

Секция «Биоинженерия и биоинформатика»

CRISPR-кассеты: построение филогении

Айдарханов Руслан Дауренович

Студент

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

E-mail: darkhan.rus@gmail.com

Половина известных прокариотических организмов имеют в своём геноме одну или несколько CRISPR-кассет – специальных структур, состоящих из прямых повторов, разделённых неповторяющимися спейсерами сходной длины. Особенности этой структуры отражены и в её названии – CRISPR является аббревиатурой от английского Clustered Regularly Interspaced Short Palindromic Repeats. Хотя CRISPR-локус (кассета и ассоциированные с ней гены) ещё недостаточно хорошо изучен, экспериментально показано, что его функцией может быть защита клетки от чужеродной ДНК. При первой атаке чужеродным ДНК-агентом, например, бактериофагом клетка может вырезать короткий фрагмент его ДНК и встроить его в CRISPR-кассету в качестве нового спейсера. При последующих инфекциях данным фагом клетка может нейтрализовать ДНК фага, разрушив её с помощью механизма, напоминающего РНК-интерференцию эукариот, используя для узнавания фрагмент последовательности ДНК фага в кассете [1].

Интересно, что новые спейсеры всегда достраивают кассету с одной стороны. Также известно, что CRISPR-кассета может терять имеющиеся спейсеры, по-видимому, в результате рекомбинации между копиями повтора [2].

Поскольку система “вирус - хозяйская клетка” очень динамична, CRISPR-локус представляет собой пример наиболее быстро меняющегося участка генома. Поэтому спейсерный состав кассет часто используют для типирования близких штаммов бактерий [3]. Кроме того, для многих систем было бы познавательным взглянуть на популяционное разнообразие родственных кассет, а также выявить родственные отношения между кассетами различных видов или представителями одного вида, которые находятся в разных условиях, в разных местах обитания. До сих пор не было предложено методов филогенетического анализа CRISPR-кассет.

Единицей эволюции кассеты становятся не отдельно взятые нуклеотиды, а целые спейсеры. То есть существующие методы построения филогении не применимы прямо для решения описанной задачи.

Целью данной работы была разработка и имплементация алгоритма построения филогении CRISPR-кассет.

Алгоритм использует принципы динамического программирования для построения парного выравнивания последовательностей. Выравниваются последовательности (цепочки) спейсеров, причем запрещена замена (вес замены задан бесконечно малым), а также используются различные штрафы за пробелы в начале выравнивания (что отражает вставку нового спейсера) и в остальной части (что означает удаление спейсеров в результате рекомбинации). Из весов выравнивания рассчитывается матрица расстояний для дальнейшей обработки с помощью алгоритма для построения филогенетических деревьев Neighbour-Joining.

Описанный алгоритм был имплементирован на языке программирования Java. На основе данной имплементации в дальнейшем планируется разработка общедоступного веб-сервиса.

Разработанный алгоритм был протестирован на базе данных CRISPR-кассет штаммов бактерии *Escherichia coli*. Результаты филогенетического анализа этих кассет будут представлены в докладе.

Литература

1. Sorek R., Kunin V., Hugenholtz P. CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea // *Nature Reviews Microbiology*. 2008. No 6. Pp 181-186.
2. Barrangou R., Fremaux C., Deveau H. CRISPR provides acquired resistance against viruses in prokaryotes // *Science*. 2007. No 5819. Pp 1709-1712.
3. Brudey K., Driscoll J.R., Rigouts L. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology // *BMC Microbiol*. 2006. No 6. P 23.

Слова благодарности

Выражаю благодарность Артамоновой Ирине Игоревне за научное руководство.