

# Отчёт: По изучению программного обеспечения MENDEL:

## Цель :

Проверка возможности использования пакета Mendel для гаплотипирования и IBD анализа различных родословных.

## Задача:

Проверка Mendel для гаплотипирования и IBD анализа для родословных:

- а) большого размера
- б) с петлями
- в) с большим количеством маркеров

## Общие возможности Mendel:

Программное обеспечение Мендель выполняет статистический анализ (основанный на вероятностном подходе), решающий целый ряд генетических проблем. Пакет включает все общие и ряд новых статистических генетических тестов. Для каждой опции предопределено, какой алгоритм в ней используется: Эльстона-Стюарта, Ландера - Грина-Кругляка или какой-нибудь более эффективный.

Пакет включает следующие опции:

1. Установление порядка и расположения маркеров на карте вдоль хромосомы
2. Локализация гена с помощью параметрического анализа сцепления
3. Локализация гена с помощью непараметрического анализа сцепления
4. Гаплотипирование
5. Идентификация потенциальных генетических ошибок
6. Тестирование аллельной ассоциации TDT- тестом
7. Определение коэффициента родства, основанное на маркерных данных
8. Оценка генетического риска
9. Тестирование отцовства или более дальнего родства
10. Оценка частоты аллеля, используя либо родословные, либо случайные выборки
11. Оценка пенетрантности
12. Идентификация отклонений от равновесия Харди- Вайнберга и равновесия по сцеплению
13. Симулирование генетических данных, используя метод gene dropping
14. Локализация полигенных локусов количественных признаков (QTLs), используя анализ компонент дисперсии
15. Анализ ассоциаций количественных признаков

## Новые характеристики последней версии, используемой для проверки:

Версия 5.5.2 была выпущена 18 августа 2004 года.

Версия 5.5.2, в отличие от версии 5.5, не содержит недостатки, сообщенные Даниелем Уексом и Анной - Луизой Леутенегер. Эти исправления не изменили какие-нибудь результаты предыдущих исследований, основанных на версии 5.5, они только увеличили скорость анализа и позволили работать с Менделем на широком множестве данных, в особенности на огромных данных. Также новая версия 5.5.2 включает метод оценки p-value, основанный на пермутации, для тестирования ассоциации количественных признаков в опции 20 (QTL ассоциации).

## ЧТО ОСТАЛОСЬ НЕЯСНЫМ ПОСЛЕ ЧТЕНИЯ ОПИСАНИЯ

### 1. Диагностика ошибок

В перечне тестируемых программой ошибок не содержалось информации о тестировании следующих ошибок :

- компиляция программы при отсутствии одного из входных файлов (Locus, Map, Ped.(формат Fortran90));
- компиляция программы при ошибках записи родословной (повтор особи; повтор особи с разными родителями; обозначение пола, неописанное в файле Control );
- компиляция программы при стандартных генетических ошибках (особь с генотипом, не описанным в файле Locus; особь с “невозможным” генотипом (в смысле, что на основе родительского генотипа невозможно построить данный генотип) );

#### Что конкретно делал:

- 1) В директории, в которой работал Mendel, по одному удалял входные файлы.
- 2) Во входном файле Ped (родословной) дублировал особь с тем же номером; с тем же номером, но разными родителями; мужской пол особи в файле Ped обозначал символом “М”, тогда как в файле Control мужской пол записан как “1”.
- 3) В файле Ped задавал генотип особи, неописанный во входном файле Locus; особи в файле Ped задавал генотип, который невозможно построить на основе родительского генотипа (например, у обоих родителей генотип a/a, а у потомка-a/b).

**Вывод:** Во всех трёх случаях Mendel типировал ошибки и выдавал в выходном файле Mendel.out тип ошибки (недостающий входной файл, номер дублированной особи и т.д. )

### 2. Структура и размер родословных

Изучение документации к последней версии Mendel 5.5.2. позволило определить алгоритм, используемый при реализации разных опций. Как известно, разные алгоритмы отличаются друг от друга по требованиям к структуре и размеру родословных и числу анализируемых маркеров.

Option	Algorithm
Mapping markers	Haldane's Poisson mapping model
Location scores	Elston-Stuart
Haplotyping	Lander-Green-Kruglyak
Allele Frequencies	Quick Expectation-Maximization algorithm
Gamete	
Competition	Bradley-Terry method
Kinship matrices	Lander-Green-Kruglyak
Genetic Equilibrium	H-W eq-m
Cases and	
Controls	Fisher's test
Pedigree Trimming	Lander-Green-Kruglyak
TDT	McNemar's test
	Depending on kind of pedigrees input (most likely
NPL	Lander-Green-Kruglyak)

#### Вывод:

Опции Haplotyping, Kinship matrices, Pedigree Trimming используют алгоритм Ландера-Кругляка. Следовательно, они имеют ограничения на размер родословной (особенно на число потомков) и мало чувствительны к структуре родословной и числу маркеров. Поскольку в описании пакета Mendel не указаны пределы ограничения, целесообразно проверить эти опции на родословных различной структуры и небольших размеров с разным числом маркеров и оценить время, затрачиваемое программой.

### Тестирование программы по опциям Haplotyping и Kinship\_matrices :

В исследовании рассматривалось восемь родословных различной структуры (см. табл):

1- Родословная 1 из шести особей содержала трёх основателей и трёх потомков. Для неё проводился анализ как одного маркера, так и двух, трёх и четырёх.

2- Родословная 2 из девяти особей включала четырёх основателей, пять потомков и одну петлю. Для неё проводился анализ от 1-го до 4-х маркеров.

3- Родословная 3, состоящая из 24 особей: 12 основателей и 12 потомков. Родословная а) содержала две петли, б) не содержала петель. Для данной родословной анализ проводился от 1-го до 4-х маркеров.

4- Родословная 4, состоящая из 25 особей: 5 основателей и 20 потомков. Родословная содержала восемь петель. Для неё проводился анализ 3-х маркеров.

5- Родословная 5, состоящая из 45 особей: 6 основателей и 39 потомков. Родословная содержала двенадцать петель. Для неё проводился анализ 3-х маркеров.

6- Родословная 6, состоящая из 74 особей: 7 основателей и 67 потомков. Родословная содержала шестнадцать петель. Для неё проводился анализ 3-х маркеров.

7- Моделированная родословная 7, состоящая из 100 особей: 8 основателей и 92 потомка. Родословная содержала восемнадцать петель. Для неё проводился анализ 3-х маркеров.

8- Родословная 8 в 50 особей состояла из 11 основателей и 39 потомков. Родословная содержала 4 петли. Для неё проводился анализ 2-х маркеров

Табл. Родословные, исследуемые программой Mendel

Родословная	Общее число	Число основателей	Число потомков	Число петель
1	6	3	3	0
2	9	4	5	1
3а	24	12	12	2
3б	24	12	12	0
4	25	5	20	8
5	45	6	39	12
6	74	7	67	16
7	100	8	92	18
8	50	11	39	4

Входные данные содержались в четырех файлах:

Control:

В нём содержались имена всех входных файлов; опция, с которой пользователь будет анализировать данные; обозначение пола.

Locus:

В этом входном файле содержалась информация о всех маркерах, используемых в анализе: на какой хромосоме расположен маркер (аутосома, X-хромосома); частота каждого аллеля для каждого маркера.

Map:

В этом входном файле содержалась информация о расстояниях между маркерами в сМ.

Ped:

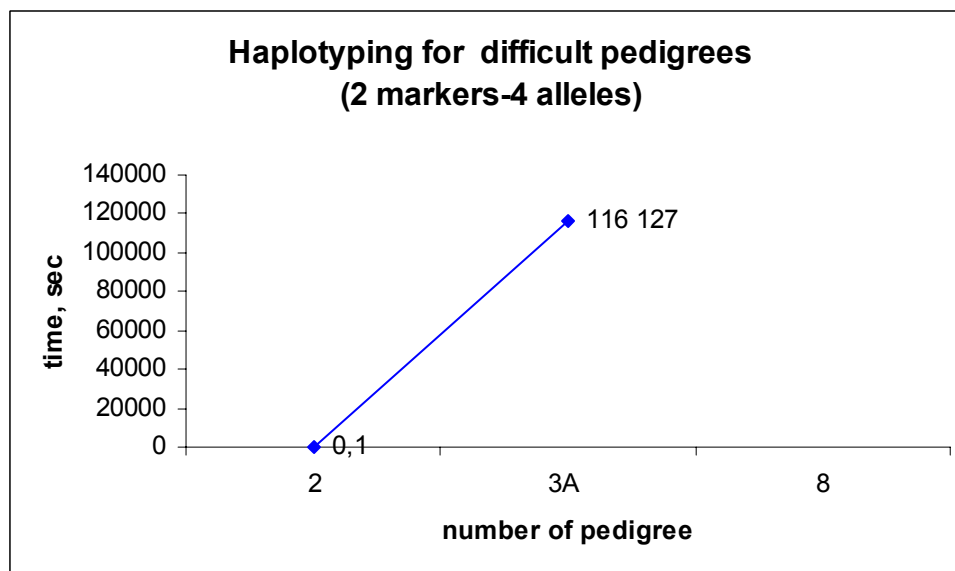
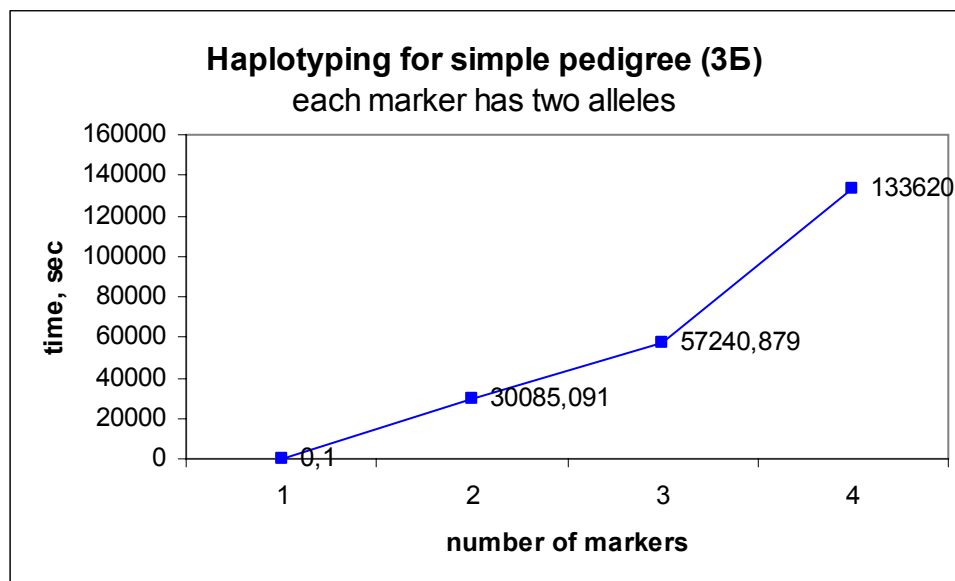
В этом входном файле содержалась родословная: её структура; пол каждой особи в родословной; известные генотипы членов родословной.

Для каждой родословной моделировались генотипы: случайные для основателей и согласно законам Менделя для потомков.

Оценивалось время, затрачиваемое программой для подсчёта коэффициента родства и оценки гаплотипов для этих родословных с разным количеством маркеров (табличные данные в [Opisanie v tabl.xls](#)). На основе полученных временных оценок строились графики зависимости.

Проверка опции `Kinship_matrices` показала, что Mendel способен анализировать родословные сложной структуры (до 100 особей) с множеством петель (до 18) и 3-мя полиаллельными маркерами за считанные секунды. Эта опция характеризует коэффициент родства по родословной, а изменение времени при изменении числа маркеров связано с увеличением времени считывания входного файла.

Проверка же опции `Haplotyping` выявила ограничения Mendel на структуру родословной, число маркеров и число аллелей:



**Вывод:** Опция Mendel `Haplotyping`, использующая алгоритм Ландера-Кругляка, может эффективно работать на родословных небольшого размера (24 особи) любой структуры (до 3 петель) с 4-мя диаллельными маркерами. Родословные большего размера, по крайней мере, начиная с 50 особей, с различным числом диаллельных маркеров (от 2-х до 4-х) Mendel не обработал и за трое суток.

Ещё одной важной особенностью или даже проблемой, возникшей в процессе проверки пакета Mendel, являлось нехватка оперативной памяти компьютера.

