

Gency (Genotype frequency):

компьютерная программа для расчёта вероятности случайного совпадения
комбинированных генотипов при исследованиях ДНК

Руководство пользователя

Версия программы: 3.1

Версия руководства 1.0



*Настоятельно рекомендуется внимательно прочитать это руководство
перед началом работы с программой.*

Оглавление

СПИСОК СОКРАЩЕНИЙ	3
НАЗНАЧЕНИЕ И КРАТКОЕ ОПИСАНИЕ ПРОГРАММЫ	5
УСЛОВИЯ ИСПОЛЬЗОВАНИЯ ПРОГРАММЫ	7
РЕДАКЦИИ ДОКУМЕНТА	7
СИСТЕМНЫЕ ТРЕБОВАНИЯ	8
ЗАПУСК ПРОГРАММЫ	8
ПАПКА ПРОГРАММЫ	9
ПЕРСОНАЛЬНЫЙ КЛЮЧ АВТОРИЗОВАННОГО ПОЛЬЗОВАТЕЛЯ	10
КАТАЛОГ <i>GENCY_LIB</i> : СЛУЖЕБНЫЕ ФАЙЛЫ JAVA	11
КАТАЛОГ <i>DATA</i> : СЛУЖЕБНЫЕ ФАЙЛЫ ДЛЯ ИСПОЛЬЗУЕМЫХ ЛОКУСОВ	12
КАТАЛОГ <i>REFERENCES</i> : ЛИТЕРАТУРНЫЕ ПЕРВОИСТОЧНИКИ	15
РАБОТА С ПРОГРАММОЙ: ВЫБОР ЭТАЛОННОЙ ПОПУЛЯЦИИ И ВВОД ГЕНОТИПОВ	16
РАБОТА С ПРОГРАММОЙ: РАСЧЁТ ПОПУЛЯЦИОННЫХ ЧАСТОТ ГЕНОТИПОВ	17
РАБОТА С ПРОГРАММОЙ: РАСЧЁТ ЧАСТОТЫ КОМБИНИРОВАННОГО ГЕНОТИПА	18
СОХРАНЕНИЕ РЕЗУЛЬТАТОВ И ЗАВЕРШЕНИЕ РАБОТЫ С ПРОГРАММОЙ	20
ИСПОЛЬЗОВАНИЕ ФАЙЛА РЕЗУЛЬТАТОВ В ТЕКСТЕ «ЗАКЛЮЧЕНИЯ ЭКСПЕРТА»	21
РАБОТА С ПРОГРАММОЙ: ИСПОЛЬЗОВАНИЕ РАЗЛИЧНЫХ ЭТАЛОННЫХ ЧАСТОТ АЛЛЕЛЕЙ	22
АДЕКВАТНЫЕ ПОПУЛЯЦИОННЫЕ ДАННЫЕ ОТСУТСТВУЮТ	22
«НЕВЫЯВЛЕННЫЕ» АЛЛЕЛИ С НУЛЕВОЙ ЧАСТОТОЙ ВСТРЕЧАЕМОСТИ	22
РАСЧЁТ <i>RMP</i> : СТАНДАРТНАЯ ПРАКТИКА В ДНК-ИССЛЕДОВАНИЯХ	23
ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ: АПРИОРНЫЕ УСЛОВИЯ, О КОТОРЫХ СЛЕДУЕТ ПОМНИТЬ	24
НЕЗАВИСИМОСТЬ ЛОКУСОВ, НЕРАВНОВЕСИЕ ПО СЦЕПЛЕНИЮ И ХРОМОСОМНАЯ ЛОКАЛИЗАЦИЯ	25
КАЧЕСТВО ПЕРВИЧНЫХ ПОПУЛЯЦИОННЫХ ДАННЫХ: ИНДЕКС ПСАЧ и РХВ	27
SNP-, INDEL- и ЛОКУСЫ «ГЕНЕТИЧЕСКОЙ ПРЕДРАСПОЛОЖЕННОСТИ»	29
ЛОКУСЫ ДЛЯ УСТАНОВЛЕНИЯ ПОЛОВОЙ ПРИНАДЛЕЖНОСТИ И МАРКЁРЫ ХРОМОСОМЫ Y	31
ЛОКУСЫ, РАСПОЛОЖЕННЫЕ НА ХРОМОСОМЕ X ЧЕЛОВЕКА	32
ДОРАБОТКА ПО	33
РЕШЕНИЕ ЗАТРУДНЕНИЙ	34
СПИСОК ЛИТЕРАТУРЫ	35
БЛАГОДАРНОСТИ	38
ОБРАТНАЯ СВЯЗЬ	38

Список сокращений

БД – база данных

ДНК – дезоксирибонуклеиновая кислота

ЗЭ – заключение эксперта

мтДНК – митохондриальная ДНК

но (н.о.) – нуклеотидное основание (ДНК)

НС – неравновесие по сцеплению

ОП – отношения правдоподобия

ОС – операционная система

ПДРФ – полиморфизм длины рестрикционных фрагментов

ПК – персональный компьютер

пн (п.н.) – пара нуклеотидов (ДНК)

по (п.о.) – пара оснований (ДНК)

ПО – программное обеспечение

ПСАЧ © – полная сумма аллельных частот

ПЦР – полимеразная цепная реакция

РВ-ПЦР – полимеразная цепная реакция в реальном времени

РП – руководство пользователя

РФ – Российская Федерация

РХВ – равновесие Харди-Вайнберга

РЦ СМЭ – Российский Центр судебно-медицинской экспертизы

снип (СНИП, СНиП) – точечный (точковый, однонуклеотидный) полиморфизм, не самая удачная русскоязычная производная от SNP

США – Соединённые Штаты Америки

BLAT – BLAST-like alignment tool, программа для выравнивания (локализации) нуклеотидных последовательностей относительно референсного генома

bp – base pair(s), пара оснований (ДНК)

CODIS – Combined DNA Index System, дословно – комбинированная система ДНК-индексов. База данных генетических профилей на уровне 13 микросателлитных локусов, разработанная и используемая в ФБР (США), а также в Канаде

CKI – combined kinship index, комбинированный индекс родства

CPI – combined paternity index, комбинированный индекс отцовства

DIP – deletion / insertion polymorphism, полиморфизм типа «делеция-вставка», синоним InDel

DNA – deoxyribonucleic acid, ДНК

DOI – digital object identifier, цифровой идентификатор объекта. Путь к документу (например, статье) в сети Интернет.

ESS – European standard set (of loci), дословно – Европейская стандартная панель (локусов). Набор из нескольких микросателлитных локусов, используемых в Евросоюзе в качестве аналога CODIS.

HWE – Hardy-Weinberg equilibrium, равновесие Харди-Вайнберга

ID – identification, идентификация (личности)

InDel (INDEL) – insertion-deletion, полиморфизм типа «вставка-делеция», синоним DIP

ISFG – International Society for Forensic Genetics, международное общество судебной генетики

KI – kinship index, индекс родства

LR – likelihood ratio, отношение (-я) правдоподобия

MAF – minimum allele frequency, минимальная частота аллеля (в популяции)

MB, Mbyte – megabyte, мегабайт, единица измерения количества информации

Mb – mega base pairs, в контексте структуры ДНК – один миллион пар нуклеотидных оснований

mtDNA – mitochondrial DNA, митохондриальная ДНК

P – probability, вероятность (в контексте расчётов – апостериорная вероятность)

PI – paternity index, индекс отцовства

PCR – polymerase chain reaction, ПЦР

PMID, PubMed ID – идентификатор статьи в Базе Данных биомедицинской литературы [PubMed](https://pubmed.ncbi.nlm.nih.gov/).

pp – prior probability, априорная вероятность

RMP – random match probability, вероятность случайного совпадения

RT-PCR – real time polymerase chain reaction, ПЦР в реальном времени

SNP – single nucleotide polymorphism, точечный полиморфизм, снип

STR(s) – short tandem repeats, короткие tandemные повторы, микросателлиты

VNTR(s) – variable number tandem repeats, переменное число tandemных повторов, минисателлитные локусы

WT – wild type, аллель «дикого типа», то есть «нормальный», не мутантный

Назначение и краткое описание программы

Программа предназначена для расчёта **вероятности случайного совпадения комбинированных генотипов** для анализируемых биологических образцов при ДНК-исследованиях и может быть использована в экспертной практике судебно-медицинских, криминалистических, клинико-диагностических, гистологических и других лабораторий, занимающихся идентификацией происхождения биологических образцов (идентификацией личности).

Программа обеспечивает выполнение следующих функций:

- ✓ выбор эталонной популяции для расчётов;
- ✓ ввод генотипов для одного обследованного лица (объекта) по отдельным локусам;
- ✓ расчёт популяционной частоты встречаемости выявленного генотипа по каждому локусу;
- ✓ расчёт популяционной частоты встречаемости комбинированного генотипа по нескольким выбранным локусам;
- ✓ сохранение результатов работы в отдельном файле;
- ✓ возможно изменение конфигурации служебных файлов по усмотрению пользователя, в том числе создание опорных (эталонных) частот аллелей для любой локальной популяции.

Для термина **вероятность случайного совпадения комбинированного (-ых) генотипа (-ов)** в русскоязычной литературе используются следующие синонимы: *вероятность случайного совпадения, вероятность случайного совпадения генетических признаков, вероятность случайного совпадения генотипов*. В англоязычной литературе – *chance of a random match, DNA profile population frequency, DNA profile probability, genotype frequency, probability of a random match, **random match probability (RMP)***.

В своей практике мы используем эту программу в том числе для численной оценки значимости выявленных совпадений генотипов при установлении индивидуальной принадлежности гистологического материала. Такая услуга часто необходима для исключения перепутывания гистологических образцов, когда пациент с подозрением на болезнь получает «чужой» диагноз (например, рак). Более подробно про «ошибку перепутывания» можно почитать у наших [партнёров](#).

На первом этапе работы с программой осуществляется ввод установленных для исследованного образца генотипов по *отдельным локусам*. На втором этапе, опять же по отдельным локусам (по произвольному числу локусов), рассчитываются популяционные частоты Q_i выявленных генотипов по следующим формулам:

$$\begin{aligned} p^2 & \text{ – для гомозиготных генотипов,} \\ 2pq = 2 \cdot p \cdot q & \text{ – для гетерозиготных генотипов,} \end{aligned}$$

где p и q – соответствующие частоты выявленных аллелей.

На завершающем этапе работы рассчитывается *частота комбинированного генотипа* (*random match probability, RMP*) путём перемножения популяционных частот установленных генотипов Q_i для всех (или произвольного числа) исследованных локусов, согласно *правилу перемножения, product rule*:

$$RMP = Q_{\text{первый локус}} * Q_{\text{второй локус}} * \dots * Q_{\text{последний локус}}$$

Дополнительно величина RMP численно выражается и как обратная:

$$\frac{1}{RMP} = 1 / RMP$$

в терминах *один из NNN (например, двух миллиардов ста миллионов пятисот тысяч трехсот) человек*.

Полученные результаты расчётов сохраняются в отдельном файле результатов формата *Microsoft Excel* и могут быть непосредственно использованы при оформлении *Заключения Эксперта (ЗЭ)*.

Условия использования программы

Программа защищена Гражданским Кодексом РФ, международными соглашениями об авторских правах, а также другими законами и договорами, регулирующими отношения авторского права.

Программа не продаётся, а предоставляется в пользование по лицензии. Принцип лицензирования – стандартный для программного обеспечения, **одна копия на один компьютер**.

При запуске продукта ознакомьтесь с условиями Лицензионного Соглашения (ЛС) в стартовом окне программы и для продолжения работы нажмите кнопку **«Принимаю»**. Нажатие этой кнопки означает полное согласие Лицензиата (пользователя) со всеми пунктами ЛС. В случае несогласия пользователя с каким-либо пунктом ЛС следует нажать кнопку **«Не принимаю»** и отказаться от использования программы.

Лицензионное соглашение содержится в файле *licence.txt* и может изменяться только правообладателями программы в одностороннем порядке в любой момент времени. Уведомление Авторизованных Пользователей о внесённых изменениях в условия лицензионного соглашения осуществляется в произвольной форме, преимущественно по электронной почте. Продолжая использовать программу после внесения таких изменений, пользователь соглашается исполнять их требования в полном объёме.

Программа и настоящее Руководство пользователя содержат охраняемые авторским правом материалы, торговые знаки и другие объекты интеллектуальной собственности, принадлежащие сторонним поставщикам. Соответственно, сторонние поставщики владеют и сохраняют за собой все права собственности и права на интеллектуальную собственность в отношении этих объектов.

Редакции документа

Версия 1.0 (06.2018). Руководство пользователя в целом написано.

Системные требования

Программа написана в системе программирования *Java*. Минимальные системные требования: процессор 1,73 GHz, оперативная память 1,0 Gbyte, операционная система *Microsoft Windows (XP и выше)*. **Обязательно наличие на рабочем компьютере предустановленного пакета *Microsoft Office Excel (2003 и выше)*.**

Альтернативные офисные пакеты (*OpenOffice* и др.) не годятся.

Для работы с программой также необходимо установить на компьютер программное обеспечение *Java* (не ниже *Version 8, Update 40*). Бесплатное программное обеспечение *Java* при необходимости следует загрузить или обновить с сайта разработчика:

<https://java.com/download>

При использовании программы *Gency* на различных версиях *Microsoft Windows* могут возникнуть случаи некорректной работы программы. Более подробная информация на эту тему приведена в разделе **Решение затруднений** настоящего Руководства.

При использовании компьютеров с *Unix*-подобными операционными системами (*UNIX system-like*) необходима установка ОС *Microsoft Windows* либо второй системой, либо виртуальной машиной (*VirtualBox*, *VMware* и т.п.).

Запуск программы

Программа поставляется либо на *USB*-накопителе, либо по электронной почте. В обоих случаях zip-архив программы содержит все необходимые для работы файлы.

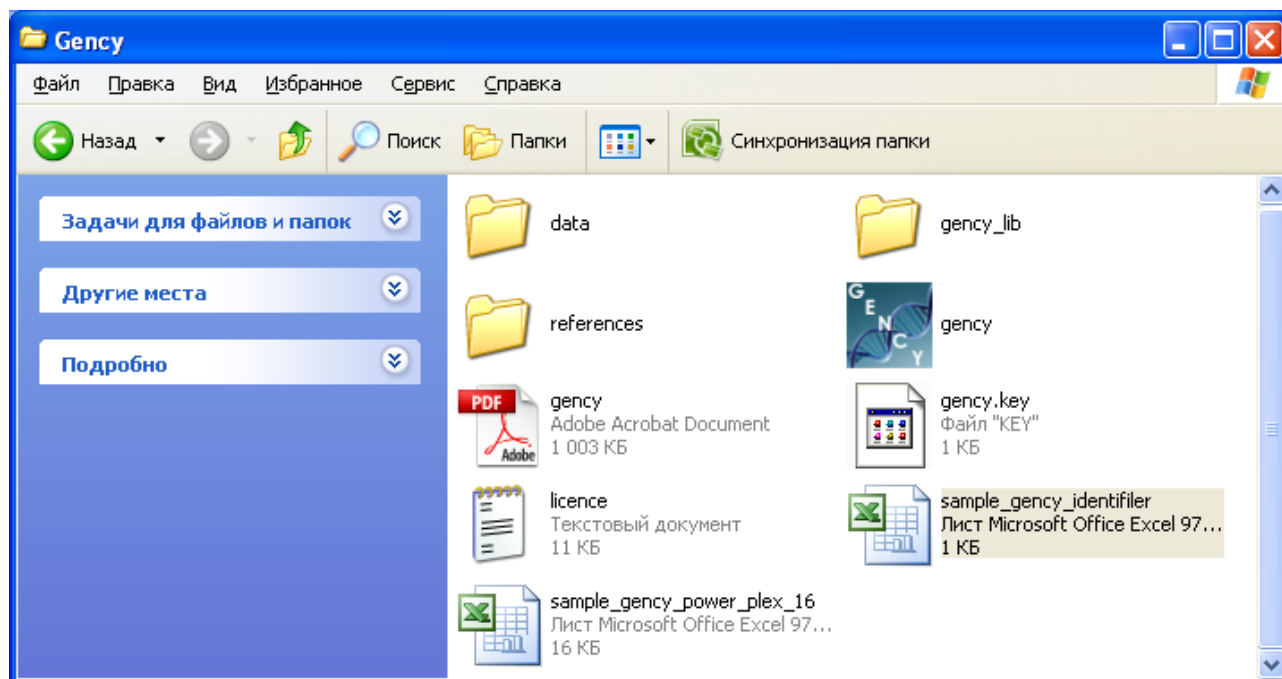
Для успешного запуска программы необходимо наличие в её каталоге трёх файлов – *gency.exe*, *gency.key* и *licence.txt*, а также двух подкаталогов: *data* и *gency_lib* со служебными файлами. Подробнее формат этих файлов и подкаталогов будет описан ниже, в соответствующих разделах.

Запуск программы осуществляется вызовом файла *gency.exe*.

Программа успешно запускается как с жёсткого диска ПК, так и с внешних *USB*-флеш-накопителей. Однако при запуске программы с внешнего диска, защищённого от записи, может возникнуть проблема с сохранением результатов работы программы (файла с расчётами). В этом случае будет выдано сообщение об ошибке.

Папка программы

На приведённом ниже рисунке изображена структура рабочей папки программы, которая включает все необходимые для работы файлы и каталоги:



Файлы *gency.exe*, *gency.key* и *licence.txt* являются исполняемым файлом программы, ключом авторизованного пользователя и текстом лицензионного соглашения, соответственно. Они обязательны для работы.

Файл *gency.pdf* является *Руководством пользователя* (Вы его сейчас читаете), из программы он вызывается кнопкой «*Справка*» и для работоспособности не критичен.

Файлы *sample_gency_identifiler.xls* и *sample_gency_power_plex_16.xls* являются примерами файлов результатов с генотипами и расчётами, они также необязательны для успешной работы. Эти файлы используются в настоящем Руководстве при описании работы программы.

В каталогах *data* и *gency_lib* находятся рабочие файлы программы, и оба этих каталога являются обязательными для работы. Каталоги обновляемые.

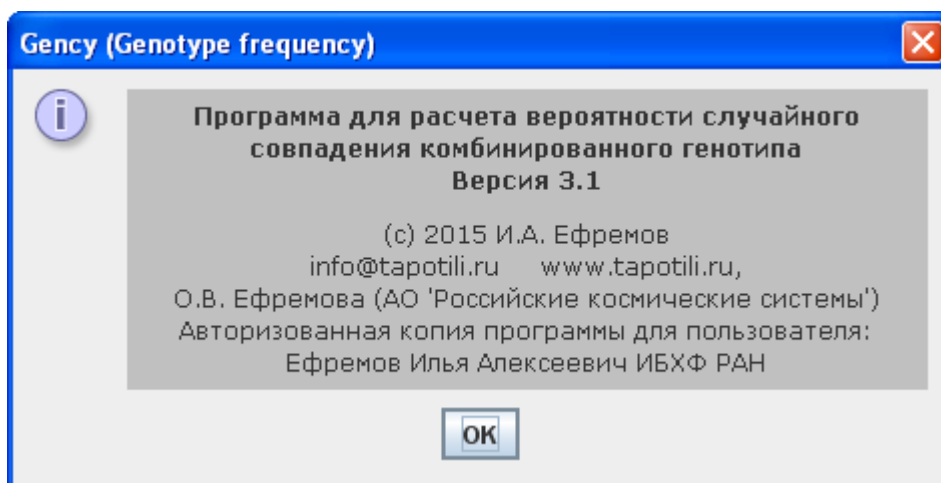
В каталоге *references* находятся файлы статей и приложений к ним, содержащие эталонные аллельные частоты для всех используемых локусов. Обновляемый каталог, является общим для программ *Pindex* и *Gency*, но для корректной работы он необязателен.

Персональный ключ авторизованного пользователя

Каждая авторизованная копия *Gency* защищена персональным ключом, файл которого *gency.key* обязательно должен находиться в одной папке с исполняемым файлом *gency.exe*.

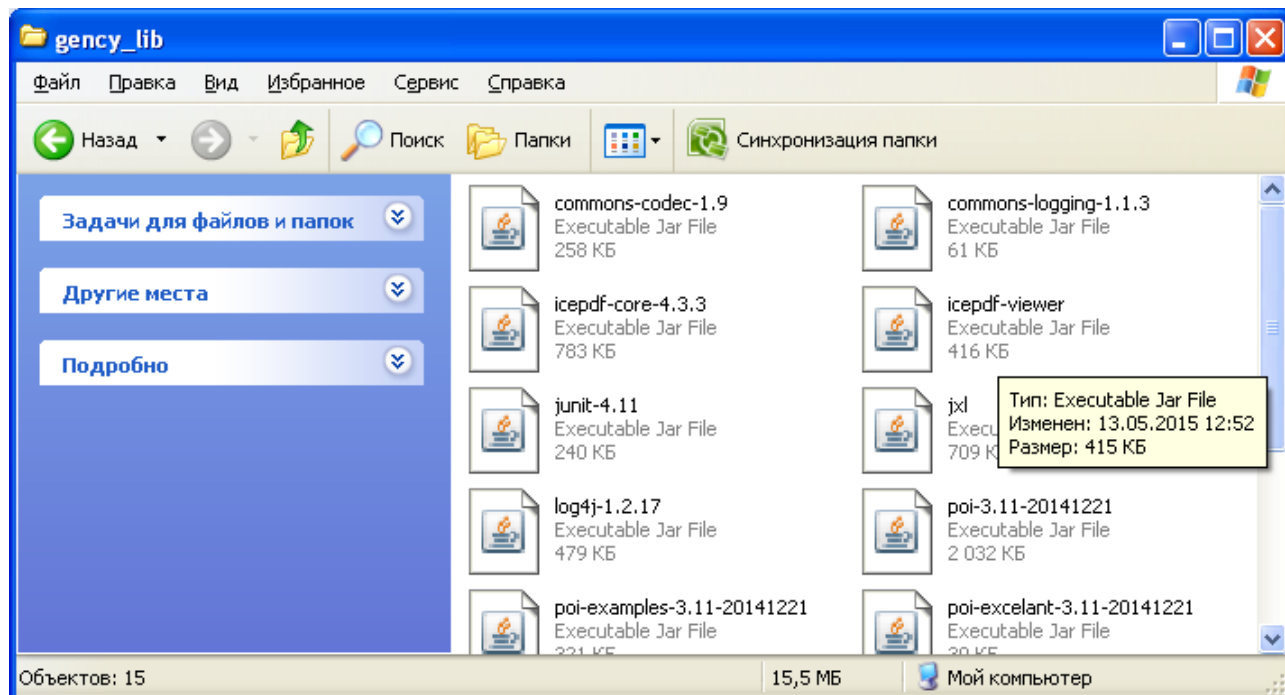
Этот ключ никак не защищает от многократного копирования программы авторизованным пользователем на произвольное число ПК. Однако наличие такого ключа в определённой степени даёт возможность разработчикам следить за возможным несанкционированным распространением программы, поскольку в нём зашифровано имя авторизованного пользователя.

Узнать имя авторизованного пользователя и дополнительную информацию можно, нажав кнопку «*О программе*», например:



Каталог *gency_lib*: служебные файлы Java

В каталоге *gency_lib* находятся *jar-файлы* (от англ. *Java ARchive*), содержащие файлы отдельных библиотек *JAR*:



Этот каталог **необходим** для успешной работы программы, помимо предварительно установленного программного обеспечения *Java*.

Содержание папки *gency_lib* обновляется для новых версий программы. Актуальный *zip*-архив этого каталога доступен на сайте (размер около 15 Mbyte):

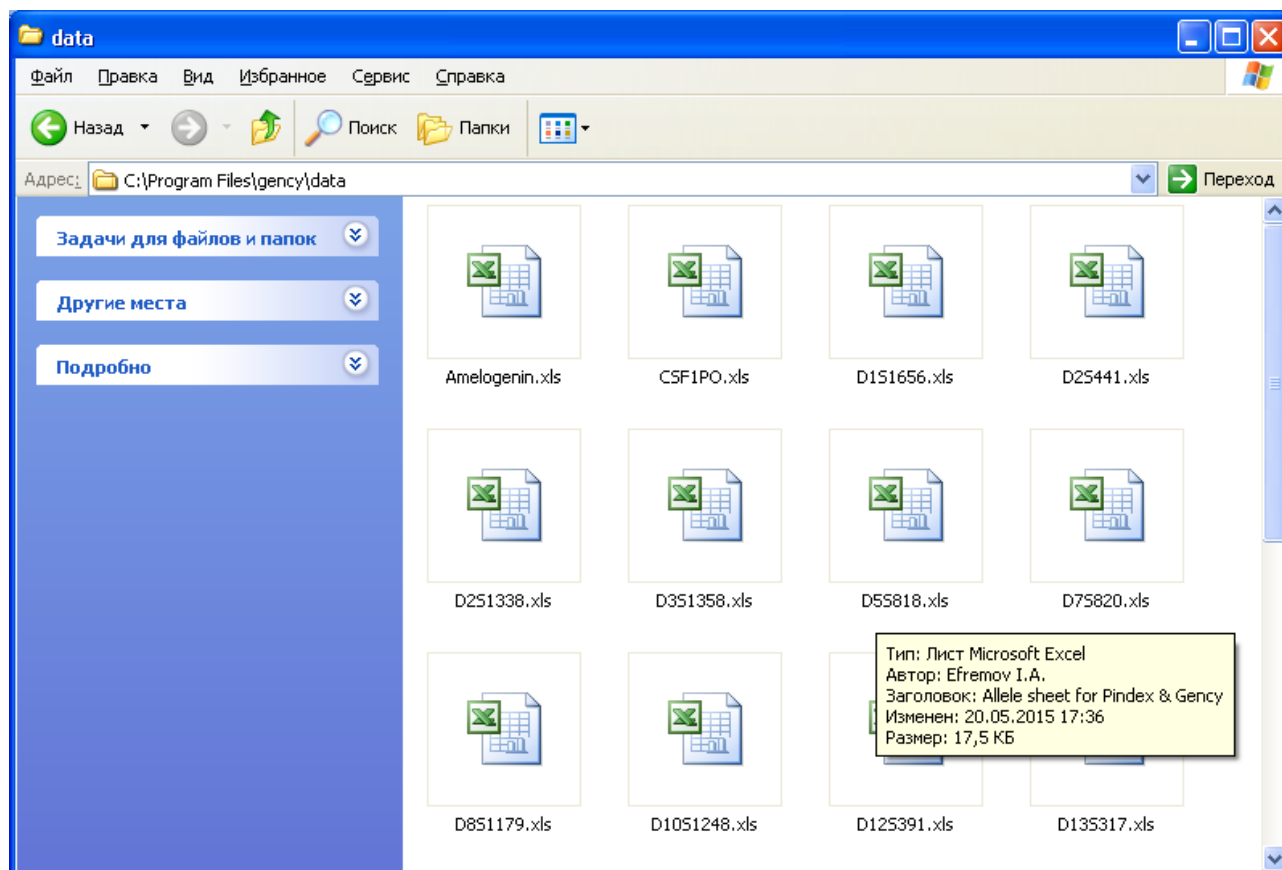
http://tapotili.ru/programs/gency_lib.zip

Для корректной работы программы в папке *gency_lib* не должно находиться никаких других (посторонних) файлов.

Каталог *data*: служебные файлы для используемых локусов

Этот каталог является общим для программ *Gency* и *Pindex*. Соответственно, описание каталога и структуры находящихся в нём служебных файлов подробно изложено в руководстве для *Пиндекса*. Ниже даны только основные существенные пояснения.

Для успешной работы программы в одной папке с исполняемым файлом *gency.exe* должен находиться непустой каталог *data*. В нём размещены служебные файлы формата *Microsoft Excel*, содержащие данные об эталонных популяционных частотах аллелей по используемым локусам:



Число рабочих файлов в каталоге *data* соответствует числу используемых пользователем локусов, по принципу *один файл – один локус*. Общее число используемых локусов не ограничено.



Для корректной работы программы в директории *data* **не должно** находиться других (посторонних) файлов формата *Microsoft Excel*.

Для максимальной совместимости программы с любыми конфигурациями *Microsoft Windows* все служебные файлы для отдельных локусов сохранены в «старом» формате «Книга Excel 97-2003», с расширением **xls*.

Названия служебных файлов соответствуют названиям локусов. Каждый файл содержит четыре листа, в которых приводятся частоты аллелей для четырёх эталонных популяций: **русские** (листы *имя_файла-rus*), **европеоиды США** (листы *имя_файла-white*), **негры США** (листы *имя_файла-black*) и **любая локальная популяция** (листы *имя_файла-other*).

Во всех служебных файлах **заливка** красным цветом использована для ячеек с нулевыми значениями частот аллелей в исследованных выборках.

Частоты аллелей для **русской популяции**, представленные в столбцах «**B**» в первых листах соответствующих файлов (листы *имя_файла-rus*), являются **консервативной оценкой** для европеоидного населения России.

Такая оценка была выполнена следующим образом:

- 1) при сравнительном анализе различных популяционных данных для населения России сделан выбор «опорной» печатной работы из нескольких доступных, по качеству представленных данных, объёму и однородности исследованной выборки;
- (i) первичные популяционные данные в «опорной» работе проверены на соответствие *ПСАЧ*® (полная сумма аллельных частот) и *PXB* (равновесие Харди-Вайнберга) – в отдельном разделе ниже это будет обсуждаться более подробно;
- (ii) частоты аллелей из выбранной «опорной» печатной работы пересчитаны с учётом рекомендаций, изложенных в работе *Gjertson et al., 2007* (PMID: [19083766](#)). При этом все нулевые значения частот аллелей из исследованных выборок получили минимальные аллельные частоты (*Minimum Allele Frequency, MAF*), уже с ненулевыми значениями;
- (iii) все полученные значения частот аллелей дополнительно округлены с точностью до трёх знаков после запятой. В единичных случаях такое округление выполнено до двух знаков: например, для таких «экзотических» локусов как *DIS111*, *CYAR04*.

Для удобства работы для большинства локусов использованы следующие дополнительные обозначения аллелей: <, >, **новый**, 0, **WT** и некоторые другие. Например, обозначение «**новый**» удобно использовать в случаях, когда по каким-либо причинам невозможно точно установить выявленный аллель или же этот действительно новый (относительно редкий) аллель ещё отсутствует в таблице для анализируемого локуса.

Обновления служебных файлов производятся по мере накопления популяционных данных. Дополнительно, при выявлении редких новых и/или «промежуточных» аллелей по отдельным локусам они также добавляются в соответствующие файлы.

К настоящему времени служебные файлы созданы более чем для **60** разных локусов генома человека. Файлы созданы для всех микросателлитных локусов, входящих в современные международные базы данных (Интерпол, *CODIS*, *ESS* и т.д.).



Конфигурирование программы возможно для работы с такими полиморфными маркерами, как *VNTRs*, *STRs*, *INDELs*, *SNPs*, под любые новые локусы.

Возможно создание служебных файлов программы для отдельных видов животных, также под любые локусы.

При работе программа считывает численные значения только из первых двух столбцов каждого листа, дополнительная служебная информация в остальных ячейках программой игнорируется. Частоты аллелей считываются программой из столбцов «*B*» с точностью до трёх знаков после запятой.



Четвёртые листы каждого файла (листы *имя_файла-other*) предназначены для задания частот аллелей в актуальной пользователю локальной популяции. По умолчанию всем аллелям в этих листах приписаны частоты, равные **0,500**, и эти ячейки выделены красной **заливкой**. Округление значений в этих ячейках установлено на точность в три знака после запятой.

Приглашаем к сотрудничеству региональные лаборатории для создания эталонных частот аллелей для локальных популяций.

Для максимально быстрой загрузки и работы программы рекомендуется хранить в директории *data* файлы только для тех локусов, которые постоянно используются в экспертной практике. Например, если рутинно исследуются 15-20 одних и тех же локусов, то можно оставить в директории *data* соответствующие этим локусам 15-20 файлов, а остальные файлы удалить или переместить в архив (в другой каталог).

Каталог references: литературные первоисточники

Этот каталог также является общим для программ *Pindex* и *Gensu* и содержит статьи и дополнительные материалы к ним (таблицы и прочие приложения в виде отдельных файлов), которые были использованы для создания служебных файлов по отдельным локусам с частотами аллелей в различных эталонных популяциях. Наличие или отсутствие этого каталога на работоспособность программы никак не влияет. Каталог обновляемый, по мере поступления новых популяционных данных. Файлы названы по номеру *DOI* или *PMID* (*PubMed ID*), фамилии первого автора и году публикации.

Каждый рабочий лист каждого служебного файла содержит краткую информацию об источнике соответствующих популяционных данных с отсылкой к соответствующей публикации. Например, для локуса *D18S51* в листе *d18s51-white* указано следующее:

Частоты аллелей для европеоидного населения США (349 неродственных человек, по данным Budowle et al., 2011, PMID: 20878415; ПСАЧ = 1,000)

Соответственно, для европеоидов США по этому локусу в программе были использованы частоты аллелей из работы *Budowle et al., 2011* (PMID: [20878415](#)). В папке *references* находится соответствующий файл: **PMID_20878415_Budowle_etal_2010.pdf**. В данном случае год публикации указан как 2011 для печатной версии, но как 2010 – для электронного варианта.

Работа с программой: выбор эталонной популяции и ввод генотипов

После запуска программы нужно выбрать в выпадающем меню окна «Популяция» нужную эталонную популяцию. Для расчётов программа будет адресоваться к соответствующим частотам аллелей:

Генотип

Аллель 1 Аллель 2

Выбрать

Рассчитать Перемножить Очистить Сохранить

Частота комбинированного генотипа, random match probability (RMP) Один из Очистить

Затем следует выбрать в выпадающем меню окна «Локус» нужный локус и указать для него установленный генотип. После нажатия кнопки «Выбрать» в рабочем поле программы появятся название локуса, установленный генотип (в формате *Аллель 1*, *Аллель 2*) и соответствующие частоты аллелей (p и q) для выбранной эталонной популяции. Пример для локуса *D8S1179* на рисунке ниже:

Популяция

Локус

Генотип

Аллель 1 Аллель 2

Выбрать

Локус	Аллель 1	Частота аллеля 1 (p)	Аллель 2	Частота аллеля 2 (q)	Формула	Частота генотипа
D8S1179	10	0.054	14	0.225		

Частота комбинированного генотипа, random match probability (RMP) Один из Очистить

Далее нужно выбрать из списка следующий локус и повторить для него описанную процедуру (ввод генотипов—> кнопка «**Выбрать**» —> визуальная проверка).

Неверно введенные данные могут быть в любой момент удалены из рабочего поля программы. Для этого необходимо навести курсор мыши на строчку с ошибочными данными, нажать правую кнопку мыши и выбрать команду «**Удалить**».

Работа с программой: расчёт популяционных частот генотипов

После ввода генотипов по всем исследованным локусам на следующем этапе работы рассчитываются индивидуальные популяционные частоты генотипов по отдельным локусам.

Для этого в рабочем поле программы следует курсором мыши выделить интересующую строку и нажать кнопку «**Рассчитать**». Для выбора нескольких локусов (строк) следует во время выделения курсором дополнительно нажать и удерживать клавишу «**Ctrl**» на клавиатуре. Формулы и результаты расчётов частот генотипов для всех выбранных локусов (они выделяются подсветкой) будут выведены в рабочем поле программы, как это показано на рисунке ниже. В этом примере для расчёта частот генотипов были выбраны только два локуса из трёх, *D8S1179* и *D7S820*.

The screenshot shows the Gency software window with the following components:

- Left Panel:**
 - Популяция:** Dropdown menu set to "РУССКИЕ".
 - Локус:** Dropdown menu set to "D7S820".
 - Генотип:** Two dropdown menus for "Аллель 1" (set to 8) and "Аллель 2" (set to 12), with a "Выбрать" button below them.
- Main Table:**

Локус	Аллель 1	Частота аллеля 1 (p)	Аллель 2	Частота аллеля 2 (q)	Формула	Частота генотипа
D8S1179	10	0.054	14	0.225	$2 \cdot p \cdot q$	0.0243
D21S11	29	0.19	33.2	0.046		
D7S820	8	0.172	12	0.145	$2 \cdot p \cdot q$	0.04988
- Bottom Panel:**
 - Buttons: "Рассчитать", "Перемножить", "Очистить", "Сохранить".
 - Fields: "Частота комбинированного генотипа, random match probability (RMP)" and "Один из".
 - Button: "Очистить" next to the "Один из" field.

Таким же образом для всех интересующих локусов рассчитываются частоты генотипов. Кнопка «**Рассчитать**» срабатывает только в отношении выбранных в рабочем поле локусов.

При необходимости на любом этапе работы можно полностью очистить рабочий лист (кнопка «**Очистить**»), однако при этом все введённые данные и результаты расчётов будут утеряны. Будьте осторожны.

На любом этапе работы можно сохранить результаты в отдельном файле формата *Microsoft Excel* (кнопка «**Сохранить**»).



Текущая версия программы пока не позволяет загружать результаты расчётов из сохранённых ранее файлов.

Также не реализована функция импорта данных из текстовых файлов результатов с генетических анализаторов (*GeneMapper* и пр.). Соответственно, на кнопке «**Импорт данных**» стоит временная заглушка.

Работа с программой: расчёт частоты комбинированного генотипа

Для аутосомных локусов такой расчёт является **простым** и **полностью корректным** только в том случае, когда наблюдается независимое наследование аллелей (отсутствие сцепления между любыми локусами). В этом случае для расчёта популяционной частоты комбинированного генотипа (*RMP*) все значения частот генотипов по отдельным аутосомным локусам могут быть перемножены (*правило перемножения, product rule*):

$$RMP = Q_{\text{первый локус}} * Q_{\text{второй локус}} * \dots * Q_{\text{последний локус}}$$

Соответственно, на этом (заключительном) этапе работы пользователь выбирает те локусы, которые он собирается использовать для расчёта *RMP*. Для этого следует, удерживая клавишу «**Ctrl**» на клавиатуре, выделить в рабочем поле программы мышью все нужные локусы, при этом выбранные строчки выделяются подсветкой. После нажатия кнопки «**Перемножить**» результаты расчёта *RMP* и обратной величины ($\frac{1}{RMP}$) для всех выбранных локусов отобразятся в нижней части окна программы, как это показано на рисунке на следующей странице.

В этом примере использованы данные о генотипе обследованного лица, полный генетический профиль которого (по 15 аутосомным локусам) указан в файле *sample_gency_identifiler.xls*. Расчёт *RMP* выполнен только по первым 10 локусам, как это видно в рабочем поле программы на этом рисунке.

The screenshot shows the Gensy software interface. On the left, there are dropdown menus for 'Популяция' (set to 'РУССКИЕ'), 'Локус' (set to 'D19S433'), and 'Генотип' (with alleles '12' and '16'). A 'Выбрать' button is below. The main area contains a table with 7 columns: 'Локус', 'Аллель 1', 'Частота аллеля 1 (p)', 'Аллель 2', 'Частота аллеля 2 (q)', 'Формула', and 'Частота генотипа'. Below the table are buttons: 'Рассчитать', 'Перемножить', 'Очистить', and 'Сохранить'. At the bottom, there are input fields for 'Частота комбинированного генотипа, random match probability (RMP)' (showing 1.024431028944986E-14) and 'Один из' (showing 9.761516117193889E13), with an 'Очистить' button.

Локус	Аллель 1	Частота аллеля 1 (p)	Аллель 2	Частота аллеля 2 (q)	Формула	Частота генотипа
D8S1179	10	0.054	14	0.225	2*p*q	0.0243
D21S11	29	0.19	33.2	0.046	2*p*q	0.01748
D7S820	8	0.172	12	0.145	2*p*q	0.04988
CSF1PO	10	0.283	12	0.307	2*p*q	0.173762
D3S1358	15	0.254	15		p*p	0.064516
TH01	6	0.223	6		p*p	0.049729
D13S317	10	0.059	11	0.353	2*p*q	0.041654
D16S539	12	0.35	13	0.192	2*p*q	0.1344
D2S1338	19	0.146	19		p*p	0.021316
D19S433	12	0.079	16	0.046	2*p*q	0.007268

Частота комбинированного генотипа, random match probability (RMP) 1.024431028944986E-14 Один из 9.761516117193889E13 Очистить

Частота комбинированного генотипа составила **1.024431028944986E-14**, можно округлить это значение просто до 10^{-14} . То есть таким генотипом в среднем обладает один из **9.761516117193889E13** человек, также если округлить полученное программой значение до $9,76 \cdot 10^{13}$, то это примерно один из 97 триллионов 600 миллиардов человек.

При необходимости можно стереть рассчитанные значения RMP и $\frac{1}{RMP}$ (нижняя кнопка «*Очистить*»), выбрать другую комбинацию локусов и заново рассчитать RMP .

Сохранение результатов и завершение работы с программой

Указанные генотипы по всем локусам и результаты расчётов, включая RMP и $\frac{1}{RMP}$, можно сохранить в отдельном файле результатов формата *Microsoft Excel*: кнопка «**Сохранить**». Этот файл может быть непосредственно использован при оформлении ЗЭ.

Для окончания работы и выхода из программы следует нажать на всем хорошо знакомую кнопку в правом верхнем углу рабочего поля:



По окончании работы настоятельно рекомендуется открыть файл результатов уже в редакторе *Microsoft Excel* и окончательно проверить указанный комбинированный генотип. Если при проверке обнаружится, что какие-то аллели были введены неправильно, то **не следует** вносить изменения непосредственно в редакторе *Microsoft Excel*. Причина в том, что для гомо- и гетерозиготных генотипов используются разные формулы расчётов, а также каждый аллель имеет собственную частоту встречаемости в популяции. Поэтому при изменении даже одного аллеля изменится соответствующее значение популяционной частоты генотипа.

Для исправления ошибочных генотипов нужно снова открыть файл результатов в программе *Gency* и удалить строчки с ошибочными данными в рабочем поле (правая кнопка мыши -> команда «**Удалить**»). После этого следует заново выбрать нужный локус из списка и ввести по нему правильный генотип.

Использование файла результатов в тексте «Заключения эксперта»

Файл результатов в общих чертах выглядит следующим образом, как это показано на примере файла *sample_gency_power_plex_16.xls*:

А	В	С	Д	Е	Ф	Г
Локус	Аллель 1	Частота аллеля 1 (р)	Аллель 2	Частота аллеля 2 (q)	Формула	Частота генотипа
D3S1358	15	0.254	15		p^2	0.064516
TH01	7	0.135	9	0.231	$2 \cdot p \cdot q$	0.06237
D21S11	31.2	0.087	32.2	0.098	$2 \cdot p \cdot q$	0.017052
D18S51	13	0.121	16	0.172	$2 \cdot p \cdot q$	0.041624
PentaE	12	0.139	16	0.083	$2 \cdot p \cdot q$	0.023074
D5S818	10	0.09	11	0.335	$2 \cdot p \cdot q$	0.0603
D13S317	11	0.353	12	0.202	$2 \cdot p \cdot q$	0.142612
D7S820	8	0.172	10	0.284	$2 \cdot p \cdot q$	0.097696
D16S539	12	0.35	13	0.192	$2 \cdot p \cdot q$	0.1344
CSF1PO	10	0.283	10		p^2	0.080089
PentaD	12	0.196	12		p^2	0.038416
Amelogenin	X	0.5	Y	0.5		
vWA	17	0.234	18	0.215	$2 \cdot p \cdot q$	0.10062
D8S1179	13	0.342	13		p^2	0.116964
TPOX	11	0.245	11		p^2	0.060025
FGA	21	0.147	24	0.122	$2 \cdot p \cdot q$	0.035868
Частота комбинированного генотипа	Один из					
2.2893893077717554E-14	4.367977069715994E13					

Можно открыть файл результатов в редакторе *Microsoft Excel*, скопировать из него ячейки с данными и вставить их в нужное место ЗЭ. Получившуюся таблицу при желании можно преобразовать в более привычную конкретному пользователю форму.

Озаглавить таблицу результатов можно следующим образом:

Расчёт вероятности случайного совпадения генетических признаков, выявленных в исследованном объекте (пятне крови, смыве с презерватива, окурке, гистологическом препарате и т.п., объект №NNN) и биологическом образце (крови, буккального эпителия и др.) г-на Пушкина-Чушкина А.С.

При этом формулировка выводов в ЗЭ может выглядеть примерно следующим образом: «Полученное значение RMP (частоты комбинированного генотипа, указать полученное значение) свидетельствует о том, что среди населения Земли теоретически только один человек обладает выявленным сочетанием генетических признаков. Полученной величины достаточно / недостаточно для формулирования вывода в категоричной форме о том, что исследованный объект произошёл от Пушкина-Чушкина А.С.».

Работа с программой: использование различных эталонных частот аллелей

Для текущей версии программы нужную эталонную популяцию следует сразу выбрать при запуске в выпадающем меню «**Популяция**». После ввода генотипов все расчёты будут выполнены по соответствующим эталонным популяционным частотам.

Объективно, существуют некоторые затруднения для корректного расчёта частот генотипов по отдельным референтным популяциям. Это касается как отсутствия внятных популяционных данных по отдельным локусам, так и нулевых значений частот для «невыявленных» аллелей в конкретных исследованных выборках. Ниже эти затруднения рассматриваются подробно.

Адекватные популяционные данные отсутствуют

Для ряда «экзотических» локусов (например, *DIS111*, *D17S5*, *CYAR04*, *PAH*, *vWFII* и некоторые другие) авторам программы в настоящее время неизвестны адекватные частоты аллелей для эталонных популяций европеоидов и негров США.

В таких случаях всем аллелям в этих популяциях по умолчанию приписана частота встречаемости **0,500** (50%, смотрите соответствующие файлы). Соответствующие ячейки в файлах выделены **заливкой** красным цветом. Частоты генотипов для этих локусов по этим популяциям будут рассчитаны с использованием аллельных частот, равных 0,500. Что, естественно, неправильно.



Русская популяция. Данные по эталонным частотам аллелей на текущий момент явно недостаточны (в разной степени вызывают вопросы) для следующих весьма актуальных локусов: *DIS1656*, *D2S441*, *D6S1043*, *D10S1248*, *D12S391*, *D22S1045*, *F13A01*, *F13B*, *FESFPS*, *LPL*, *SE33*. Смотрите примечания в соответствующих файлах и указанные в них литературные первоисточники.

«Невыявленные» аллели с нулевой частотой встречаемости

Для всех локусов в двух референтных популяциях из четырёх (для негров и европеоидов США) отдельным «невыявленным» аллелям приписана нулевая частота встречаемости в столбцах «В».

Например, для локуса *D16S539* аллели <5, 5 и 6 не были выявлены ни у европеоидов России, ни у европеоидов США. У европеоидов США не был выявлен также и аллель 7.

Экспериментальные частоты таких «невыявленных» аллелей в исследованных выборках во всех этих случаях равны **0**. В служебных файлах для ячеек с такими нулевыми экспериментальными частотами аллелей применена красная **заливка**.

Однако для европеоидов России дополнительно выполнена консервативная оценка частот аллелей по первичным данным. Поэтому для «нулевых» аллелей <5, 5 и 6 локуса *D16S539* «актуальные» частоты для русской популяции указаны равными 0,001 (*MAF*, в столбце *B*). В этом случае значения частот генотипов будут рассчитаны корректно, поскольку программа использует эталонные частоты аллелей именно из столбцов «*B*» рабочих листов.

А при попытке расчёта частот генотипов по эталонным частотам аллелей для других популяций может происходить операция «умножение на 0». При этом частота генотипа, естественно, окажется равной нулю.



Будьте внимательны при расчёте частот генотипов по эталонным частотам аллелей для разных популяций.

Для европеоидов и негров США именно **консервативная оценка частот аллелей** по первичным данным пока выполнена только для следующих локусов: *D3S1358*, *D8S1179*, *D18S51*, *D21S11*, *FGA*, *TH01*. То есть всем «невьявленным» аллелям приписана минимальная (ненулевая) популяционная частота *MAF*.

Расчёт *RMP*: стандартная практика в ДНК-исследованиях

Расчёт *частоты комбинированного генотипа* является стандартной международной практикой при оценке значимости выполненных молекулярно-генетических экспертиз идентификации личности (например, *NRC, 1992; NRC, 1996; Перепечина, Гришечкин, 1996*).

Использованные алгоритмы расчётов к настоящему времени углублённо проработаны, в том числе с учётом подразделённости референтной популяции, гипотез *донорства в ДНК-профиль* различных родственников подозреваемого лица, эффекта «нулевых» аллелей и пр. (например, *Brenner, 1994; Evett, Weir, 1998; Forensic DNA evidence interpretation, 2005; Lee et al., 2001; Walsh, Buckleton, 1991*): смотрите список литературы и статьи в папке *references*.

Интерпретация результатов: априорные условия, о которых следует помнить

В реализованном алгоритме расчётов за основу взяты следующие априорные условия:

- 1) для гомозиготных генотипов не предполагается «ложной гомозиготности»;
- 2) полагается полная невозможность мутаций;
- 3) частота комбинированного генотипа (RMP) рассчитывается путём перемножения популяционных частот генотипов по отдельным локусам (Q_i), исходя из предположения о том, что все эти локусы наследуются независимо. *Независимость локусов* обсуждается подробнее в соответствующем разделе;
- 4) частота комбинированного генотипа рассчитывается для любого случайного человека из той же самой референтной популяции, то есть для лица, не являющегося каким-либо родственником обследуемого.

Исследователю, выполняющему такие расчёты, необходимо совершенно отчётливо понимать, что независимо от полученной величины RMP , 100%-ная идентификация личности в случаях подтверждения (полного совпадения генетических профилей) недостижима в принципе. Хотя бы по той причине, что у однояйцевых близнецов ДНК абсолютно одинаковая, и с таким же успехом донором следа биологического происхождения может быть брат (сестра) -близнец обследованного лица: эксперт в общем случае не знает, есть на самом деле такой близнец у обследуемого лица или нет.

Дополнительно, независимо от числа исследованных локусов, сохраняется вероятность того, что где-нибудь на островах Океании или в соседнем городе проживает человек с точно таким же набором аллелей по всем этим локусам, как у обследованного лица, и именно этот «сосед» является истинным донором следа биологического происхождения.

Однако все эти рассуждения имеют в *большой* степени риторический характер, и достоверность ДНК-исследований является действительно очень высокой.

В этой связи необходимо отметить, что корректность формулирования выводов именно в категоричной форме, как *«исследованный объект произошёл от Пушкина-Чушкина А.С.»* при достижении какого-либо определённого «порогового» значения RMP является предметом отдельной дискуссии. Тем не менее, значимость выполненных расчётов имеет смысл оценивать в сравнении с численностью общемировой популяции. Оценку населения Земли на текущий момент можно смотреть, например, тут:

<http://countrysmeters.info/ru/World>

Независимость локусов, неравновесие по сцеплению и хромосомная локализация

При численной оценке результатов молекулярно-генетических экспертиз популяционная частота встречаемости мультилокусного генотипа определяется как произведение соответствующих частот для всех исследованных независимых (аутосомных) локусов. Таким образом, для правомерного использования «*правила перемножения*» частот генотипов принципиально важным является выполнение условия о том, что анализируемые локусы являются генетически независимыми, то есть отсутствует межлокусное *неравновесие по сцеплению* (НС).

Под НС понимают любую неслучайную ассоциацию (взаимозависимость) между аллелями разных локусов. Степень НС можно численно оценить разными методами, и такой анализ используют при картировании генов наследственных заболеваний, а также для упорядочивания генетических маркёров относительно друг друга на физических картах хромосом. При этом для нейтральных локусов величина НС является мерой генетической дистанции: чем больше значение НС, тем меньше физическое расстояние между маркёрами и ниже частоты рекомбинации и мутирования.

Соответственно, отсутствие НС (независимое наследование аллелей) по умолчанию подразумевается только для маркёров, расположенных или на разных хромосомах, или на разных плечах одной хромосомы. Во всех остальных случаях расположение локусов относительно близко друг к другу на одном плече хромосомы может рассматриваться как сигнал, указывающий на их возможную взаимозависимость. При выявлении достоверного НС между двумя локусами (группами таких локусов) применение правила перемножения частот (ОП) при экспертных вероятностных расчётах **неприемливо**.

Отличным примером некорректности многих расчётов является пара локусов *CSF1PO* и *D5S818*. Оба маркёра входят в национальную базу данных США **CODIS** (**C**ombined **D**NA **I**ndex **S**ystem) и, соответственно, во все современные мультиплексные наборы. Однако они локализованы на длинном плече хромосомы 5: *5q32* и *5q23.2*, соответственно. Более того, физическое расстояние между ними составляет около 26-27 Mb (миллионов нуклеотидных оснований): по данным [BLAT](#) (*Kent, 2002*; версия Feb.2009, GRCh37/hg19). Соответственно, эти локусы сцеплены между собой, чему имеются многочисленные литературные подтверждения, например *Wu et al., 2014*. Таким образом, использование «*правила перемножения*» для этих двух маркёров некорректно.

Тема сцепленных локусов становится всё более актуальной в связи с увеличением числа маркёров в мультиплексных наборах последних поколений. Как быть в таких ситуациях?

Прежде всего нужно чётко понимать хромосомную локализацию всех используемых локусов. Например, в [Инструкции](#) к наборам *Tapotили* приведена таблица хромосомной локализации около 100 аутосомных маркёров, использовавшихся ранее и используемых в настоящее время для идентификации личности и установления спорного родства (стр. 64 и далее). Эта же таблица вынесена в папку *references* отдельным файлом:

Ефремов_2017_Таблица_возможного_неравновесия_по_сцеплению_между_полиморфными_локусами_генома_человека.pdf.

Использование индивидуальных коэффициентов неравновесия по сцеплению сильно усложняет расчётную часть и требует аргументированных опорных экспериментальных данных для разных пар таких локусов.

Сильно упрощая, НС «по умолчанию» следует полагать для всех аутосомных маркёров, локализованных на одном плече хромосомы, физическое расстояние между которыми меньше 50 миллионов нуклеотидных оснований (50 Mb). И для расчётов комбинированных индексов *CPI* и *RMP* использовать соответствующие значения только для одного локуса из каждой группы сцепления.

Более подробно на тему НС и «*правила перемножения*» можно почитать в разных статьях из списка литературы, например: *Иванов и др., 2005; Buckleton et al., 2001.*

Качество первичных популяционных данных: индекс ПСАЧ и РХВ

Корректность выполняемых расчётов напрямую связана с корректным выбором соответствующей «эталонной» популяционной выборки с частотами аллелей. Существенны объём и однородность исследованной выборки.

Оценку качества первичных популяционных данных можно пояснить на примере локуса *D18S51*. Из таблицы в статье *Budowle et al., 2011* видно, что частоты аллелей для европеоидов и негров США авторами указаны именно с точностью до трёх знаков после запятой, как это и перенесено в соответствующие листы рабочего файла *D18S51.xls*.

Проверка *Полной Суммы Аллельных Частот* в рабочем листе файла программы для европеоидов США даёт *ПСАЧ* = 1,000. А вот для негров США получаем *ПСАЧ* = 0,998. Такие выявленные «авторские небрежности округления» при публикации популяционных данных (аллельных частот) во всех случаях отмечены выделением красным цветом для соответствующих значений *ПСАЧ*. Таким образом, *индекс ПСАЧ*® является хорошим показателем качества самой «опорной» публикации.

В настоящее время в папке *Data* для 17 микросателлитных локусов в качестве «опорной» для русской популяции использована работа *Zhivotovsky et al., 2009* (PMID: [19647694](#)). Однако для локусов *FGA*, *D13S317*, *D18S51*, *D2S1338*, *D21S11*, *vWA* в этой работе в разной степени *ПСАЧ* ≠ 1,00000 (см. соответствующие файлы).

При этом авторы приводят значения частот аллелей с точностью именно до пяти (!) знаков после запятой при объёме исследованной выборки порядка 500-600 неродственных человек. Такая «точность» частот аллелей видится явно ошибочной, поскольку с точки зрения популяционной генетики это означает, что опорная выборка должна была включать не менее 50 000 (пятидесяти тысяч) неродственных человек: $1/(2N) \geq 0,00001$.

Соответственно, в программах *Pindex* и *Gency* для расчётов используются данные по частотам аллелей для русской популяции из работы *Zhivotovsky et al., 2009*, но эти значения округлены с точностью до трёх знаков после запятой.

Помимо *индекса ПСАЧ*®, другим важным критерием качества «опорной» публикации является соответствие наблюдавшегося распределения генотипов *равновесию Харди-Вайнберга (РХВ)*. Про РХВ можно почитать, например, [здесь](#) или в специальной литературе (см. раздел **Список литературы**).

Для выбора эталонных частот аллелей были использованы литературные данные, соответствующие равновесию РХВ. В редких случаях (при отсутствии других доступных и более адекватных публикаций) несоответствие исследованной выборки РХВ отмечено в соответствующих листах. Например, для локуса *D6S1043* в листе *d6s1043-white* указано:

Внимание! Для исследованной выборки отмечено несоответствие распределения генотипов равновесию Харди-Вайнберга (РХВ, $p=0,0198$).

Важно учитывать, что в случаях несоответствия распределения генотипов РХВ использование соответствующих частот аллелей для последующих расчётов (индекса отцовства, родства, вероятности случайного совпадения и др.) не является корректным. Степень этой «некорректности» оценить затруднительно, поэтому в качестве практического экспертного решения предложено использование «консервативной оценки аллельных частот». Однако наилучшим решением является отслеживание новых публикаций по актуальным локусам и популяциям и обновление соответствующих служебных файлов.

SNP-, InDel- и локусы «генетической предрасположенности»

К настоящему времени автоматизированный фрагментный анализ, основанный на использовании полиморфных микросателлитных локусов (*STRs*), является стандартом *де-факто* при исследованиях по установлению спорного родства и идентификации личности. Однако этот метод уже подошёл к пределу своих возможностей и, дополнительно, исходно обладает рядом недостатков.

Исследование полиморфных маркёров другого рода, с однонуклеотидными заменами (*SNPs*), для целей судебно-медицинских и криминалистических экспертиз началось в 2000-х годах. В 2003 г. был основан международный консорциум для всестороннего анализа возможностей *SNP*-маркёров для идентификации личности и установления родства (*SNP for ID consortium*). Использование для генотипирования именно снипов имеет следующие преимущества:

- низкая частота мутаций для *SNPs* (около 10^{-7} – 10^{-8} по сравнению с 10^{-3} – 10^{-5} для микросателлитных локусов);
- возможность упрощения и полной автоматизации процедуры генотипирования при высокой пропускной способности лаборатории;
- возможность использования очень коротких целевых фрагментов ДНК (вплоть до 40-50 п.н.), что актуально для анализа сильно деградированных образцов (при последствиях техногенных и природных катастроф, воздействия высоких температур, при большом возрасте ископаемых останков и т.д.);
- однозначность установления генотипа, на уровне конкретного нуклеотидного основания в заданной позиции цепи ДНК.

Здесь нужно также отметить, что подавляющее большинство снипов являются простыми двух-аллельными полиморфизмами, что исключает ряд возможных технических ошибок и неоднозначностей, присущих генотипированию микросателлитных локусов.

Дополнительно, достаточно большое количество снипов (и некоторых других полиморфных маркёров) сейчас уже рутинно исследуется в рамках медико-генетических тестирований, в том числе семейных. На слуху самые разнообразные «панели генетических предрасположенностей»: репродуктивное здоровье и патологии плода, риски нарушений системы свертывания крови, болезни сердца и сосудов, онкологические заболевания, алкогольная и наркотическая зависимость, нарушения обмена веществ, синдром Жильбера, спортивная успешность, ожирение, фармакогенетика, и пр.

Анализ уровня «полное генетическое обследование для мужчин (женщин)» обычно включает не менее 40 различных генов. В случае такого обследования вполне резонно использовать результаты (> 40 полиморфных маркёров) для оценки «уникальности» установленного генотипа в локальной популяции.

Если говорить о маркёрах *InDel* (*INDEL*, *insertion / deletion*, также *DIP*, *deletion / insertion polymorphism*, полиморфизм типа делеция-вставка), то они также являются простыми двух-аллельными системами. Для идентификации личности и установления спорного родства ещё в конце 2000-х гг. появились готовые решения в виде коммерческих наборов для таких локусов, например *Pereira et al.*, 2009; *Investigator DIPplex Kit* (*Qiagen*, Германия). Несколько особняком в этой же группе находятся *Alu*-повторы (мобильные элементы генома), повторяющаяся последовательность которых имеет относительно большой размер, около 300 п.о. Однако и для них были предложены готовые решения для идентификации личности (32 маркёра, *Mamedov et al.*, 2010).

Более подробную информацию на тему использования двух-аллельных маркёров в идентификации личности и установлении спорного родства можно найти, например, в обзоре *Schneider*, 2012 или по ссылке: <http://strbase.nist.gov/SNP.htm>.

Для программ *Pindex* и *Gency* рабочие файлы уже созданы для ряда двух-аллельных локусов, а также для отдельных мини- и микросателлитных маркёров, актуальных в первую очередь в медико-генетическом консультировании, например: *CCR5-delta32*, *D7S8*, *GYP A* (*glycophorin A*), *HLA_DQA1*, *IL1RN-VNTR*, *LDLR* (*low density lipoprotein receptor*), *NOS3-VNTR*, *UGT1A1-STR*. Возможно создание рабочих файлов под любые новые локусы.

Локусы для установления половой принадлежности и маркёры хромосомы Y

Практически во все коммерческие мультиплексные наборы, используемые в настоящее время для установления спорного родства, включён локус *амелогенина* (*amelogenin*, *Amel*, *AMEL*) – для определения половой принадлежности как обследуемых лиц, так и биологических следов человеческого происхождения. Более того, в связи с узкими местами амелогенинового теста, в мультиплексные наборы «последнего поколения» включены дополнительные маркёры, расположенные на хромосоме Y человека (*DYS391*, *DYS570*, *SRY*, *Y-indel* и др).

В целом, результаты установления половой принадлежности обследованного лица (объекта) не влияют на значения *RMP*, поскольку конкретный комбинированный генотип встречается, грубо говоря, с одинаковой частотой как среди мужчин, так и среди женщин. Однако установление половой принадлежности удваивает значимость расчётов частоты встречаемости совпадающего генотипа. Например, вывод может быть сформулирован уже следующим образом: «установленный комбинированный генотип ожидаемо встречается у одного из *NNN мужчин*».

Если говорить именно о маркёрах хромосомы Y (без локуса *амелогенина*), то результаты исследования следует интерпретировать в соответствии с тем, что все такие локусы являются сцепленными между собой и формируют достаточно уникальный мультилокусный *гаплотип*. Поэтому для нескольких маркёров хромосомы Y правило перемножения частот аллелей (генотипов) **неприменимо**. В этом плане уместно говорить о частоте выявленного комбинированного гаплотипа именно для совокупности всех таких локусов.

Таким образом, использовать для расчётов *RMP* «генотипы» по маркёрам хромосомы Y и локусу *амелогенина* **не следует**. Предупреждения об этом добавлены непосредственно в соответствующие файлы (*AMEL.xls* и др.).

Для программ *Pindex* и *Gency* по локусам *Амелогенин*, *DYS391*, *Y-indel*, *DYS570*, *DYS576* также созданы рабочие файлы. Это сделано в первую очередь для того, чтобы сохранять в файле результатов полные данные обо всех установленных генотипах обследованных лиц (объектов).

Локусы, расположенные на хромосоме X человека

Полиморфные микросателлитные локусы, расположенные на хромосоме X (икс) человека, в последние годы всё шире используются в молекулярно-генетических экспертизах спорного родства и идентификации личности.

У женщин в норме присутствуют две хромосомы X. Тогда как у мужчин в норме присутствует только одна такая хромосома, то есть у них все локусы на X-хромосоме находятся в *гемизиготном* состоянии. Это является принципиальным различием X- и аутосомных маркеров.

При расчётах частот генотипов следует иметь в виду, что популяционная частота гемизиготного по аллелю *A* мужчины равна собственно популяционной частоте этого аллеля $p(a)$. Тогда как популяционная частота гомозиготных *AA* и гетерозиготных *AB* женщин равна $p(a)^2$ и $2p(a)p(b)$, соответственно (как для аутосомных локусов).



В текущей версии программы *Gency* алгоритм расчёта частоты гемизиготного генотипа не реализован. Поэтому расчёт частот генотипов по локусам X-хромосомы будет выполнен корректно только в отношении женщин.

В настоящее время готовый к использованию служебный файл создан только для такого локуса X-хромосомы, как *HPRTB* (файл *HPRTB.xls*). К сожалению, адекватные данные о частотах аллелей в локусе *HPRTB* для каких-либо российских популяций разработчикам программы в настоящее время неизвестны. В единственной найденной работе [Асеев и др., 1995] частотное распределение аллелей для этого маркера было изучено на весьма малочисленной выборке (50 неродственных человек), чего явно недостаточно для использования этих данных в качестве эталонных для русской популяции.

В настоящее время в рабочем файле для локуса *HPRTB* приведена «консервативная оценка» частот аллелей для русской популяции. В качестве опорных данных были использованы частоты аллелей в популяции Германии из работы [Becker et al., 2008]. Исследованная выборка включала 717 неродственных человек, в том числе 439 мужчин и 278 женщин. Эталонные частоты аллелей для европейцев и негров США также внесены в рабочий файл, для этого были использованы данные из описания к наборам *GenePrint STR Systems* (Promega Corporation, США).

При расчётах следует также иметь в виду, что все X-маркеры локализованы на одной хромосоме. Некоторые расположенные близко друг к другу локусы сцеплены между собой, и поэтому получаемые для них значения частот генотипов при расчёте комбинированного

значения RMP нельзя просто перемножать (некорректно). Для таких сцепленных локусов при расчётах RMP следует оперировать не частотами отдельных аллелей по каждому маркёру, а частотой гаплотипа, то есть набора аллелей, локализованных на одной из парных хромосом по всем этим маркёрам. При этом формулы расчётов при использовании гаплотипов вместо отдельных аллелей остаются такими же. Если анализ гаплотипов для таких сцепленных локусов по каким-либо причинам невозможен или затруднителен, то для расчётов RMP следует использовать **только одно** значение Q_i (для одного маркёра) из каждой группы сцепления.

Доработка ПО

Подразумевается (возможна) следующая доработка программы:

- ✓ загрузка сохранённых ранее файлов с результатами расчётов для их редактирования (пересчёта по другим эталонным частотам);
- ✓ импорт данных из текстовых файлов результатов с генетических анализаторов (*GeneMapper* и пр.);
- ✓ вывод в файл результатов хромосомной локализации всех исследованных локусов;
- ✓ возможность выбора более консервативных формул расчёта частот гомозиготных генотипов (в частности, использование формулы $2p$ в дополнение к p^2 для нивелирования эффекта «ложной гомозиготности»);
- ✓ возможность использования специальных корректирующих популяционных коэффициентов (коэффициент инбридинга, F_{ST} , θ);
- ✓ создание рабочих файлов для дополнительных локусов X-хромосомы и реализация алгоритма расчёта частот гемизиготных генотипов по этим маркёрам.

Решение затруднений

Данный раздел будет наполняться по мере выявления каких-либо затруднений при использовании программы. Пока это лист для замечаний и пожеланий конечных пользователей.

Список литературы

1. Асеев М.В., Скакун В.Н., Баранов В.С. (1995) Анализ аллельного полиморфизма четырех коротких tandemных повторов в популяции северо-западного региона России. – *Генетика*, 31 (6), 839-845. PMID: [7635321](#).
 2. Гаврилей Ю.К., Шепелов И.Е., Гузунова М.А., Лапенков М.И., Богданов В.М. (2007) Программа идентификации личности на основе обработки результатов ДНК-анализа «DNA-expert Lab». – Свидетельство об официальной регистрации программы для ЭВМ №2007613189 от 27.07.2007. (-)
- знаком (-) здесь и далее в списке литературы помечены публикации, полный текст которых пока отсутствует в папке references.
3. Ефремов И.А., Серегин Ю.А. (2009) Программа для расчёта индекса и вероятности отцовства (материнства) при исследованиях ДНК – Pindex: Paternity Index. – Свидетельство о государственной регистрации программы для ЭВМ №2009610058 от 11.01.2009. – *Программы для ЭВМ / RU ОБПБТ* №2 (67), 20.06.2009.
 4. Иванов П.Л., Земскова Е.Ю., Туракулов Р.И., Ефремов И.А. (2005) Изучение потенциально сцепленных вариантов полиморфизма хромосомной ДНК в аспекте судебно-экспертного применения молекулярно-генетических индивидуализирующих систем CD4, vWA и VWFI. – *Судебно-медицинская экспертиза*, 48 (2), 29-34. PMID: [15881140](#).
 5. Перепечина И.О., Гришечкин С.А. (1996) Вероятностные расчёты в ДНК-дактилоскопии: Методические рекомендации. – Москва, ЭКЦ МВД России. (-)
 6. Пименов М.Г., Кондрашов С.А., Стороженко И.В., Культин А.Ю., Бакун К.В. (2004) Исследование частот встречаемости аллелей STR-локусов среди жителей России: информационное письмо. – Москва, ЭКЦ МВД России. (-)
 7. Becker D., Rodig H., Augustin C., Edelmann J., Götz F., Hering S., Szibor R., Brabetz W. (2008) Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit. – *Forensic Sci Int Genet.*, 2 (1), 69-74. PMID: [19083792](#).
 8. Brenner C.H. (1994) Evaluation of the Product Rule. – in *Advances in Forensic Haemogenetics*, vol. 5, Eds. Bär W., Fiori A., Rossi U. Springer, Berlin, Heidelberg, 453-455. https://doi.org/10.1007/978-3-642-78782-9_123 (+/-)
 9. Buckleton J.S., Walsh S., Harbison S.A. (2001) The fallacy of independence testing and the use of the product rule. – *Sci. Justice*, 41 (2), 81-84. PMID: [11393945](#).

10. Budowle B., Ge J., Chakraborty R., Eisenberg A.J., Green R., Mulero J., Lagace R., Hennessy L. (2011) Population genetic analyses of the NGM STR loci. – *Int. J. Legal Med.*, 125 (1), 101-109. PMID: [20878415](#).
11. Evett I. W., Weir B.S. (1998) Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. – Sunderland, Massachusetts, Sinauer Associates Inc.
12. *Forensic DNA evidence interpretation*. (2005) – CRC Press, Boca Raton, London, New York, Washington, D.C. Edited by John Buckleton, Christopher M. Triggs, Simon J. Walsh.
13. GenePrint™ STR Systems (Silver Stain Detection). Technical manual. – Promega Corporation (USA), revised 7/99, 04/01, 7/06. (-)
14. Gjertson D.W., Brenner C.H., Baur M.P., Carracedo A., Guidet F., Luque J.A., Lessig R., Mayr W.R., Pascali V.L., Prinz M., Schneider P.M., Morling N. (2007) ISFG: Recommendations on biostatistics in paternity testing. — *Forensic Sci Int Genet.*, 1 (3-4), 223-231. PMID: [19083766](#).
15. Kent W.J. (2002) BLAT – the BLAST-like alignment tool. – *Genome Res.*, 12 (4), 656-664. PMID: [11932250](#).
16. Lee J.W., Lee H.S., Park M., Hwang J.J. (2001) Evaluation of DNA match probability in criminal case. – *Forensic Sci Int.*, 116 (2-3), 139-148. PMID: [11182265](#).
17. Mamedov I.Z., Shagina I.A., Kurnikova M.A., Novozhilov S.N., Shagin D.A., Lebedev Y.B. (2010) A new set of markers for human identification based on 32 polymorphic Alu insertions. – *Eur. J. Hum. Genet.*, 18 (7), 808-814. PMID: [20179741](#).
18. NRC, 1992: National Research Council. *DNA Technology in Forensic Science*. – Washington, D.C., National Academy Press, 1992.
19. NRC, 1996: National Research Council (US) Committee on DNA Forensic Science: An Update. *The Evaluation of Forensic DNA Evidence*. – Washington, D.C., National Academy Press, 1996. <https://doi.org/10.17226/5141>.
20. Pereira R. Phillips C., Alves C., Amorim A., Carracedo A., Gusmao L. (2009) Insertion/deletion polymorphisms: A multiplex assay and forensic applications. – *Forensic Science International: Genetics Supplement Series*, 2, 513-515.
21. Ruitberg C.M., Reeder D.J., Butler J.M. (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. – *Nucleic Acids Res.*, 29 (1), 320-322. PMID: [11125125](#).
22. Schneider P.M. (2012) Beyond STRs: The Role of Diallelic Markers in Forensic Genetics. – *Transfus. Med. Hemother.*, 39 (3), 176-180. PMID: [22851932](#).
23. Walsh K.A., Buckleton J.S. (1988) – A discussion of the law of mutual independence and its application to blood group frequency data. – *J. Forensic Sci Soc.*, 28 (2), 95-98. PMID: [3404111](#).

24. Walsh K., Buckleton J.S. (1991) Calculating the frequency of occurrence of a blood type for a 'random man'. – *J. Forensic Sci Soc.*, 31 (1), 49-58. PMID: [1856674](#).
25. Weir B.S. (1992) Population genetics in the forensic DNA debate. – *Proc Natl Acad Sci USA*, 89 (24), 11654-11659. Review. PMID: [1465380](#).
26. Wu W., Hao H., Liu Q., Han X., Wu Y., Cheng J., Lu D. (2014) Analysis of linkage and linkage disequilibrium for syntenic STRs on 12 chromosomes. – *Int. J. Legal Med.*, 128 (5), 735-739. PMID: [24929640](#).
27. Zhivotovsky L.A., Malyarchuk B.A., Derenko M.V., Wozniak M., Grzybowski T. (2009) Developing STR databases on structured populations: the native South Siberian population versus the Russian population. – *Forensic Sci Int Genet.*, 3 (4), e111-116. PMID: [19647694](#).

*Список литературы является далеко не полным и включает только ссылки, упомянутые в настоящем руководстве. В папке **references** содержатся дополнительные публикации, как по теории расчётов, так и по частотам аллелей в отдельных популяциях.*

Благодарности

Нидерландскому художнику периода Северного Возрождения *Иерониму Босху* (*Hieronymus Bosch*, около 1450-1516) за триптих «*Сад земных наслаждений*», фрагмент которого использован на обложке настоящего Руководства.

Надеемся, что эта иллюстрация наглядно поясняет отдельным неподготовленным читателям тот факт, что возможно *случайное совпадение* комбинированного генотипа между двумя разными людьми, проживающими на Земле. Полностью этот триптих в высоком разрешении можно посмотреть на сайте Музея Прадо (Мадрид):

<https://www.museodelprado.es/coleccion/obra-de-arte/el-jardin-de-las-delicias/02388242-6d6a-4e9e-a992-e1311eab3609>

Эксперту ФГКУ «111 Главный государственный центр судебно-медицинских и криминалистических экспертиз Министерства обороны Российской Федерации» (г. Москва) Харламову Сергею Глебовичу – за техническое задание на программу, тестирование первых рабочих версий и конструктивное обсуждение.

Проекту [Sci-Hub](#) и лично [Александр Элбакян](#) – за устранение преград в поисках полных текстов статей.

Обратная связь

Со всеми вопросами, замечаниями, пожеланиями и предложениями – обращайтесь!

Ефремов Илья Алексеевич, к.б.н., с.н.с.

Моб. телефон, Viber: +7 (903) 786-4-789

Эл. почта: info@tapotili.ru	Сайт: www.tapotili.ru	Facebook
---	--	--------------------------

Лаборатория химической физики и биоаналитических процессов

Института биохимической физики им. Н.М. Эмануэля Российской академии наук (ИБХФ РАН). 119334, Россия, г. Москва, ул. Косыгина, д. 4.

Контакты авторов также доступны непосредственно в самой программе нажатием кнопки «*О программе*».

Обновления служебных файлов высылаются авторизованным пользователям по запросам, присланным по электронной почте.