

Построение генетических карт по неполным и зашумленным данным

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

Санкт-Петербургский государственный университет

18 мая 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)
- LM_MAP
(Экспоненциальная сложность от количества маркеров)
- Прототип алгоритма Сергея Сысоева
(Степенная сложность от количества особей и маркеров)

Постановка задачи

Сравнить существующие алгоритмы генетического картирования и улучшить алгоритм Сергея Сысоева

Задачи:

- Сравнение алгоритмов генетического картирования
 - реализовать алгоритм генерации тестовых данных
 - сравнить существующие алгоритмы на реализованной базе
- Усовершенствование алгоритма Сысоева
 - выявить недостатки алгоритма путём
 - реализовать новую версию алгоритма
 - выложить в открытый доступ новую версию

Генерация тестовых данных

- Моделирование мейоза и скрещивания
- Генерация родословных по заданным параметрам
- Симуляция мутаций
- Генерация несоответствующих модели данных

Пример родословной

Результаты сравнения алгоритмов

Особи/Маркеры	CRIMAP	LM_MAP	genmap.py
192/35	120 секунд	111 секунд	17 секунд
2011/146	1 день	1 день	11 минут
100000/100	10+ дней	9+ дней	37 минут

Результаты тестирования алгоритма С.Сысоева

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

Учёт кратности кроссинговера

Результаты

- Произведено сравнение алгоритмов
 - реализован алгоритм генерации тестовых данных
 - произведено сравнение алгоритмов с помощью реализованной базы
- Улучшения алгоритма С.Сысоева
 - произведено тестирование алгоритма на синтетических данных
 - учёт кратных рекомбинаций
 - скрипт выложен в открытый доступ (URL)
- Выступление на конференции СПИСОК-2014