

МФТИ

ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА

6 СЕМЕСТР

**Алгоритмы выравнивания
последовательностей и построение
филогенетического дерева**

Выполнили: Шарафутдинов Эмиль
Блохина Яна

1 Аннотация

Целью проекта является определение степени родства различных видов вирусов. В данной работе изучаются и анализируются необходимые 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

В основе этого проекта лежит интерес авторов к **молекулярной филогенетике** — науке устанавливающей и исследующей родственные связи между живыми организмами на основании изучения структуры полимерных макромолекул — ДНК, РНК и белков. Результатом молекулярно-филогенетического анализа является построение филогенетического дерева живых организмов. Близкое родство между живыми организмами обычно сопровождается большой степенью сходства в строении тех или иных макромолекул, а молекулы не родственных организмов сильно различаются между собой. Молекулярная филогения использует такие данные для построения филогенетического дерева, которое отражает гипотетический ход эволюции исследуемых организмов.

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.

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.

3 Методическая часть

Выравнивание последовательностей проводилось с использованием трёх алгоритмов: алгоритма Нидлмана-Вунша, Fasta и BLAST

3.1 Алгоритм Нидлмана-Вунша

	A	C	T	G	A	T	T	C	A	
A	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
C	-2	2	0	-2	-4	-6	-8	-10	-12	-14
G	-4	0	4	2	0	-2	-4	-6	-8	-10
C	-6	-2	2	1	4	2	0	-2	-4	-6
A	-8	-4	0	-1	2	1	-1	-3	0	-2
T	-10	-6	-2	-3	0	4	2	0	-2	2
T	-12	-8	-4	0	-2	2	6	4	2	0
C	-14	-10	-6	-2	-4	0	4	2	6	4
A	-16	-12	-8	-4	-5	-2	2	1	4	8

Рис. 1: Принцип работы алгоритма Нидлмана-Вунша

3.2 Алгоритм FastA

На входе имеется некоторое количество строк-последовательностей, этот алгоритм попарно сравнивает все эти последовательности, находя самую короткую последовательность он ищет участки похожие участки этой последовательности в других последовательностях. Далее идет выборка наиболее длинных диагоналей(которую мы можем варьировать) и соединяет все эти диагонали. Таким образом, происходит процесс нахождения участков идентичных исходной последовательности.

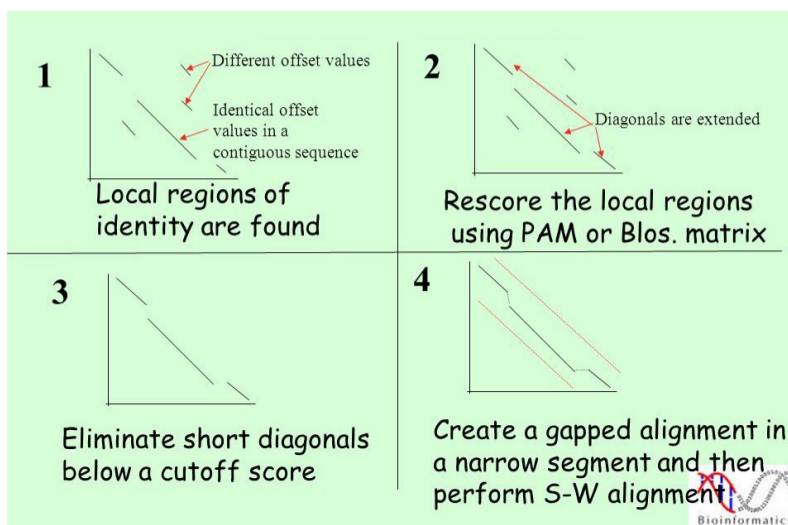


Рис. 2: Принцип работы FastA

3.3 Алгоритм BLAST

BLAST (англ. Basic Local Alignment Search Tool — средство поиска основного локального выравнивания). Используя BLAST, можно сравнить имеющуюся последовательность с другой и найти последовательности предполагаемых гомологов. Программы серии BLAST производят локальные выравнивания, что связано с наличием в различных последовательностях. В случае глобального выравнивания обнаруживается меньшее сходство последовательностей.

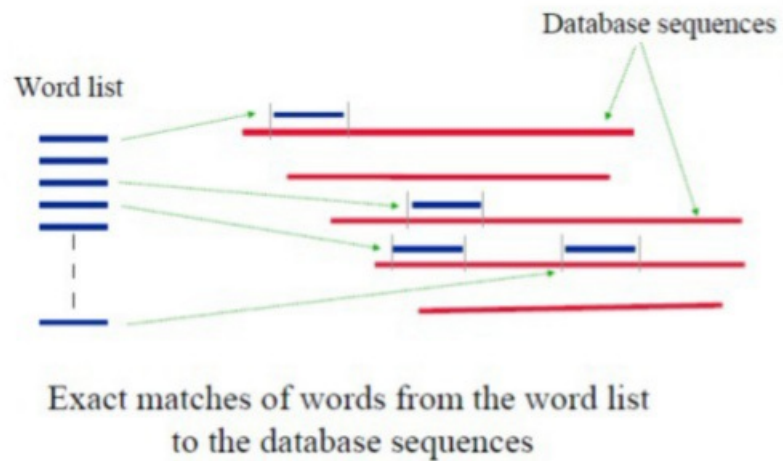


Рис. 3: Принцип работы BLAST