

Pindex (Paternity INDEX): программа для расчета индекса и вероятности отцовства (материнства) при исследованиях ДНК в случаях спорного родства
Настоятельно рекомендуется прочитать эту справку перед началом работы с программой.

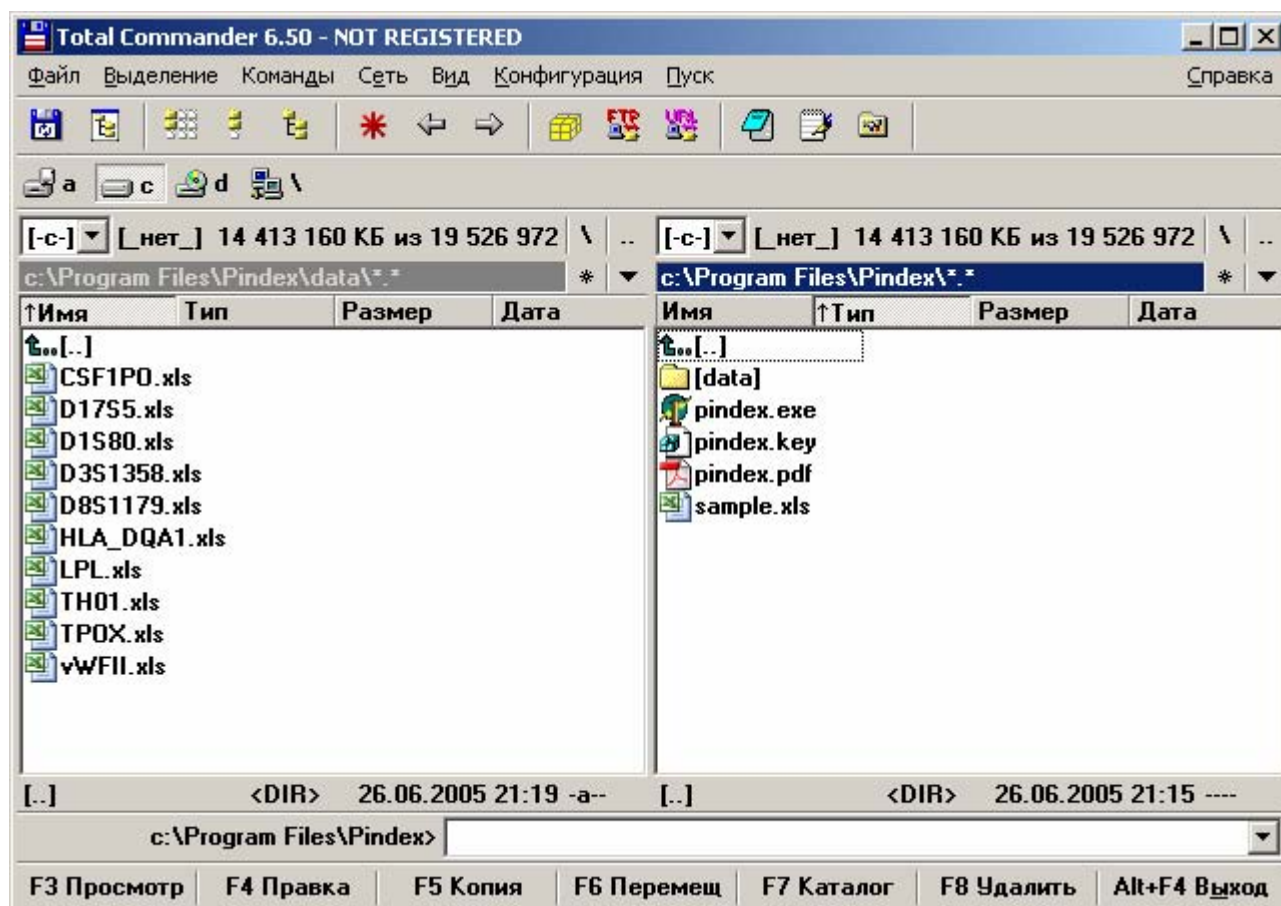
Системные требования

Программа написана в системе программирования *Inprise Delphi 5.0*. Минимальные системные требования: процессор 486MHz, RAM 16Mb, *Microsoft Windows 95/98/Me*, *Microsoft Excel 95*.

При тестировании *Pindex'a* на различных версиях *Windows* и *Microsoft Office* отмечены отдельные конфликты, которые могут вызывать сбой или зависание программы. Со всеми замечаниями и пожеланиями обращайтесь к разработчикам по указанным координатам (кнопка «О программе»).

Рабочий каталог программы

На приведенном ниже рисунке изображена структура рабочего каталога программы (окно справа), а также рабочие файлы, содержащиеся в каталоге *data* (окно слева).



Файлы *pindex.exe* и *pindex.key* - это исполняемый файл программы и ключ авторизованного пользователя. Файл *pindex.pdf*, который Вы сейчас читаете, из программы

вызывается кнопкой "Справка" (он необязателен для работы). Файл *sample.xls* является примером расчетов (также необязателен для работы).

В директории *data* находятся рабочие файлы для отдельных локусов. Представленный на рисунке список рабочих файлов является, естественно, неполным. Число рабочих файлов (соответствующих используемым локусам) не ограничено и формируется пользователями программы индивидуально.

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*. Наличие этого ключа позволяет авторизованному пользователю многократно копировать программу на любое количество компьютеров. Ключ авторизованного пользователя не защищает, конечно, *Pindex* от взлома, но в определенной степени дает возможность разработчикам следить за возможным несанкционированным распространением программы: в нем зашифровано имя авторизованного пользователя.

Узнать используемую версию программы и имя авторизованного пользователя можно, нажав кнопку "О программе". В настоящее время последней рабочей версией является *Pindex_1.0*.

Директория *data*: служебные файлы для используемых локусов

Для успешной работы программы в одном каталоге с исполняемым файлом *pindex.exe* обязательно должна находиться папка *data*. В этой папке размещаются служебные файлы формата *Excel*, содержащие данные об эталонных частотах аллелей по используемым локусам. **В директории *data* не должно находиться никаких посторонних файлов формата *Excel*!**

Названия служебных файлов соответствуют названиям локусов. Каждый файл содержит четыре листа, в которых приводятся частоты аллелей для четырех эталонных популяций: русские (листы *имя_файла-rus*), европеоиды США (листы *имя_файла-white*), негры США (листы *имя_файла-black*) и любая локальная популяция (листы *имя_файла-other*). Пример служебного файла для локуса *D16S539* приводится на рисунке на следующей странице, на котором видны два листа из четырех (данные для европеоидов России и США).

Частоты аллелей для русской популяции, уже представленные в первых листах соответствующих файлов (листы *имя_файла-rus*), являются консервативной оценкой для европеоидного населения России. Эта оценка проводилась в ходе сравнительного анализа различных популяционных данных для населения России (печатные работы и персональные

сообщения различных авторов). В своей практике мы используем именно указанные частоты аллелей.

По мере накопления популяционных данных производятся обновления всех служебных файлов, которые можно найти здесь:

http://www.tapotili.ru/programs/pindex_data.zip.

В служебных файлах для каждого локуса приводится также дополнительная информация, непосредственно не используемая программой. Например, указаны соответствующие литературные первоисточники для референтных частот аллелей, размеры аллелей при использовании различных пар праймеров, порядок нумерации аллелей, состав аллельных "лестниц".

The screenshot shows two worksheets in Microsoft Excel. The top worksheet, titled 'D16S539.xls:2', contains the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	5	0,01	94	0											
2	6	0,01	98	0											
3	7	0,01	102	0											
4	8	0,02	106	0,014											
5	9	0,13	110	0,106											
6	10	0,06	114	0,047											
7	11	0,31	118	0,274											
8	12	0,37	122	0,331											
9	13	0,21	126	0,176											
10	14	0,07	130	0,051											
11	15	0,01	134	0,002											
12	Аллельные частоты по данным Пушкарев и др., 2004; популяционная выборка 148 неродственных человек.														
13	Размер аллелей, в парах нуклеотидов (Tapotili-set, 2004)														
14	Консервативная оценка частот аллелей для европеоидного населения России (по данным Пушкарев и др., 2004; верхняя граница 95%-ного довери														
15	Нумерация аллелей согласно числу тандемных повторов. Аллельная "лестница" к наборам ТАПОТИЛИ включает 7 аллелей: 8, 9, 10, 11, 12, 13, 14.														

The bottom worksheet, titled 'D16S539.xls:1', contains the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	5	0	264												
2	6	0	268												
3	7	0	272												
4	8	0,026	276												
5	9	0,107	280												
6	10	0,079	284												
7	11	0,319	288												
8	12	0,269	292												
9	13	0,167	296												
10	14	0,031	300												
11	15	0,002	304												
12	Размер аллелей, в парах нуклеотидов (Promega's set, GenePrint® STR Systems, 04/01)														
13	Частоты аллелей для европеоидного населения США (210 человек, Promega, GenePrint® STR Systems, 04/01)														
14	Нумерация аллелей согласно числу тандемных повторов														
15															

При работе программа считывает численные значения только из первых двух столбцов каждого листа (в приведенном примере для локуса *D16S539* задействуются только ячейки A1-B11), вся остальная информация программой игнорируется. Частоты аллелей считываются программой из столбцов "B" с точностью до трех знаков после запятой.

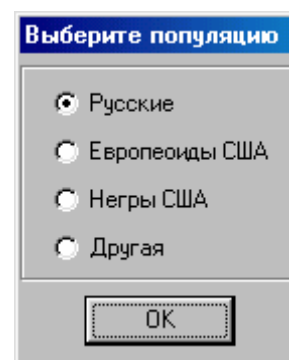
Пользователь может по своему усмотрению менять нумерацию аллелей и эталонные частоты для всех локусов, вводить и редактировать любую дополнительную информацию, а также создавать новые аналогичные файлы для любых других локусов.

Четвертые листы каждого файла (листы *имя_файла-other*) как раз и предназначены для задания частот аллелей самим пользователем: например, в этих листах можно указать частоты аллелей в отдельной локальной популяции. В затруднительных случаях обращайтесь к разработчикам. Смотрите также раздел [Работа с программой: использование различных эталонных частот аллелей](#).

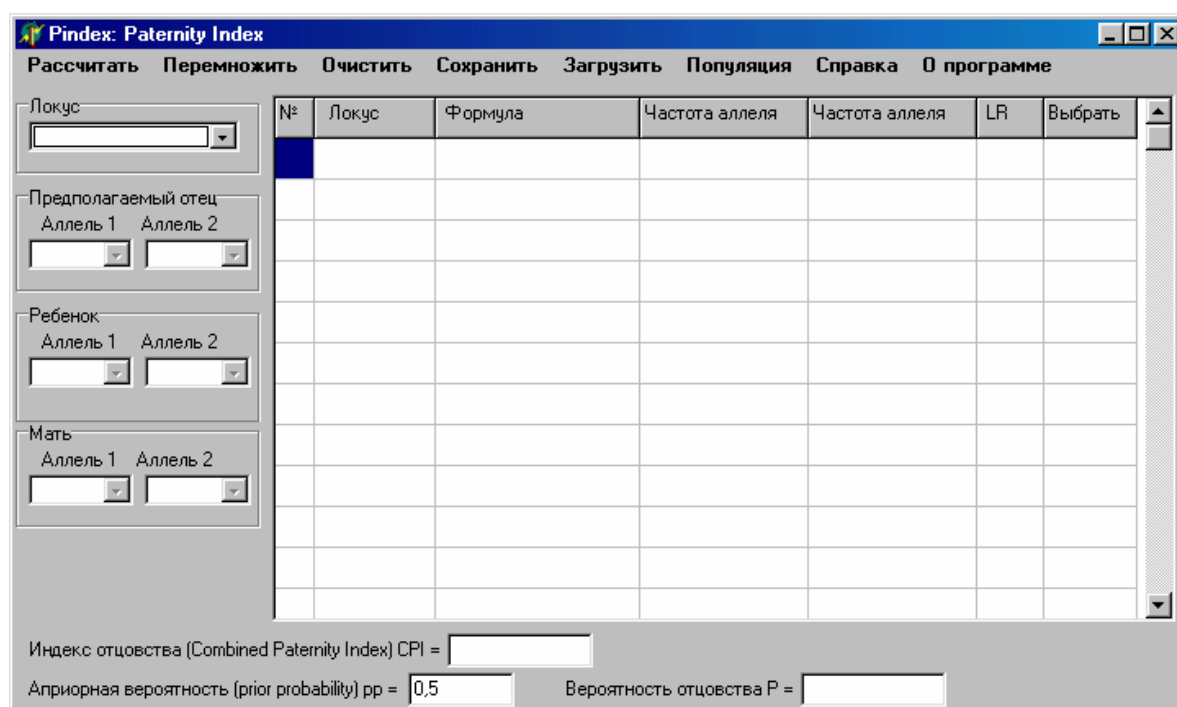
Для максимально быстрой загрузки и работы программы рекомендуется хранить в директории *data* только файлы для тех локусов, которые постоянно используются в экспертной практике. Например, если рутинно исследуются только 10 локусов, оставьте в директории *data* соответствующие этим локусам 10 файлов, остальные файлы удалите или переместите в другой каталог.

[Работа с программой: выбор эталонной популяции и ввод генотипов](#)

После запуска программы (вызов *Pindex.exe*) нужно выбрать в стартовом меню эталонную популяционную базу данных. Для расчетов программа будет адресоваться к соответствующим частотам аллелей (см. рисунок справа):



После появления рабочего листа (см. рисунок внизу) следует выбрать из списка нужный локус и указать для него наблюдаемые комбинации генотипов ребенка и заявленных родителей (или только одного из родителей).



Когда для выбранного локуса указаны все генотипы, следует еще раз проверить эти данные и, если все верно, нажать кнопку "Рассчитать". При этом в рабочем поле появятся определенные значения и формулы.

Далее нужно выбрать из списка следующий локус и повторить для него описанную процедуру (ввод генотипов – проверка – нажатие кнопки "Рассчитать"). Таким образом поэтапно для всех проанализированных локусов рассчитываются отношения правдоподобия (*likelihood ratios*, столбец «LR» рабочего поля). На рисунке ниже изображено рабочее поле программы при работе с файлом *sample.xls*.

№	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	

При необходимости на любом этапе работы можно очистить рабочий лист (кнопка "Очистить") или сохранить рабочие данные в формате *Excel* (кнопка "Сохранить"). Данные можно сохранять с произвольным именем в любом каталоге, при этом имя и путь к текущему рабочему файлу указываются на синем фоне в верхней части окна программы. В приведенном выше примере файл сохранен с именем *sample.xls* в папке *C:\Program Files\Pindex*.

Для открытия ранее созданного рабочего файла нужно выбрать команду "Загрузить". Программа может обрабатывать одновременно только один рабочий файл.

Неверные данные для отдельных локусов могут быть удалены из рабочего листа. Для этого необходимо поместить курсор мыши на строку с ошибочными данными в рабочем поле программы и, нажав правую кнопку мыши, выбрать команду "Удалить".

Указанные пользователем комбинации генотипов по отдельным локусам не выводятся на рабочем поле программы, но сохраняются в рабочем файле!

Работа с программой: расчет индекса и вероятности отцовства

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

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

Соответственно, на следующем этапе работы пользователь выбирает те локусы, которые будут использоваться для расчета индекса отцовства. Для этого нужные строчки помечаются левой кнопкой мыши в поле "Выбрать" (в этом столбце появляются символы "+"), а затем нажимается кнопка "Перемножить":

The screenshot shows the 'Paternity Index' software window. It features a menu bar with options: 'Расчитать', 'Перемножить', 'Очистить', 'Сохранить', 'Загрузить', 'Популяция', 'Справка', and 'О программе'. On the left, there are input fields for 'Локус' (set to 'PAH'), '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 area contains a table with 10 rows of loci data. The 'Выбрать' column has '+' symbols for rows 2 through 10. At the bottom, the 'Индекс отцовства (Combined Paternity Index) CPI' is displayed as 1706, and the 'Вероятность отцовства P' is 99.94%.

№	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 = 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 приведенном примере значение комбинированного индекса отцовства составляет 1706, и при этом расчете использовались данные для 8 из 10 локусов (не учитывались значения LR для локусов *IL1RN* и *CD4*).

Для того, чтобы исключить какой-либо локус из расчетов, следует убрать для него пометку "+" в соответствующих ячейках левой кнопкой мыши. Теперь после повторного

нажатия кнопки "Перемножить" значения индекса и вероятности отцовства будут пересчитаны с учетом внесенных изменений.

Сохранение результатов и завершение работы с программой

Указанные генотипы по всем локусам и результаты расчетов индекса и вероятности отцовства сохраняются в отдельном файле результатов формата *Excel* (кнопка "Сохранить"). В файле результатов автоматически указывается использованная при расчетах эталонная популяция. Этот файл может быть непосредственно использован при оформлении заключения эксперта.

Для выхода из программы следует нажать закрывающую кнопку "x" в правом верхнем углу окна программы. Вам будет предложено сохранить изменения (или не делать этого) в последнем рабочем файле.

Настоятельно рекомендуем по окончании работы с программой открыть файл результатов уже в редакторе *Excel* и проверить все указанные генотипы. Если какие-то генотипы были указаны неправильно, то **не следует** вносить изменения в редакторе *Excel*. Лучше снова открыть файл результатов в программе *Pindex* и удалить строчки с ошибочными данными в рабочем поле программы (правая кнопка мыши и 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»

Если все генотипы в файле результатов указаны верно, то можно из *Excel* скопировать ячейки с данными и вставить их в текст документа *Word*. Для файла *sample.xls* получится:

№	Локус	Генотип отца	Генотип ребенка	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									

Эту таблицу при желании можно преобразовать в более наглядную форму. Например, убрать избыточные столбцы и ячейки, расставить локусы в порядке увеличения значений LR. Частоты аллелей лучше округлить до второго знака после запятой. (Программа работает со

значениями с точностью до трех знаков). Озаглавить таблицу результатов можно следующим образом:

*Выявленные генотипы у обследуемых лиц
и расчет отношений правдоподобия (LR) для исследованных локусов.*

№	Локус	Генотип отца	Генотип ребенка	Генотип матери	Формула	Частота аллеля	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	CSF1PO	12/13	11/13	11/11	$1/2p(13)$	$p(13)=0,08$	6,25
Индекс отцовства (Combined Paternity Index, CPI)							1706
Вероятность отцовства (prior probability $pp=0,5$)							99,94%
Использовались референтные частоты аллелей для русских							
(-) Значения LR, не использованные для расчета 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, приведенные в служебных файлах, по каким-либо причинам изменяются 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 пересчета значений LR с учетом внесенных изменений в эталонные частоты следует нажать кнопку "Популяция" и выбрать популяцию.

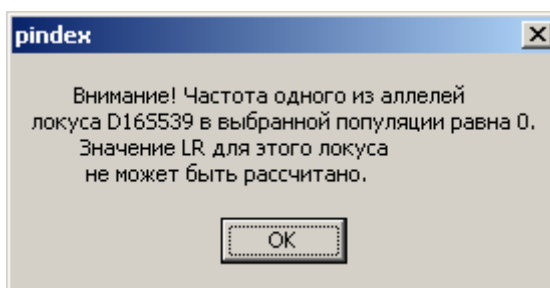
Если же в выпадающем меню кнопки "Загрузить" выбрать "генотипы и рассчитать LR", то при загрузке файла результатов все значения LR будут автоматически пересчитаны, с учетом сделанных последних изменений в эталонных аллельных частотах для всех локусов.

[Использование различных эталонных частот аллелей: затруднение №1](#)

Для ряда локусов (например, *IL1RN*, *D17S5*, *DIS111*, *CYAR04*, *vWFI*, *PAH* и др.) разработчикам в настоящее время неизвестны адекватные частоты аллелей для отдельных эталонных популяций. В таких случаях всем аллелям в этих популяциях приписана частота встречаемости 0,5 (смотрите соответствующие файлы). Значения отношений правдоподобия будут пересчитаны с использованием значений 0,5.

[Использование различных эталонных частот аллелей: затруднение №2](#)

Для всех локусов в трех референтных популяциях из четырех (за исключением российской популяции) отдельным аллелям приписана нулевая частота встречаемости. Например, в приведенном на стр. 3 примере для локуса *D16S539* аллели 5, 6 и 7 не были выявлены у европеоидов России и США (экспериментальные частоты аллелей в исследованных выборках равны 0). Тем не менее, консервативная оценка частот этих трех аллелей для европеоидного населения России указана равной 0,01 (см. соответствующий файл). В этом случае значения LR корректно рассчитываются только для русской популяции. А при попытке пересчета значений LR по другим эталонным частотам аллелей может происходить деление на 0. При этом выводится сообщение об ошибке (на отдельных платформах возможно зависание или сбой программы):



Будьте внимательны при пересчете отношений правдоподобия по эталонным частотам для разных популяций!

[Отношения правдоподобия: стандартная практика в ДНК-исследованиях](#)

Величина LR показывает, во сколько раз более вероятно, что наблюдаемое совпадение аллелей у заявленного отца и ребенка 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 отец является отцом ребенка (числитель) к вероятности того, что отцом этого ребенка является другой мужчина (знаменатель).

Индекс отцовства CPI можно интерпретировать аналогичным образом: исходя из полученных результатов исследования ДНК, в CPI раз более вероятно, что именно

обследуемый мужчина и передал ребенку «отцовский набор аллелей», чем такие результаты можно объяснить случайным совпадением генотипов.

Расчет отношений правдоподобия является **стандартной практикой** при оценке значимости выполненных экспертиз спорного отцовства (*Morling et al., 2002*). Подробно использованные в настоящей программе алгоритмы описаны в работах *Walsh & Buckleton, 1989, Evett & Weir, 1998*, а также можно найти подробную информацию в интернете (например, <http://www.dna-view.com>). Смотрите также список литературы в конце этого описания.

Аналогичные компьютерные программы

К настоящему времени известно несколько аналогичных компьютерных программ, в которых реализован алгоритм отношений правдоподобия для вероятностных расчетов в ДНК-исследованиях:

- *Brenner C.* DNA-VIEW: an integrated software package for DNA identification.
- *Fung W.K.* (2003)
- *Гаврилей и др.* (2002)

Интерпретация результатов: априорные условия, о которых следует помнить

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

- 1) для выявленных гомозиготных генотипов не предполагается «ложной гомозиготности»;
- 2) материнство рассматривается как бесспорное;
- 3) полагается полная невозможность мутаций;
- 4) помимо заявленного отца, истинным отцом ребенка может быть любой случайный мужчина из этой же эталонной популяции, не подвергнутый ДНК-исследованию и не являющийся родственником обследуемого мужчины.

Работа с программой: какие случаи можно анализировать?

Использование программы полностью корректно в случаях, когда обследуются один предполагаемый отец и один ребенок (с матерью или без нее).

Если в отдельном случае одновременно обследуются два или более детей от одной матери, то расчет индекса и вероятности отцовства в отношении каждого из них по отдельности является некорректным.

Предположим, в отдельном случае одновременно обследуются два или более предполагаемых отцов в отношении одного и того же ребенка. Тогда для того мужчины, который не исключился в ходе экспертизы, можно, конечно, рассчитать вероятность

отцовства, но это не будет полностью корректно (поскольку не выполняется п.4 априорных условий предыдущего раздела). Смотрите об этом также раздел [Вероятность отцовства при различных значениях априорной вероятности](#).

Программу можно использовать для анализа случаев спорного материнства (пара ребенок – 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.

Вероятность отцовства: что это такое?

Это другой и более спорный показатель, 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 простой формулой:

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

Как общее правило, в мировой практике при расчете 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, более корректно использование индекса отцовства, как менее спорного показателя.

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 (а в компетенции суда). Возможность непредвзятого выбора априорной вероятности отдельными авторами ставится под сомнение либо вообще отвергается. С точки зрения эксперта, выполняющего исследование, выбор $pp=0,5$ говорит о том, что в половине таких 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, приведенную в конце предыдущего раздела:

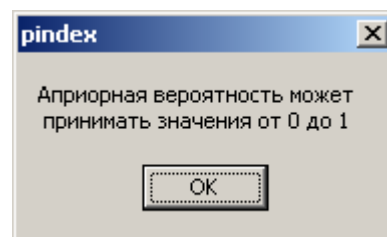
$$P = (0,5 * CPI) / (0,5 * CPI + 1 - 0,5) = 0,5 * CPI / (0,5 * CPI + 0,5) = CPI / (1 + CPI)$$

Естественно, величины вероятности отцовства находятся в диапазоне от 100% (абсолютное подтверждение отцовства, при $pp=1$) до 0% (абсолютное отрицание отцовства, при $pp=0$). Соответственно, диапазон допустимых практических значений априорной вероятности составляет $0 \leq pp \leq 1$.

В *Pindex'e* реализован расчет вероятности отцовства при произвольных значениях априорной вероятности, на усмотрение эксперта.

При запуске программы априорная вероятность по умолчанию составляет 0,5. Это значение приводится в левом нижнем поле рабочего листа программы (для файла *sample.xls* смотрите рисунки на страницах 4-6 этого описания).

При желании в это поле можно ввести любое значение pp с любым количеством десятичных знаков. **Программа воспринимает значения априорной вероятности только в корректном диапазоне (от 0 до 1 включительно). В противном случае выводится сообщение об ошибке (см. рисунок справа).**



После нажатия кнопки "Перемножить" значение 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 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 $CPI=1706$ (см. файл *sample.xls*) вероятность отцовства при таком значении априорной вероятности будет рассчитываться следующим образом:

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

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

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

Предположим, что в отдельном случае одновременно обследуются несколько предполагаемых отцов в отношении одного и того же ребенка (эта ситуация уже упоминалась в разделе [Работа с программой: какие случаи можно анализировать?](#)). Пусть в

результате исследования выявлено, что для всех, кроме одного, обследованных мужчин отцовство исключается. В этой ситуации корректный расчет вероятности отцовства для этого мужчины затруднителен, поскольку сложно выбрать адекватную априорную вероятность. Использование $p=0,5$ в этой ситуации неправомерно, 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.

Отношения правдоподобия: пример расчета

Предположим, по отдельному локусу выявлена следующая комбинация генотипов:

мать	ребенок	предполагаемый отец
pp	pq	pq

Решетка Пюнетта для такой комбинации генотипов выглядит следующим образом:

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

Мать с вероятностью 100% передаст ребенку аллель p, а предполагаемый отец может передать аллель q с вероятностью 50%. Соответственно числитель для LR равен $1 \cdot 1/2 = 1/2$. То есть, вероятность рождения у таких родителей ребенка с таким генотипом равна 50%.

Знаменатель рассчитывается следующим образом.

От матери по-прежнему ребенок с вероятностью 100% получает аллель p. Тогда аллель q он должен унаследовать от "случайного" отца, который может быть или гомозиготой qq, или гетерозиготой pq (или qp, неважно).

Частота встречаемости в популяции гомозиготного мужчины равна q^2 , и с вероятностью 100% он передаст ребенку аллель q. Вероятность рождения ребенка с заданным генотипом pq 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 ребенку аллель q. У такой пары родителей вероятность рождения ребенка с заданным генотипом составит $1 \cdot 1/2 \cdot 2q(1-q) = q(1-q)$.

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 расписываются все остальные возможные комбинации генотипов.

Используя алгоритм «локальных статистических частот» (Иванов, 1999), для этой же комбинации генотипов имеем следующее. Аллелем отцовского происхождения у ребенка является q . Частота встречаемости в популяции «случайного» мужчины с этим аллелем (гомо- или гетерозиготного) составляет: $r = (2q - q^2)$.

При сравнении двух методов расчета получаем $LR < 1/r$ (поскольку $2q > (2q - q^2)$), и при таких комбинациях генотипов оценка CPI и P через отношения правдоподобия **более консервативна**.

Литература

1. Brenner C.H. (1983) Calculation of paternity index. - In *Inclusion Probabilities in Paternity Testing*, American Association of Blood Banks, Arlington, 632-638.
2. Brenner C.H. (1994) Evaluation of the Product Rule. - *Advances in Forensic Haemogenetics*, 5, Springer Verlag, 453-455.
3. Brenner C.H., Morris J.W. (1990) Paternity Index Calculations in Single Locus Hypervariable DNA Probes: Validation and Other Studies. - *Proceedings for The International Symposium on Human Identification 1989*, Promega Corporation, 21-53.
4. 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.
5. Evett I. W., Weir B.S. (1998) Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. - Sunderland, Massachusetts, Sinauer Associates Inc.
6. Fung W.K. (2003) User-friendly programs for easy calculations in paternity testing and kinship determinations. - *Forensic Science International*, 136, 22-34.
7. 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 Science International*, 129, 148-157.
8. Walsh K.A., Buckleton J.S. (1988) - A discussion of the law of mutual independence and its application to blood group frequency data. - *J. Forensic Sci. Soc.*, 28 (2), 95-98.
9. Walsh K.A., Buckleton J.S. (1991) Calculating the frequency of occurrence of a blood type for a 'random man'. - *J. Forensic Sci. Soc.*, 31 (1), 49-58.
10. Гаврилей Ю. К., Корниенко И. В., Щербаков В. В., Иванов П. Л. (2002) Применение компьютерных программных средств при решении задач непрямой молекулярно-генетической идентификации неопознанных тел. - *Суд.-мед. Экспертиза*, 45 (2), 11-16.
11. Ефремов И.А., Серегин Ю. А. (2002) Расчеты индекса и вероятности отцовства в судебно-медицинских экспертизах случаев спорного родства. - *Одесский Медицинский Журнал*, 4 (72), 11-16.

12. *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.
13. *Иванов П.Л.* (1999) Использование индивидуализирующих систем на основе полиморфизма длины амплифицированных фрагментов (ПДАФ) ДНК в судебно-медицинской экспертизе идентификации личности и установления родства: Метод. Указания (утверждены Минздравом России 19.01.99). – *Суд.-мед. Экспертиза*, 42 (4), 35-41.