

Построение генетических карт по неполным и зашумленным данным

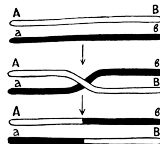
Алина Крамар
группа 545
руководитель Сысоев С.С
рецензент Добрынин П.В.

Санкт-Петербургский государственный университет

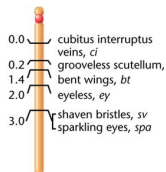
4 июня 2014 г.

Необходимые определения

- геном
- маркер
- кроссинговер
- генетическая карта



Кроссинговер



Генетическая карта

Существующие решения

- Алгоритм Ландера-Грина и CRIMAP
Экспоненциальная сложность от количества особей в родословной.
- Алгоритм Элстона-Стюарта и FASTLINK
Экспоненциальная сложность от количества маркеров.
- Алгоритм Сергея Сысоева и **genmap**
Неправильный ответ в случае кратного кроссинговера.

Постановка задачи

Разработать новый алгоритм генетического картирования на базе метода прямого извлечения данных **genmap**.

Задачи:

- Доработка алгоритма прямого извлечения данных **genmap**.
- Верификация полученного алгоритма.
 - Моделирование и генерация тестовых данных.
 - Сравнение с существующими алгоритмам.

Улучшение **genmap**

Недостатки:

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

Что улучшили:

- Работа с “плохими” данными.
- Уточнение карты.
- Учёт кратных кроссинговеров.

Генерация тестовых данных

Синтетические vs. реальные данные.

Много параметров:

- Стратегия кроссинговера.
- Численные параметры родословной.
- Ошибки секвенирования.

Результаты сравнения алгоритмов

Особи/Маркеры	Genmap2	FASTLINK	CRIMAP
200/30	0.3 секунд	10 минут	1 час
200/50	0.3 секунды	10 минут	∞
400/40	6 секунд	∞	3 дня

Количество ошибок в результатах

Тестовое множество	Genmap	Genmap2
nocross	0	0
singlecross	0	0
manycross	12	10
cats	1	1

Результаты

Реализован более хороший алгоритм на основе **genmap**. Проведена его верификация.

- Использованы естественные и синтетические данные.
- Экспериментально показаны его сильные стороны.
- Произведено сравнение с аналогами.

Результаты работы были представлены на конференции **СПИСОК-2014**.