

Определение наиболее вероятного генотипа и этнической принадлежности индивида исходя из генотипа его потомков

Исаев Василий
Институт биоинформатики
студент

Руководитель:
Татьяна Татарина
University of La Verne
Лаборатория вычислительной биологии

Цель

Получать информации об одном из родителей по геномам второго родителя и ребенка, например:

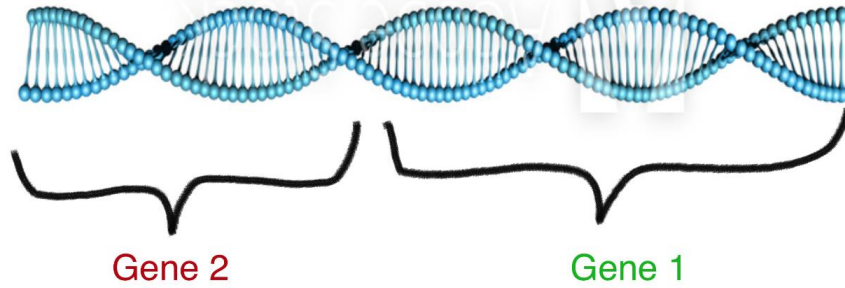
- Этническая группа
- Родственность (если ребенок появился в результате инцеста)
- Внешние особенности (вероятный цвет глаз)

Задачи

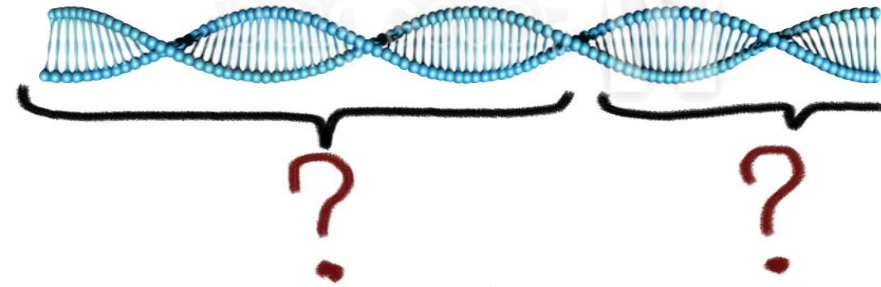
1. Разработка метода нахождения наиболее вероятного генотипа отца
2. Определения этнической группы
3. Тест на данных проекта “1000 геномов “
4. Улучшение качества предсказания с помощью гапlobлоков

Методы

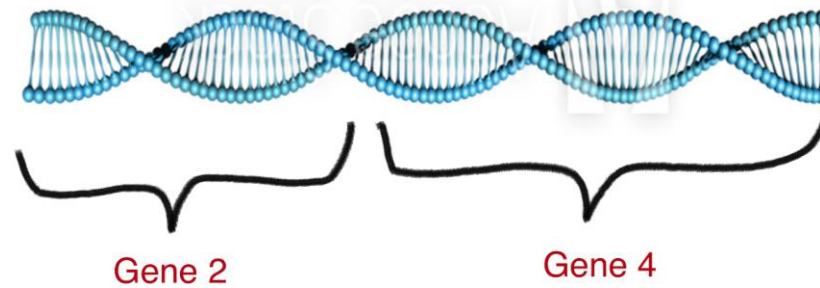
Known parent



Unknown parent



*Gene 4 in
Unknown parent*



Gene 2

Gene 1

Gene 2

Gene 4

Метод восстановления вероятностей генотипа одного родителя по генотипам второго родителя и ребенка.

Родитель1 Родитель2	AA	AB	BB
AA	$P(AA) = 1$	$P(AA) = 0,5$ $P(AB) = 0,5$	$P(AB) = 1$
AB	$P(AA) = 0,5$ $P(AB) = 0,5$	$P(AA) = 0,25$ $P(AB) = 0,5$ $P(BB) = 0,25$	$P(BB) = 0,5$ $P(AB) = 0,5$
BB	$P(AB) = 1$	$P(BB) = 0,5$ $P(AB) = 0,5$	$P(BB)=1$

Родитель1 – генотип известен
Родитель2 - генотип неизвестен

$$P(\text{Родитель2} \mid \text{Родитель1, ребенок}) = \frac{P(\text{Родитель2, Родитель1, ребенок})}{P(\text{Родитель1}) * P(\text{Родитель2})}$$

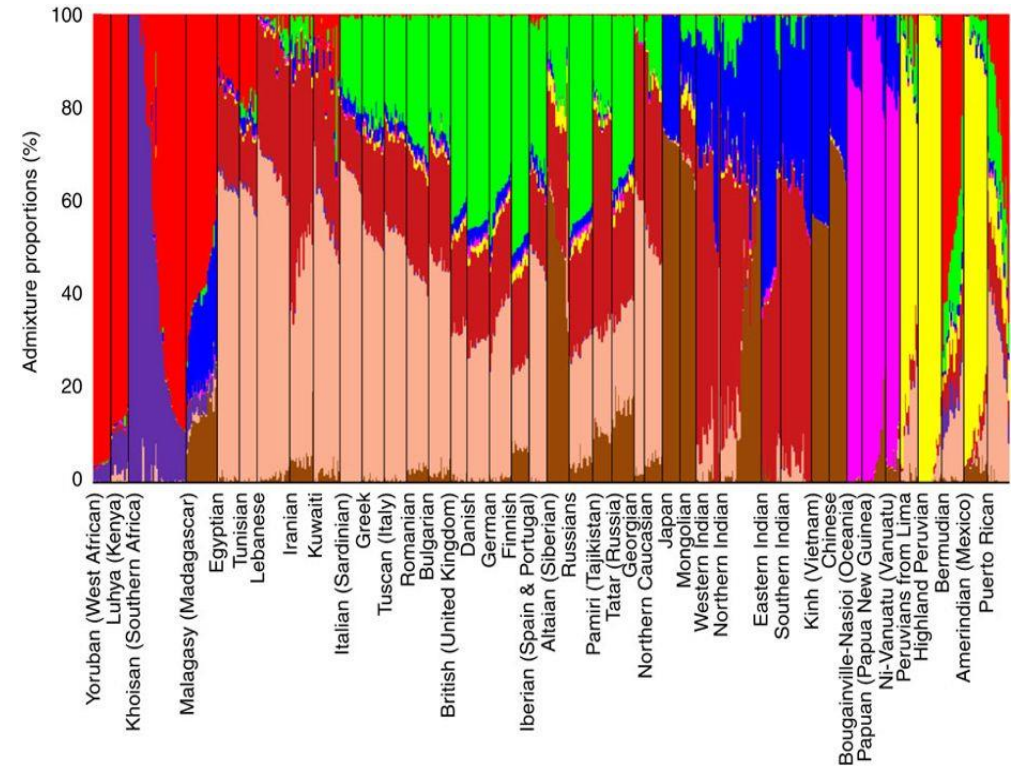
Этапы

- Знакомство со статьями по теме*
- Ознакомление с работой уже написанного алгоритма
- Подбор(генерация) геномов для тестов
- Добавление фазирования(Beagle, PHASE) и сравнения по гапблокам
- Тестирование на подобранных геномах

Определение этнической принадлежности

- Получаем генотип неизвестного родителя
- Получаем admixture vectors из генотипа
- Geographic population structure analysis

статья: <https://www.nature.com/articles/ncomms4513>



Определение этнической принадлежности

1. Получение наиболее вероятного генотипа отца
(Основной скрипт)
2. Admixture - оценки максимального правдоподобия отдельных предков из множественных наборов данных генотипов (запуск с $K = 9$)
3. Geographic population structure analysis * - прогнозирование места происхождения человека (**struct2geno**)

*<http://software.genetics.ucla.edu/admixture/>

*<https://www.nature.com/articles/ncomms4513>

Прогресс

- Знакомство со статьями по теме* Готово
- Ознакомление с работой уже написанного алгоритма Готово
- Подбор(генерация) геномов для тестов Готово
- Pipeline для определения этнической принадлежности Готово
- Добавление фазирования(Beagle, PHASE) и сравнения по гаплоглокам В работе
- Тестирование на подобранных геномах Почти готово

Спасибо за внимание!