

Pindex (Paternity INDEX):

компьютерная программа для расчёта индекса и вероятности отцовства
(материнства) при исследованиях ДНК

Руководство пользователя

Версия документа 10.0



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

Оглавление

СПИСОК СОКРАЩЕНИЙ	3
НАЗНАЧЕНИЕ И КРАТКОЕ ОПИСАНИЕ ПРОГРАММЫ	5
УСЛОВИЯ ИСПОЛЬЗОВАНИЯ ПРОГРАММЫ	6
РЕДАКЦИИ ДОКУМЕНТА	7
ДЕМО-ВЕРСИЯ ПРОГРАММЫ	8
СИСТЕМНЫЕ ТРЕБОВАНИЯ	9
ЗАПУСК ПРОГРАММЫ	9
ПАПКА ПРОГРАММЫ	10
ПЕРСОНАЛЬНЫЙ КЛЮЧ АВТОРИЗОВАННОГО ПОЛЬЗОВАТЕЛЯ	11
КАТАЛОГ <i>DATA</i> : СЛУЖЕБНЫЕ ФАЙЛЫ ДЛЯ ИСПОЛЬЗУЕМЫХ ЛОКУСОВ	12
КАТАЛОГ <i>REFERENCES</i> : ЛИТЕРАТУРНЫЕ ПЕРВОИСТОЧНИКИ	16
РАБОТА С ПРОГРАММОЙ: ВЫБОР ЭТАЛОННОЙ ПОПУЛЯЦИИ И ВВОД ГЕНОТИПОВ	17
РАБОТА С ПРОГРАММОЙ: РАСЧЁТ ИНДЕКСА И ВЕРОЯТНОСТИ ОТЦОВСТВА	20
СОХРАНЕНИЕ РЕЗУЛЬТАТОВ И ЗАВЕРШЕНИЕ РАБОТЫ С ПРОГРАММОЙ	22
ИСПОЛЬЗОВАНИЕ ФАЙЛА РЕЗУЛЬТАТОВ В ТЕКСТЕ «ЗАКЛЮЧЕНИЯ ЭКСПЕРТА»	23
РАБОТА С ПРОГРАММОЙ: ИСПОЛЬЗОВАНИЕ РАЗЛИЧНЫХ ЭТАЛОННЫХ ЧАСТОТ АЛЛЕЛЕЙ	25
КАК ССЫЛАТЬСЯ НА ПРОГРАММУ ПРИ ОФОРМЛЕНИИ ЗАКЛЮЧЕНИЯ ЭКСПЕРТА?	26
ОТНОШЕНИЯ ПРАВДОПОДОБИЯ: СТАНДАРТНАЯ ПРАКТИКА В ДНК-ИССЛЕДОВАНИЯХ	27
ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ: АПРИОРНЫЕ УСЛОВИЯ, О КОТОРЫХ СЛЕДУЕТ ПОМНИТЬ	28
КАКИЕ СЛУЧАИ МОЖНО КОРРЕКТНО АНАЛИЗИРОВАТЬ?	29
ВЕРОЯТНОСТЬ ОТЦОВСТВА: ЧТО ЭТО ТАКОЕ?	30
ВЕРОЯТНОСТЬ ОТЦОВСТВА ПРИ РАЗЛИЧНЫХ ЗНАЧЕНИЯХ АПРИОРНОЙ ВЕРОЯТНОСТИ	31
ОТНОШЕНИЯ ПРАВДОПОДОБИЯ: ПРИМЕР РАСЧЁТА	33
НЕЗАВИСИМОСТЬ ЛОКУСОВ, НЕРАВНОВЕСИЕ ПО СЦЕПЛЕНИЮ И ХРОМОСОМНАЯ ЛОКАЛИЗАЦИЯ	34
КАЧЕСТВО ПЕРВИЧНЫХ ПОПУЛЯЦИОННЫХ ДАННЫХ: ИНДЕКС ПСАЧ И РХВ	36
SNP-, INDEL- и ЛОКУСЫ «ГЕНЕТИЧЕСКОЙ ПРЕДРАСПОЛОЖЕННОСТИ»	38
ЛОКУСЫ ДЛЯ УСТАНОВЛЕНИЯ ПОЛОВОЙ ПРИНАДЛЕЖНОСТИ И МАРКЕРЫ ХРОМОСОМЫ Y	40
ЛОКУСЫ, РАСПОЛОЖЕННЫЕ НА ХРОМОСОМЕ X ЧЕЛОВЕКА	41
ПРОГРАММА GENCY	44
РЕШЕНИЕ ЗАТРУДНЕНИЙ	45
ПРОБЛЕМЫ С КОДИРОВКОЙ КИРИЛЛИЦЫ ПРИ ЗАПУСКЕ ПРОГРАММЫ	45
КЛЮЧ ОТСУТСТВУЕТ ИЛИ ПОВРЕЖДЁН	48
ФАЙЛЫ НЕ НАЙДЕНЫ (FILES NOT FOUND)	48
НЕВОЗМОЖНО ПРОЧИТАТЬ ДАННЫЕ В МАССИВ	48
НЕВОЗМОЖНО СОХРАНИТЬ РЕЗУЛЬТАТЫ РАБОТЫ ПРОГРАММЫ	49
АДЕКВАТНЫЕ ПОПУЛЯЦИОННЫЕ ДАННЫЕ ОТСУТСТВУЮТ	51
«НЕВЫЯВЛЕННЫЕ» АЛЛЕЛИ С НУЛЕВОЙ ЧАСТОТОЙ ВСТРЕЧАЕМОСТИ	52
НЕСКОЛЬКО «ПРЕДПОЛАГАЕМЫХ ОТЦОВ»	53
АНАЛОГИЧНЫЕ КОМПЬЮТЕРНЫЕ ПРОГРАММЫ	53
СПИСОК ЛИТЕРАТУРЫ	54
БЛАГОДАРНОСТИ	58
ОБРАТНАЯ СВЯЗЬ	58
СВИДЕТЕЛЬСТВО О РЕГИСТРАЦИИ ПРОГРАММЫ	59

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, снп

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 индекса и вероятности отцовства (материнства) при ДНК-исследованиях и может быть использована в экспертной практике судебно-медицинских, криминалистических и клинико-диагностических лабораторий.

Программа обеспечивает выполнение следующих функций:

- ✓ выбор эталонной популяции для расчётов;
- ✓ ввод генотипов для двух или трёх обследованных лиц (ребенок и заявленные родители) по отдельным локусам;
- ✓ расчёт отношений правдоподобия для отдельных локусов;
- ✓ расчёт комбинированного индекса отцовства (материнства) по нескольким выбранным локусам;
- ✓ расчёт комбинированной вероятности отцовства (материнства), в том числе при различных значениях априорной вероятности;
- ✓ 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.

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**.

Перед использованием продукта ознакомьтесь с условиями Лицензионного Соглашения, актуальная версия которого доступна на сайте в виде отдельного файла (*End-User License Agreement*, [EULA](#)). Для нового Авторизованного Пользователя лицензионное соглашение вступает в силу с момента установки (запуска) продукта.

Лицензионное соглашение может изменяться правообладателями программы в одностороннем порядке в любой момент времени. Уведомление Авторизованных Пользователей о внесённых изменениях в условия лицензионного соглашения осуществляется в произвольной форме, в том числе 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 10 (08.2022). Созданы новые служебные файлы для 50 SNP-локусов. Доработан ряд разделов описания. Внесены дополнения в отдельные служебные файлы по частотам аллелей. В папку «References» добавлены полезные публикации.

Версия 9 (09.2020). Доработаны разделы «Проблемы с кодировкой кириллицы при запуске программы» и «Невозможно сохранить результаты работы программы». Созданы новые служебные файлы для ряда минисателлитных локусов. Внесены дополнения в отдельные служебные файлы по частотам аллелей. В папку «References» добавлены полезные публикации.

Версия 8 (06.2018). Доработан более подробно ряд разделов, с учётом пожеланий авторизованных пользователей, а также в связи с анонсированием программы *Gency* (*Genotype Frequency* – смотрите соответствующий раздел про эту программу). Внесены изменения в ряд служебных файлов по частотам аллелей для европеоидов и негров США. В папку «References» добавлены полезные публикации.

Версия 7 (09.2017). Добавлены разделы «Условия использования программы», «Список сокращений», «SNP-, InDel- и локусы генетической предрасположенности».

Версия 6 (08.2017). Добавлена папка «References», содержащая основные статьи как по референтным частотам аллелей, так и по теории расчётов с использованием отношений правдоподобия. Созданы новые разделы «Руководства», а также внесены дополнения в большинство остальных разделов. Внесены изменения в служебные файлы для всех основных используемых локусов. Созданы служебные файлы для ряда новых локусов: *Penta C*, *D6S1043*, *UGT1A1-STR* и др.

Версия 5 (02.2015). Добавлены разделы «Решение затруднений» и «Локусы для установления половой принадлежности», изменена структура документа. Созданы служебные файлы для ряда новых локусов.

Версия 4 (04.2011). Внесены изменения в большинство разделов документа. Созданы служебные файлы для ряда новых локусов.

Версия 3 (03.2009). Добавлены разделы «Свидетельство о регистрации программы» и «Расчёт индекса отцовства при исследовании локусов, расположенных на хромосоме X», переработана структура документа. Созданы служебные файлы для ряда новых локусов.

Версия 2 (01.2005). Выпущена первая стабильная версия программы с «Руководством пользователя» в виде отдельного pdf-файла.

Версия 1 (10.2002). Выпущена первая тестовая версия программы Pindex с кратким текстовым файлом помощи (*Help.txt*).

Демо-версия программы

Демо-версия предназначена для целей самостоятельного ознакомления, оценки и проверки потенциальным новым Пользователем функциональных возможностей программы. Это версия с ограниченными функциональными возможностями:

- ✓ работает с ограниченным числом локусов (не более пяти);
- ✓ не позволяет сохранять результаты расчётов;
- ✓ не позволяет загружать сторонние файлы с результатами расчётов;
- ✓ отключены функции расчётов по локальной популяции и выбору априорной вероятности.

Демо-версия предоставляется на любой необходимый Пользователю ознакомительный период времени без выплаты вознаграждения (безвозмездно). Высылается по запросу, присланному по электронной почте (предоставляется в виде *zip*-архива, размер около 3,2 Mbyte).

Пользователю не разрешается использовать Демо-версию программы для извлечения коммерческой выгоды.

Использование Демо-версии программы подразумевает принятие Пользователем всех условий лицензионного соглашения ([EULA](#)).

Достаточно редко (у новых пользователей с нерусифицированной конфигурацией ПК) Демо-версия программы может первоначально запускаться с некорректной кодировкой кириллицы. Пример такой ошибки и как это исправить – смотрите раздел «Решение затруднений» → «Проблемы с кодировкой кириллицы при запуске программы».

Системные требования

Программа написана в системе программирования *Inprise Delphi 5.0*. Реализована технология *OLE (Object Linking and Embedding)*. Минимальные системные требования: процессор 486 MHz, оперативная память 16 Mbyte, операционная система *Microsoft Windows (95 / 98 / 2000 / Me / XP / Vista / 7 / 8 / 10)*, *Microsoft Office (95 / 97 / 2000 / XP / 2003 / 2007 / 2010 / 2013) Excel (7 / 97 / 2000 (9) – 2013 (15))*.

Наличие на рабочем компьютере предустановленного офисного пакета *Microsoft Office* обязательно. Альтернативные пакеты (*OpenOffice* и др.) **не годятся**.

При тестировании *Pindex'a* на различных версиях *Microsoft Windows* были отмечены отдельные редкие 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.

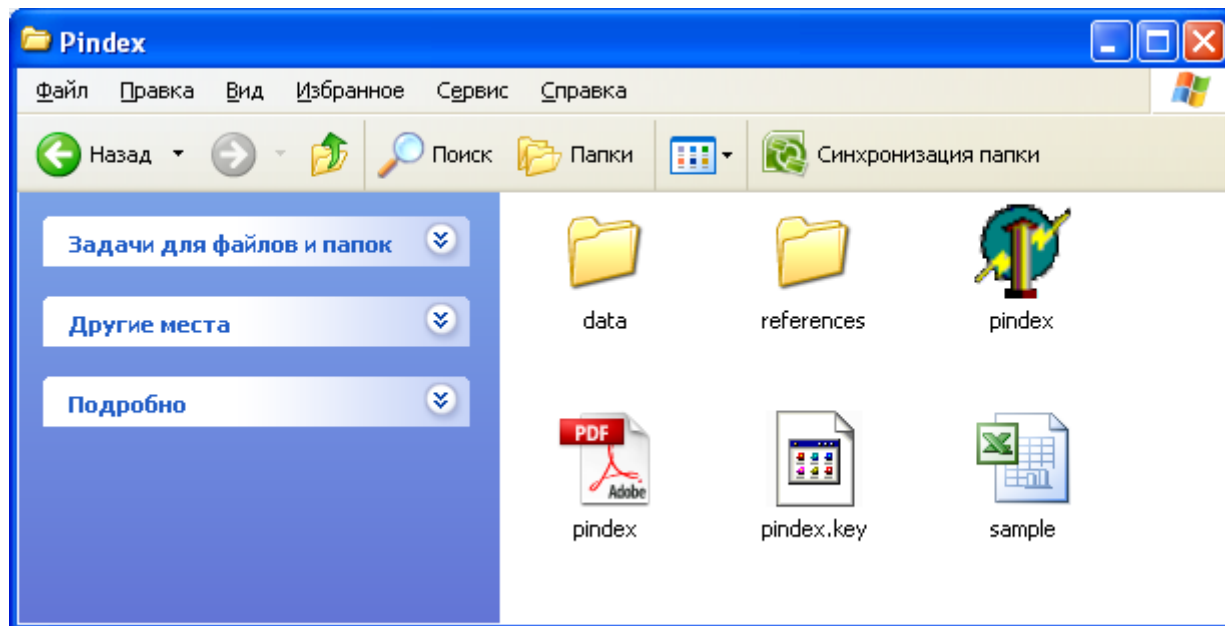
Для успешного запуска программы необходимо наличие в её каталоге двух файлов – *pindex.exe* и *pindex.key*, а также не-пустого подкаталога *data* со служебными файлами формата *Microsoft Excel*. Подробнее формат этих файлов будет описан ниже, в соответствующем разделе.

Запуск программы осуществляется вызовом файла *pindex.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, результаты работы программы (файл с расчётами) могут быть не сохранены. В этом случае будет выдано сообщение об ошибке. Более подробно это обсуждается в разделе **Решение затруднений** настоящего Руководства.

Папка программы

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



Файлы *pindex.exe* и *pindex.key* являются исполняемым файлом программы и ключом авторизованного пользователя, соответственно. Они обязательны для работы.

Файл *pindex.pdf* является *Руководством пользователя* (Вы его сейчас читаете), из программы он вызывается кнопкой «*Справка*» и на работоспособность не влияет.

Файл *sample.xls* является примером файла результатов с генотипами и расчётами, он также необязателен для работы. Результаты, содержащиеся в этом файле, используются в настоящем Руководстве при описании работы программы.

В каталоге *data* находятся рабочие файлы формата *Microsoft Excel* для отдельных локусов, и этот каталог является обязательным для работы. Число рабочих файлов, находящихся в нём, определяется числом используемых пользователем полиморфных локусов. Число таких локусов (файлов) не ограничено и формируется индивидуально. Обновляемый каталог.

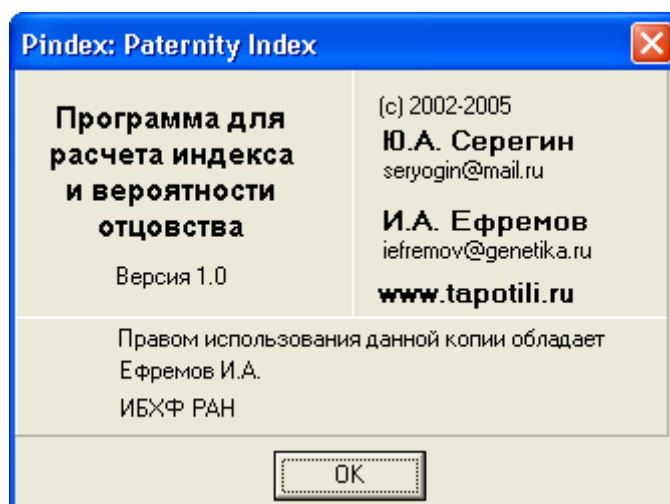
В каталоге *references* находятся файлы статей и приложений к ним, содержащие эталонные аллельные частоты для всех используемых локусов. Обновляемый каталог, но для корректной работы программы он необязателен.

Персональный ключ авторизованного пользователя

Каждая авторизованная копия *Pindex'a* защищена персональным ключом, файл которого *pindex.key* обязательно должен находиться в одной папке с исполняемым файлом *pindex.exe*.

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

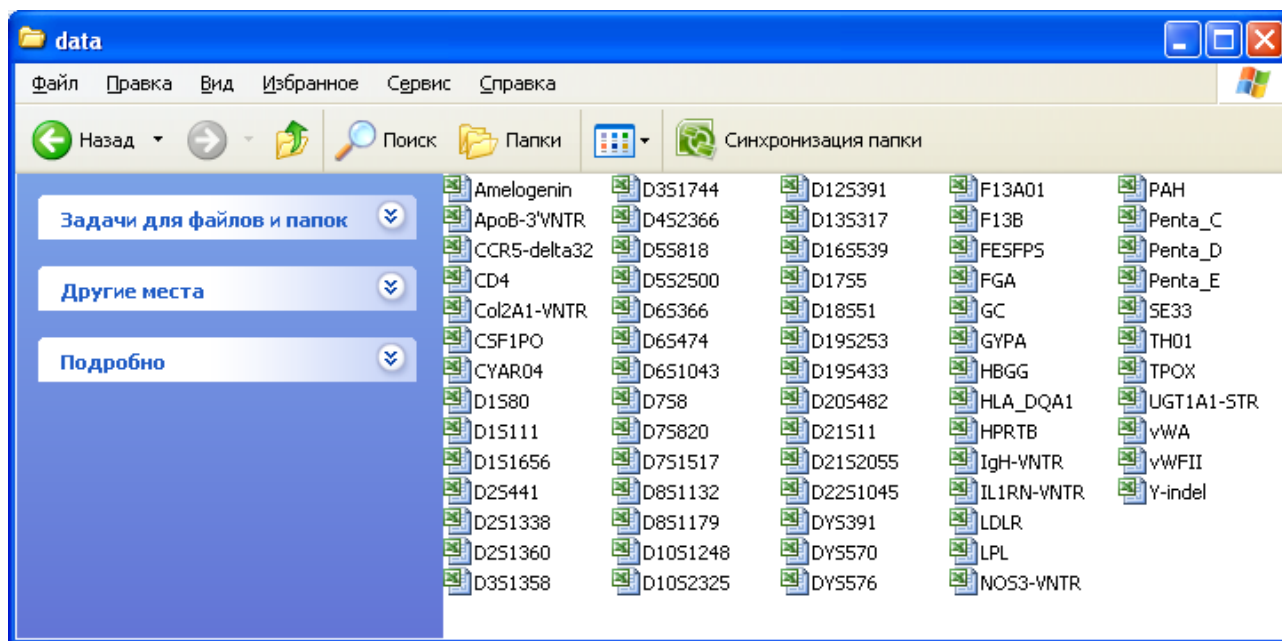
Узнать имя авторизованного пользователя и дополнительную информацию можно, нажав кнопку «*О программе*», например:



Каталог *data*: служебные файлы для используемых локусов

Для успешной работы программы в одном каталоге с исполняемым файлом *pindex.exe* обязательно должен находиться непустой каталог (папка) *data*. Допустимы такие его названия, как *Data*, *DATA*.

Все служебные файлы в этом каталоге являются общими для программ *Pindex* и *Gency*. Это файлы формата *Microsoft Excel*, содержащие данные об эталонных частотах аллелей по используемым локусам. Пример файлов в папке *Data* показан на рисунке ниже:



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*), *европеиды США* (листы *имя_файла-white*), *негры США* (листы *имя_файла-black*) и *любая локальная популяция* (листы *имя_файла-other*).



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

Также для корректной работы программы названия служебных файлов не должны превышать 16 символов.

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

Пример служебного файла для локуса *D16S539* приведён на рисунке ниже. На этом рисунке виден первый лист из четырёх, с частотами аллелей для русской популяции.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	<5	0,001	<94	0,00000											
2	5	0,001	94	0,00000											
3	6	0,001	98	0,00000											
4	7	0,003	102	0,00135											
5	8	0,016	106	0,01482											
6	9	0,085	110	0,08356											
7	10	0,054	114	0,05256											
8	11	0,284	118	0,28302											
9	12	0,350	122	0,34906											
10	13	0,192	126	0,19137											
11	14	0,024	130	0,02291											
12	15	0,003	134	0,00135											
13	>15	0,001	>134	0,00000											
14	новый	0,001		0,00000											
15	Аллельные частоты по данным Zhivotovsky et al., 2009, PMID: 19647694; популяционная выборка 371 неродственный человек. Суммарный размер аллелей, в парах нуклеотидов (наборы ТАПОТИЛИ)														
16	«Консервативная» оценка частот аллелей проведена для исследованной выборки (последний столбец таблицы) согласно рекомендациям Gjertson et al.														
17	Нумерация аллелей согласно числу tandemных повторов														
18	Аллельная «лестница» к наборам ТАПОТИЛИ включает 8 аллелей: 7, 8, 9, 10, 11, 12, 13, 14. Эти аллели выделены цветом в первом столбце.														

Обозначение аллелей для подавляющего большинства локусов является стандартным международным: для микро- и мини- сателлитных локусов указано, что «*нумерация аллелей согласно числу tandemных повторов*». В редких случаях альтернативного обозначения аллелей это также указано в соответствующих файлах: например, для локуса *PAN*.

Дополнительно **заливка** красным цветом использована для ячеек с нулевыми значениями частот аллелей в исследованных выборках.

В служебных файлах для каждого локуса находится некоторая дополнительная информация, непосредственно не используемая при работе программы. В частности, указаны соответствующие литературные первоисточники для референтных частот аллелей (на уровне *DOI* или *PubMed ID*), размеры аллелей при использовании отдельных пар праймеров, порядок нумерации аллелей, состав аллельных «лестниц» (для локусов в наборах *Tanotили*).

Например, для локуса *D16S539* на представленном выше рисунке видно, что частоты аллелей для русской популяции были взяты из работы *Zhivotovsky et al., 2009* (PMID: [19647694](https://pubmed.ncbi.nlm.nih.gov/19647694/)) с размером популяционной выборки в 371 неродственного человека. Эти значения аллельных частот (из самой статьи) приведены в четвёртом столбце («D»). При желании полный текст этой статьи и дополнительные материалы к ней можно найти в папке *References* по указанному номеру *PMID*.

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:

- (i) 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;
- (ii) 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;
- (iii) 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;
- (iv) 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*.

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.

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.

К настоящему времени служебные файлы созданы для более **100** разных локусов генома человека, в том числе 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 (в приведенном выше примере для локуса *D16S539* задействуются только ячейки ***A1-B14***), информация в остальных ячейках программой игнорируется. Частоты аллелей считываются программой из столбцов «*B*» с точностью до трёх знаков после запятой.

Пользователь может по своему усмотрению менять нумерацию аллелей и эталонные частоты для всех локусов, вводить и редактировать любую дополнительную информацию, а также создавать новые аналогичные файлы для любых других локусов. Смотрите раздел **Работа с программой: использование различных эталонных частот аллелей**.



Четвёртые листы каждого файла (листы *имя_файла-other*) как раз и предназначены для задания частот аллелей самим пользователем (для актуальной ему 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 *Gency* 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.

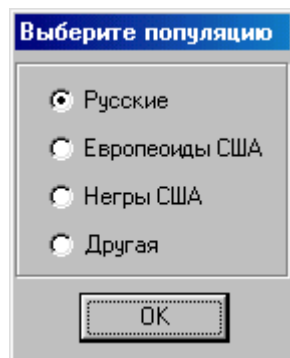
Каждый рабочий лист каждого служебного файла содержит краткую информацию об источнике соответствующих популяционных данных с отсылкой к соответствующей публикации. Например, для локуса *D18S51* в листе *d18s51-white* указано следующее:

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 стартовом меню нужную эталонную популяцию. Для расчётов программа будет адресоваться к соответствующим частотам аллелей:



После появления рабочего листа следует выбрать из списка нужный локус и указать для него наблюдаемые комбинации генотипов ребёнка и заявленных родителей (или только одного из родителей):

Когда для выбранного локуса указаны все генотипы, следует на всякий случай проверить указанные данные и, если всё верно, нажать кнопку **«Расчитать»**. При этом в рабочем поле программы появятся определённые значения и формулы (см. рисунок на следующей странице).

Пaternity Index - C:\Program Files\Pindex\sample.xls

Рассчитать Перемножить Очистить Сохранить Загрузить Популяция Справка О программе

Локус: **PAH**

Предполагаемый отец:
Аллель 1: **4** Аллель 2: **4**

Ребенок:
Аллель 1: **4** Аллель 2: **6**

Мать:
Аллель 1: **3** Аллель 2: **6**

№	Локус	Формула	Частота аллеля	Частота аллеля	LR	Выбрать
1	IL1RN	$1/(p(2)+p(4))$	$p(2)=0,300$	$p(4)=0,630$	1,08	
2	IgH	$1/2p(10)$	$p(10)=0,380$		1,32	
3	CD4	$1/2p(5)$	$p(5)=0,370$		1,35	
4	LPL	$1/2p(10)$	$p(10)=0,480$		1,04	
5	D3S1358	$1/2p(18)$	$p(18)=0,140$		3,57	
6	CYAR04	$1/p(6)$	$p(6)=0,400$		2,50	
7	CSF1PO	$1/2p(13)$	$p(13)=0,080$		6,25	
8	TH01	$1/p(6)$	$p(6)=0,220$		4,55	
9	vWFII	$1/2p(12)$	$p(12)=0,340$		1,47	
10	PAH	$1/p(4)$	$p(4)=0,300$		3,33	

Индекс отцовства (Combined Paternity Index) CPI =

Априорная вероятность (prior probability) pp = Вероятность отцовства P =

Далее нужно выбрать из списка следующий локус и повторить для него описанную процедуру: ввод генотипов → визуальная проверка → нажатие кнопки «**Рассчитать**».

Аналогичным образом для всех проанализированных локусов поэтапно рассчитываются *отношения правдоподобия* (*Likelihood Ratios*, столбец «**LR**» рабочего поля). На приведённом выше рисунке изображено рабочее поле программы при работе с файлом *sample.xls*.



Указанные пользователем комбинации генотипов по отдельным локусам не выводятся на рабочем поле программы, но сохраняются в оперативной памяти и будут сохранены в файле результатов.

При необходимости на любом этапе работы можно полностью очистить рабочее поле программы (кнопка «**Очистить**», при этом все введённые данные будут утеряны) или сохранить результаты работы в отдельном файле формата *Microsoft Excel* (кнопка «**Сохранить**»).

Неверные данные для отдельных локусов могут быть удалены из рабочего листа программы (выборочно построчно). Для этого необходимо поместить курсор мыши на строку с ошибочными данными в рабочем поле, нажать правую кнопку мыши и выбрать команду «**Удалить**».

Результаты работы можно сохранить в файле с произвольным именем в любом каталоге. Для открытия (редактирования) ранее созданного рабочего файла нужно выбрать команду **«Загрузить»**. Программа может обрабатывать только один рабочий файл. Имя и путь к уже открытому рабочему файлу указываются на синем фоне в верхней части окна программы. В приведённом выше примере файл сохранён под именем *sample.xls* в папке *C:\Program Files\Pindex*.

Работа с программой: расчёт индекса и вероятности отцовства

Предположим, что для всех проанализированных локусов, исходя из их хромосомной локализации, предполагается независимое наследование аллелей (отсутствие физического сцепления между локусами). В этом случае все значения отношений правдоподобия (LR) для отдельных локусов могут быть перемножены и рассчитан *комбинированный индекс отцовства* (*Combined Paternity Index, CPI*). Такой алгоритм расчёта называется *правило произведения*, или *multiplication rule*:

$$CPI = LR (\text{локус } 1) * LR (\text{локус } 2) * * LR (\text{локус } n)$$

Соответственно, на следующем этапе работы пользователь выбирает те локусы, которые будут использоваться для расчёта индекса отцовства. Для этого нужные строки в рабочем поле программы помечаются левой кнопкой мыши в столбце «**Выбрать**» (при этом в соответствующих ячейках появляются символы ++), а затем в верхнем меню нажимается кнопка «**Перемножить**», как это можно видеть на рисунке ниже:

The screenshot shows the Pindex software interface. The main window is titled "Paternity Index - C:\Program Files\Pindex\sample.xls". The menu bar includes "Рассчитать", "Перемножить", "Очистить", "Сохранить", "Загрузить", "Популяция", "Справка", and "О программе". On the left, there are input fields for "Локус" (set to "РАН"), "Предполагаемый отец" (Allele 1: 4, Allele 2: 4), "Ребенок" (Allele 1: 4, Allele 2: 6), and "Мать" (Allele 1: 3, Allele 2: 6). The main table has columns: №, Локус, Формула, Частота аллеля, Частота аллеля, LR, and Выбрать. The table lists 10 loci. The "Выбрать" column shows "++" for loci 2 through 10, indicating they are selected for calculation. At the bottom, the "Индекс отцовства (Combined Paternity Index) CPI" is calculated as 1706. The "Априорная вероятность (prior probability) pp" is 0,5, and the "Вероятность отцовства P" is 99,94% (pp=0,5).

№	Локус	Формула	Частота аллеля	Частота аллеля	LR	Выбрать
1	IL1RN	1/(p(2)+p(4))	p(2)=0,300	p(4)=0,630	1,08	
2	IgH	1/2p(10)	p(10)=0,380		1,32	++
3	CD4	1/2p(5)	p(5)=0,370		1,35	
4	LPL	1/2p(10)	p(10)=0,480		1,04	++
5	D3S1358	1/2p(18)	p(18)=0,140		3,57	++
6	CYAR04	1/p(6)	p(6)=0,400		2,50	++
7	CSF1PO	1/2p(13)	p(13)=0,080		6,25	++
8	TH01	1/p(6)	p(6)=0,220		4,55	++
9	vWFII	1/2p(12)	p(12)=0,340		1,47	++
10	РАН	1/p(4)	p(4)=0,300		3,33	++

Индекс отцовства (Combined Paternity Index) CPI = 1706
 Априорная вероятность (prior probability) pp = 0,5 Вероятность отцовства P = 99,94% (pp=0,5)

В результате в нижней части окна программы выводятся значения индекса и вероятности отцовства. Так, в приведённом примере значение комбинированного индекса отцовства составляет **1 706**, и для расчёта CPI использовались данные по 8 из 10 локусов (не учитывались значения LR для локусов *IL1RN* и *CD4*).

Для того, чтобы дополнительно исключить (включить) какой-либо локус из расчётов, следует убрать (добавить) для него пометку ++ в соответствующей ячейке левой кнопкой мыши. В этом случае после повторного нажатия кнопки «*Перемножить*» значения индекса и вероятности отцовства будут пересчитаны с учётом внесённых изменений.

По умолчанию, *апостериорное* (после молекулярно-генетического исследования) значение вероятности отцовства (P , *probability*) рассчитывается из CPI , полагая значение *априорной* вероятности (до начала молекулярно-генетического исследования) равным 50%. Более подробно это обсуждается в разделе **Вероятность отцовства при различных значениях априорной вероятности**.

Сохранение результатов и завершение работы с программой

Указанные генотипы по всем локусам и результаты расчётов индекса и вероятности отцовства сохраняются в отдельном файле результатов формата *Microsoft Excel* (кнопка «*Сохранить*»). В файл результатов также записывается информация об использованной для расчётов эталонной популяции. Этот файл может быть непосредственно использован при оформлении заключения эксперта (ЗЭ).

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



При этом будет предложено сохранить изменения (или не делать этого) в текущем рабочем файле.

По окончании работы настоятельно рекомендуется открыть файл результатов уже в редакторе *Microsoft Excel* и окончательно проверить все указанные генотипы.



Если при такой проверке обнаружится, что какие-то генотипы были введены неправильно, то **не следует** вносить изменения в редакторе *Microsoft Excel*. Это обусловлено тем, что для определённых комбинаций генотипов (их сочетания) используются разные соответствующие формулы расчётов. Поэтому вполне вероятно, что изменение даже одного генотипа повлечёт изменение соответствующей формулы расчёта *LR*. ***Такие ошибки уже были отмечены в экспертной практике.***

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

Использование файла результатов в тексте «Заключения эксперта»

Можно открыть файл результатов в редакторе *Microsoft Excel*, скопировать из него ячейки с данными и вставить их в нужное место документа *Microsoft Word*. Для файла *sample.xls* получится примерно так:

№	Локус	Генотип отца	Генотип ребёнка	Генотип матери	Формула	Частота аллеля	Частота аллеля	LR	Выбрать
1	IL1RN	2/4	2/4	2/4	$1/(p(2)+p(4))$	$p(2)=0,300$	$p(4)=0,630$	1,08	
2	IgH	10/12	10/10	10/10	$1/2p(10)$	$p(10)=0,380$		1,32	++
3	CD4	5/12	5/6	6/10	$1/2p(5)$	$p(5)=0,370$		1,35	
4	LPL	10/11	10/10	10/10	$1/2p(10)$	$p(10)=0,480$		1,04	++
5	D3S1358	17/18	16/18	16/16	$1/2p(18)$	$p(18)=0,140$		3,57	++
6	CYAR04	6/6	6/11	11/11	$1/p(6)$	$p(6)=0,400$		2,50	++
7	CSF1PO	12/13	11/13	11/11	$1/2p(13)$	$p(13)=0,080$		6,25	++
8	TH01	6/6	6/6	6/9.3	$1/p(6)$	$p(6)=0,220$		4,55	++
9	vWFII	11/12	11/12	11/11	$1/2p(12)$	$p(12)=0,340$		1,47	++
10	PAH	4/4	4/6	3/6	$1/p(4)$	$p(4)=0,300$		3,33	++
Индекс отцовства (Combined Paternity Index) CPI=1706									
Вероятность отцовства P=99,94% (prior probability pp=0,5)									
Использовались референтные частоты аллелей для русских									

Затем эту таблицу при желании можно преобразовать в более наглядную форму. Например, убрать избыточные (пустые) ячейки, расставить локусы в порядке увеличения значений *LR*, выбрать необходимый стиль форматирования таблицы.

Окончательно, таблица результатов будет выглядеть примерно так, как это показано на рисунке на следующей странице.

**Генотипы, установленные для обследованных лиц,
и результаты расчётов отношений правдоподобия (LR), индекса и вероятности
отцовства для исследованного случая спорного отцовства**

№	Локусы	Генотип отца	Генотип ребёнка	Генотип матери	Формулы расчёта LR	Частоты аллелей	Значения LR
1	LPL	10/11	10/10	10/10	$1/2p(10)$	$p(10)=0,48$	1,04
2	IL1RN	2/4	2/4	2/4	$1/(p(2)+p(4))$	$p(2)=0,30$ $p(4)=0,63$	1,08 (-)
3	IgH	10/12	10/10	10/10	$1/2p(10)$	$p(10)=0,38$	1,32
4	CD4	5/12	5/6	6/10	$1/2p(5)$	$p(5)=0,37$	1,35 (-)
5	vWFII	11/12	11/12	11/11	$1/2p(12)$	$p(12)=0,34$	1,47
6	CYAR04	6/6	6/11	11/11	$1/p(6)$	$p(6)=0,40$	2,50
7	PAH	4/4	4/6	3/6	$1/p(4)$	$p(4)=0,30$	3,33
8	D3S1358	17/18	16/18	16/16	$1/2p(18)$	$p(18)=0,14$	3,57
9	TH01	6/6	6/6	6/9.3	$1/p(6)$	$p(6)=0,22$	4,55
10	CSFIPO	12/13	11/13	11/11	$1/2p(13)$	$p(13)=0,08$	6,25
Индекс отцовства (Combined Paternity Index, CPI)							1 706
Вероятность отцовства (prior probability pp=0,5)							99,94%
Использовались референтные частоты аллелей для русских							
(-) Значения LR, которые не были использованы для расчёта CPI							

Работа с программой: использование различных эталонных частот аллелей

Нужную эталонную популяцию можно сразу выбрать при очередном запуске программы в стартовом меню. В этом случае после указания генотипов и нажатия кнопки **«Рассчитать»** значения LR будут рассчитаны по соответствующим эталонным частотам.

При необходимости пересчитать значения LR по другим референтным частотам аллелей следует нажать кнопку **«Популяция»** и выбрать нужную эталонную популяцию. Значения LR будут пересчитаны для всех локусов.

Предположим, что частоты аллелей, приведённые в служебных файлах, по каким-либо причинам были изменены пользователем. Тогда при очередном запуске программы ранее созданный рабочий файл можно загрузить двумя способами (кнопка **«Загрузить»**).

В выпадающем меню этой кнопки можно выбрать **«генотипы и старые значения LR »**. В этом случае программа просто загрузит ранее сохраненные данные из соответствующего файла (будут приведены значения LR , рассчитанные по старым частотам аллелей). Для пересчёта значений LR с учётом внесенных изменений в эталонные частоты следует нажать кнопку **«Популяция»** и выбрать популяцию.

Если же в выпадающем меню кнопки **«Загрузить»** выбрать **«генотипы и рассчитать LR »**, то при загрузке файла результатов все значения LR будут автоматически пересчитаны, с учётом последних актуальных изменений в эталонных аллельных частотах для всех локусов.

Наглядно в этом убедиться можно, поэкспериментировав с файлом *sample.xls*. Сам этот файл был создан в 2007 г, и к настоящему времени для многих локусов эталонные частоты в соответствующих файлах уже изменены.

Объективно, существуют некоторые затруднения для корректного расчёта значений LR по отдельным референтным популяциям. Это касается как отсутствия внятных популяционных данных по отдельным локусам, так и нулевых значений частот для «невывявленных» аллелей в конкретных исследованных выборках. Более подробно этот вопрос обсуждается в разделе **Решение затруднений**.

Как ссылаться на программу при оформлении Заключения эксперта?

Нужно активно ссылаться! Ибо давно уже пора изгнать мракобесие (заодно – и бесовщину) в расчётах, явно и исподволь насаждаемых известным идеологом из РЦ СМЭ (г. Москва). Особо осторожным экспертам, не понаслышке внимавшим *ивановской* (не путать с *сурковской*) пропаганде, можно посоветовать использовать в тексте ЗЭ четыре ссылки ниже, обосновывающие использованные в *Пиндексе* алгоритмы расчётов:

- *Морлинг и др., 2003*: международные рекомендации по стандартам проверки отцовства, перевод на русский язык опубликован в журнале «Проблемы стандартизации в здравоохранении»;
- *Gjertson et al., 2007*: международные рекомендации по биостатистике при производстве молекулярно-генетических исследований случаев спорного отцовства (опубликованы только на английском);
- *Forensic DNA evidence interpretation, 2005*: в главе 10 непосредственно приведены формулы расчётов для различных комбинаций генотипов (английский);
- *Ефремов И.А., Серегин Ю.А., 2009*: свидетельство о регистрации *Pindex'a* регистратором.

Получены также «Акты внедрения» на программу:

- ФГКУ «111 Главный государственный центр судебно-медицинских и криминалистических экспертиз Министерства обороны Российской Федерации», г. Москва, 2015.
- ГБУЗ «Самарское областное бюро судебно-медицинской экспертизы», г. Самара, 2015.
- Филиал №2 ФГКУ «111 Главный государственный центр судебно-медицинских и криминалистических экспертиз Министерства обороны Российской Федерации», г. Ростов-на-Дону, 2018.

Дополнительно смотрите список литературы в соответствующем разделе этого РП и папку *references*.

Отношения правдоподобия: стандартная практика в ДНК-исследованиях

Естественно, и у двух неродственных между собою человек по нескольким исследованным локусам могут быть выявлены одинаковые (совпадающие) аллели. Поэтому значимость выявленных совпадений при проведении молекулярно-генетических экспертиз спорного отцовства (родства) обязательно подлежит численной оценке.

Величина LR показывает, во сколько раз более вероятно, что наблюдаемое совпадение аллелей у заявленного отца и ребёнка обусловлено тем, что именно этот мужчина является биологическим отцом, нежели истинным отцом является некто другой (случайно выбранный из популяции мужчина). Другими словами, LR является отношением вероятности того, что заявленный мужчина является отцом ребёнка (числитель) к вероятности того, что отцом этого ребёнка является другой мужчина (знаменатель).

Комбинированный индекс отцовства CPI (перемноженные значения LR по всем локусам) можно интерпретировать аналогичным образом:

исходя из полученных результатов исследования ДНК, в CPI раз более вероятно, что именно обследуемый мужчина является биологическим отцом (передал ребёнку «отцовский набор аллелей»), чем такие результаты объясняются случайным совпадением отдельных аллелей между генотипами обследованных лиц.

Расчёт отношений правдоподобия в настоящее время является стандартной международной практикой при оценке значимости выполненных молекулярно-генетических экспертиз спорного родства. Это прямо указано в Рекомендациях Международного общества по судебной генетике, *ISFG: Morling et al., 2002* (русский перевод *Морлинг и др., 2003*); *Gjertson et al., 2007*. Используемые в настоящей программе алгоритмы описаны во многих (преимущественно – англоязычных) источниках: смотрите список литературы и статьи в папке *references*.

Интерпретация результатов: априорные условия, о которых следует помнить

В реализованном алгоритме расчёта индекса и вероятности отцовства за основу взяты следующие априорные условия:

- 1) для гомозиготных генотипов не предполагается «ложной гомозиготности»;
- 2) материнство рассматривается как бесспорное;
- 3) полагается полная невозможность мутаций;

4) помимо заявленного мужчины, истинным отцом ребёнка может быть любой случайный мужчина из этой же эталонной популяции, не подвергнутый ДНК-исследованию и не являющийся родственником обследуемого мужчины.

Исследователю, выполняющему такие расчёты, необходимо совершенно отчётливо понимать, что независимо от полученной величины *CPI*, 100% вероятности отцовства в случаях подтверждения (не-исключения) недостижимы в принципе. Хотя бы по той причине, что у однояйцевых близнецов ДНК абсолютно одинаковая, и с таким же успехом отцом ребёнка может быть брат-близнец предполагаемого отца: эксперт в общем случае не знает, есть на самом деле такой брат-близнец у обследуемого мужчины или нет.

Дополнительно, независимо от числа исследованных локусов, сохраняется вероятность того, что где-нибудь на островах Океании или в соседнем городе проживает *мужчинка* с точно таким же набором аллелей по всем этим локусам, как у обследованного нами мужчины, и именно этот «сосед» является истинным отцом обследованного ребёнка.

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

Какие случаи можно корректно анализировать?

Использование программы полностью корректно в случаях, когда обследуются один предполагаемый отец и один ребёнок (с биологической матерью или без неё). Это подавляющее большинство практических случаев.

Если в отдельном случае одновременно обследуются **двое или более детей** от одной биологической матери, с одним предполагаемым отцом, то расчёт индекса и вероятности отцовства в отношении каждого из детей по отдельности **является некорректным**. Однако эта тема достаточно ёмкая и требует отдельного обсуждения.

Также программа требует более осмысленного использования в тех случаях, когда одновременно обследуются два или более предполагаемых отцов в отношении одного и того же ребёнка.

Предположим, из двоих «предполагаемых отцов» один мужчина исключился. Тогда для второго мужчины, который не исключился в ходе исследования, можно, конечно, рассчитать индекс и вероятность отцовства, но возникает вопрос корректного выбора значения априорной вероятности: 50%, 33% или другая величина? В целом, расчёт *CPI* для этого мужчины в программе *Пиндекс* не является полностью корректным: не выполняется п. 4 априорных условий предыдущего раздела. Дополнительно смотрите также раздел **Вероятность отцовства при различных значениях априорной вероятности**.

Программу также можно использовать и для анализа случаев **спорного материнства**: когда обследуются ребёнок и предполагаемая мать. В этом случае генотипы предполагаемой матери нужно указывать в поле «*Предполагаемый отец*», а в поле «*Мать*» выбирать значения «*Нет*» для аллелей 1 и 2. Формулы расчётов ОП для индекса родительства (отцовства или материнства) при обследовании пары родитель-ребёнок являются одинаковыми. Такие исследования актуальны при возможном перепутывании детей в родильных домах.

Вероятность отцовства: что это такое?

По сравнению с индексом отцовства, вероятность отцовства – это гораздо более спорный показатель, но также определяющий значимость выполненной экспертизы. В настоящее время в мире достаточно широко распространена практика преобразования индекса отцовства (соотношение двух вероятностей, «за» и «против») в вероятность отцовства (*Probability of Paternity*, P – вероятность «за»). Она стандартно оценивается на основании *формулы Байеса для условной вероятности*.

В случае принятия *априорной вероятности (prior probability)* $pp=0,5$ (50% «за» и 50% «против»), соотношение априорных шансов 1:1) вероятность отцовства P связана с индексом отцовства CPI достаточно простой формулой:

$$P = CPI / (1 + CPI)$$

Как общее правило, в мировой практике при расчёте вероятности отцовства обязательно требуется прямое указание соответствующего значения априорной вероятности. Допустимо также указание нескольких значений вероятности отцовства при различных величинах априорной вероятности.

С нашей точки зрения, наиболее корректно использование только индекса отцовства (необходимо и достаточно).

Вероятность отцовства при различных значениях априорной вероятности

Чаще всего при указании значений вероятности отцовства *априорная вероятность* (*prior probability, pp*) в явной или неявной форме полагается равной **0,5**. Однако такой выбор является достаточно условным, да и в целом определение априорной вероятности в каждом конкретном случае находится вне компетенции эксперта / специалиста (а находится, скорее, в компетенции суда). Возможность непредвзятого выбора априорной вероятности отдельными авторами ставится под сомнение, либо вообще отвергается. С практической точки зрения лаборатории, выполняющей такие исследования, выбор $pp=0,5$ говорит о том, что в половине исследуемых случаев отцовство исключается.

Тем не менее, вероятность отцовства может рассчитываться при любых значениях априорной вероятности по общей формуле:

$$P = (pp * CPI) / (pp * CPI + 1 - pp)$$

Например, полагая $pp=0,5$, мы получим формулу, приведённую в предыдущем разделе:

$$P = (0,5 * CPI) / (0,5 * CPI + 1 - 0,5) = 0,5 * CPI / (0,5 * CPI + 0,5) = CPI / (1 + CPI)$$

Естественно, величины априорной вероятности отцовства находятся в диапазоне от 100% (абсолютное подтверждение отцовства, $pp=1,0$) до 0% (абсолютное отрицание отцовства, $pp=0,0$). Таким образом, диапазон допустимых значений априорной вероятности составляет $0,0 \leq pp \leq 1,0$. В Pindex'e реализован расчёт вероятности отцовства при произвольных значениях априорной вероятности в этом диапазоне, на усмотрение пользователя.

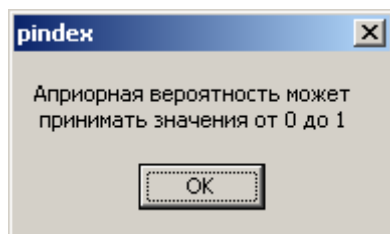
Значение априорной вероятности приводится в левом нижнем углу рабочего листа программы. При запуске по умолчанию $pp=0,5$:

Мать
Аллель 1 Аллель 2
[] []

Индекс отцовства (Combined Paternity Index) CPI = []

Априорная вероятность (prior probability) pp = 0,5 Вероятность отцовства P = []

При желании в поле для априорной вероятности можно ввести любое значение pp с любым количеством десятичных знаков. Естественно, программа воспринимает значения априорной вероятности только в корректном диапазоне (от 0 до 1 включительно). В противном случае выводится следующее сообщение об ошибке:



После ввода нового значения *априорной* вероятности и нажатия кнопки «Перемножить» значение *апостериорной* вероятности отцовства будет пересчитано. Величина индекса отцовства останется, естественно, неизменной. При нажатии кнопки «Сохранить» в файл результатов запишется последнее заданное значение априорной вероятности и соответствующая ему величина *апостериорной* вероятности отцовства.

С точки зрения эксперта, выбор $pp=0,1$ (1/10) говорит о том, что из 10 исследуемых аналогичных случаев только в одном случае отцовство подтверждается (*априори* в 10 раз более вероятно исключение). Для индекса отцовства $CPI=1\,706$ (см. файл *sample.xls*) вероятность отцовства при таком значении априорной вероятности будет рассчитываться следующим образом:

$$P = (0,1 * 1706) / (0,1 * 1706 + 1 - 0,1) = 170,6 / (170,6 + 0,9) = 170,6 / 171,5 = 0,99475 = \mathbf{99,48\%}$$

Наоборот, полагая $pp=0,9$ (9/10, из 10 исследуемых аналогичных случаев в девяти случаях отцовство подтверждается), вероятность отцовства составит:

$$P = (0,9 * 1706) / (0,9 * 1706 + 1 - 0,9) = 1535,4 / 1535,5 = 0,99993 = \mathbf{99,99\%}$$

В целом – разница между значениями для апостериорной вероятности невелика. Но следует помнить, что точность – это вежливость королей!

Отношения правдоподобия: пример расчёта

Осмысленное использование расчётов подразумевает их полное понимание. Поэтому пояснение по формулам на одном примере.

Предположим, по отдельному локусу в результате выполненного исследования установлена следующая комбинация генотипов:

мать	ребёнок	предполагаемый отец
pp	pq	pq

Решётка Пуннетта (*Punnett square*) для такой комбинации генотипов рисуется следующим образом:

		Аллели, наследуемые от матери	
		p	p
Аллели, наследуемые от отца	p	pp	pp
	q	pq	pq

Легко видеть, что мать с вероятностью 100% передаёт ребёнку аллель p , а обследуемый мужчина (предполагаемый отец с установленным генотипом pq) может передать «недостающий» аллель q лишь с вероятностью 50%. Соответственно, числитель для LR равен $1 \cdot 1/2 = 1/2$. Иными словами, вероятность рождения у таких родителей ребёнка с заданным генотипом pq равна 50%.

Знаменатель рассчитывается следующим образом. От матери по-прежнему ребёнок с вероятностью 100% получает аллель p . Тогда аллель q он должен унаследовать от «случайного» отца, который может быть или гомозиготой qq , или гетерозиготой qr (или даже pq , в данном случае это уже неважно).

Частота встречаемости в популяции гомозиготного мужчины равна q^2 , и с вероятностью 100% он передаст ребёнку аллель q . Вероятность рождения ребёнка с заданным генотипом pq в этом случае составит $1 \cdot 1 \cdot q^2 = q^2$.

Частота встречаемости в популяции гетерозиготного по аллелю q мужчины равна $2q(1-q)$, и с вероятностью 50% он передаст ребёнку аллель q . У такой пары родителей вероятность рождения ребёнка с заданным генотипом составит $1 \cdot 1/2 \cdot 2q(1-q) = q(1-q)$.

Окончательно, знаменатель для LR будет равен сумме этих двух вероятностей: $q^2 + q(1-q) = q$.

Окончательно, отношение правдоподобия для такой комбинации генотипов составит:

$$LR = 1 / (2q)$$

Аналогичным образом могут быть выведены формулы для всех остальных возможных комбинаций генотипов.

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. Аналогично, перемножение отношений правдоподобия для получения суммарного индекса родства (отцовства) полностью корректно только для несцепленных между собой маркёров. Таким образом, для правомерного использования «*правила перемножения*» частот генотипов (или ОП) 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** (**Combined DNA Index System**) 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.

Анализы уровня «полное генетическое обследование для мужчин (женщин)» обычно включают не менее 40 различных генов. В случае такого семейного обследования (*папа-мама-ребёнок*) вполне резонно (допустимо) использовать результаты (> 40 полиморфных

маркёров) и для подтверждения родства обоих родителей с ребёнком, тем более если при таком медико-генетическом консультировании речь идёт о планировании последующих детей. Подтверждение родства в этом случае подразумевает численную оценку значимости выявленных совпадений генотипов в терминах ОП и индексов отцовства (материнства).

Однако исключение родительства (ожидаемо – отцовства) по результатам такого теста переводит вопрос уже в поле биомедицинской этики, поскольку заявителями исследования (родителями) такой вопрос изначально не ставился, и это сохраняет их «*право не знание*». В этом случае для врача-генетика исключающие родство результаты свидетельствуют или о технической ошибке тестов по отдельным локусам или же о необходимости скорректировать сам вектор генетической консультации.

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: *ACE-SNP*, *ACTN3-SNP*, *CCR5-delta32*, *D7S8*, *GYPA* (*glycophorin A*), *HFE-SNP*, *HLA_DQA1*, *IL1RN-VNTR*, *LDLR* (*low density lipoprotein receptor*), *MTHFR-SNP*, *NOS3-VNTR-SNP*, *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).

Очевидно, что результаты установления именно половой принадлежности обследованных лиц никоим образом не влияют на значения индекса отцовства (материнства).

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. И расчёт ОП выполняется как индекс именно *патрилинейного* родства.

Таким образом, использовать для расчётов *LR* «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. По этой причине при расчётах индекса отцовства необходимо использовать соответствующие корректные формулы.

Следует отметить, что ребёнок мужского пола не наследует от отца хромосому X. Поэтому исследование полиморфных локусов этой хромосомы в экспертизах спорного отцовства уместно лишь в отношении детей-девочек. При этом биологические сёстры получают от общего отца идентичные копии X-хромосомы. Поэтому у женщин, имеющих такого общего отца, во всех локусах хромосомы X будет присутствовать как минимум один общий (одинаковый) аллель.

Исследование маркёров хромосомы X при детях любого пола эффективно для разрешения вопросов именно спорного материнства. Однако такие расчёты являются совершенно отдельной темой.



Таким образом, для полиморфных маркёров, расположенных на хромосоме X человека, программу *Pindex* (версия 1.0) вообще **не следует использовать для расчётов индекса материнства**. Использование программы *Pindex* уместно только для случаев спорного отцовства в отношении детей-девочек (при наличии матери или без неё).

Соответствующие формулы расчётов индекса отцовства (PI) для локусов, расположенных на хромосоме X, подробно разобраны в работах [Ayres & Powley, 2005; Ефремов, 2010; Ефремов, Кожемяко, 2011]. Для мужчин (предполагаемых отцов) вероятность отцовства P рассчитывается из значений PI точно таким же образом, как и для аутосомных локусов. Аналогично, допустимо использование и различных значений априорной вероятности.

При расчётах индекса отцовства 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).

Версию 1.0 программы *Pindex* можно использовать для корректных расчётов индекса отцовства по любым полиморфным локусам, расположенным на хромосоме X человека. В настоящее время готовый к использованию служебный файл создан только для локуса *HPRTB* (файл *HPRTB.xls*). В служебном файле *HPRTB.xls* на листе 1 указано следующее важное предупреждение:

Внимание! Этот маркёр локализован на хромосоме X человека. Для правильного расчёта индекса отцовства гемизиготный генотип предполагаемого отца следует вводить в «бинарном» формате: 6/6, 7/7 и т.д.

Приведённый ниже рисунок поясняет, как следует вводить в программу в «бинарном» формате гемизиготный генотип отца, у которого выявлен аллель 12 по локусу *HPRTB*:

После переноса сводной таблицы генотипов и расчётов из файла результатов в текст «Заключения эксперта» генотипы отца по маркёрам хромосомы X необходимо отредактировать вручную. Это будет выглядеть примерно так:

Неправильный генотип, указанный в таблице,

№	Локус	Генотип отца	Генотип ребёнка	Генотип матери	Формула	Частота аллеля	Частота аллеля	LR
1	HPRTB	12/12	12/12	12/12	1/p(12)	p(12)=0,360		2,78

исправляется вручную до **правильного** гемизиготного:

№	Локус	Генотип отца	Генотип ребёнка	Генотип матери	Формула	Частота аллеля	Частота аллеля	LR
1	HPRTB	12	12/12	12/12	1/p(12)	p(12)=0,360		2,78

В следующей версии программы *Pindex* (если она, конечно, выйдет) этот недостаток будет устранён: исходно программа создавалась для расчётов только по аутосомным локусам.

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 европеоидов 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 *PI* при расчёте комбинированного *CPI* нельзя просто перемножать (некорректно). Для таких сцепленных локусов при расчётах *CPI* следует оперировать не частотами отдельных аллелей по каждому маркеру, а частотой гаплотипа, то есть набора аллелей, локализованных на одной из парных хромосом по всем этим маркерам. При этом формулы расчётов индекса отцовства 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 *CPI* следует использовать **только одно** значение *PI* (для одного маркера) из каждой группы сцепления.

Программа Gency

Программа **Gency** (*Genotype frequency*) – это отдельная компьютерная программа для расчёта вероятности случайного совпадения комбинированного генотипа при исследованиях ДНК. Она использует в своей работе те же самые рабочие файлы, находящиеся в папке *data*, что и *Pindex*.

Gency предназначена для расчёта вероятности случайного совпадения комбинированного генотипа для анализируемого биологического образца (ДНК-профиля). В русскоязычной литературе для этого термина используются следующие синонимы: *вероятность случайного совпадения, вероятность случайного совпадения генетических признаков, вероятность случайного совпадения генотипов*. В англоязычной литературе – *random match probability (RMP), DNA profile probability, DNA profile population frequency, probability of a random match, chance of a random match, genotype frequency...*

Программа может быть использована в экспертной практике криминалистических, судебно-медицинских, гистологических и других лабораторий, занимающихся идентификацией происхождения биологических образцов (идентификацией личности).

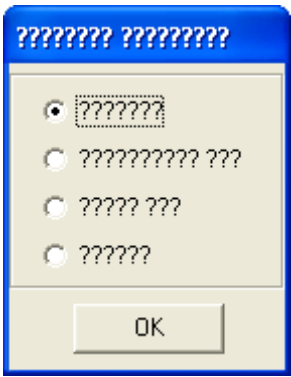
В своей практике мы используем эту программу для численной оценки значимости выявленных совпадений генотипов при установлении индивидуальной принадлежности гистологического материала. Такая услуга часто необходима для исключения перепутывания гистологических образцов, когда пациент с подозрением на болезнь получает «чужой» диагноз (например, рак).

Gency поставляется как отдельная независимая программа.

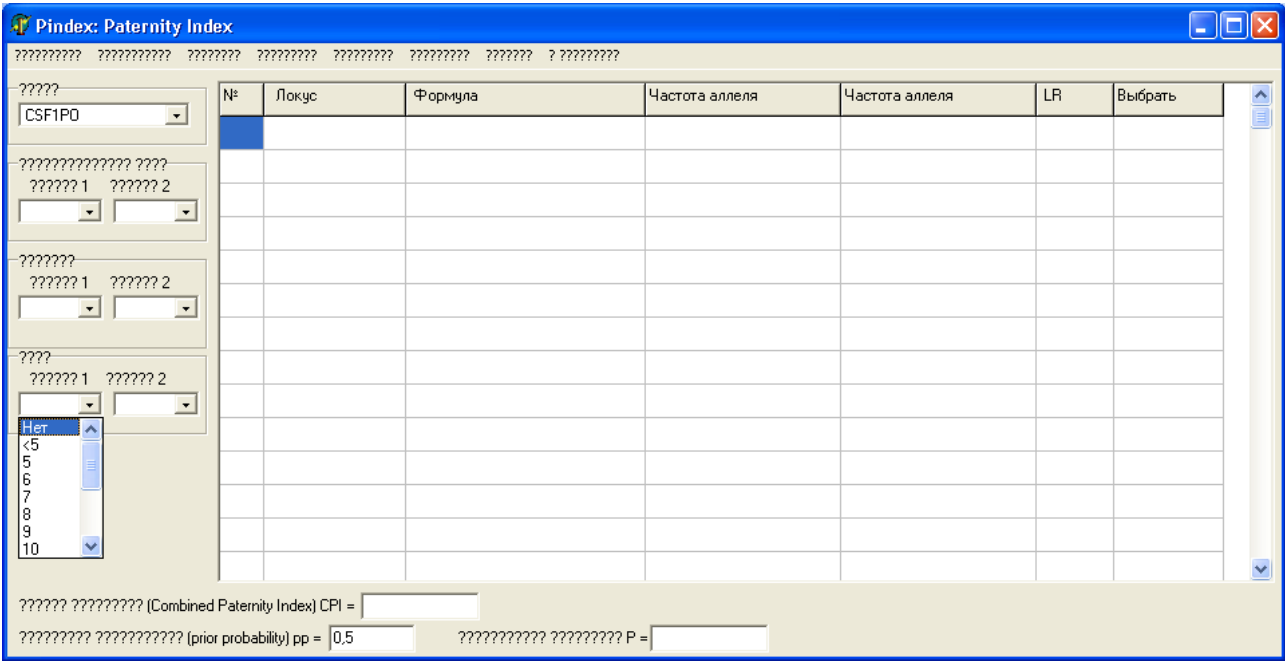
Решение затруднений

Проблемы с кодировкой кириллицы при запуске программы

Достаточно редко (у новых пользователей с нерусифицированной конфигурацией ПК или же после обновления ОС) программа (демо-версия в том числе) может запускаться следующим образом, с кракозябрами:



Затем, после нажатия «OK», рабочее поле выглядит примерно так:

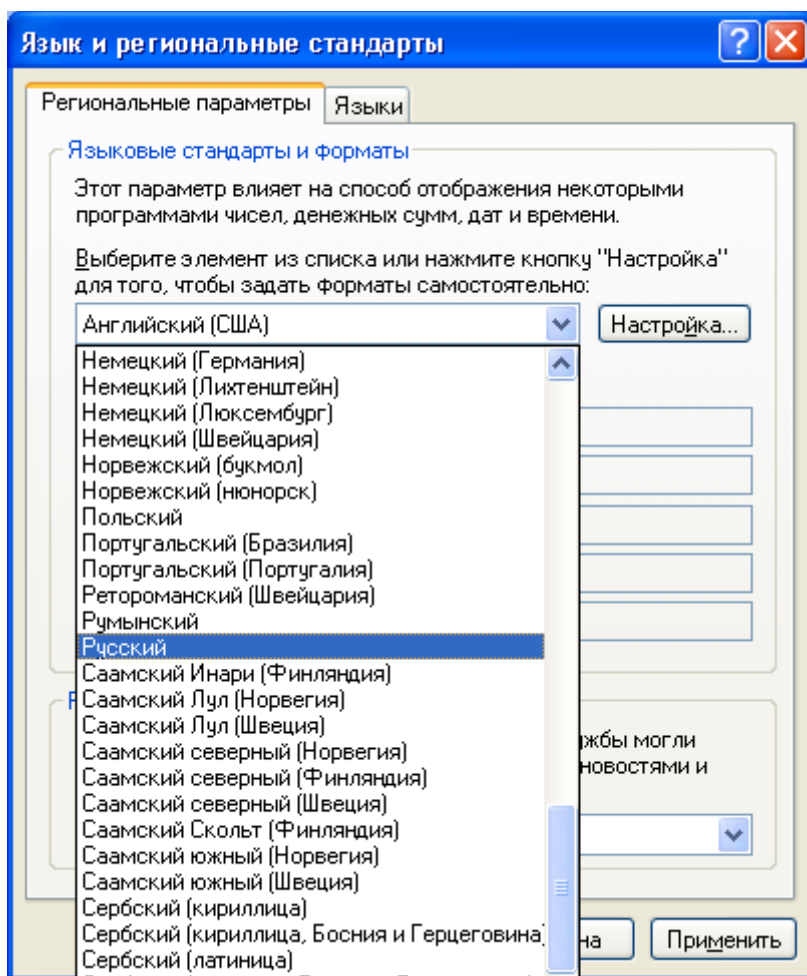


Несомненно, это неудобно и неправильно! Кирилл, даже вместе с Мефодием, такой текст прочесть не смогут.

Решается эта проблема через настройки *Windows*, на примере *Microsoft Windows XP* следующим образом:

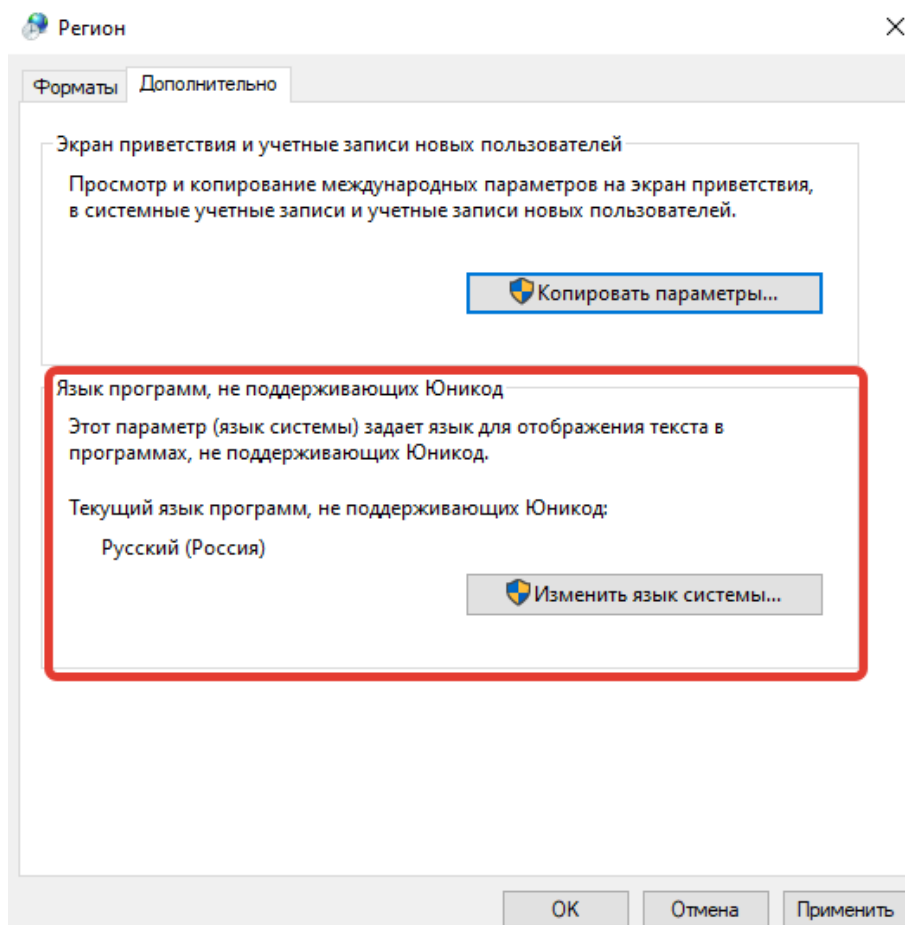
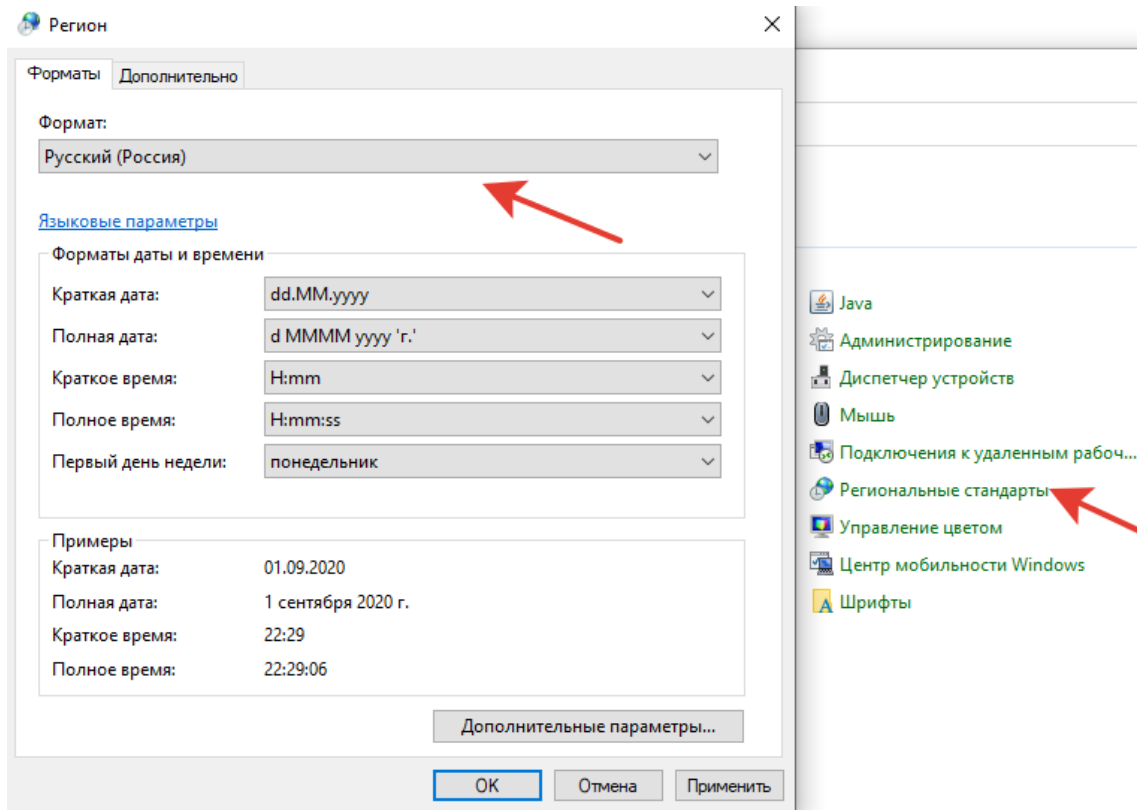


Мой компьютер -> Панель управления -> Язык и региональные стандарты -> закладка «Региональные параметры» -> выбрать в выпадающем меню «русский». Эта закладка изображена на рисунке ниже:



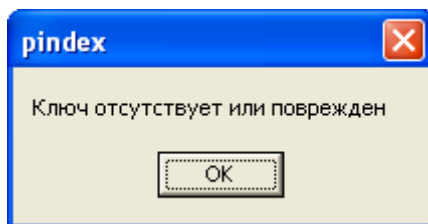
ОС может потребовать установочный диск, а также последующую перезагрузку компьютера.

Для других версий *Windows* в данном случае будут актуальны настройки, изображённые на скриншотах на следующей странице.



Ключ отсутствует или повреждён

При запуске программы в отсутствие файла *pindex.key* выдаётся следующее сообщение об ошибке:



Если этого файла действительно нет, то программа не запустится.

Иногда подобный конфликт может возникнуть в тех случаях, когда программа копировалась на жёсткий диск под одной учётной записью, а запускается под другой. Следует скопировать программу на диск компьютера под той учётной записью, под которой планируется работать. Также может помочь перезагрузка компьютера.

Файлы не найдены (Files not found)

Если папка *Data* вообще не содержит служебных файлов формата **.xls*, то при запуске программы выдаётся следующее сообщение об ошибке:



Следует поместить в папку *Data* необходимые для работы служебные файлы для отдельных локусов.

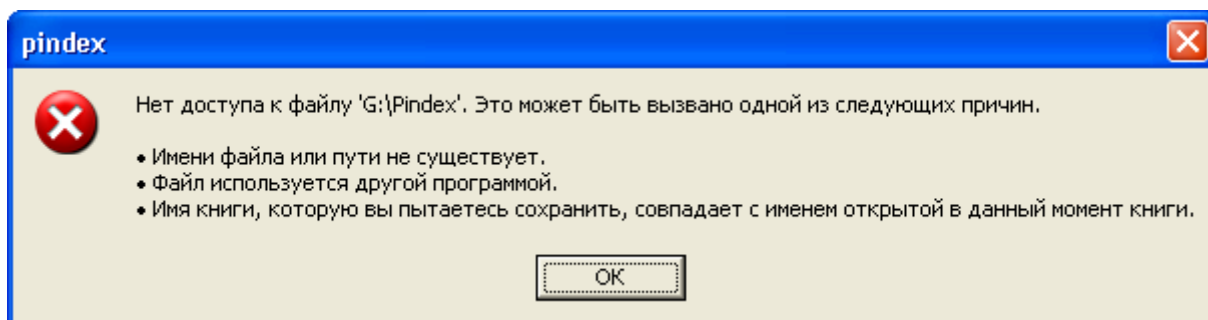
Невозможно прочитать данные в массив

Это сообщение об ошибке выдаётся в случаях:

- если названия отдельных служебных файлов, созданных пользователем, содержат более 16 символов. Следует переименовать эти файлы (и рабочие листы).
- если названия отдельных служебных файлов, созданных пользователем, не соответствуют названиям рабочих листов этих файлов. Следует согласовать названия файлов и рабочих листов.

Невозможно сохранить результаты работы программы

При запуске программы с внешнего носителя, защищённого от записи, результаты работы программы (файл с расчётами) не могут быть сохранены. При попытке сохранения результатов в этом случае выдаётся следующее сообщение об ошибке:

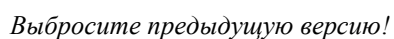


Причина в том, что программа пытается сохранить файл результатов в том же каталоге, из которого была запущена. Для устранения ошибки следует или снять защиту от записи с внешнего носителя или скопировать программу на жёсткий диск компьютера и запускать оттуда.

Для отдельных конфигураций ОС и *Microsoft Office* отмечена непонятная (пока) ошибка невозможности сохранения результатов расчётов. В частности, для версий *Excel-2016-Professional* (русско- и англоязычные). В этом случае при попытке сохранения результатов выдаётся следующее сообщение об ошибке: «**Неверный индекс / Invalid index**». И файл результатов сохраняется некорректно, без формул расчётов. Два скриншота этой ошибки приведены на следующей странице.

Ошибка пока устраняется только переустановкой другого дистрибутива *Excel-2016*.

В этой связи следует отметить, что Пиндекс корректно сохраняет результаты расчётов при использовании, например, *Excel-2007*, *Excel-365* (под *Windows-10*).



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. Значения отношений правдоподобия 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 актуальных микросателлитных локусов: *F13A01*, *F13B*, *FESFPS*, *LPL*, *SE33*. Смотрите примечания в соответствующих файлах и указанные в них литературные первоисточники.

«Невыявленные» аллели с нулевой частотой встречаемости

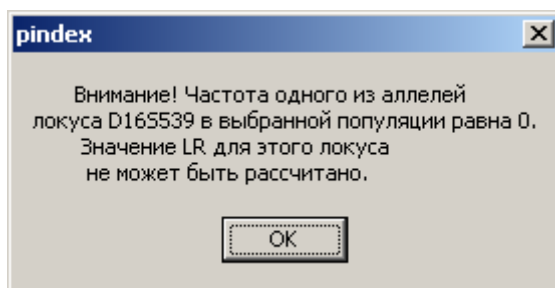
Для всех локусов в двух референтных популяциях из четырёх (для негров и европеоидов США) отдельным «невыявленным» аллелям приписана нулевая частота встречаемости в столбцах «В».

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 (в столбце В). В этом случае значения *LR* будут рассчитаны корректно, поскольку программа использует эталонные частоты аллелей именно из столбцов «В» рабочих листов.

А при попытке пересчёта значений *LR* по эталонным частотам аллелей для других популяций может происходить некорректная операция «деление на 0». При этом выводится сообщение об ошибке, также возможно зависание или сбой программы:



Важно!

Будьте внимательны при пересчёте отношений правдоподобия 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 «невыявленным» аллелям приписана минимальная (ненулевая) популяционная частота *MAF*.

Несколько «предполагаемых отцов»

Достаточно часто бывает (особенно – для беременных женщин), что в конкретном случае одновременно обследуются несколько «предполагаемых отцов» в отношении одного и того же ребёнка. Эта ситуация уже обсуждалась в разделе **«Какие случаи можно корректно анализировать?»**

Предположим, что в результате молекулярно-генетического исследования выявлено, что для всех (кроме одного-единственного) обследованных мужчин предполагаемое отцовство исключилось. В этой ситуации корректный расчёт вероятности отцовства для *неисключённого мужчины* затруднителен, поскольку сложно выбрать адекватную априорную вероятность. Использование $pp=0,5$ в этой ситуации однозначно некорректно, поскольку такой выбор подразумевает в качестве «альтернативного истинного отца» некоего случайного человека из той же референтной популяции. А на самом деле выбор происходит между ограниченным числом «предполагаемых отцов», подвергнутых ДНК-исследованию. При этом сохраняется и ненулевая вероятность того, что на самом деле истинный отец ребёнка всё-таки и не входит в круг *2-3-так далее* обследованных «предполагаемых отцов».

Аналогичные компьютерные программы

К настоящему времени известно достаточно много компьютерных программ, в которых реализован алгоритм отношений правдоподобия для расчётов в молекулярно-генетических экспертизах по идентификации личности и установлению спорного родства. Вот далеко неполный список:

- ❑ *Гаврилей и др.* (2002-2007): DNA-expert Lab;
- ❑ *Курочкин А.В., Кораблин И.И.* (2010): ДНК-Крим Родство;
- ❑ *Brenner C.* (1997) DNA-View: <http://dna-view.com>;
- ❑ *Egeland et al.* (2000): Familias;
- ❑ *Fung W.K.* (2003): EasyDNA;
- ❑ *Qualitytype GmbH* (2006): GenoProof.

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., Корниенко И.В., Щербаков В.В., Иванов П.Л. (2002) Применение компьютерных программных средств при решении задач непрямо́й молекулярно-генетической идентификации неопознанных тел. – *Судебно-медицинская экспертиза*, 45 (2), 11-16. PMID: [12063791](#). (-)
- знаком (-) здесь и далее в списке литературы помечены публикации, полный текст которых пока отсутствует в папке references.
3. Ефремов И.А. (2007) Молекулярно-генетические экспертизы семей с двумя детьми: особенности расчёта индекса и вероятности отцовства. – Тезисы докладов *Международной научно-практической конференции судебных медиков*, Украина, г. Одесса, 7-8 июня 2007 г., стр. 77-80.
 4. Ефремов И.А. (2010) Расчёт индекса отцовства при исследовании локусов, расположенных на хромосоме X человека. – Материалы VI Съезда Российского общества медицинских генетиков, г. Ростов-на-Дону, 14-18 мая 2010 г., журнал *Медицинская генетика*, стр. 61.
 5. Ефремов И.А., Кожемяко В.Б. (2011) Молекулярно-генетические экспертизы спорного родства по полиморфным маркёрам хромосомы X человека: особенности интерпретации результатов и расчётов индекса отцовства. – *Сибирский медицинский журнал*, 106 (7), 49-55.
 6. Ефремов И.А., Серегин Ю. А. (2002) Расчёты индекса и вероятности отцовства в судебно-медицинских экспертизах случаев спорного родства. – *Одесский Медицинский Журнал*, 4 (72), 11-16.
 7. Ефремов И.А., Серегин Ю. А. (2003) Компьютерная программа для численной оценки результатов судебно-медицинских генетических экспертиз случаев спорного отцовства. – *Альманах Судебной Медицины*, Санкт-Петербург, 4, 77-81.
 8. 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.

9. 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 VWFII. – *Судебно-медицинская экспертиза*, 48 (2), 29-34. PMID: [15881140](#).
10. Культин А.Ю., Стороженко И.В., Никитаев В.Г., Проничев А.Н., Власов В.А. (2011) Экспертная оценка и вероятностно-статистическая обработка результатов исследования ДНК при установлении биологического родства. Учебное пособие. – Москва, ЭКЦ МВД России. <https://elibrary.ru/item.asp?id=19503327> (-)
11. Морлинг Н., Аллен Р.В., Каррачедо А., Жеада Е., Гуде Ф., Нелленберг Ш., Мартин В., Майр В.Р., Олайсен Б., Паскали В., Шнайдер П.М. (2003) Международные рекомендации по стандартам проверки отцовства. – *Проблемы стандартизации в здравоохранении*, 7, 8-20.
12. Перепечина И.О., Гришечкин С.А. (1996) Вероятностные расчёты в ДНК-дактилоскопии: Методические рекомендации. – Москва, ЭКЦ МВД России. (-)
13. 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. (-)
14. Ayres K.L., Powley W.M. (2005) Calculating the exclusion probability and paternity index for X-chromosomal loci in the presence of substructure. – *Forensic Sci. Int.*, 149 (2-3), 201-203. PMID: [15749362](#).
15. 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](#).
16. Børsting C., Sanchez J.J., Hansen H.E., Hansen A.J., Bruun H.Q., Morling N. (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing. – *Forensic Sci Int Genet.*, 2 (4), 292-300. PMID: [19083839](#). (-)
17. Brenner C.H. (1983) Calculation of paternity index. – in *Inclusion Probabilities in Parentage Testing*. Arlington, Va: American Association of Blood Banks, 632-638. ISBN: 978-3-8055-4133-6. (-)
18. 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 (+/-)
19. Brenner C.H. (1997) Symbolic kinship program. – *Genetics*, 145 (2), 535-542. Erratum in: *Genetics*, 1997, 147 (1), 398. PMID: [9071605](#).

20. Brenner C., Morris J.W. (1990) Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. – In: *International symposium on human identification 1989: proceedings*. Madison, WI, Promega, 21-53.
21. 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](#).
22. 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](#).
23. Drábek J. (2009) Validation of software for calculating the likelihood ratio for parentage and kinship. – *Forensic Sci Int Genet.*, 3 (2), 112-118. PMID: [19215880](#).
24. Egeland T., Mostad P.F., Mevåg B., Stenersen M. (2000) Beyond traditional paternity and identification cases. Selecting the most probable pedigree. – *Forensic Sci Int.*, 110 (1), 47-59. PMID: [10802200](#).
25. Evett I. W., Weir B.S. (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. – Sunderland, Massachusetts, Sinauer Associates Inc.
26. *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.
27. Fung W.K. (2003) User-friendly programs for easy calculations in paternity testing and kinship determinations. – *Forensic Science International*, 136 (1-3), 22-34. PMID: [12969617](#).
28. *GenePrint™ STR Systems (Silver Stain Detection)*. Technical manual. – Promega Corporation (USA), revised 7/99, 04/01, 7/06. (-)
29. 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](#).
30. Hallenberg C., Morling N. (2002) A report on the 2000 and 2001 paternity testing workshops of the English speaking working group of the International Society for Forensic Genetics. – *Forensic Sci. Int.*, 129 (1), 43–50. PMID: [12230996](#).
31. Kent W.J. (2002) BLAT – the BLAST-like alignment tool. – *Genome Res.*, 12 (4), 656-664. PMID: [11932250](#).
32. Lee H.S., Lee J.W., Han G.R., Hwang J.J. (2000) Motherless case in paternity testing. – *Forensic Sci. Int.*, 114 (2), 57–65. PMID: [10967247](#).
33. 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](#).

34. Morling N., Allen R.W., Carracedo A., Geada H., Guidet F., Hallenberg C., Martin W., Mayr W.R., Olaisen B., Pascali V.L., Schneider P.M. (2002) Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. – *Forensic Sci Int.*, 129 (3), 148–157. PMID: [12372685](#).
35. NCBI, *Allele Frequency Aggregator (The ALFA project)*: Current Build 155, Released April 9, 2021, <https://www.ncbi.nlm.nih.gov/snp/>
36. 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.
37. 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](#).
38. Schneider P.M. (2012) Beyond STRs: The Role of Diallelic Markers in Forensic Genetics. – *Transfus. Med. Hemother.*, 39 (3), 176-180. PMID: [22851932](#).
39. 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](#).
40. 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

Неизвестному художнику, чей комикс про [*Cinture di Castità*](#) (пояс верности) использован на обложке настоящего Руководства. Надеемся, что эта иллюстрация облегчает отдельным неподготовленным читателям понимание основной сути обсуждаемых вопросов...

Другу одного из авторов – Сан Санычу [Клевитову](#) за техническую поддержку, тестирование программы и помощь советами.

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

Всем авторизованным пользователям – за то, что выбрали нашу программу, рекомендовали её знакомым, а также за все высказанные замечания и пожелания по работе программы. Список авторизованных пользователей обновляется на [сайте](#).

Обратная связь

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

Ефремов Илья Алексеевич, к.б.н.

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

Эл. почта: info@tapotili.ru

www.tapotili.ru

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

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

Свидетельство о регистрации программы

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ

№ 2009610058

**Программа для расчета индекса и вероятности отцовства
(материнства) при исследованиях ДНК - Pindex: Paternity Index**

Правообладатель(ли): **Ефремов Илья Алексеевич (RU),
Серегин Юрий Александрович (RU)**

Автор(ы): **Ефремов Илья Алексеевич,
Серегин Юрий Александрович (RU)**

Заявка № **2008615246**
Дата поступления **13 ноября 2008 г.**
Зарегистрировано в Реестре программ для ЭВМ
11 января 2009 г.



Руководитель Федеральной службы по интеллектуальной
собственности, патентам и товарным знакам



Б.П. Симонов