

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Санкт-Петербургский политехнический университет Петра Великого»

УДК

№

Инв. №

У Т В Е Р Ж Д А Ю

Зав. НИЛ «Математическая биология
и биоинформатика», ИПММ
ФГАОУ ВО «СПбПУ»,
д.б.н.

_____ М. Г. Самсонова
«_____» _____ 2016 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

по теме:
«Стохастическое моделирование экспрессии генов»

Выполнил студент гр. №53601/4

_____ Д. В. Яковлев
«_____» _____ 2016 г.

Руководитель НИР, к.ф.-м.н.

_____ В.В. Гурский
«_____» _____ 2016 г.

Санкт-Петербург 2016

РЕФЕРАТ

Отчёт 21 стр., 1 часть, 8 рис., 1 таблица, 5 источников

РАЗРАБОТКА И АНАЛИЗ МЕТОДОВ УСКОРЕНИЯ ПАРНОГО ВЫРАВНИВАНИЯ РЕСТРИКЦИОННЫХ КАРТ

В данной работе были рассмотрены два подхода ускорения алгоритма парного выравнивания на основании апостериорной информации о смещении рестрикционных карт относительно друг друга. Каждый из подходов был запрограммирован и проведён их сравнительный анализ. Сравнение подходов проведено по таким критериям, как временная сложность. Предложены критерии выбора алгоритмов выравнивания при различных условиях.

СОДЕРЖАНИЕ

Введение	4
1 Постановка задачи	5
2 Экспериментальные расчёты	6
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	8

Введение

В наше время биоинформатика является неотъемлемой частью практически всех областей биологии. Наиболее значим вклад биоинформатики в области генетики. С помощью введения методов биоинформатики гораздо упростилась работа с большим количеством информации о структуре и последовательности генетического материала.

Одним из инструментов исследования ДНК стало использование ферментов рестрикции (рестрикционное картирование). Они позволяют превращать молекулы ДНК очень большого размера в набор фрагментов длиной от нескольких сотен до нескольких тысяч оснований.

С помощью рестрикционного картирования генома можно идентифицировать на ДНК биологически важные участки. Метод рестрикционного картирования позволяет выделить в геноме крупные генетические изменения, такие как хромосомные перестройки (делеция или инсерция). Метод использования рестрикционных карт рассматривается как скоринговый метод анализа структурных вариаций и в целом для анализа регионов генома не поддающихся сборке с помощью методов основанных на нуклиотидных последовательностях, так как значительно менее трудозатратен и менее информативен, поэтому он менее изучен. Природа рестрикционных карт отличается от таковой для нуклиотидных последовательностей, поэтому необходимо учитывать особенность модели при рассмотрении методов их анализа.

В рамках проблемы карты генома WGM Assembly [1], основанного на множественном выравнивании рестрикционных карт, время попарного выравнивания составляет существенную часть общего времени работы алгоритма [2]. Поэтому особое значение приобретает задача его ускорения. В работе рассматриваются модификации метода парного выравнивания, предложенного в статьях Смита-Ватермана [3], за счет использования информации об относительном смещении карт друг относительно друга, позволяющие существенно ускорить время его работы.

1 Постановка задачи

Время работы алгоритма попарного выравнивания, основанного на динамическом заполнении матрицы выравниваний можно оценить как $O(n^2)$, где n - максимально возможное количество разрезов на карте. Целью данной работы является описание и исследование двух эвристических методов, которые позволяют его ускорить, и добиться оценки среднего времени работы $O(n)$ для каждого из них. В данной работе будет проведено сравнение двух этих методов между собой и с исходным алгоритмом. Будет показано, что использование этих методов предпочтительно в случае определенной модели структуры множества исходных рестрикционных карт, а так же будут оценены границы применимости данных эвристик.

2 Экспериментальные расчёты

ВЫВОДЫ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Whole Genome Mapping Assembly. — <http://opgen.com/genomic-services/what-is-whole-genome-mapping>. — Accessed: 2015-30-02.
- 2 *Erin Newburn Bin Zhu Jia Liu Emily Zentsand Rich Moore*. Accurate Construction of the Genomic Architercture of Plants and Animals Using Whole Genome Mapping(WGM) Technology.
- 3 *Smith T. F., Waterman M. S.* Identification of Common Molecular Subsequences. — 1980.