

Построение генетических карт по неполным и зашумленным данным

Алина Крамар
группа 545
руководитель Сысоев С.С
рецензент Добрынин П.В.

Санкт-Петербургский государственный университет

3 июня 2014 г.

Необходимые определения

Генетические карты

- При получении генетического материала невозможно дать ответ о расположении маркеров на хромосоме
- Кроссинговер вносит возмущение в этот порядок
- Порядок маркеров и формирование гамет происходит во время мейоза

Существующие решения

- CRIMAP
(Экспоненциальная сложность от количества особей в родословной)
- FASTLINK
(Экспоненциальная сложность от количества маркеров)
- genmap
(Степенная сложность от количества особей и маркеров)

Постановка задачи

Написать новый алгоритм генетического картирования на базе метода прямого извлечения данных (genmap)

Задачи:

- Доработка алгоритма прямого извлечения данных
- Верификация полученного алгоритма
 - моделирование и генерация тестовых данных
 - сравнение существующих алгоритмов с полученной версией

Недостатки genmap

- Неустойчивость к вариациям входных данных
- Неоптимальная реализация
- Неправильный результат в случае кратного кроссинговера

Генерация тестовых данных

- Стратегия кроссовера
- Численные параметры родословной
- Симуляция ошибок секвенирования

Пример родословной

Результаты сравнения алгоритмов

Особи/Маркеры	CRIMAP	FASTLINK	Genmap2
200/30	0.3 секунд	10 минут	1 час
200/50	0.3 секунды	10 минут	∞
4000/40	6 секунд	∞	3 дня

Учёт кратности кроссинговера

Результаты

- Реализован более хороший алгоритм на основе гентар
- Проведена его верификация
 - путём моделирования и генерации тестовых данных
 - экспериментально показав его сильные стороны
- Выступление на конференции СПИСОК-2014