

МФТИ

ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА

6 СЕМЕСТР

**Алгоритмы выравнивания
последовательностей и построение
филогенетического дерева**

Выполнили: Шарафутдинов Эмиль
Блохина Яна

1 Аннотация

Целью проекта является построение филогенетического дерева, которое в идеале позволяет точно понять и отследить происхождение тех или иных видов и их родственные связи. Это хороший инструмент для уточнения механизмов эволюции, создания новых лекарств против болезнетворных организмов, улучшения селекции сельскохозяйственных растений и животных. В данной работе изучаются и анализируются необходимые при построении филогенетического дерева алгоритмы выравнивания при решении задачи поиска общих генов. Используются последовательности нуклеотидов вирусов, так как количество нуклеотидов у этих организмов минимальное из всех существующих. Производится сравнение генов различных вирусов.

2 Введение

Филогенетическое дерево представляет гипотезу о эволюционном происхождении набора генов, видов или других таксонов. В данной работе мы используем нуклеотидные последовательности, кодирующие белки, в качестве основы для классификации. Молекулярная филогенетика широко использует выравнивание последовательностей при построении и уточнения филогенетических деревьев, которые используются для классификации эволюционных взаимоотношений между гомологичными генами, представленными в геномах расходящихся видов.

Выравнивание последовательностей — метод, основанный на размещении двух или более последовательностей мономеров ДНК, РНК или белков друг под другом таким образом, чтобы легко увидеть сходные участки в этих последовательностях. Сходство первичных структур двух молекул может отражать их функциональные, структурные или эволюционные взаимосвязи. Выровненные последовательности оснований нуклеотидов или аминокислот обычно представляются в виде строк матрицы. Добавляются разрывы между основаниями таким образом, чтобы одинаковые или похожие элементы были расположены в следующих друг за другом столбцах матрицы. Существует несколько видов выравнивания. Парное выравнивание используется для нахождения сходных участков двух последовательностей. Различают глобальное и локальное выравнивание. Глобальное выравнивание предполагает, что последовательности гомологичны по всей длине. В глобальное выравнивание включаются обе входные последовательности целиком. Локальное выравнивание применяется, если последовательности содержат как родственные (гомологичные), так и неродственные участки. Результатом локального выравнивания является выбор участка в каждой из последовательностей и выравнивание между этими участками. Для получения парного выравнивания используются разновидности метода динамического программирования.

Глобальное выравнивание использует полную длину обеих последовательностей, и может быть использовано для проверки последовательностей на

гомологию (общность происхождения) по всей длине. Однако, если последовательности имеют мало участков гомологии (или просто схожести), то не всегда можно хорошо определить эти участки. В приведённом примере алгоритм зацепился за четыре совпадающих аминокислоты, так что длинный участок гомологии не виден. На основании этого можно предположить, что последовательности целиком не гомологичны между собой.

Локальное выравнивание использует части последовательностей, на которых прогнозируется максимальная гомология. Оно отлично подходит, если лишь части последовательностей похожи, например в ходе рекомбинации или конвергентной эволюции. Всегда стоит аккуратно относиться к небольшим участкам имеющим низкое сходство, особенно при выравнивании больших последовательностей, так как повышается вероятность встречи случайного схожего участка.

Однако, не всегда в локальное выравнивание может попасть интересующий участок последовательности. Это можно обойти, если обрезать последовательность по границам интересующего участка. Также возможны и другие комбинации глобального и локального выравниваний.