

Определение родственных связей по данным генотипирования

Абрамова Анастасия Николаевна

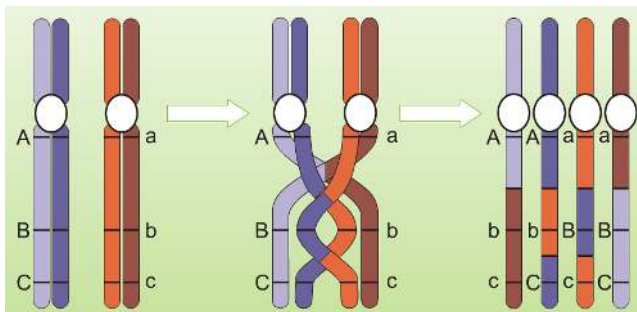
Руководитель: Александр Ракитько, Генотек

Институт биоинформатики

Санкт-Петербург, 2015

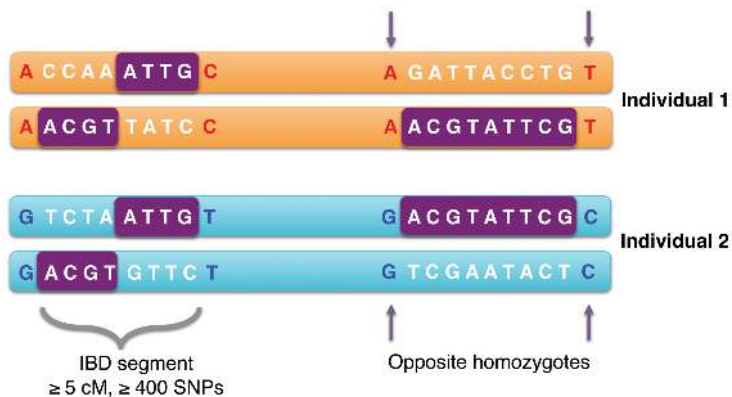
- Научиться моделировать родословную
- Реализовать алгоритм, предложенный *23&me*, вычисляющий степень родства между двумя индивидами на основе количества и суммарной длины IBDhalf сегментов
- Подобрать распределения для количества и длин сегментов в каждом поколении
- На основе теста отношения правдоподобий научиться вычислять степень родства двух индивидов

Кроссинговер

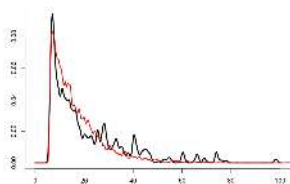


Для моделирования будем считать, что число пар оснований, в которых произошел кроссинговер имеет распределение Пуассона с параметром $\lambda = \frac{\text{длина хромосомы в сМ}}{100}$

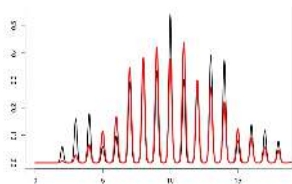
IBDhalf сегменты



Распределение длин и количества IBDhalf сегментов



Плотность длины сегментов



Плотность количества сегментов

Сложная гипотеза против сложной альтернативы:

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_0^c$$

Статистика имеет вид

$$\Lambda(x) = \frac{\sup\{L(\theta | x) : \theta \in \Theta_0\}}{\sup\{L(\theta | x) : \theta \in \Theta\}},$$

где L – функция правдоподобия, x – выборка.

Таблица: Сравнение результатов по первым пяти поколениям

Generation	distribution	length/number
1	100%	100%
2	91%	100%
3	90%	89%
4	71%	49%

- Реализованы описанные алгоритмы
- Проведено сравнение результатов

