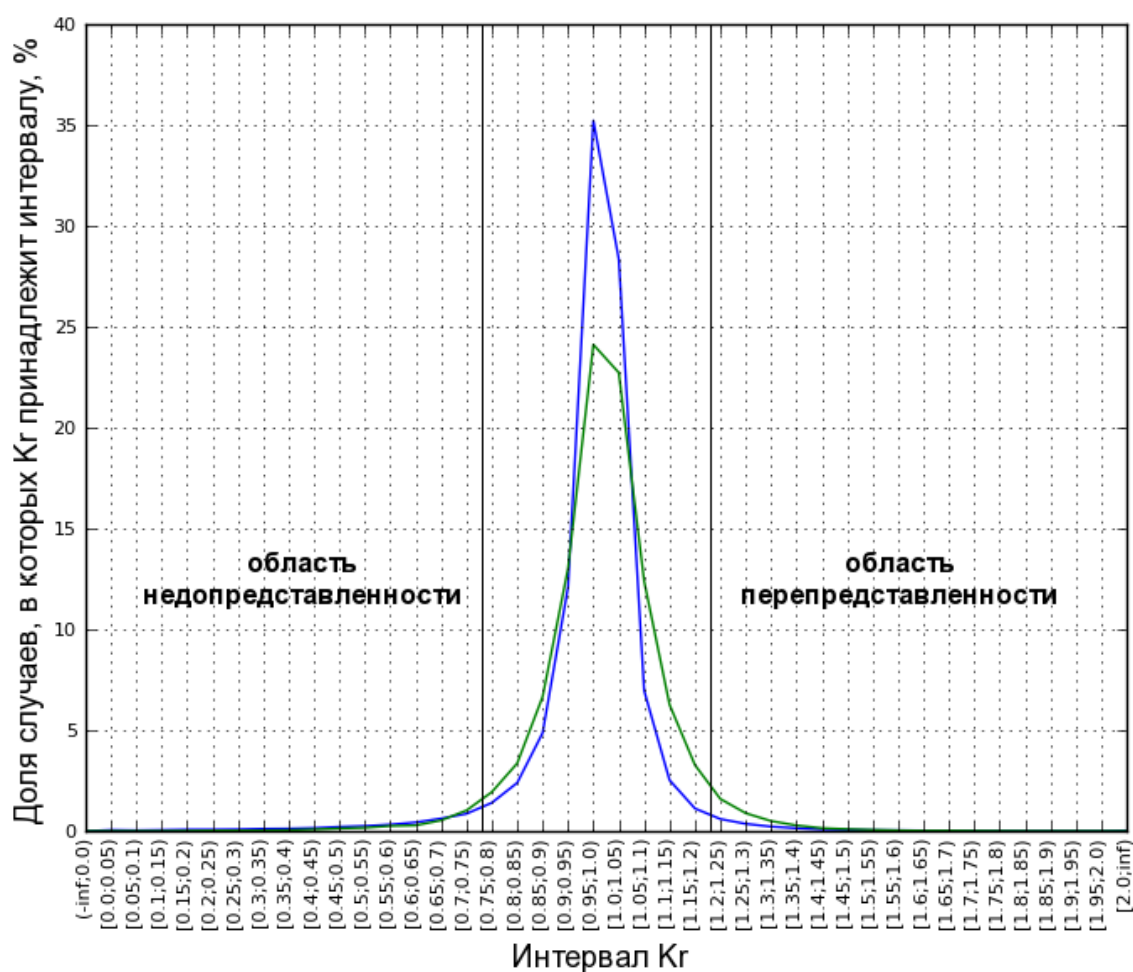


Рис. 1: Распределение встречаемости сайтов рестрикции. Kt - характеристика встречаемости сайта, равная отношению числа сайтов в геноме к ожидаемому числу сайтов, вычисленному методом Карлина и соавторов. Синий график соответствует прокариотическим геномам, зеленый - геномам эукариотических вирусов.