Tutorial de Familias

Por: Daniel Kling, Lourdes Prieto and Thore Egeland

Actualizado el 29/Sept/2022

Prefacio

Este tutorial describe las características básicas de Familias. La presente versión del tutorial se basa en la versión 3.3 de Familias, pero debe funcionar razonablemente bien para las versiones 3.2.2 y posteriores

Para una descripción más detallada y para entender el soporte teórico del software consulte el manual disponible en http://www.familias.no, y las publicaciones que aparecen listadas en el mismo link.

La versión en inglés de este documento (Daniel Kling, Lourdes Prieto and Thore Egeland) se encuentra en http://familias.no/espana/

Puede consultar videos que incluyen soluciones a ejercicios en Egeland, Kling, Mostad (2015), ver https://familias.name/VideosBook.pdf

Contenidos

1. Lo básico. Un caso de paternidad en cuatro pasos
2. Complicaciones
3. R Familias, paramlink, y plotting
4. Simulación
5. Búsqueda ciega
6. DVI (Identificación de víctimas en desastres)
7. Búsqueda familiar (familial searching)
8. Archivos de entrada (formatos)
9. Archivos e informes que genera Familias
10. Avanzado
11. Miscelánea

1. Lo básico. Un caso de paternidad en cuatro pasos

Input file: tutorial-Ch1.fam

Ejemplo para presentar Familias

Considere

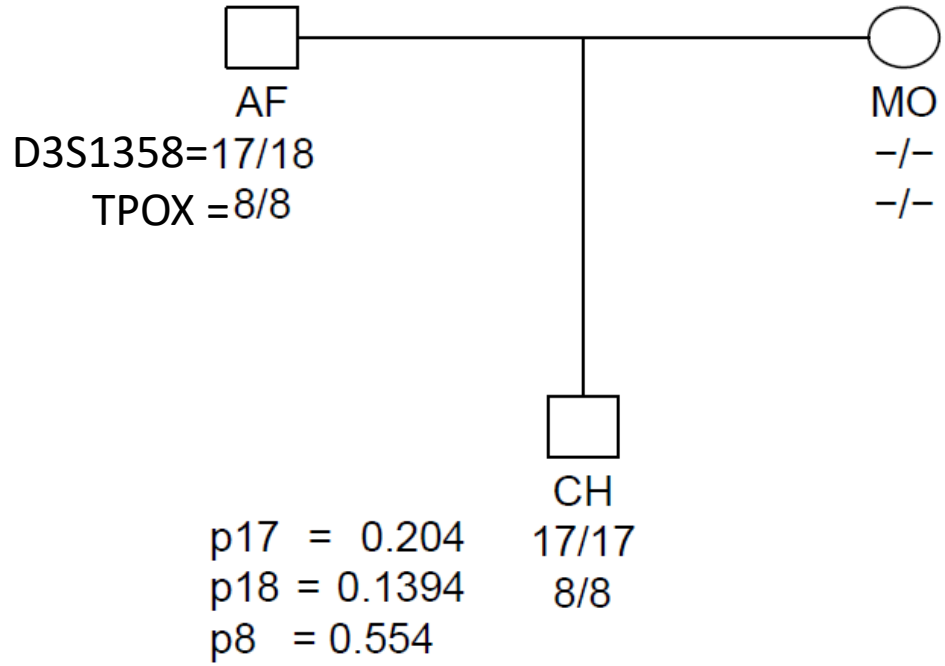
- H1: AF padre de CH (figura)
- H2: AF y CH no relacionados

Calculamos los LRs por marcador:

$$LR_1 = \frac{\frac{1}{2}p_{17}}{p_{17}^2} = \frac{1}{2 \times 0.204} = 2.45$$

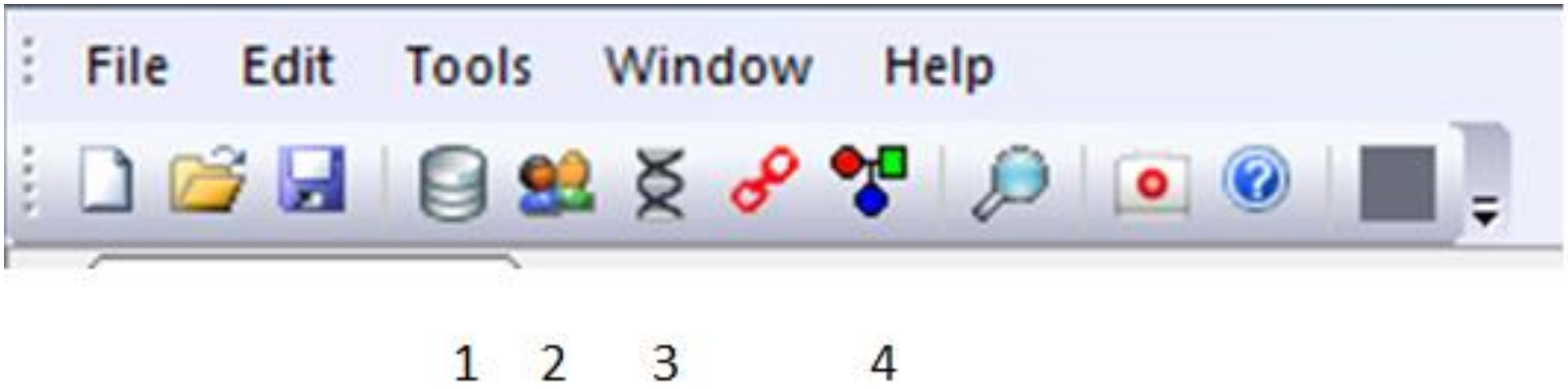
$$LR_2 = \frac{p_8}{p_8^2} = \frac{1}{0.554} = 1.81$$

$$LR = LR_1 \times LR_2 = 2.45 \times 1.81 = 4.4.$$



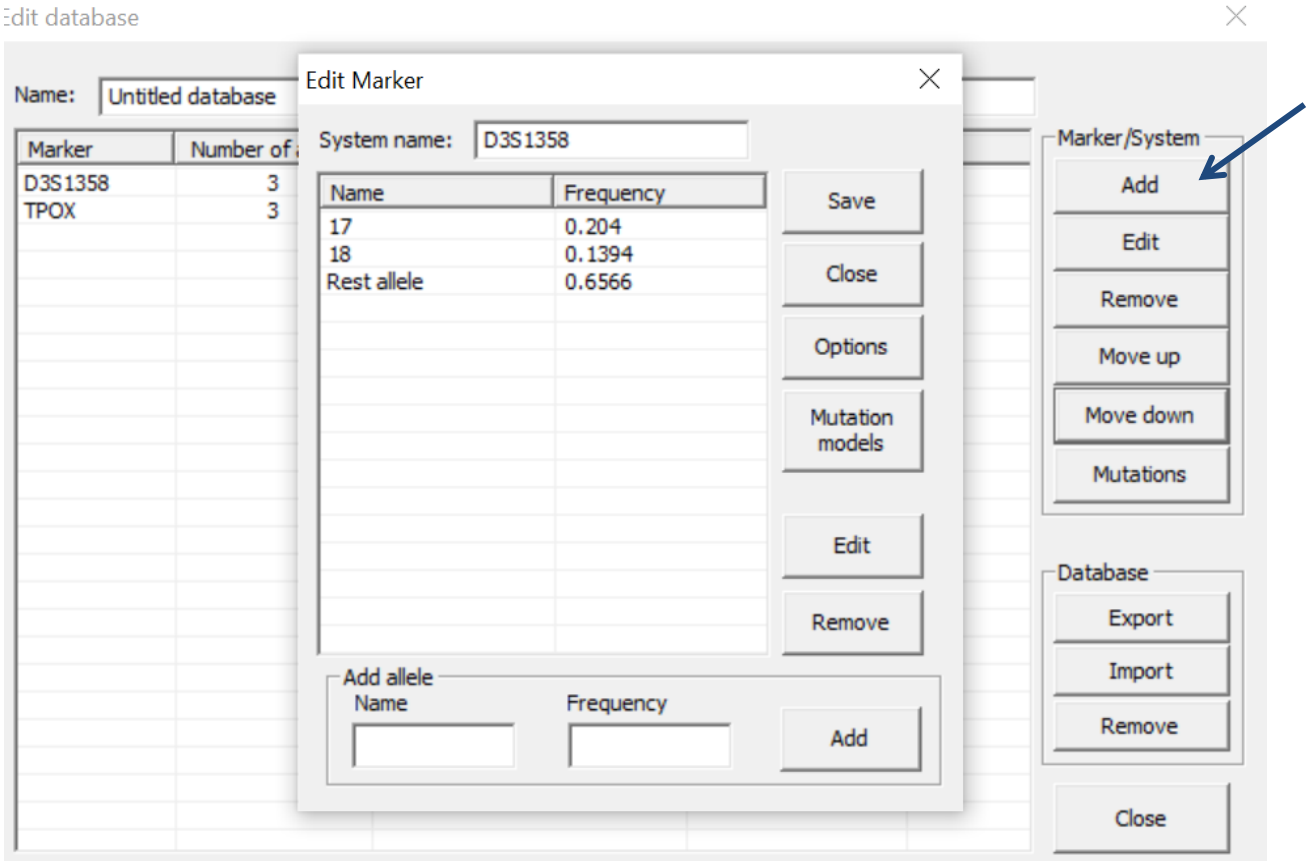
Ahora explicaremos los cuatro pasos necesarios para verificar los cálculos usando Familias

Cuatro pasos básicos



1. **Datos Generales de ADN (General DNA data)**. Cargar la base de datos, es decir, las frecuencias alélicas, etc.
2. **Personas (Persons)**. Individuos necesarios para definir los pedigríes.
3. **Datos de ADN del caso (Case DNA data)**. Genotipos
4. **Pedigríes (Pedigrees)**. Definición de las hipótesis y cálculo

Paso1: Datos Generales de ADN



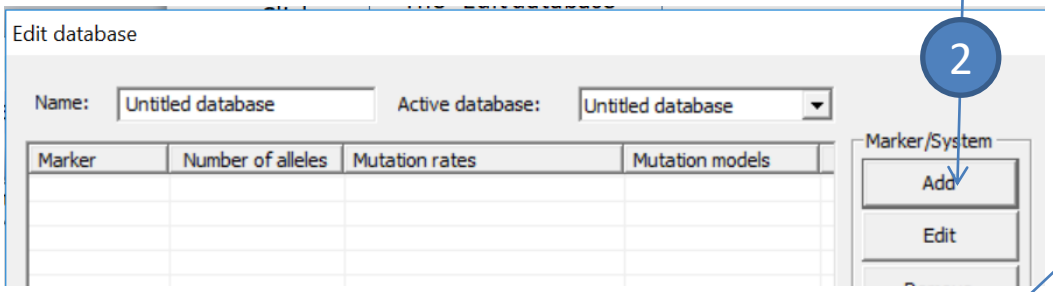
- Clic en Add (Añadir) para definir un nuevo marcador. Se abrirá una ventana donde podrá introducir los datos como se muestra en la figura.
- Haga lo mismo para el siguiente marcador (TPOX) que presenta los alelos 8 y 9, con frecuencias de 0.554 y 0.104, respectivamente.
- Se puede modificar el orden de los marcadores usando los botones “Move up” o “Move down”.

Paso 1 en detalle: datos Generales de ADN

Clic

1

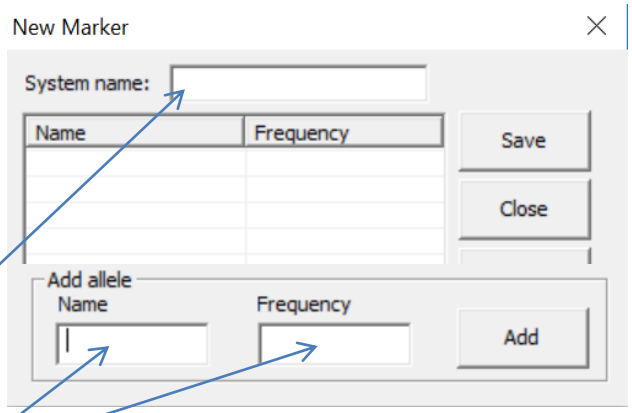
La ventana “Edit database” se abre



Clic

2

La ventana “New marker” se abre

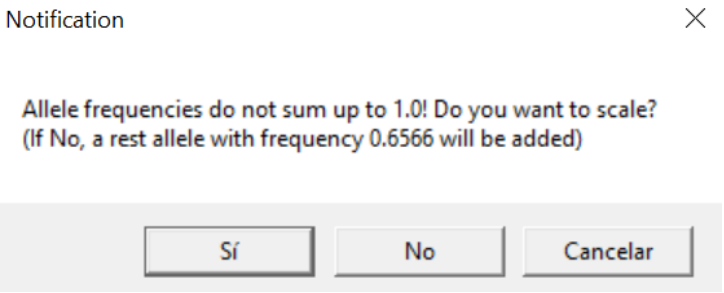


3 Añadir el nombre del primer marcador (D3S1358 en el ejemplo)

4 Añadir el primer alelo (17), frecuencia (0.204) y presionar ‘Add’

5 Hacer lo mismo con el segundo alelo (18, frec. 0.1394)

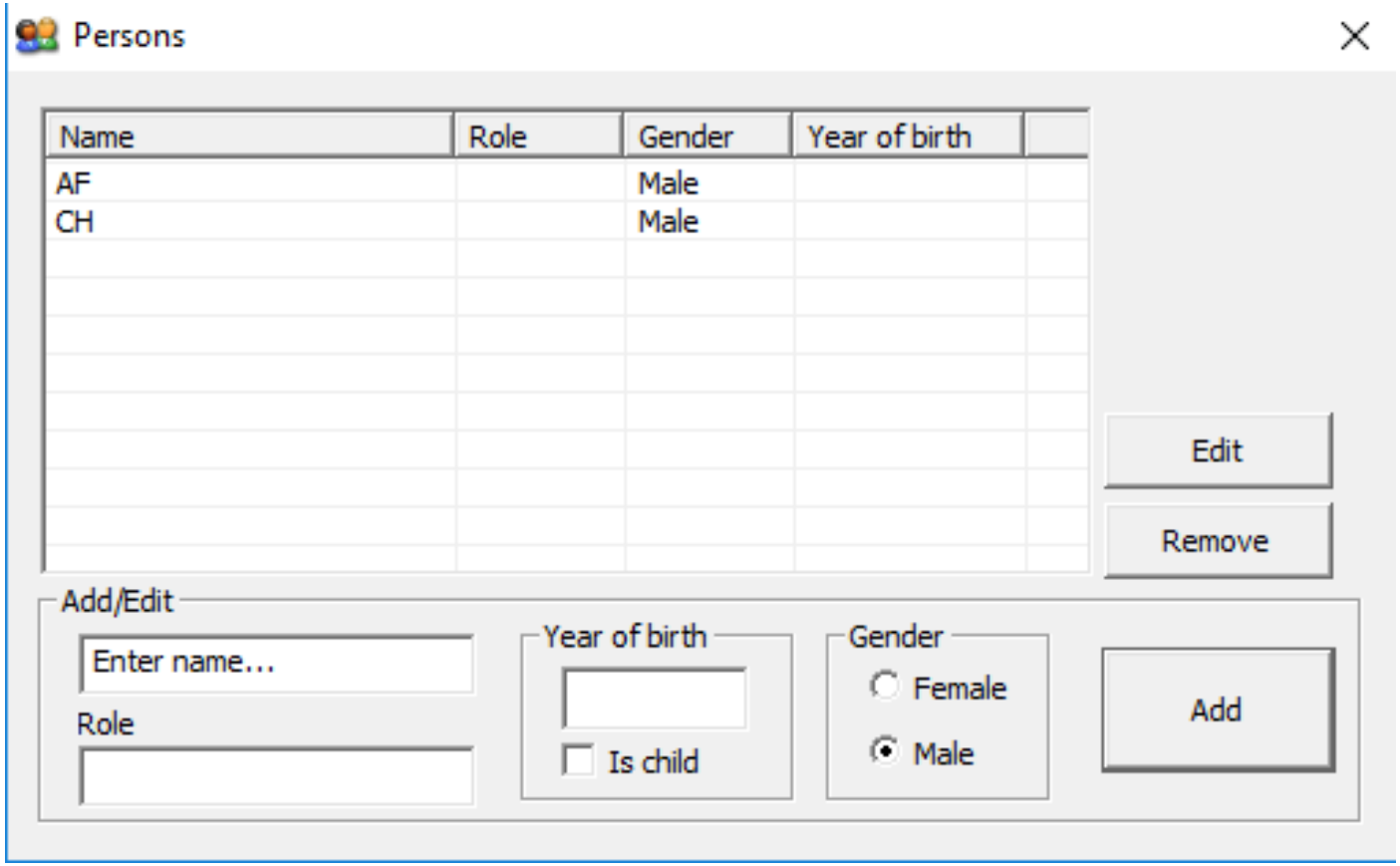
6 Clic ‘Save’ y aparecerá:



7 Para “Escalar”, clic Sí (la frec. de los dos alelos cambiará para que sumen 1)
Para “añadir un alelo más”, clic No (se añadirá un alelos más, rest allele, con frec. 0.6566)

8 Repetir el proceso para el segundo marcador (TPOX, en el ejemplo)

Paso 2. Personas



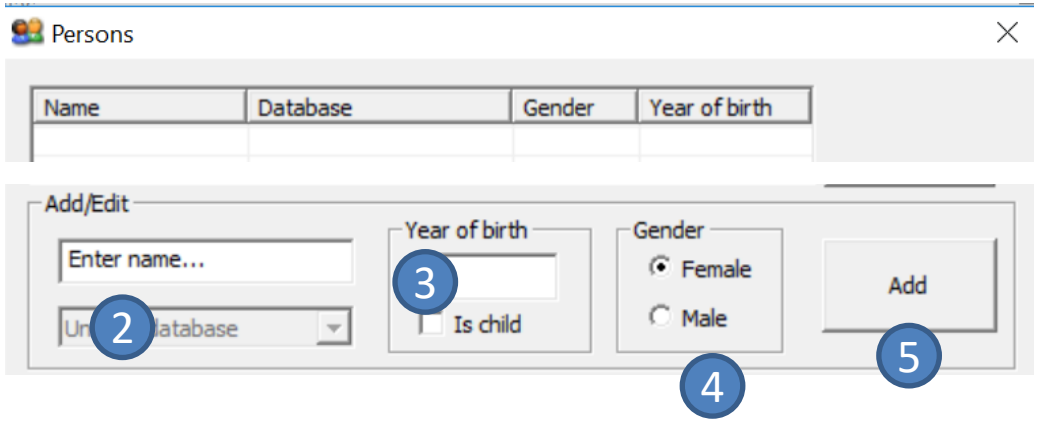
- Introduzca las personas: AF (presunto padre), y CH (hijo) como se muestra en la figura
- La información sobre el ‘Role’, ’Year of birth’, ‘Is child’ no se necesita ni se usa normalmente.

Paso 2 en detalle: personas

Clic

1

Se abre la ventana “Persons”



2 Introducir el nombre de la primera persona (AF)

3 Opcional: introducir el año de nacimiento

4 Introducir el género

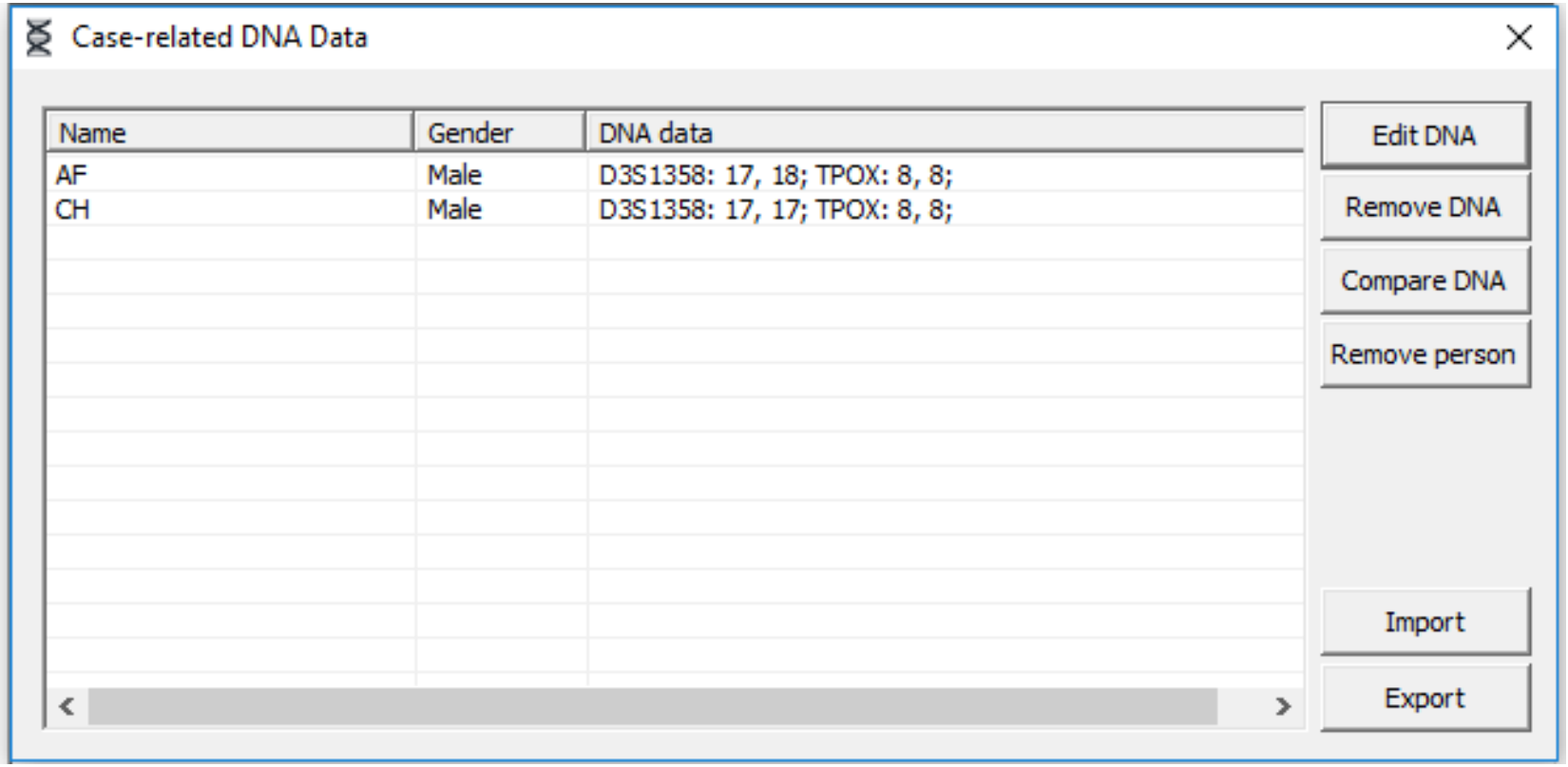
5 Clic ‘Add’

6 Hacer lo mismo con la siguiente persona (CH)

<u>Comentario 1</u>
Si introduce el año de nacimiento, se asegura de que una persona joven no puede ser el progenitor de una más vieja (útil para evitar errores en los pedigríes y en pedigríes imposibles)

<u>Comentario 2</u>
Si hace clic in la casilla “Is child”, se asegurará de que esta persona no puede tener hijos (útil por las mismas razones que anteriormente)

Paso 3. Datos del caso



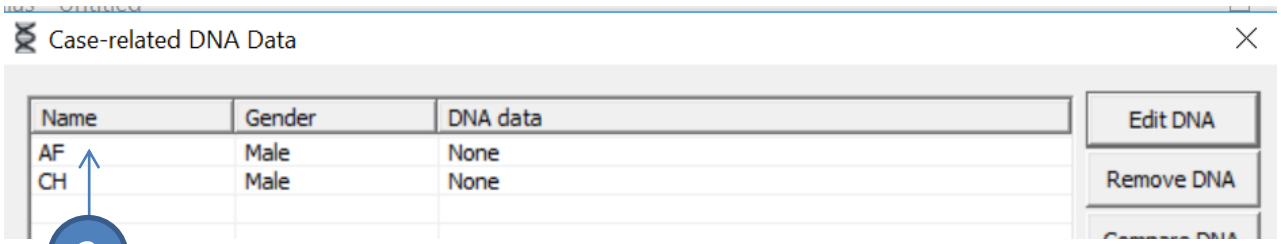
- Haga doble clic en cada persona. En la nueva ventana, seleccione el marcador presionando en el menú, seleccione los alelos, presione Add y OK. Obtendrá los datos como se muestra en la figura.

Paso 3 en detalle: Datos del caso

Clic

1

Se abre la ventana “Case-related DNA data”

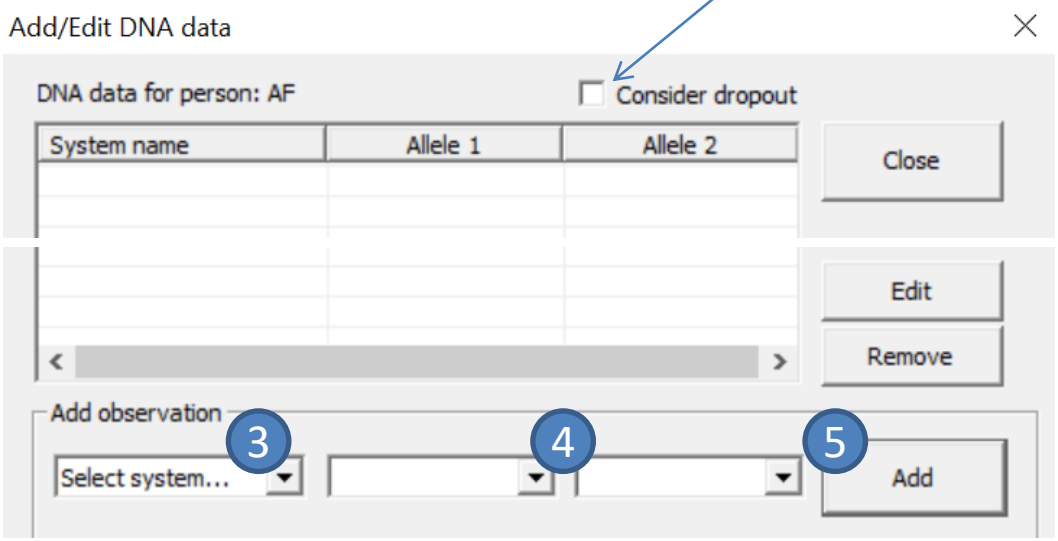


2

Doble Click (AF)

Comentario
Si quiere considerar drop-out en un individuo específico, clic en la casilla. Vea la diapositiva 18 para introducir la probabilidad de drop-out

Se abre la ventana “Add/Edit DNA data”



3 Seleccione el marcador (D3S1358)

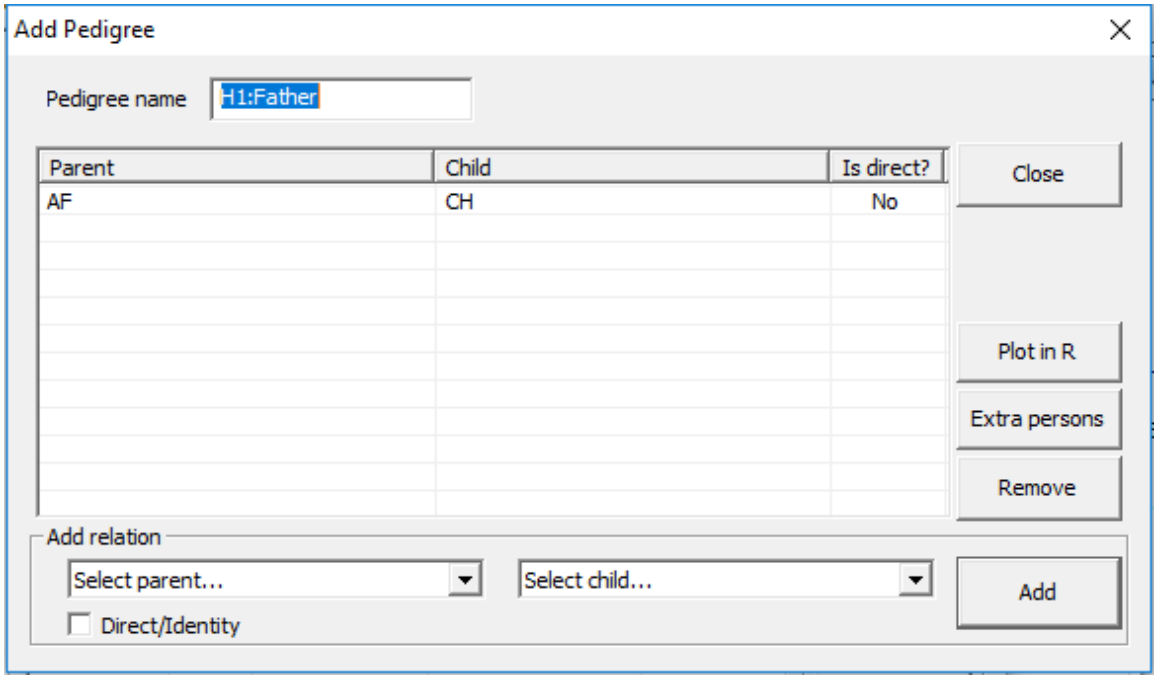
4 Introduzca el genotipo (17-18)

5 Clic ‘Add’

6 Haga lo mismo con TPOX y “close”

7 Repita el proceso desde 2 para CH

Paso 4. Pedigríes



Pedigree	Prior	Posterior	Likelihood Ratio	Ln likelihood
H1:Father	0.5	0.815639396	4.424152332	-6.92145
H2: Unrela...	0.5	0.184360604	1	-8.408529

- Clic en Add para introducir las hipótesis “H1: Padre” como se muestra en la figura (panel superior).
- Introducir el pedigrí “H2: no relacionado”. En este caso no hay que añadir ninguna relación.
- Haga clic en Calculate (calcular) para obtener lo que se muestra en la figura (panel inferior).

Paso 4 en detalle: Pedigríes

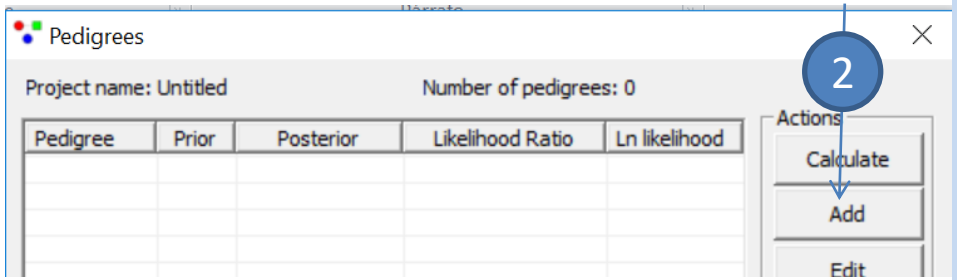
Clic

1

Se abre la ventana “Pedigrees”

Clic

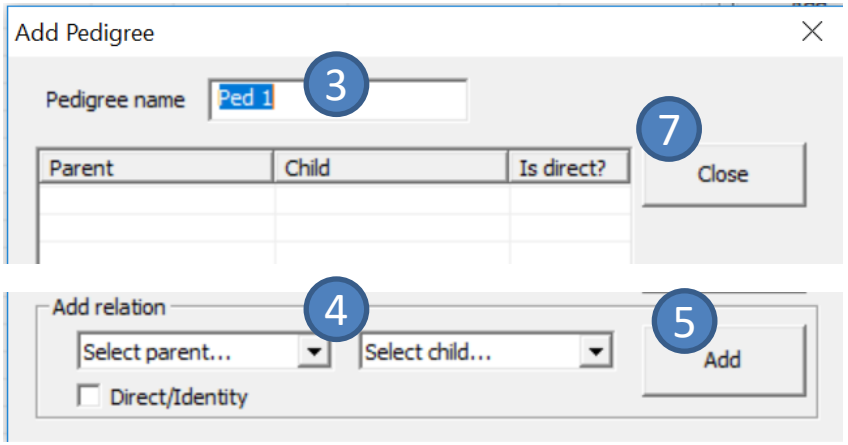
2



Comentario 1
Familias solo permite definir relaciones progenitor-hijo/a. Si quiere definir la relación “2 hermanos”, tiene que introducir una madre y un padre, y establecer que ambos son sus hijos.

Comentario 2
Para definir un mach directo, ej. gemelos, usar la opción Direct/Identity

Se abre la ventana “Add Pedigree”



3 Nombre el pedigrí (ej., H1: Father)

4 Defina la relación (AF padre de CH)

5 Clic ‘Add’

6 Nombre el segundo pedigrí (H2: unrelated). En este caso no tiene que definir ninguna relación

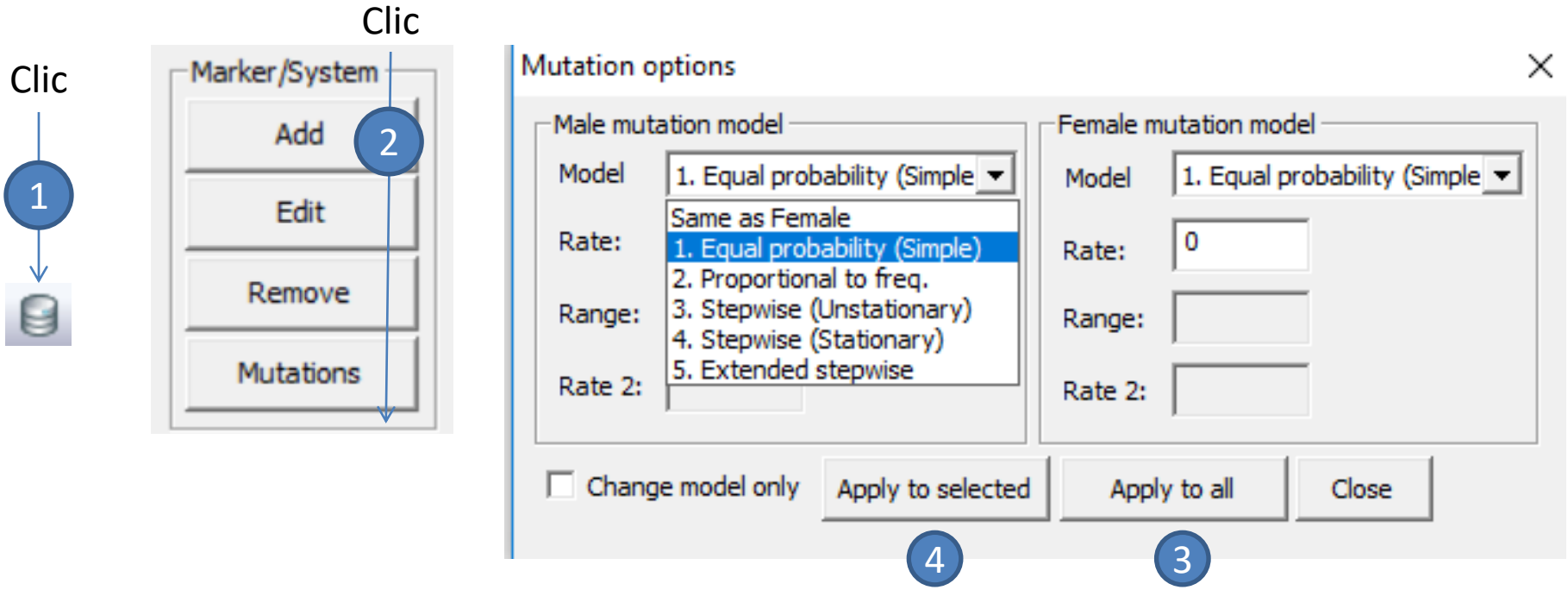
7 Clic ‘Close’

8 Ahora clic en ‘Calculate’ de la ventana Pedigrees para obtener:

Pedigree	Prior	Posterior	Likelihood Ratio	Ln likelihood
H1:Father	0.5	0.815639396	4.424152332	-6.92145
H2: Unrela...	0.5	0.184360604	1	-8.408529

2. Complicaciones

Mutaciones

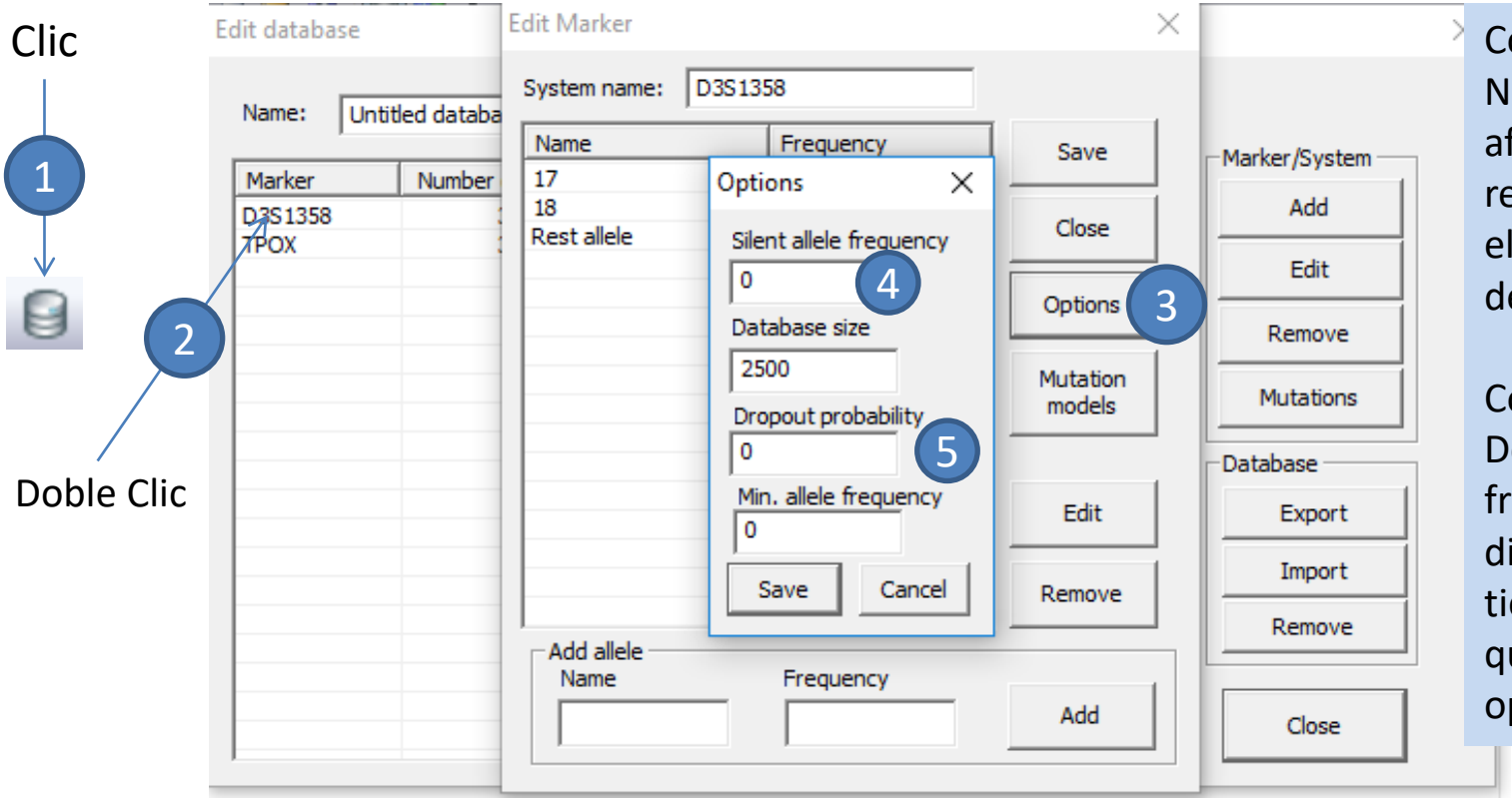


- Vaya al paso 1 (1), la ventana de la base de datos y presiones 'Mutations' (2) para obtener la ventana "Mutation options".
- Hay cinco modelos de mutación (ver el manual para más información)
- Se puede seleccionar el modelo para todos los marcadores (3) o sólo para los seleccionados (4), como se muestra en la figura.

Modelos de mutación en Familias

Modelo Simple
Todas las mutaciones ocurren con igual probabilidad. Apropiado cuando se necesita que los cálculos sean rápidos (ej. Pedigríes complejos y DVI) y también para marcadores tipo SNP.
Proporcional a la frecuencia
Las mutaciones ocurren con probabilidad proporcional a la frecuencia del nuevo alelo generado. Si la frecuencia es baja, la probabilidad de mutación será baja. Apropiado para probar propiedades estadísticas de los cálculos.
Modelo Stepwise
El modelo tradicional en el que se tiene en cuenta el número de steps entre el alelo origen y la versión mutada, decreciendo la probabilidad para mutaciones que implican más repeats. Apropiado para marcadores STR sin microvariantes.
Modelos Stepwise estable
El modelo Stepwise tradicional, en el cual las probabilidades de mutación se han ajustado para crear una matriz de mutación estable. Apropiado para pruebas de propiedades estadísticas de los cálculos.
Modelo Stepwise extendido (RECOMENDADO)
El modelo Stepwise más completo, donde se tienen en cuenta tanto repeats exactos como microvariantes. Apropiado para todos los marcadores STR.

Alelos silentes y Dropout



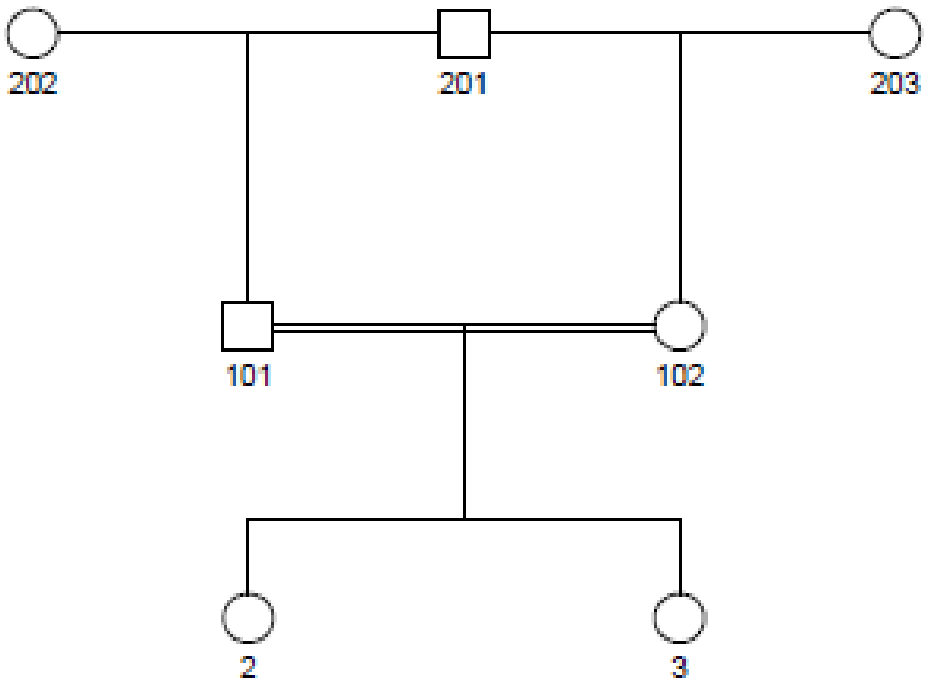
Comentario 1
Nótese que no afecta a los resultados cambiar el tamaño de la base de datos.

Comentario 2
Definir una frecuencia mínima distinta de cero sólo tiene efecto una vez que se chequea la opción Advanced

- Entrar en la ventana Database (1), doble clic en el marcador para editarlo (2.
- Clic en Options (3 y aparece la ventana en la que se puede introducir la frecuencia del alelo silente (4) y/o la probabilidad de Dropout (5).
- En el caso del Dropout , se requiere hacer algo más en la pestaña de Case DNA para especificar en qué individuos debemos tener en cuenta el dropout (ver diapo 12).

Endogamia: Pedigríes complejos

- Dibuje el pedigrí
- Identifique las personas necesarias para definir el pedigrí: personas que serán analizadas y personas que no serán analizadas (extra persons).
- Introducir todas las personas requeridas (Paso 2) y los pedigríes (Paso 4).



Pedigree name 2: Sisters

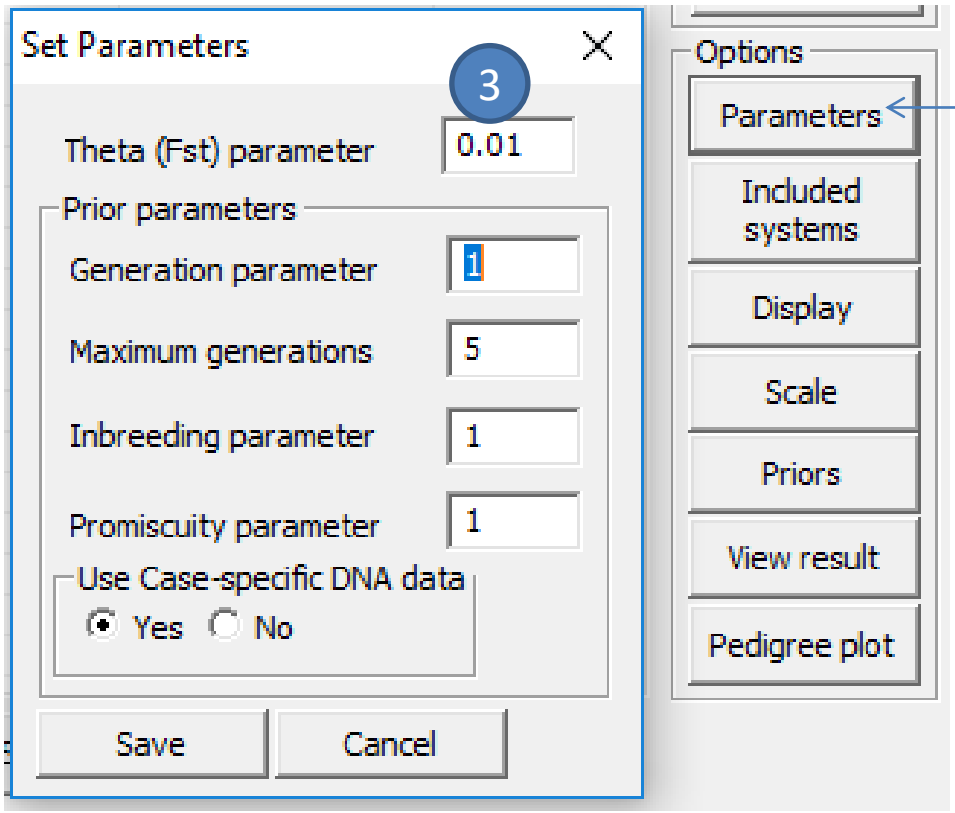
Parent	Child
202	101
201	101
201	102
203	102
101	2
101	3
102	2
102	3

Comentario: Las persons 202 y 203 no son necesarias, pero la función de dibujo (ploting) las introduce . La línea doble que conecta 101 y 102 indica endogamia, es decir, que los dos progenitores están relacionados familiarmente, dentro del propio pedigrí.

Corrección Theta

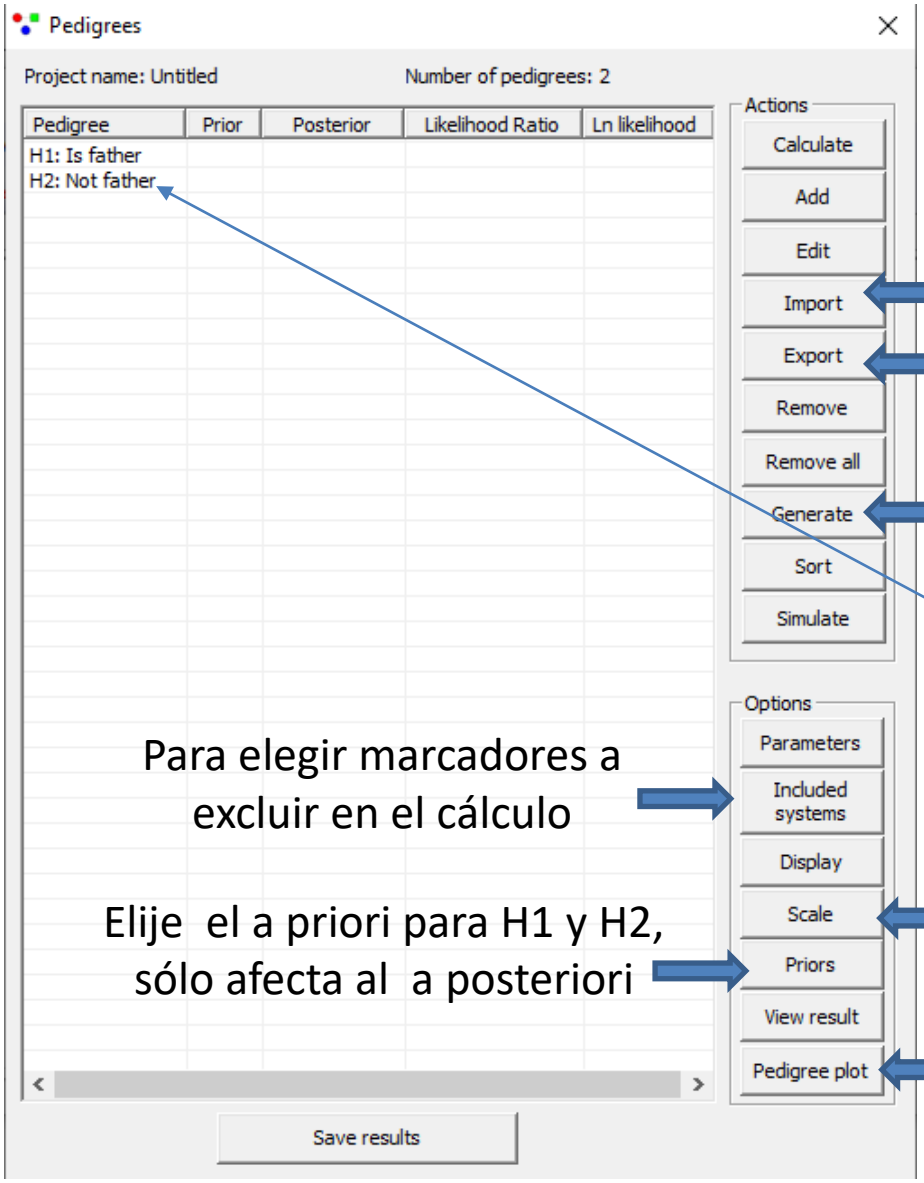
Clic

1



- Click en 'Parameters' (parámetros) en la ventana de Pedigríes (2). Hemos introducido el valor 0.01 (3).
- Los parámetros a priori ('Prior parameters') prácticamente nunca se cambian: no tienen impacto en el LR, sólo lo tienen en la probabilidad a priori (y por lo tanto en la probabilidad a posteriori).

Más opciones en la ventana “Pedigree”



Importa pedigríes, ver diapositiva siguiente

Exporta pedigríes

Genera pedigríes, ver dispositivas siguientes

Para elegir marcadores a excluir en el cálculo

Elije el a priori para H1 y H2, sólo afecta al a posteriori

Selecciona el denominador del LR: marcando H1 y presionando “Scale” obtiene el LR para H2 vs. H1

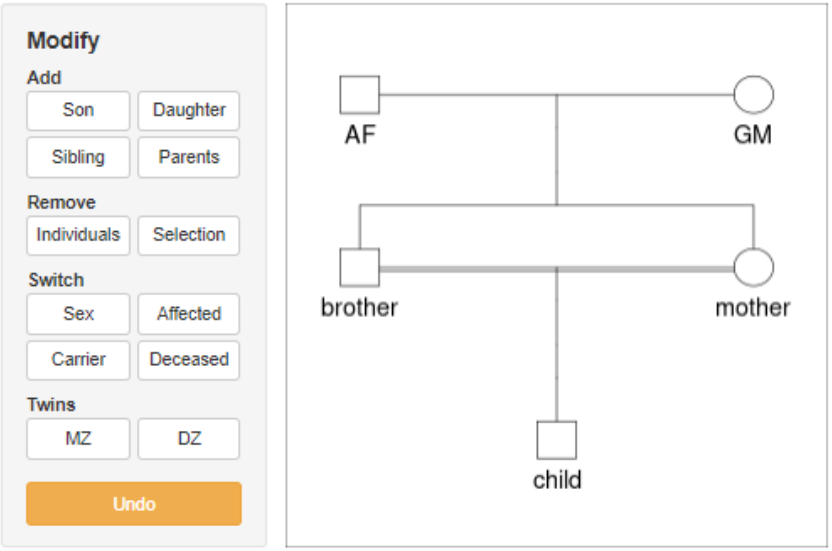
Dibuja el pedigrí, ver Sección 3

Creación de pedigríes en QuickPed, importación a Familias

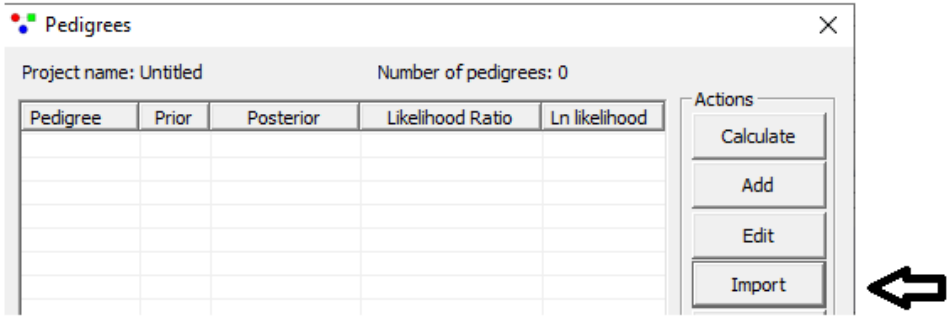
- La siguiente diapositiva muestra la app QuickPed (Vigeland, 2022), una herramienta online para dibujar pedigríes y analizar parentesco, que también se puede usar para crear pedigríes.
- Como se explica a continuación, los pedigríes se pueden importar a Familias

QuickPed: An Interactive Pedigree Creator

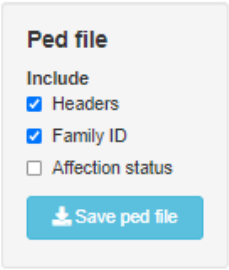
1



4



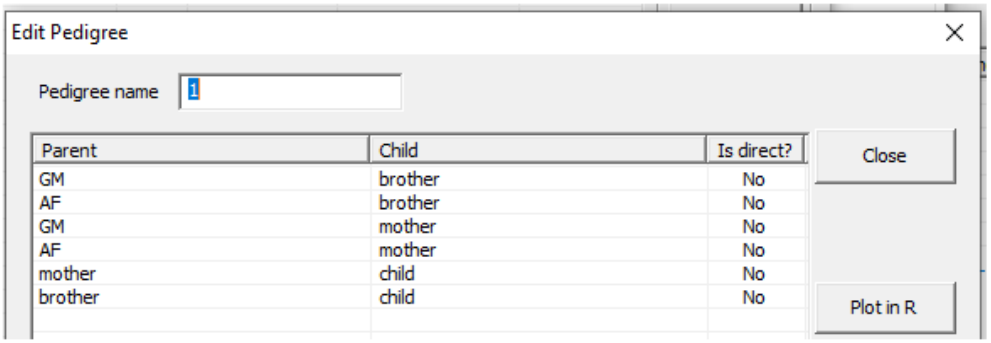
2



3

```
famid→id→fid→mid→sex
1→AF→0→0→1
1→GM→0→0→2
1→brother→AF→GM→1
1→mother→AF→GM→2
1→child→brother→mother→1
```

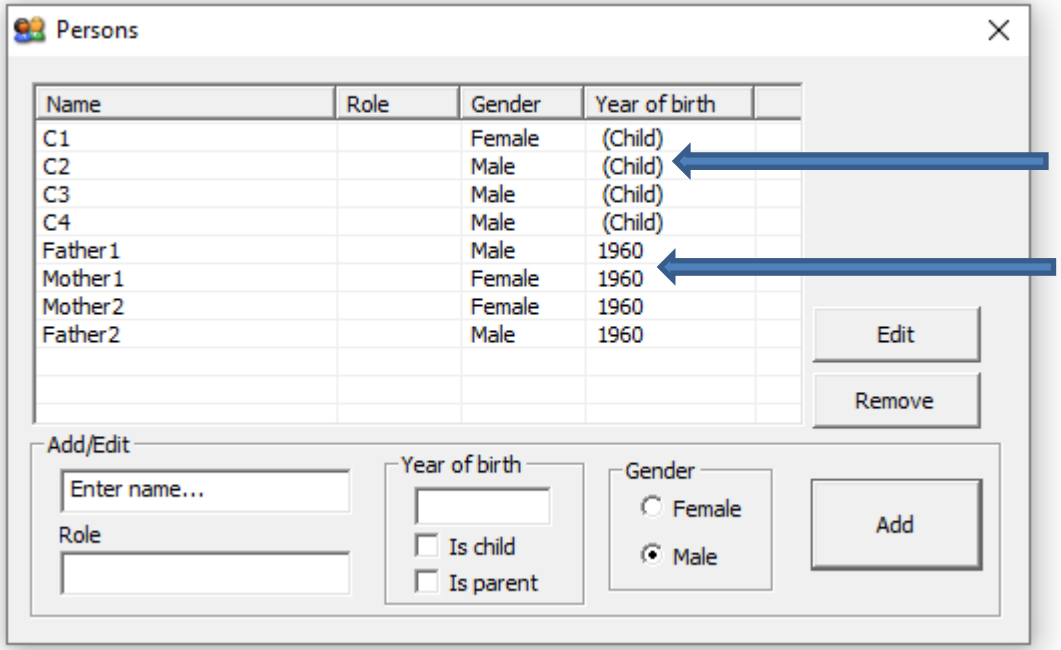
5



1. Cree el pedigrí por ejemplo en QuickPed: https://magnusdv.shinyapps.io/quickped/
2. Salve el archivo .ped de QuickPed
3. Opcional: puede ver el archivo .ped en tutorial.ped
4. Importe el archivo tutorial.ped en Familias
5. Ya tiene el archivo .ped incorporado en Familias

Generar pedigríes

Cuestión de interés: ¿Cuál es la relación entre los 4 niños C1, C2, C3 and C4? ¿Hermanos, medio hermanos o no están relacionados?



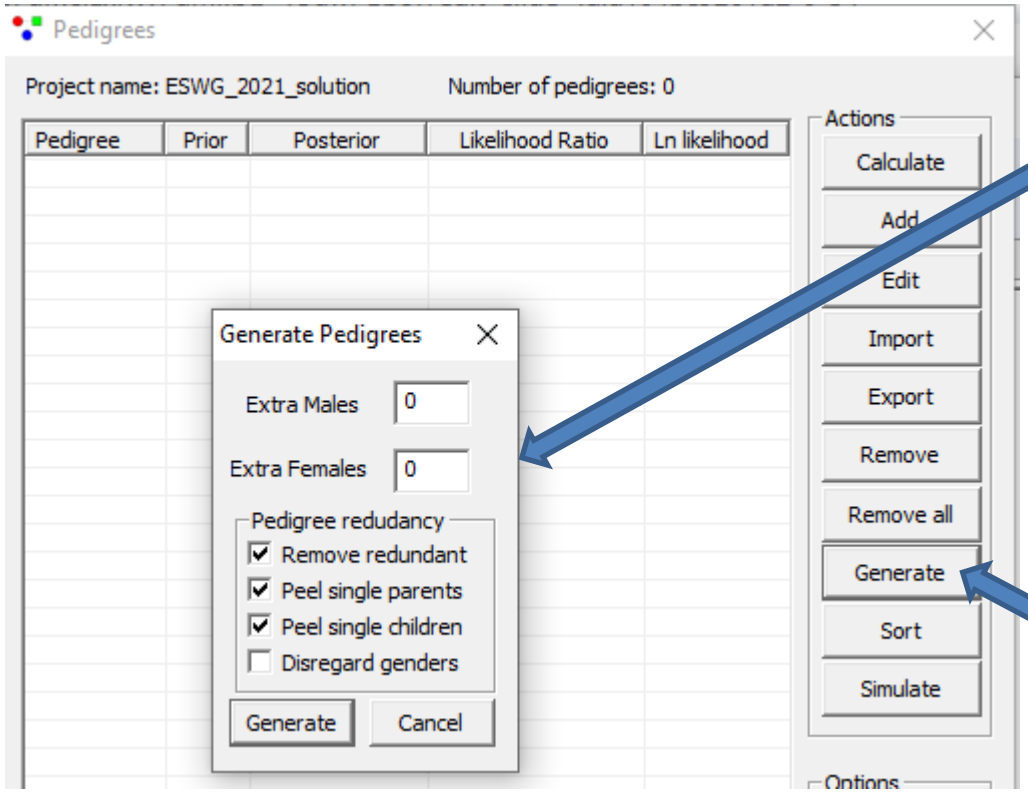
Definir que son niños, es decir, no pueden ser padres

Definir fechas de nacimiento de los padres (nacidos el mismo año para que no puedan ser padres unos de otros, evitando así varias generaciones)

Comentario
Ver el test de suficiencia ESWG de 2021: ejercicio con video y presentación

Generar pedigríes

Cuestión de interés: ¿Cuál es la relación entre los 4 niños C1, C2, C3 and C4? ¿Hermanos, medio hermanos o no están relacionados?



<u>Opciones para generar pedigríes</u>

a. Incluir hombres/ mujeres extra (no recomendado)
b. Eliminar pedigríes redundantes (recomendado)
c. No generar pedigríes con padres solteros (recomendado)
d. No generar pedigríes con hijos únicos (recomendado)
e. No tener en cuenta el género (opcional)

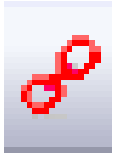
Generar pedigríes

Generar pedigríes

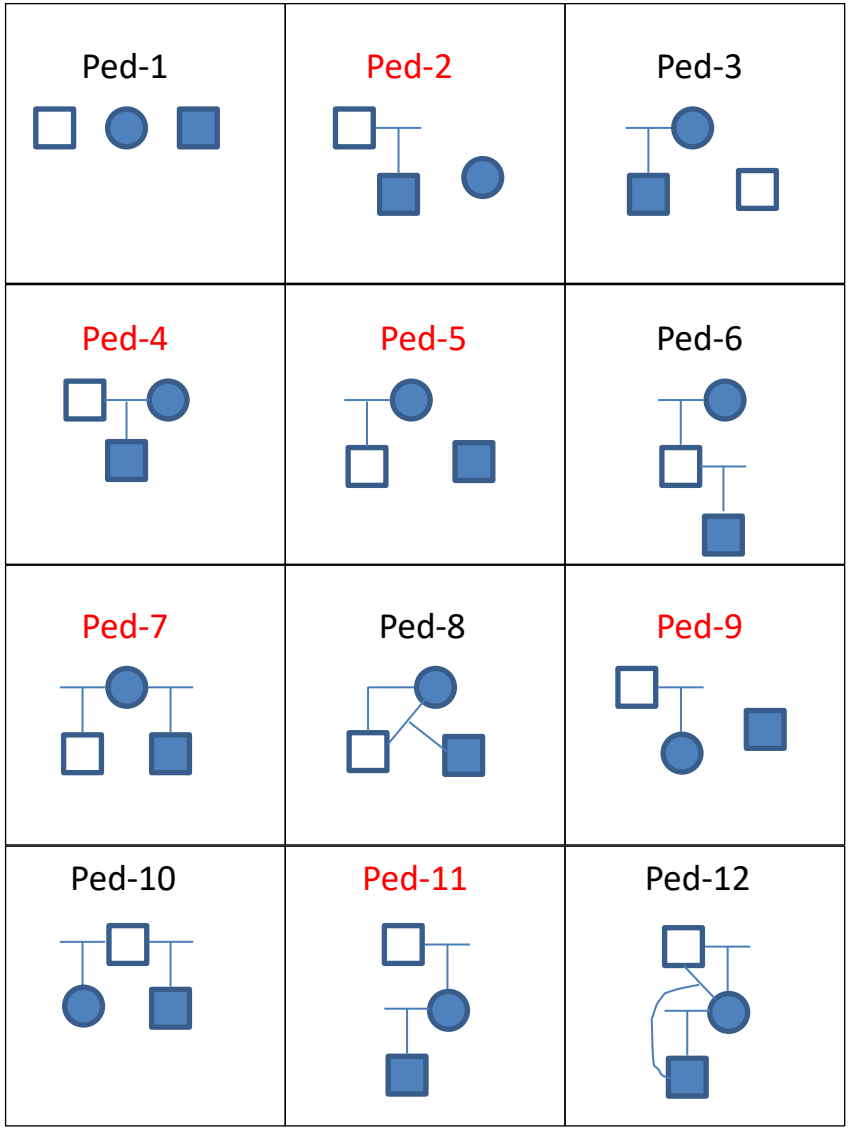
Cuestión de interés: ¿Cuál es la relación entre los 4 niños C1, C2, C3 and C4? ¿Hermanos, medio hermanos o no están relacionados?

Familias generará todos los pedigríes posibles con las restricciones impuestas por las personas y las opciones definidas. También generará, por ejemplo, pedigríes donde un individuo tenga un único progenitor y en los que ni el individuo ni el progenitor estén genotipados. Estos pedigríes no se generan si se han seleccionado las opciones "peel". La opción "Disregard genders" considerará que los pedigríes "medio hermanos paternos" y "medio hermanos maternos" son idénticos y por lo tanto el software los considerará como duplicados.

Puede restringir el número de pedigríes generados si algunas de las relaciones de parentesco son conocidas y lo define así en la ventana 'Known relationship':



Generar pedigríes con restricciones: ejemplo



Frec. 1 = 0.1; Frec. 2 = 0.9; sin mutación

La función “Generate” por defecto (sin restricciones) ofrece los 12 pedigríesde la izquierda (archivo peel.fam)

- Si se aplican las siguientes restricciones:
  - Remove redundant
  - Peel single parents
  - Peel single children

  Se obtienen los pedigríes 1, 3, 6, 8, 10 ,12
- Si el parámetro Inbreeding se pone en 0 se obtiene un LR=0 para los pedigríes 8 y 12.

3. R Familias, paramlink, plotting y blog Parentela

¿Qué es R ?

- Un marco para realizar cálculos estadísticos y numéricos
  - calculadora
  - Dibujo de figuras y gráficos flexible
  - Gran núcleo de funciones para el manejo de datos y el análisis numérico
  - Lenguaje de programación
  - Paquetes (conjuntos de funciones) externos
    - Cualquiera puede hacer uno
    - Hay miles!

- Sobre R:
  - Se encuentra disponible gratuitamente en https://cran.r-project.org/
  - Ampliamente utilizado
  - Puede hacer cualquier cosa (pero puede no ser fácil)

Instalación y carga de la librería Familias de R

Para acceder a las funciones de una librería externa en R, tomando como ejemplo la librería de Familias, se debe:

- Instalar el paquete, esto se hace sólo una vez. Lo siguiente funcionará en la mayoría de plataformas (más detalles aquí):

  - **install.packages("https://familias.name/RFamilias/Familias_2.5.zip", repos = NULL, type = "win.binary")** #(se necesita conexión a internet)

- Cargar la librería en R, se hace cada vez que se inicia una sesión en R:
  - **library(`Familias`)**

Para obtener ayuda, teclee

- **help(Familias)**

Ejemplo de sesión desde help("FamiliasPosterior"): copiar y pegar en R!

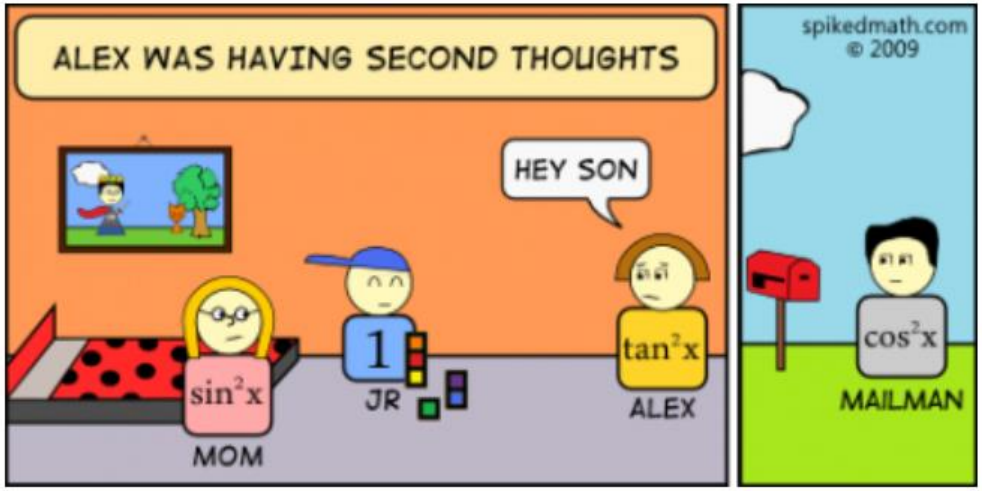
```
library(Familias)
persons <- c("mother", "daughter", "AF")
ped1 <- FamiliasPedigree(id=persons, dadid=c(NA, "AF", NA), momid=c(NA, "mother", NA),
                    sex=c("female", "female", "male"))
ped2 <- FamiliasPedigree(id=c(persons, "TF"), dadid=c(NA, "TF", NA, NA),
                    momid=c(NA, "mother", NA, NA),
                    sex=c("female", "female", "male", "male"))
ped3 <- FamiliasPedigree(id=c(persons, "TF", "gf", "gm"), dadid = c(NA, "TF", "gf", "gf", NA, NA),
                    momid=c(NA, "mother", "gm", "gm", NA, NA),
                    sex=c("female", "female", "male", "male", "male", "female"))
X11()
plot(ped1);title("ped1, i.e., AF is father")
mypedigrees <- list(isFather = ped1, unrelated=ped2, isUncle = ped3)
locus1 <- FamiliasLocus(frequencies=c(0.1, 0.2, 0.3, 0.4),
                   allelenames= c("A", "B", "C", "D"), name="locus1")
locus2 <- FamiliasLocus(c(0.2, 0.3, 0.5), c(17, 18, 19), "loc2", femaleMutationRate = 0.05)
myloci <- list(locus1, locus2)
datamatrix <- data.frame(locus1.1=c("A", "A", "A"), locus1.2=c("B", "B", "C"),
                    locus2.1=c(17, 19, 19), locus2.2=c(18, 18, 18))
rownames(datamatrix) <- persons
result = FamiliasPosterior(mypedigrees, myloci, datamatrix,ref=2)
result
```

Código R generado desde Familias (modo windows)

- Familias exporta a **R** el código necesario para:
  - ✓ Dibujar pedigríes
  - ✓ Hacer simulaciones condicionadas
  - ✓ Calcular la probabilidad de exclusión
  - ✓ …
- La exportación se hace desde Familias (modo Windows):
  - ✓ File > Export to R-Familias (se exporta todo)
  - ✓ 'Plot in R' en la ventana Edit Pedigree (para dibujar el pedigrí)
  - ✓ Simulaciones condicionadas en el módulo DVI

Blog Parentela

- Genética forense y probabilidad
- https://parentela.familias.name/category/category5/

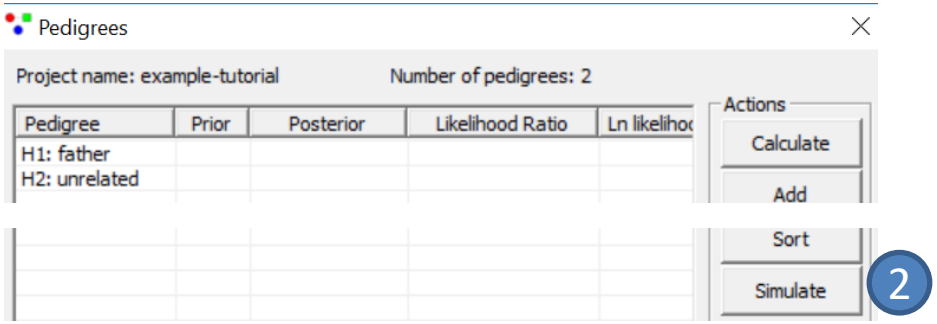


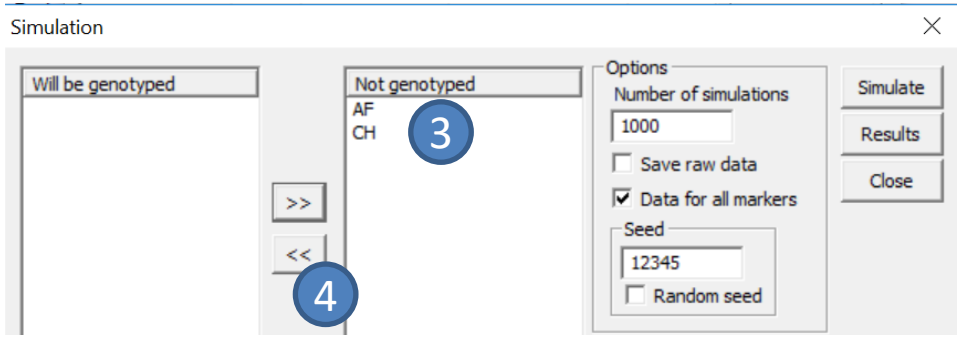
¿No os acordáis?
sen^2 x + cos^2 x = 1
:))))))

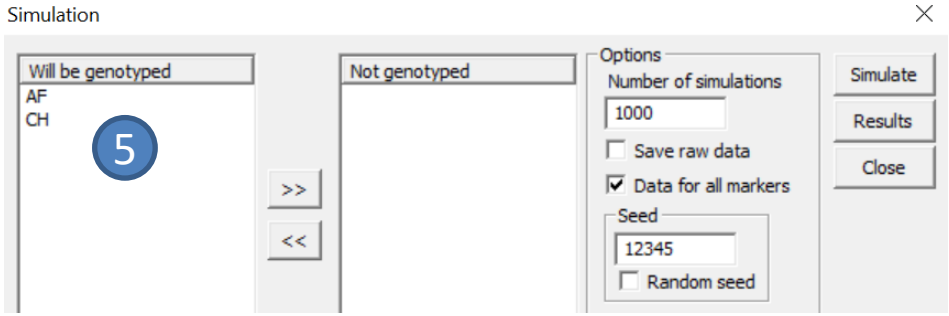
4. Simulaciones

Simulaciones

Clic

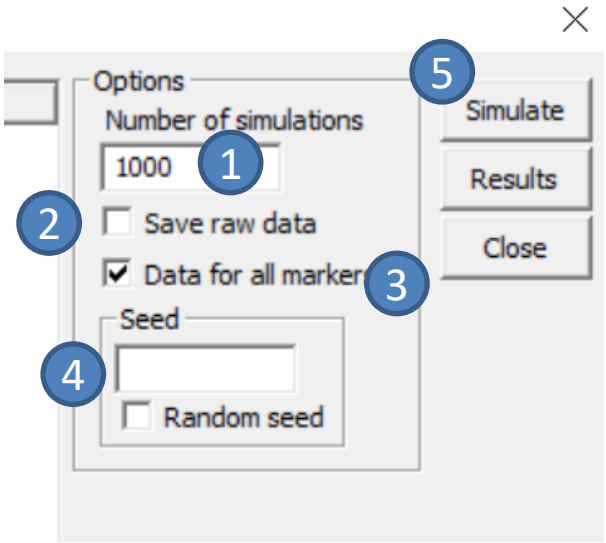




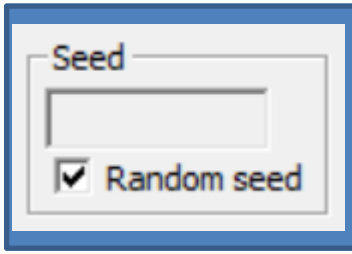


- En la ventana Pedigrees (1), una vez definidas las hipótesis que quiere simular, clic ‘Simulate’ (2). Se abrirá la ventana Simulation.
- Seleccione los individuos que serán genotipados (3) haciendo clic sobre ellos y después sobre las flechas (4). Esos individuos se moverán a la casilla “will be genotyped” (5).

Simulaciones - Opciones



1. Puede elegir el número de simulaciones (1000 por defecto)
2. Puede salvar el dato bruto (genotipos y/o LRs) en un archivo .txt
3. Puede realizar simulaciones con todos los marcadores de la base de datos de frec. o sólo con los seleccionados en el diálogo de la ventana Pedigrees (Included systems)
4. Puede especificar dónde comenzará la simulación (Seed).



Si usa “Random seed”, se obtendrán diferentes simulaciones cada vez que lleve a cabo una simulación.

Si no usa “Random seed”, tiene que introducir un número, y obtendrá exactamente el mismo resultado tras diferentes simulaciones (si todos los demás valores/parámetros no se cambian y si usa siempre el mismo número en Seed).

5. Clic ‘Simulate’

Simulaciones - Resultados

Simulation results

Numerator	Denominator	Median	Mean	95%	5%	Stdev
H1: father	H2: unrelated (TRUE)	0.7715	0.99	2.746	0.0003981	0.9648
H1: father (TRUE)	H2: unrelated	1.54	1.874	4.008	0.6872	1.218

LR limit

Save data

Report

Significado de la primera fila de la figura (H2: "AF y CH are unrelated" ES CIERTA):
No se esperan valores de LR = 0 ya que en el ejemplo se ha tenido en cuenta la mutación (tasa 0.001 y rango 0.1). Sin embargo, se esperan valores de LR bajos ya que la hipótesis cierta es la de "no relación" entre AF y CH.
La media de los valores de LR obtenidos es 0.99
El 50% de las simulaciones están por debajo (y por encima) de la mediana 0.7715
El 95% de las simulaciones da un valor de LR por debajo de 2.746
El 5% de las simulaciones da un valor de LR por debajo de 0.0003981

Significado de la segunda fila de la figura (H1: "AF is the father of CH" ES CIERTA):
La media de los valores de LR obtenidos es 1.874
50% de las simu. está por debajo de la mediana 1.54
95% de las simulaciones está por debajo de 4.008
5% de las simulaciones está por debajo de 0.6872

Puede usar este archivo para reproducir estos resultados

Comentario
Los valores de LR bajo ambas situaciones (H1 cierta y H2 cierta) se solapan. Así que no seremos capaces de llegar a ninguna conclusión sobre la paternidad si sólo analizamos estos 2 marcadores (con 3 y 2 alelos).

Simulaciones – Un ejemplo “más real”

Mismas hipótesis, mismas opciones de simulación (1000 simus, todos los marcadores, Seed 12345), pero 16 marcadores aSTR con frecuencias alélicas reales (PP ESX17)

Simulation results

Numerator	Denominator	Median	Mean	95%	5%	Stdev
H1: father	H2: unrelated (TRUE)	4.051e-031	1.018e-009	8.159e-017	2.145e-047	3.045e-008
H1: father (TRUE)	H2: unrelated	4.418e+006	6.317e+008	1.101e+009	3.319e+004	5.737e+...

LR limit

Save data

Significado de la primera fila de la figura (H2: “AF and CH are unrelated” ES CIERTA):
No se esperan valores de LR = 0 ya que en el ejemplo se ha tenido en cuenta la mutación (tasa 0.001 y rango 0.1). Sin embargo, se esperan valores de LR bajos ya que la hipótesis cierta es la de “no relación” entre AF y CH.
La media de los valores de LR obtenidos es 1.018e-09
El 50% de las simulaciones están por debajo de la mediana 4.051e-31
El 95% de las simulaciones da un valor de LR por debajo de 8.159e-17
El 5% de las simulaciones da un valor de LR por debajo de 2.145e-47

Puede usar este archivo para reproducir los resultados

Significado de la segunda fila en la figura (H1: “AF is the father of CH” ES CIERTA):
La media de los valores de LR obtenidos es 6.317e+08
50% de las simus. Está por debajo de la mediana 4.418e+06
95% de las simulaciones está por debajo de 1.101e+09
5% de las simulaciones está por debajo de 3.319e+04

Comentario
Los valores de LR bajo ambas situaciones (H1 cierta y H2 cierta) NO se solapan. Podremos alcanzar una conclusión sobre la paternidad

Simulaciones – LR limit

Útil para ver los resultados de una forma más fácil de entender
El botón “LR limit” se utiliza para encontrar la fracción de simulaciones que excede un umbral de LR

Simulation results

Numerator	Denominator	Median	Mean	95%	5%	Stdev
H1: father	H2: unrelated (TRUE)	4.051e-031	1.018e-009	8.159e-017	2.145e-047	3.045e-008
H1: father (TRUE)	H2: unrelated	4.418e+006	6.317e+008	1.101e+009	3.319e+004	5.737e+...

LR limit (1)

Save data

Simulation limits

Settings

LR threshold: 1000 (2)

Pedigree: H1: father

Versus: H2: unrelated

Update (3)

Close

Statistic information (4)

Simulating the true relationship (H1: father), 99.8000% of the 1000 simulations were equal to or above the LR limit

Simulating the alternative hypothesis (H2: unrelated) yielded 0.0000% of false positives

- Clic en LR limit (1). Se abre la ventana “Simulation limits”.
- Seleccione su umbral de LR (2), 1000 por ejemplo, y clic “Update” (3).
- Obtendrá la información en 4 .

Significado
Si su umbral de LR = 1000, no se esperan falsos positivos si analiza 16 marcadores. Los falsos negativos son muy pocos (ya que el 99.8% de las simulaciones dio LRs ≥ 1000)

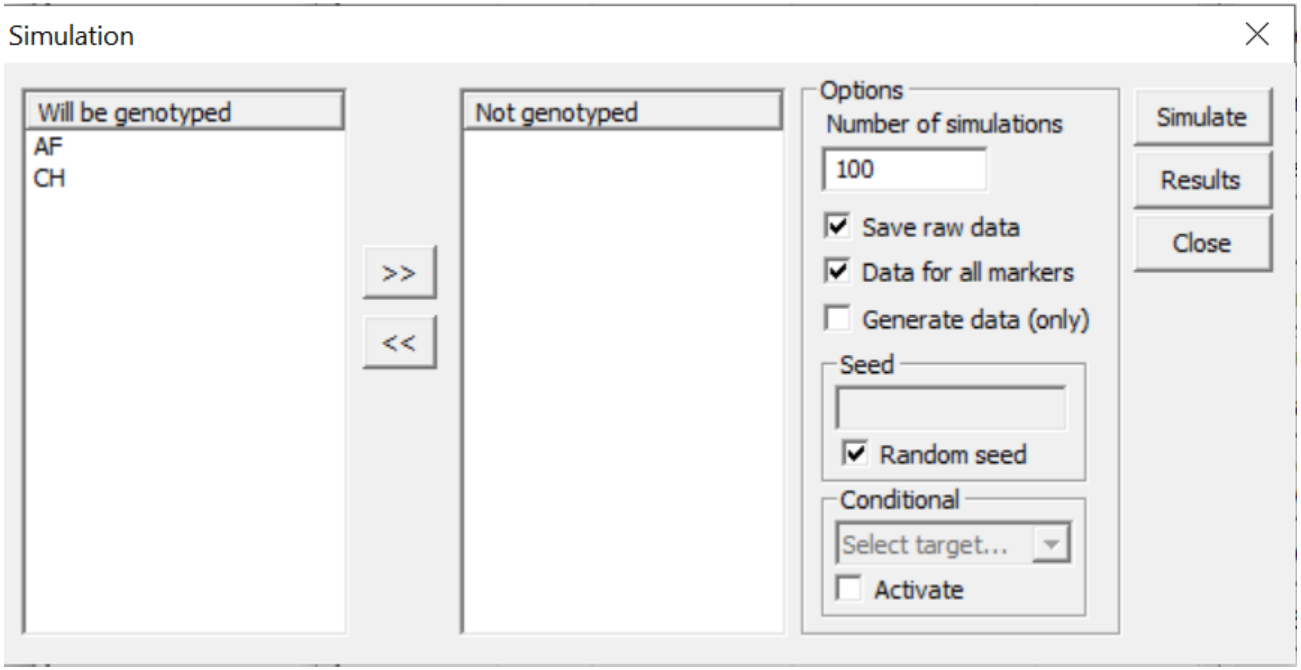
Falso positivo significa LR ≥ 1000 si H2 es cierta
Falso negativo significa LR ≤ 1000 si H1 es cierta

Simulaciones – Más información

- Recomendamos al menos 1000 simulaciones para obtener una idea de la distribución de LRs. Sin embargo, para explorar posibilidades más extremas, se debe realizar un número mucho mayor de simulaciones, por ejemplo 100,000.
- Recomendamos usar «random seed» como norma general, mientras que se debe usar «fixed seed» (ej., 12345) cuando necesitamos que los resultados sean reproducibles, por ejemplo en las publicaciones.
- Consejo: El dato bruto exportado con la información de los genotipos, puede usarse como entrada (input) en Familias. Por ejemplo, si simulamos datos para un escenario DVI (ver Sección 6), los datos de las familias pueden simularse y usarse como punto de partida.
- El dato bruto de los LRs obtenidos se puede exportar desde la ventana de resultados. Recomendamos usar el logaritmo del LR ya que la distribución de LRs resultante se aproxima a una distribución Normal. <u>Ver ejemplos en las páginas siguientes.</u>

Gráficos de las simulaciones I. Lo que tiene que hacer en Familias

- Descargue el archivo localizado en
  http://familias.name/tutorial/Basics-sim.fam
  y realice 100 simulaciones usando las siguientes opciones



Gráficos de las simulaciones II. Excel

- Obtendrá un archivo que podrá visualizar en Excel. Si es experto en Excel también podrá obtener la representación gráfica de los datos (si no es experto, ver las páginas siguientes).

Simulation raw data, number of simulations: 100

#Simu	Marker	TRUE PED	logLikelihc	logLikelihc	AF	Allele 1	Allele2	CH	Allele 1	Allele2
1	D3S1358	H1:Father	-2.16555	-2.63434		17	99		17	99
1	D3S1358	H2: Unrela	-3.94669	-3.01511		99	18		99	17
2	D3S1358	H1:Father	-2.63076	-3.39588		99	18		99	18
2	D3S1358	H2: Unrela	-3.94669	-3.01511		18	99		99	17

- Puede organizar los datos en Excel (Datos > las opciones de filtrado pueden ser útiles)

Gráficos de las simulaciones III. Formato de los datos para introducir en R.

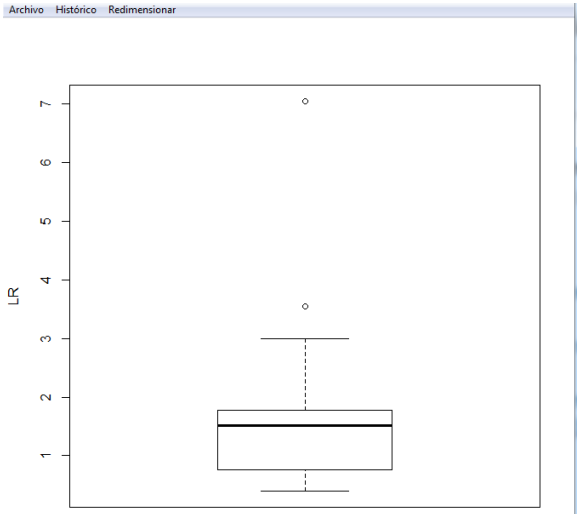
1	Sim	Marker	TRUEPED	logH1	logH2
2	1	D3S1358	H1	-2.17	-2.63
3	1	D3S1358	H2	-3.95	-3.02
4	2	D3S1358	H1	-2.63	-3.4
5	2	D3S1358	H2	-3.95	-3.02

- Puede insertar gráficos y hacer más análisis en Excel (no se explica en este Tutorial).
- Si prefiere usar R, guarde el archivo de Excel como archivo de texto delimitado por tabulaciones, con el formato que se indica en la figura (puede llamarlo Sim.txt). Tenga en cuenta que:
  - Los nombres de las Variables tienen que cumplir los requisitos de R (por ejemplo, 'TRUEPED', no 'TRUE PED')
  - Los Valores de las variables también tienen que cumplir los requisitos de R: Podemos usar por ejemplo 'H1.father', pero no 'H1:father' o 'H1 Father'.

Gráficos de las simulaciones IV. Usando R

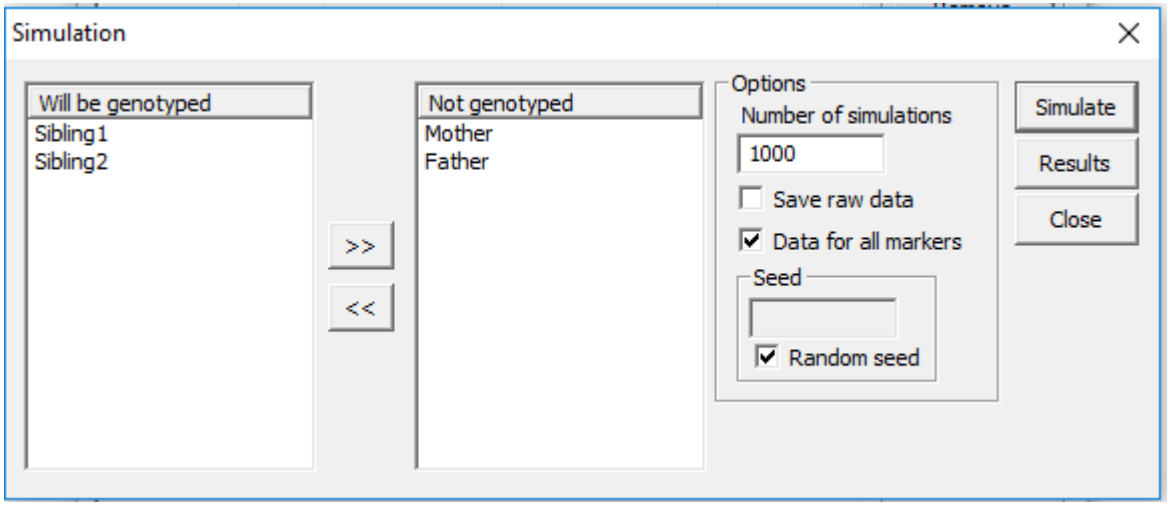
- Proporcionamos un ejemplo para hacer gráficas en R, con el archivo ya en su formato correcto.
- Puede copiar y pegar estos comandos en R, pero se requiere conexión a internet, ya que el archivo en cuestión se encuentra en el link de la primera línea de comandos (http://familias.name/tutorial/Sim.txt):

```
dat = read.table("http://familias.name/tutorial/Sim.txt", header = TRUE)
summary(dat) # Check data
dat = dat[dat$TRUEPED == "H1", ] #Extracts data for hypothesis H1
summary(dat)
logLR = dat$logH1 - dat$logH2 # Calculates log (=ln , i.e., base=e) of LR
hist(logLR, xlab = "logLR", ylab = "count",  main ="logLR (lnLR)")
boxplot(logLR, ylab ="logLR")
LR = exp(logLR)
hist(LR, xlab = "LR", ylab = "count ", main ="LR")
boxplot(LR, ylab ="LR")
```



Gráficos de simulaciones V. Alternativa

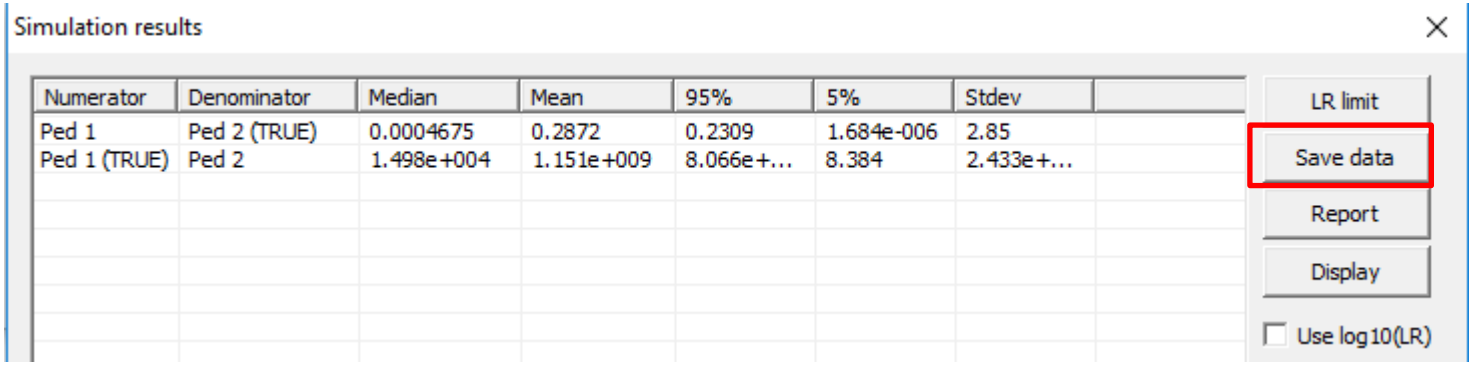
- Descargue el archivo localizado en
  http://familias.name/tutorial/Siblings-sim.fam
  y realice 1000 simulaciones usando las siguientes opciones



Observe que la opción ”Save raw data” NO está seleccionada!

Gráficos de simulaciones VI. Alternativa

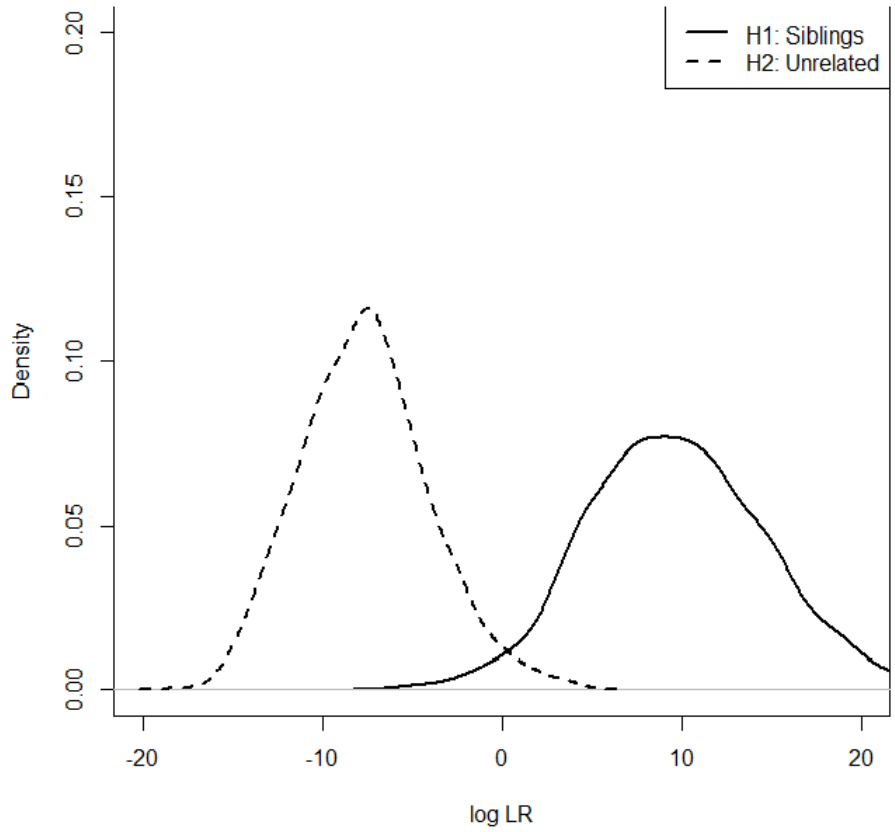
- Se pueden guardar los LRs obtenidos desde el diálogo de resultados para luego procesarlos como se muestra abajo




```
# Se mostrará un gráfica de función de densidad (distribución de los valores de logLR). Para usar este código se necesita conexión a internet
dat = read.table("http://familias.name/tutorial/Sim2.txt", header = FALSE, skip=5)
plot(density(log(dat$V1)), xlab= "log LR", ylab= "Density", col=1, lwd=2, main="", xlim=c(-20,20), ylim=c(0,0.2))
points(density(log(dat$V2)), col=1, lty=2, lwd=2, type= "l")
legend("topright", c("H1: Siblings", "H2: Unrelated"), lwd=2,lty=1:2)
# Next exceedance plots, i.e. the probability to exceed each treshold given each hypothesis
dev.new()
threshold <- exp(seq(log(0.01), log(20), length.out=1000))
eh1 = eh2 = threshold
for (i in 1:1000) { eh1[i] = sum(log10(dat$V1)>threshold[i])/1000; eh2[i]= sum(log10(dat$V2)>threshold[i])/1000; }
plot(x=threshold, y=eh1, lwd=2, ylab= "Exceedance probability" ,xlab= "log Threshold" , type= "l")
points(x=threshold, y=eh2, lwd=2, lty=2 , type= "l")
grid(); legend("topright", c("H1: Siblings", "H2: Unrelated"), lwd=2,lty=1:2)
```

Gráficos de simulaciones VII

Con los datos del archivo http://familias.name/tutorial/Sim2.txt , obtendrá:



Significado

Tras 1000 simulaciones

- los valores de LR si H2 es cierta van desde 10^{-20} a 10^{5} (más o menos)
- los valores de LR si H2 es cierta van desde 10^{-5} a $>10^{20}$

Afortunadamente, la probabilidad de que el LR (para no relacionados) sea mayor que el LR (para hermanos) es pequeña, con R se estima que es de 0.002 (2 de cada 1000 simulaciones):

```
sum(dat$V2 > dat$V1)/1000
```

Simulaciones condicionadas I

- Se tienen en cuenta los genotipos de los individuos analizados
- Muy útiles en escenarios DVI, ya que los perfiles genéticos de referencia determinan cuán informativo es un pedigrí
- Ejemplo: la referencia es el padre y la persona desaparecida es el hijo. Si el perfil genético del padre es raro, los valores de los LRs simulados tienden a ser más elevados que si el genotipo del padre fuera común.
- Sólo se simulan perfiles genéticos compatibles con el perfil genético del padre

	Father
D3S1338	14-16

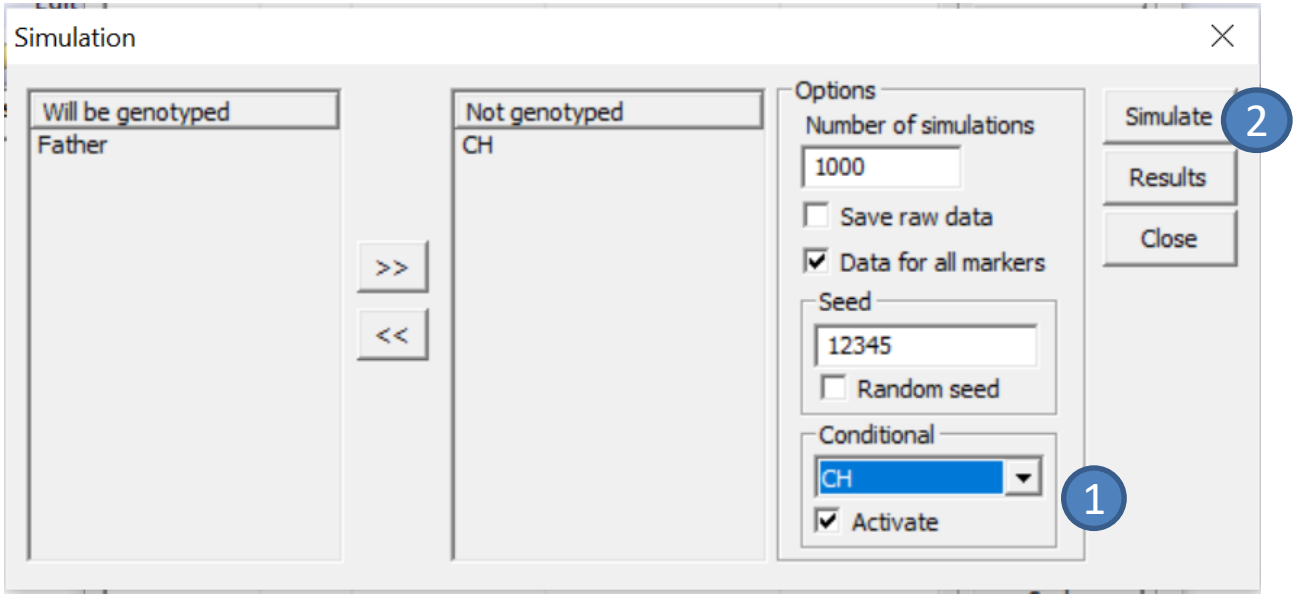
D3S1338	Simulated child
Simu1	14-14
Simu2	14-16
Simu2	16-17
…	…

Sólo se simularán genotipos que tengan el alelo 14, o el 16, o ambos
El software calcula el LR para cada genotipo simulado

Simulaciones condicionadas II

Use los datos del archivo http://familias.name/tutorial/cond-simu-father-child.fam

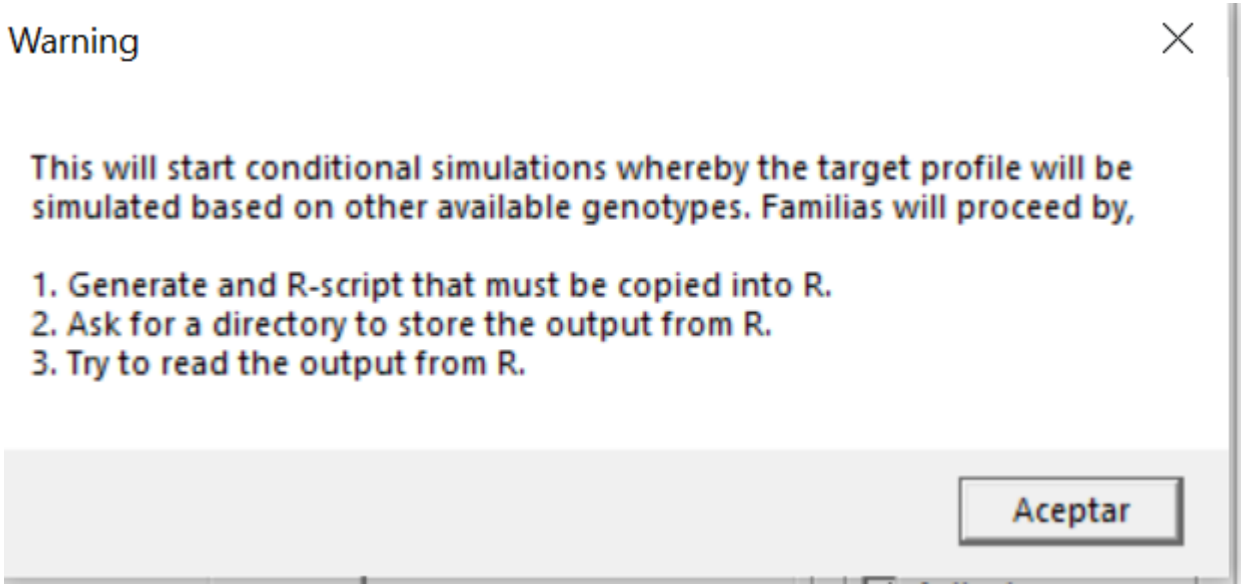
Abrir la ventana Simulation desde la ventana Pedigrees
Activar "Conditional" y seleccionar la diana a simular (en este caso, el hijo CH) 1
Clic en Simulate 2



Nota: Familias sólo simulará datos para la persona diana, teniendo en cuenta los genotipos ya analizados! No está disponible la opción para incluir/excluir personas (las personas que aparecen en "Will be genotype/Not genotyped" están en gris)

Simulaciones condicionadas III

Se genera un R-script.
Copiar y pegar el script en la consola R



Simulaciones condicionadas IV

Los resultados aparecerán en la pantalla de Familias

Simulation results

Numerator	Denominator	Median	Mean	95%	5%	Stdev
H1: father	H2: unrelated (TRUE)	0	0	0	0	0
H1: father (TRUE)	H2: unrelated	1.968e+008	1.4e+010	2.844e+010	1.058e+006	1.796e+...

Simulation limits

Settings

LR threshold: 1000

Pedigree: H1: father

Versus: H2: unrelated

Update

Close

Statistic information

Simulating the true relationship (H1: father), 100.0000% of the 1000 simulations were equal to or above the LR limit

Simulating the alternative hypothesis (H2: unrelated) yielded 0.0000% of false positives

5. Blind search («búsqueda a ciegas»)

Input file: tutorial-Ch5.fam

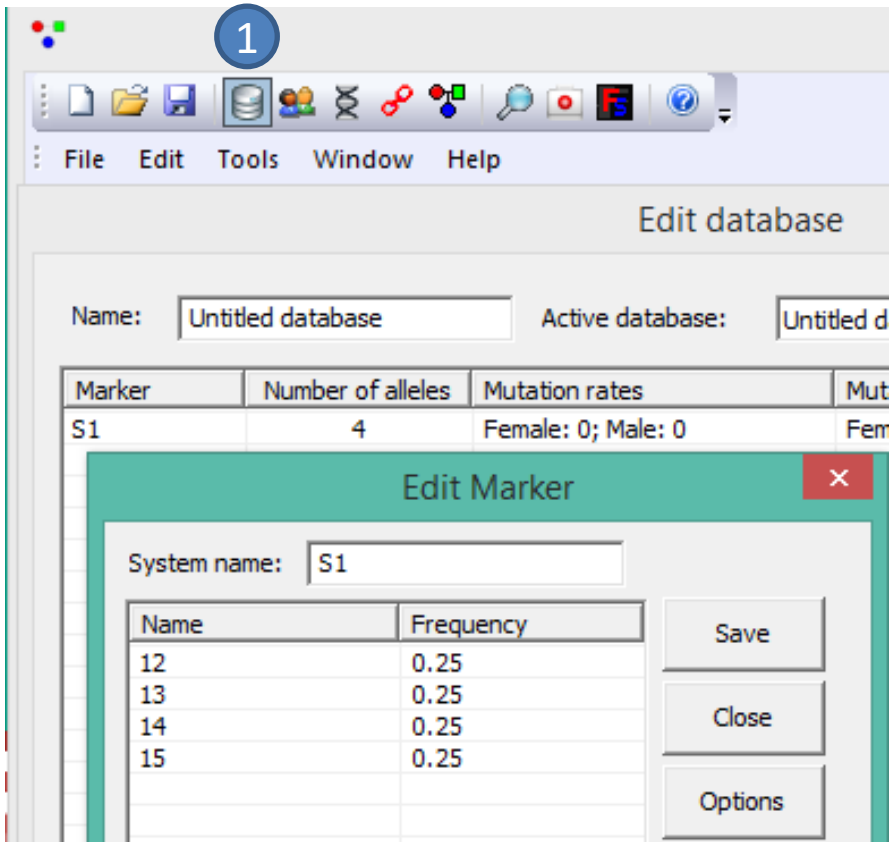
Propósito

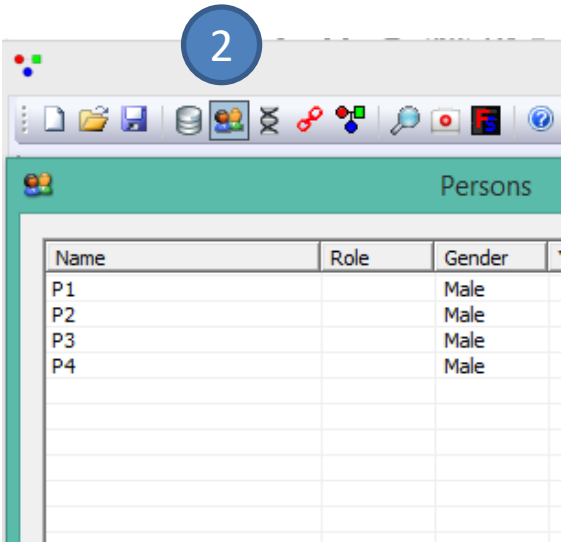
- Encontrar relaciones entre pares de perfiles genéticos.
  - Caso simple:

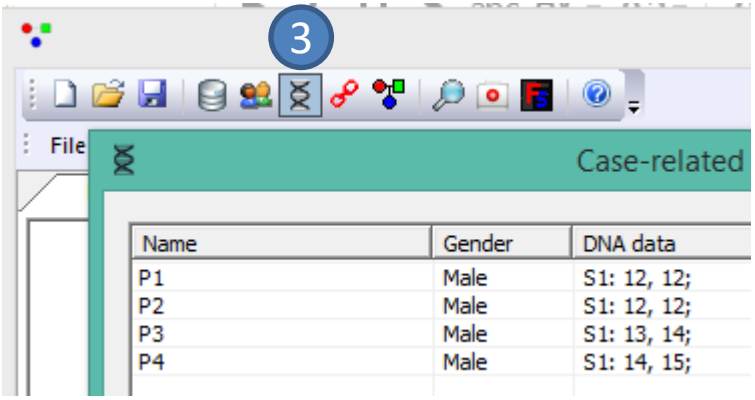
Name	Gender	DNA data
P1	Male	S1: 12, 12;
P2	Male	S1: 12, 12;
P3	Male	S1: 13, 14;
P4	Male	S1: 14, 15;

- Podemos hacer las siguientes 6 comparaciones con los perfiles de la figura:
  - P1-P2, P1-P3, P1-P4, P2-P3, P2-P4, P3-P4
- 'Blind' implica que se van a buscar relaciones fijas (no pedigríes explícitos) que incluyen:
  - Progenitor-hijo (Parent-Child), Hermanos (Siblings), Medio hermanos (Half-siblings), Coincidencia directa (idéntico perfil, Direct-match), Primos (Cousins), primos segundos (2nd cousins)
- Por ejemplo, podríamos suponer que P1-P2 es muy probablemente un match directo (son la misma persona). De todas formas, también son posibles otras relaciones, aunque menos probables.

Cómo hacer Blind Search - Preliminares

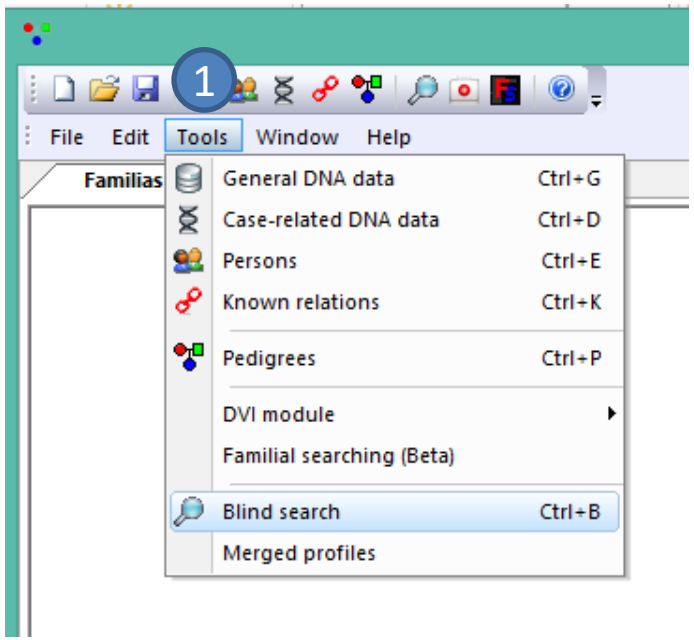


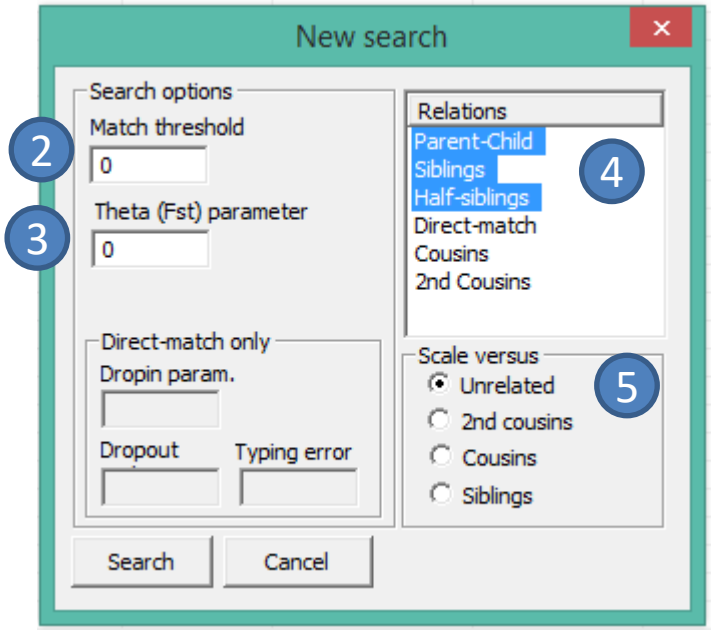




- Definir General DNA data (marcadores y frecuencias) 1, Personas 2 y Case DNA data (genotipos del caso) 3

La búsqueda





- Se inicia la búsqueda como se muestra en 1 (también se puede hacer Blind search en los módulos DVI y Familial Searching; en estos casos, sólo hay que hacer el paso 1 de la página anterior antes de entrar en estos módulos)

- Configuración de los parámetros:
  - ✓ Se mostrarán los LRs mayores que 0 (el valor por defecto es 1) 2
  - ✓ Theta, por defecto se usa 0 3
  - ✓ Comparaciones que se desean realizar 4. En la figura: Parent-Child, Siblings, Half-siblings
  - ✓ Comparar con “no relacionado” (denominador del LR) 5

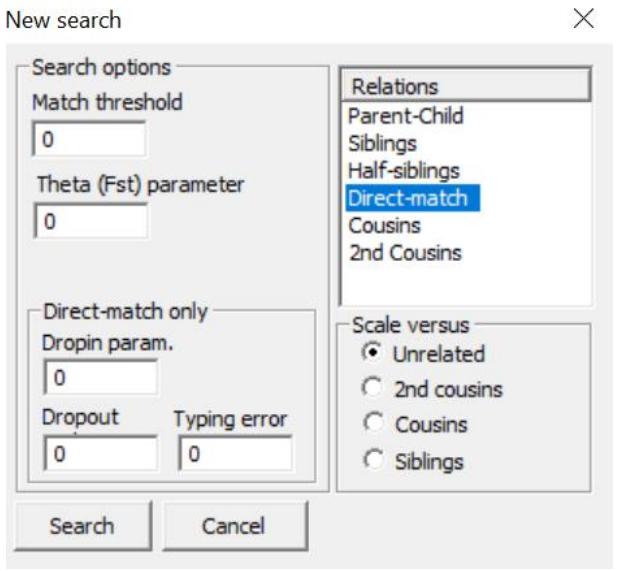
Resultado

Person 1	Perso...	Relationship	LR	Inconsistencies
P1	P2	Siblings	6.25	NA
P1	P2	Parent-Child	4	0
P1	P2	Half-siblings	2.5	NA
P3	P4	Half-siblings	1	NA
P3	P4	Parent-Child	1	0
P3	P4	Siblings	0.75	NA
P2	P4	Half-siblings	0.5	NA
P2	P3	Half-siblings	0.5	NA
P1	P4	Half-siblings	0.5	NA
P1	P3	Half-siblings	0.5	NA
P2	P4	Siblings	0.25	NA
P2	P3	Siblings	0.25	NA
P1	P4	Siblings	0.25	NA
P1	P3	Siblings	0.25	NA
P2	P4	Parent-Child	0	1
P2	P3	Parent-Child	0	1
P1	P4	Parent-Child	0	1
P1	P3	Parent-Child	0	1

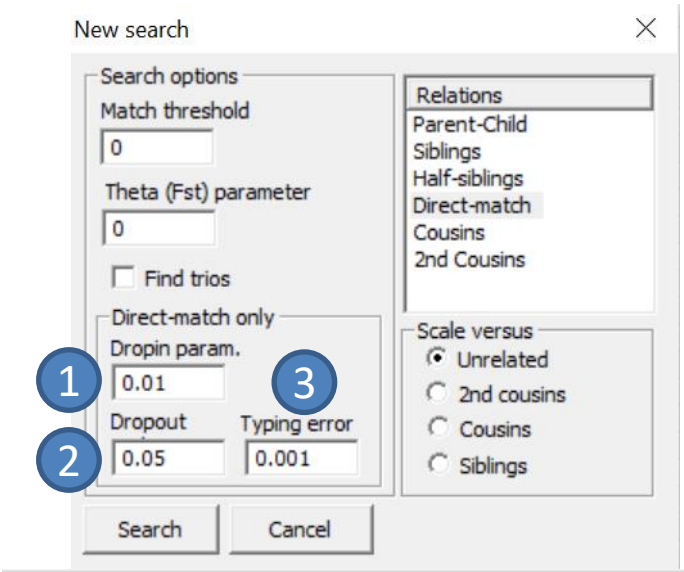
- Se han estudiado 3 tipos de relaciones, y hay 6 comparaciones de parejas de perfiles (ver página 42), por tanto, la lista contiene 3*6=18 comparaciones.
- La lista se ordena de acuerdo al valor del LR.
- Sólo es posible establecer inconsistencias (marcadores que excluyen) en el caso de 'Parent-Child' y se éstas se indican en la última columna;
- En el Blind Search del módulo DVI, se pueden tener en cuenta mutaciones sólo en la relación de parentesco padre-hijo.

Direct match

- En el caso de match directo, se pueden configurar más parámetros
  - Dropin param. 1, Dropout 2 y Typing error 3
  - Ahora, el LR para P1-P2 LR se ha reducido a 9.6, mientras que otras comparaciones ahora dan LR>0



Person 1	Perso...	Relationship	LR	Inc...
P1	P2	Direct-match	16	0
P1	P3	Direct-match	0	1
P1	P4	Direct-match	0	1
P2	P3	Direct-match	0	1
P2	P4	Direct-match	0	1
P3	P4	Direct-match	0	1



Person 1	Perso...	Gender ma...	Relationship	LR	Inconsist...
P1	P2	Yes	Direct-match	16	0
P3	P4	Yes	Direct-match	0	1
P2	P4	Yes	Direct-match	0	1
P2	P3	Yes	Direct-match	0	1
P1	P4	Yes	Direct-match	0	1
P1	P3	Yes	Direct-match	0	1

Opciones avanzadas e información obtenida

- Hay más opciones, la mayoría son intuitivas y no se necesitan para las aplicaciones básicas.
- Se puede obtener además información avanzada sobre alelos IBS (Idénticos por estado) y alelos IBD (Idénticos por descendencia). Esta información no es relevante para aplicaciones estándar.

6. DVI (Identificación de víctimas en desastres)

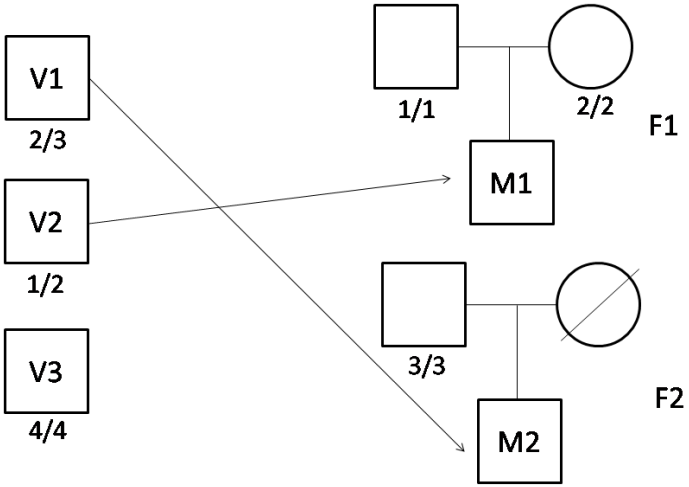
Puede encontrar información sobre características más avanzadas del módulo DVI, incluyendo simulaciones y análisis con varias bases de datos de frecuencias alélicas en un mismo proyecto en este link.

Archivo necesario para la mayoría de los siguientes ejemplos en: Exercise.3.1.fam. La apariencia de algunas de las capturas de pantalla que mostramos en este tutorial pueden diferir ligeramente si utiliza las versiones más recientes de Familias.

Pasos Básicos

1. **General DNA data**. Carga de la base de datos, es decir frecuencias alélicas, etc. Descrito en detalle en la sección 1 de este tutorial.
2. **Unidentified samples (Muestras no identificadas)**. Muestras que se quieren identificar, en un escenario DVI es el conjunto de restos humanos sin identificar mientras que en un escenario de personas desaparecidas pueden ser individuos identificados que se quieren reencontrar con sus familias biológicas.
3. **Reference families (Familias de referencia)**. Datos de referencia que se usarán para identificar y reagrupar familias con personas desaparecidas.
4. **Search (Búsqueda)**. Comparación e interpretación de resultados.

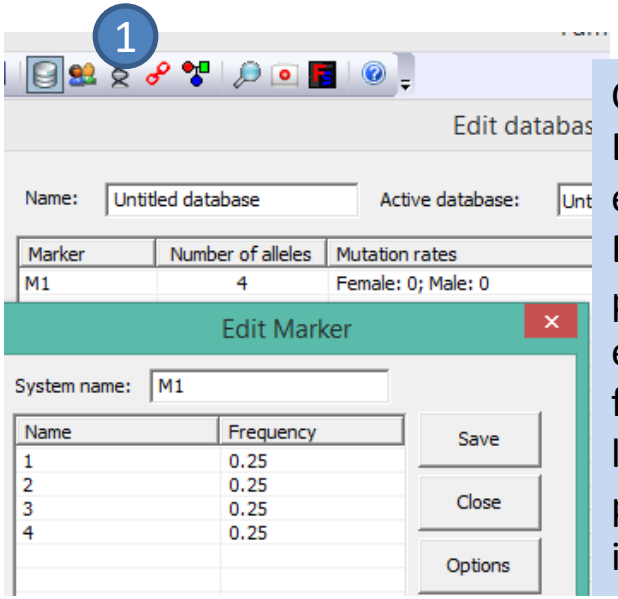
Veremos el módulo DVI con un ejemplo



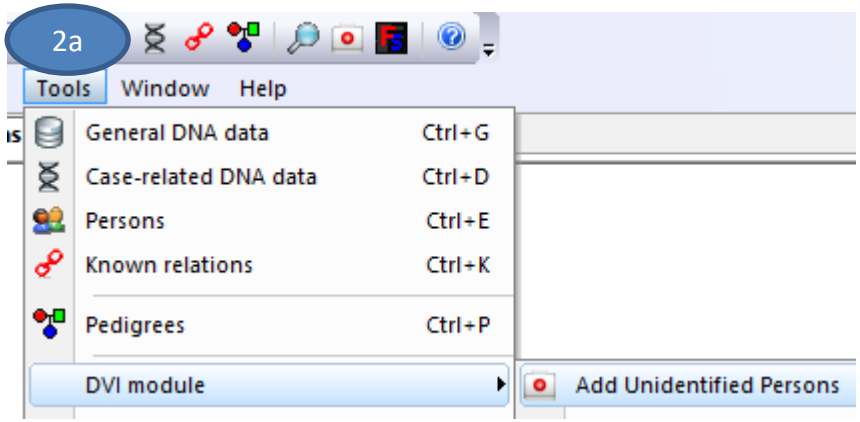
1. **General DNA data:**
   Un marcador con alelos 1 ,2, 3, 4 (todos frec = 0.25)
2. **Unidentified samples:**
   Tres muestras V1, V2 y V3 y sus genotipos.
3. **Reference families:**
   Dos familias de referencia F1, F2 con sus individuos desaparecidos (hijos en este caso) M1 y M2.
   Datos genéticos de los familiares disponibles (madre y padre en F1, sólo el padre en F2.)

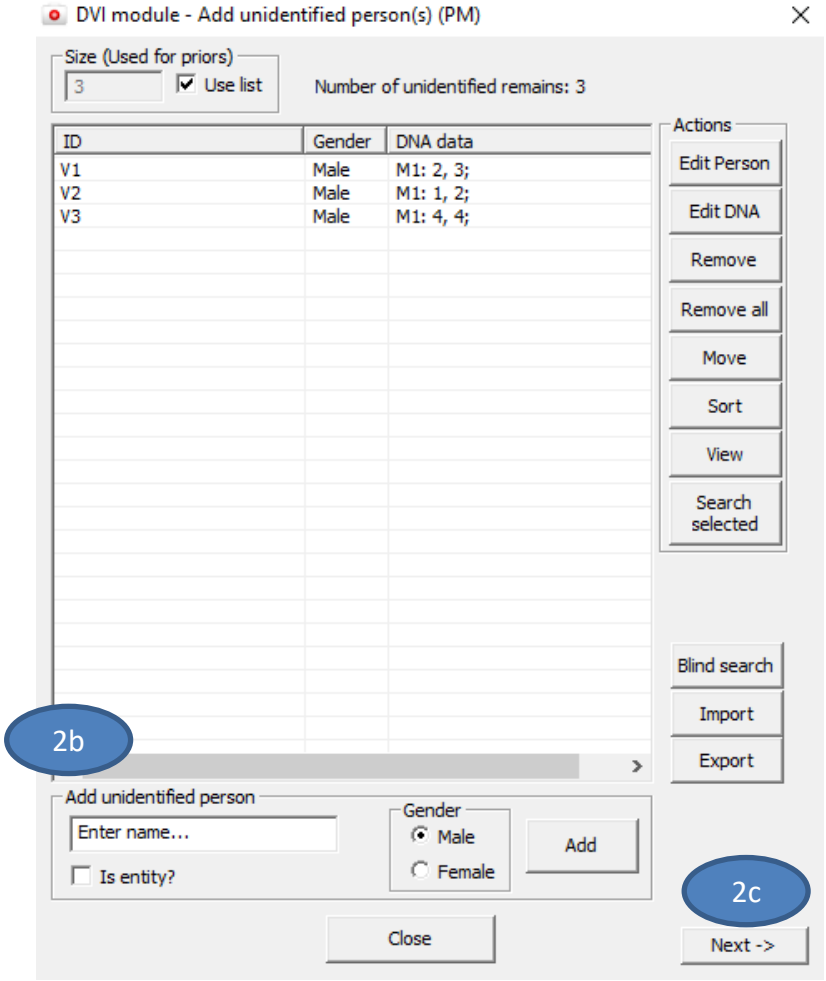
- En este ejemplo “de juguete” ilustramos cómo se realiza la entrada de datos manualmente; es un ejercicio útil.
- Para casos reales, los datos se pueden importar desde archivos. Esto se explica al final de esta sección.

Cómo proceder en DVI



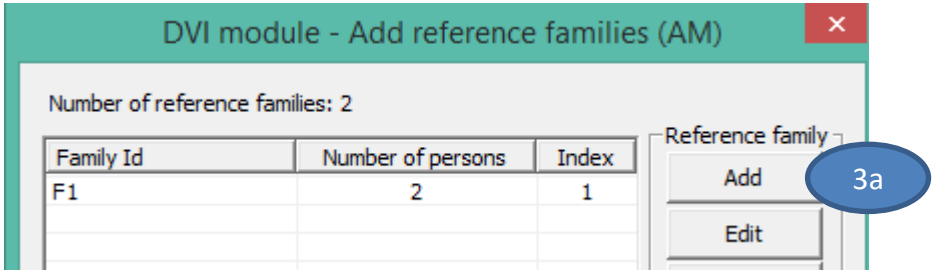
Comentario
La opción "Is entity?" está en Beta testing y permite al usuario especificar una familia/pedigrí en lugar de una persona sin identificar

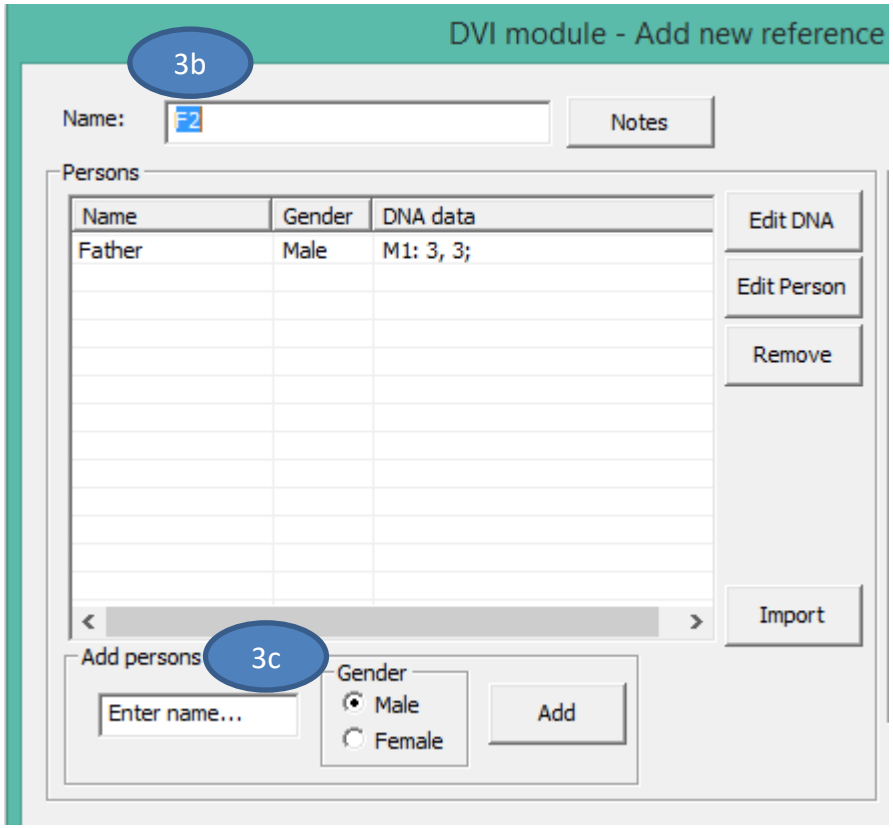




Defina el marcador 1 y entre en el módulo DVI module > Add 2a
Introduzca las muestras no identificadas 2b . Presione Next 2c

Definición de las familias de referencia



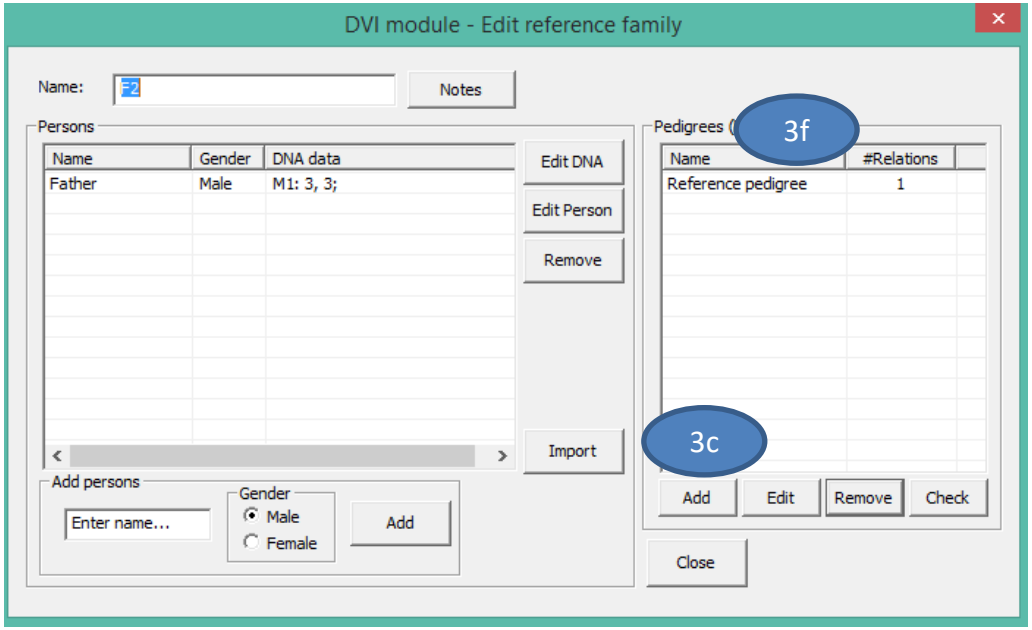


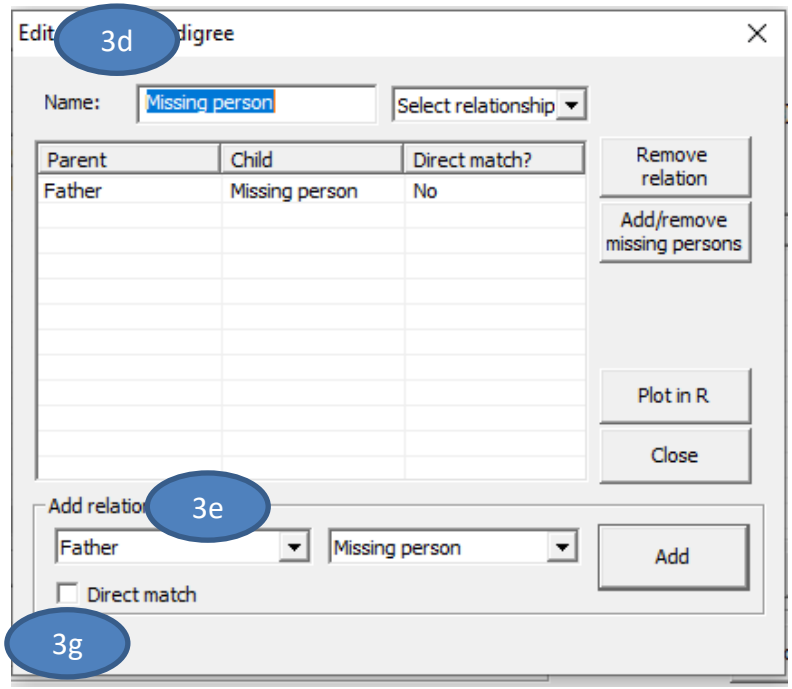
Comentario
En estas ventanas tiene que definir el nombre de la familia, los miembros que la forman y los perfiles genéticos (las relaciones entre estos miembros y con la persona desaparecida no se definen aquí)

- Se ha definido la primera familia, presionamos 'Add' 3a para introducir una nueva familia.

- Nombramos la familia (F2) 3b
- Introducimos las personas necesarias para definir el pedigrí posteriormente e introducimos los genotipos (de las personas disponibles) 3c

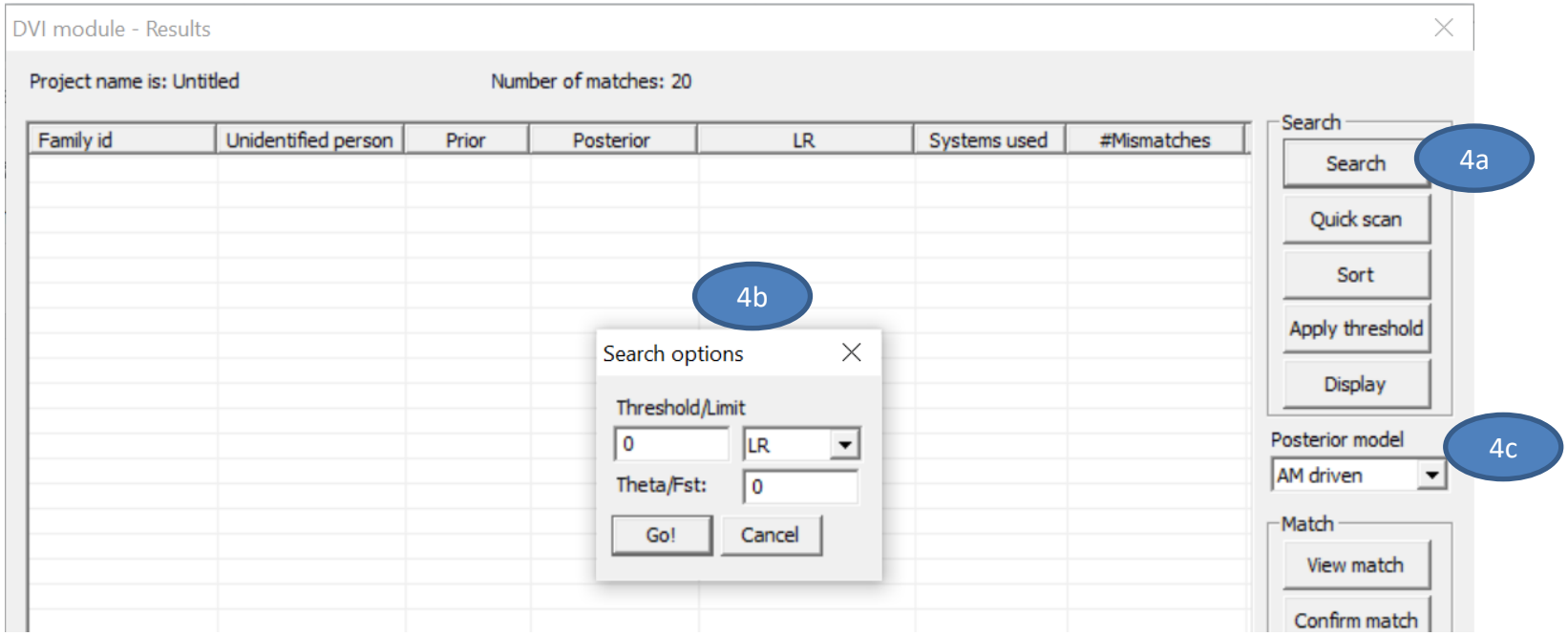
Definición de pedigríes, es decir, qué parentesco se tiene con la persona desaparecida (MP)





- Vamos a Add 3c para definir el pedigrí
- Nombramos el pedigrí 3d
- Definimos cómo está relacionada a MP con esta familia 3e
- NO tocar 'Reference pedigree', 3f ,
  está ahí porque una MP puede no pertenecer a ninguna de las familias de referencia.
- Cuando haya definido todo, presione 'Next' para hacer la búsqueda ….
- Nota: para definir una muestra de referencia directa (ej., objeto personal), presione "Direct match" 3g cuando añada la relación

La búsqueda en DVI



- Presione Search 4a
- Elija las opciones 4b. Aquí, el umbral/límite es 0, así que veremos todos los resultados
- Modelos de Posterior: 4c

  AM driven: se tiene en cuenta a cada familia por separado
  PM driven: se tienen en cuenta a cada víctima por separado
  One-to-one: sólo dos hipótesis

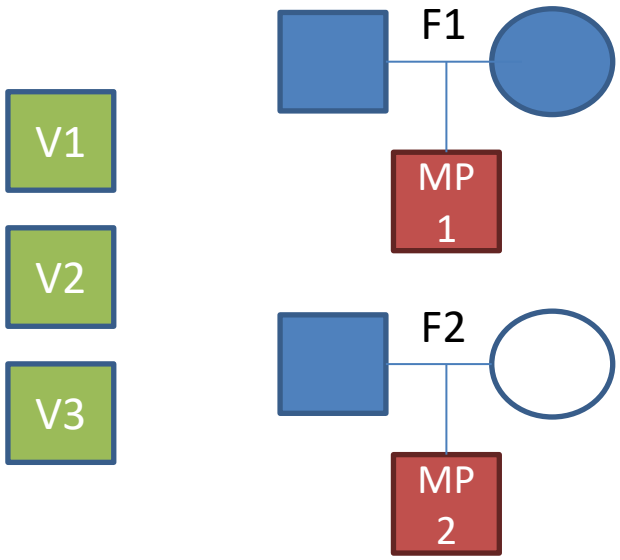
  } Ver diapositiva siguiente

Nota: priors y posteriors no se aplican cuando se usa la herramienta Quick Scan, aunque aún se muestren en la pantalla.

Modelos de Posterior en el ejemplo

AM driven

H_1 = la familia F1 está relacionada con V1
H_2 = la familia F1 está relacionada con V2
H_3 = la familia F1 está relacionada con V3
H_4 = la familia F1 no está relacionada con ninguna V
H_5 = la familia F2 está relacionada con V1
H_6 = la familia F2 está relacionada con V2
H_7 = la familia F2 está relacionada con V3
H_8 = la familia F2 no está relacionada con ninguna V



PM driven

H_1 = V1 está relacionada con la familia F1
H_2 = V1 está relacionada con la familia F2
H_3 = V1 no está relacionada con ninguna familia
H_4 = V2 está relacionada con la familia F1
H_5 = V2 está relacionada con la familia F2
H_6 = V2 is no está relacionada con ninguna familia
H_7 = V3 está relacionada con la familia F1
H_8 = V3 está relacionada con la familia F2
H_9 = V3 no está relacionada con ninguna familia

One to one

H_1 = Vn está relacionada con Fn
H_2 = Vn no está relacionada con Fn
Vn = cualquier víctima
Fn = cualquier familia

Resultado del modelo de posterior AM driven

DVI module - Results

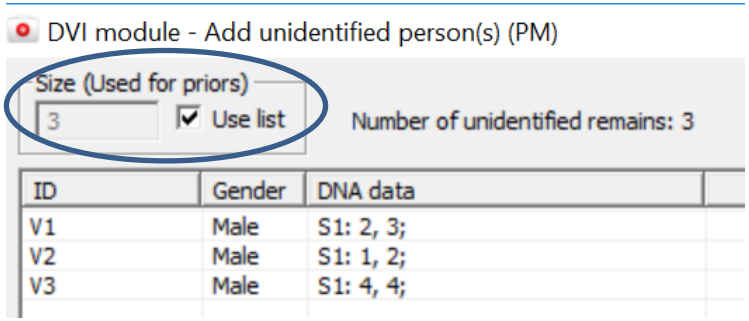
Project name is: Untitled

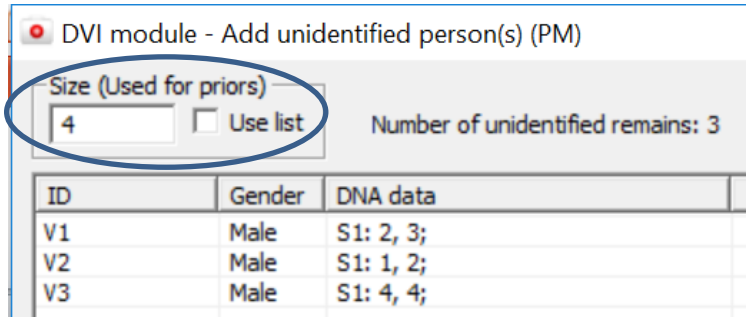
Family id	Unidentified person	Prior	Posterior	LR	Systems used	#Mismatches	Pedigree
F1	V1	0.25	0	0	1	1	Missing son1
F1	V2	0.25	0.888889	8	1	0	Missing son1
F1	V3	0.25	0	0	1	1	Missing son1
F2	V1	0.25	0.666667	2	1	0	Missing son2
F2	V2	0.25	0	0	1	1	Missing son2
F2	V3	0.25	0	0	1	1	Missing son2

- Se obtienen primero los resultados de la primera familia (F1).
- Se muestra el LR = P(datos ADN| MP pertenece a F1)/ P(datos ADN| MP no relacionada con F1) para todas las muestras no identificadas.
  Para V1 y V3, LR=0, mientras que para V2, LR = 1/(2*0.25*0.25) = 8.
- Para el cálculo del posterior se necesita un prior. En el ejemplo suponemos que a *priori,* V1, V2 y V3 tienen un prior de 0.25 de pertenecer a F1; la persona desaparecida puede ser alguien que no hemos encontrado con un prior de 0.25. La probabilidad *a posteriori*, P(V2 pertenece a F1| datos ADN)=0.89.
- Las familias restantes, aquí sólo F2, se tratan de manera similar.

Cómo se definen priors en Familias

En la ventana “Add unidentified person(s) (PM)”





- Por defecto (use list), Familias tiene en cuenta el número de muestras sin identificar (3 en este caso) y además añade la posibilidad de que aún no hayamos encontrado a la persona desaparecida: 1 / 4 = 0.25 (como se mostró en la página anterior)

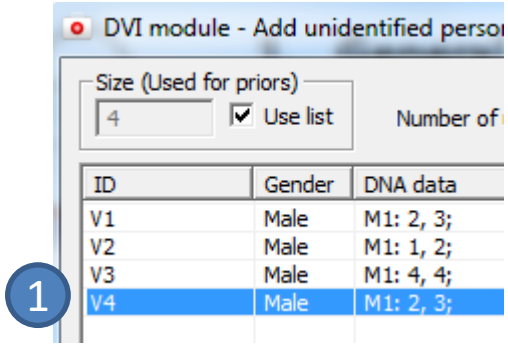
- Se puede no seleccionar “use list” y añadir el tamaño deseado. Por ejemplo, si en este caso supiéramos que hay al menos 4 víctimas (aunque sólo encontráramos 3), podemos añadir “4” en la casilla “size”. Así, el prior será 1/5= 0.20, para cada víctima.

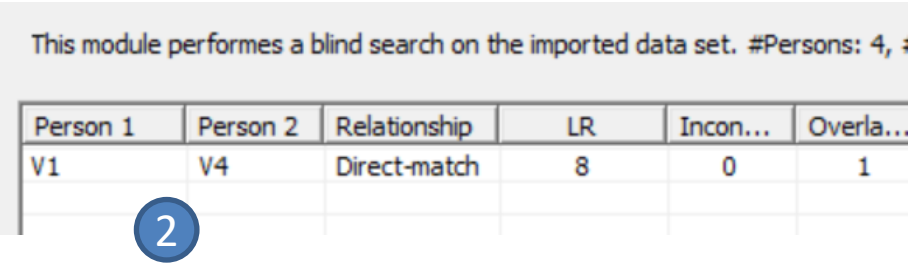
Más: personas sin identificar – comparaciones

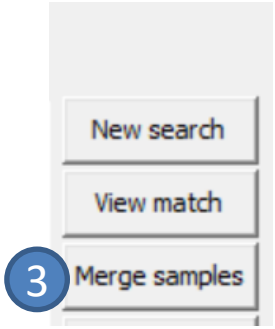
Muy útil para re-asociar restos humanos (del mismo individuo) y/o detectar posibles relaciones entre las víctimas

Ejemplo:

Añadir un nuevo resto humano V4 con el mismo perfil que V1 (M1= 2-3) en la ventana "Add unidentified persons" 1 y click en "Blind search". Presionar "New search" y seleccionar la relación "Direct match" vs. "unrelated" para obtener 2 . Seleccionar las muestras y Click en "Merge samples" 3 (fusionar muestras)



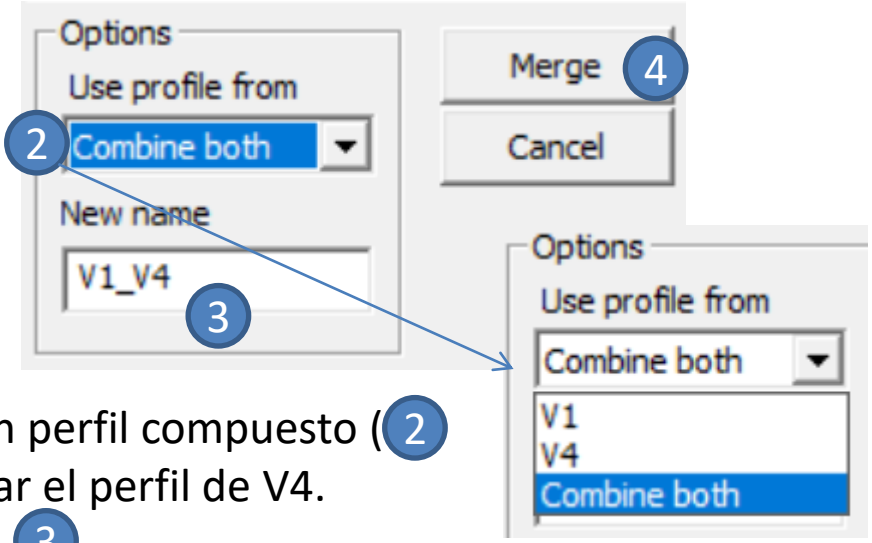




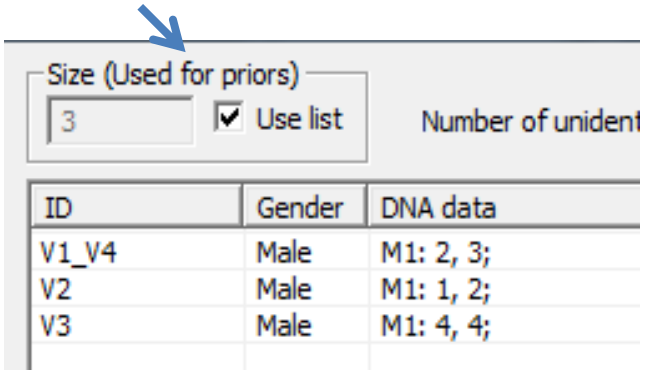
Fusionando muestras

1

Marker	LR	V1	V4	Merged profile
M1	8	2, 3	2, 3	2, 3

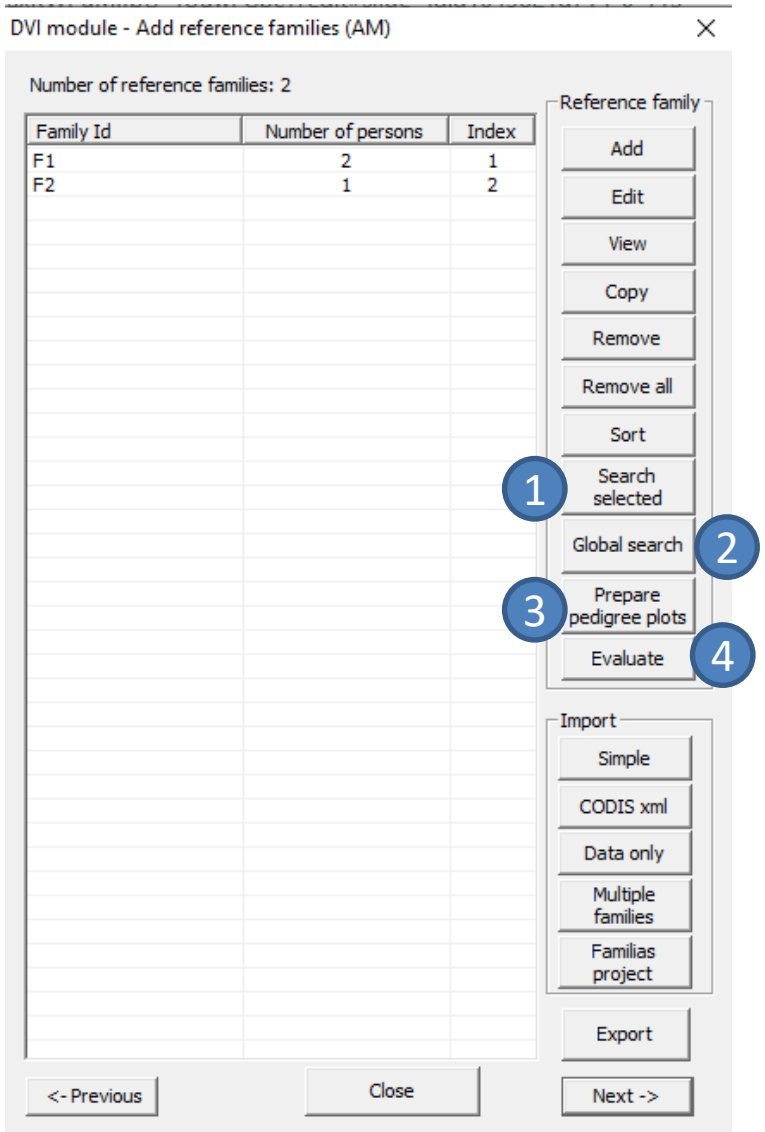


Obtendrá un resumen de las muestras. 1
En el menú “Options” puede seleccionar: crear un perfil compuesto (2 opción “combine both”), usar el perfil de V1 o usar el perfil de V4.
También puede renombrar este perfil combinado 3 .
Click en “Merge” 4 y la muestra fusionada V1_V4 aparecerá ahora en la ventana “Unidentified persons”



Comentario
En este ejemplo V1 y V4 tienen el mismo perfi, pero en los casos reales es posible que una de las muestras no tenga el perfil completo. Este es el motivo por el que la opción “Combine both” está disponible

Más – Familias de referencia



1 Se puede buscar sólo con las familias seleccionadas

Para búsquedas globales (aún no disponible) 2

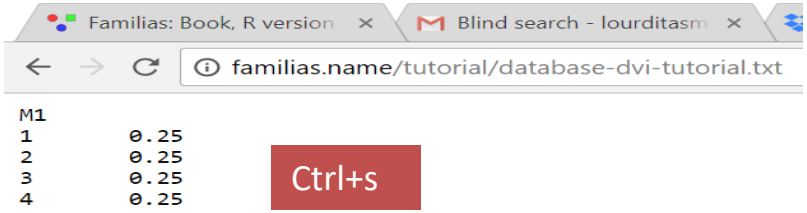
3 Prepara en R los dibujos de lo pedigríes seleccionados

Evalúa las familias de referencia, hace simulaciones y otras estadísticas 4

DVI: importando datos desde archivos

- Abra Familias e importe la base de datos de frecuencias:
  - **Tools > General DNA data** formato de archivo: http://familias.name/tutorial/database-dvi-tutorial.txt
- Para importar las muestras sin identificar:
  - **Tools > DVI module > Add Unide…** formato de archivo:
  http://familias.name/tutorial/unidentified-dvi-tutorial.txt
- Opción 1: para importar personas de referencia con los datos genotípicos y luego introducir los pedigríes manualmente
  - **Tools > DVI module > Add refe … Data only**
    formato de archivo: http://familias.name/tutorial/ReferenceFamilies-dvi-tutorial.txt
  - Definir los pedigríes manualmente
- Opción 2: para importar todos los datos de las familias de referencia (pedigríes incluidos)
  - **Tools > DVI module > Add refe … Multiple families**
    formato de archivo: http://familias.name/tutorial/pedigrees-dvi-tutorial.txt
- La búsqueda puede hacerse después, obteniéndose los resultados anteriores. Se proporcionan más detalles sobre la importación de archivos en la sección 8 de este manual.

Nota: puede copiar y pegar los links en su buscador de internet y después guardar los archivos usando Ctrl+s en su teclado

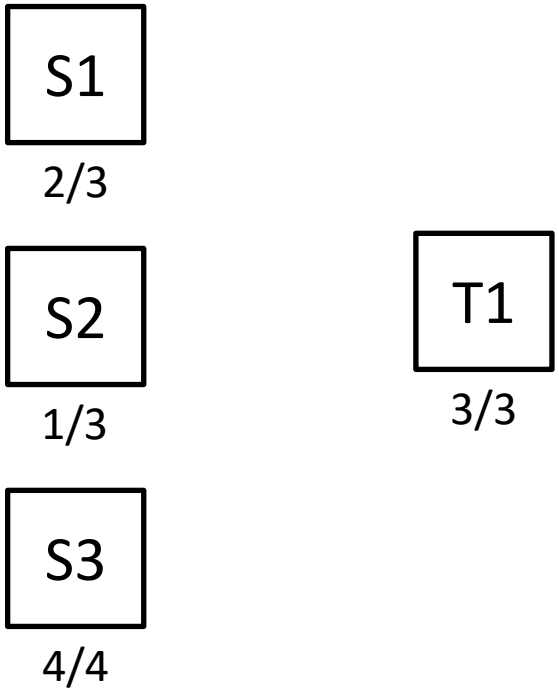


7. Familial searching (búsqueda familiar)

Pasos Básicos

1. **General DNA data**. Carga de la base de datos, es decir, frecuencias alélicas, etc.
2. **Profile database (base de datos de perfiles)**. Individuos/muestras dubitadas en la base de datos de sospechosos.
3. **Traces (muestras dubitadas)**. Muestras sobre las que realizar la búsqueda familiar
4. **Search (búsqueda)**. Realizar la búsqueda e interpretar los resultados

Ejemplo de *Familial Searching*



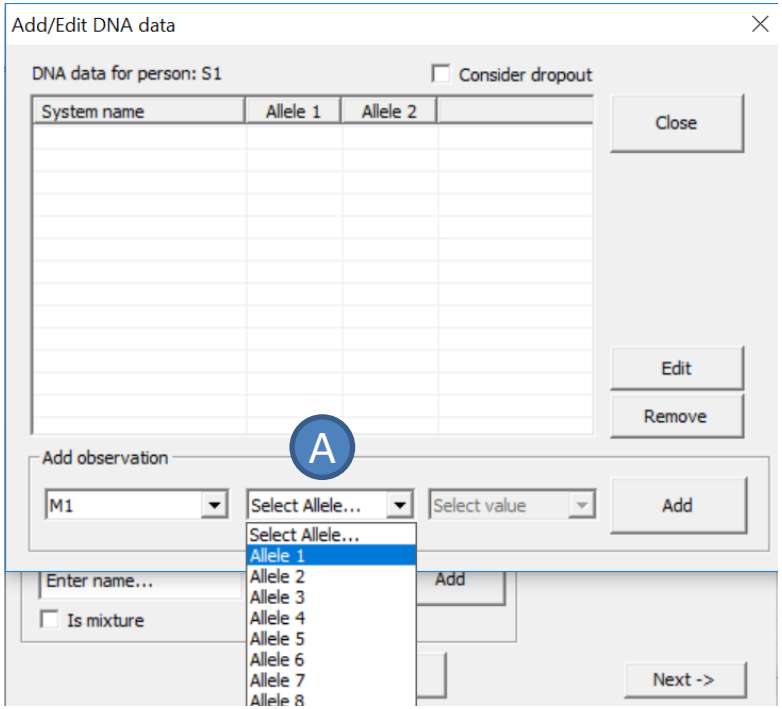
1. ***General DNA data***:
   Un marcador con alelos 1 ,2, 3, 4 (todos con Frec.=0.25).
2. ***Offender database***:
   Tres muestras S1, S2 y S3, y sus genotipos.
3. ***Target profile***:
   Una muestra dubitada T1 con un genotipo que no coincide plenamente con ninguno de los sospechosos.

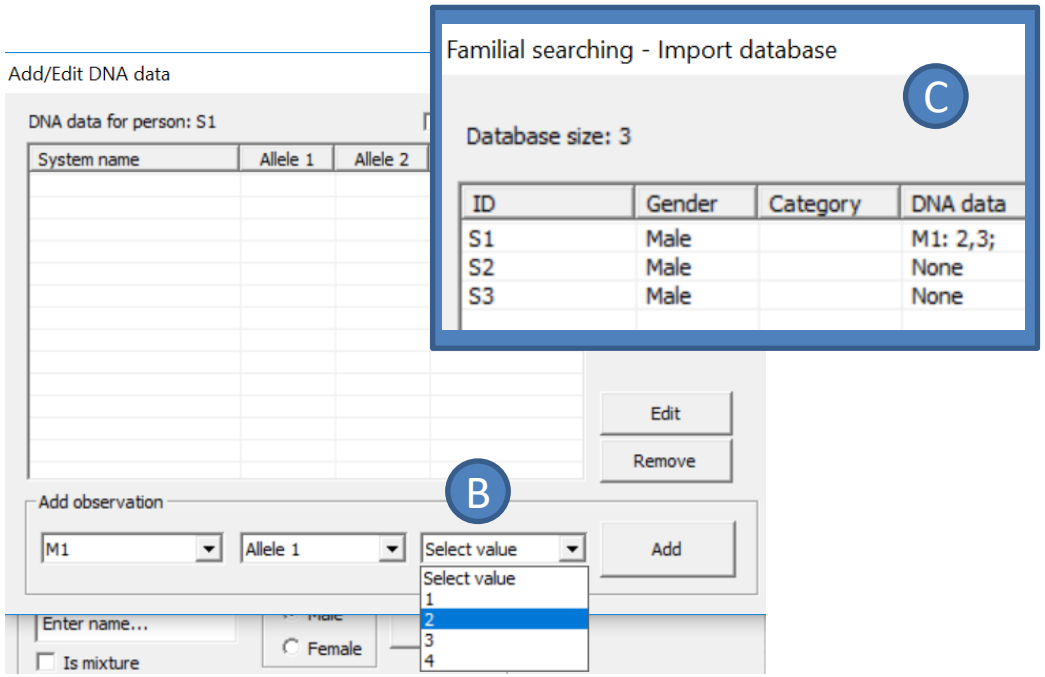
- En este ejemplo “de juguete” ilustramos cómo se realiza la entrada de datos manualmente; es un ejercicio útil.
- Para casos reales, los datos se importan desde archivos, como se explica después. Familias puede manejar bases de datos de >300,000 muestras.

Cómo introducir perfiles en este módulo

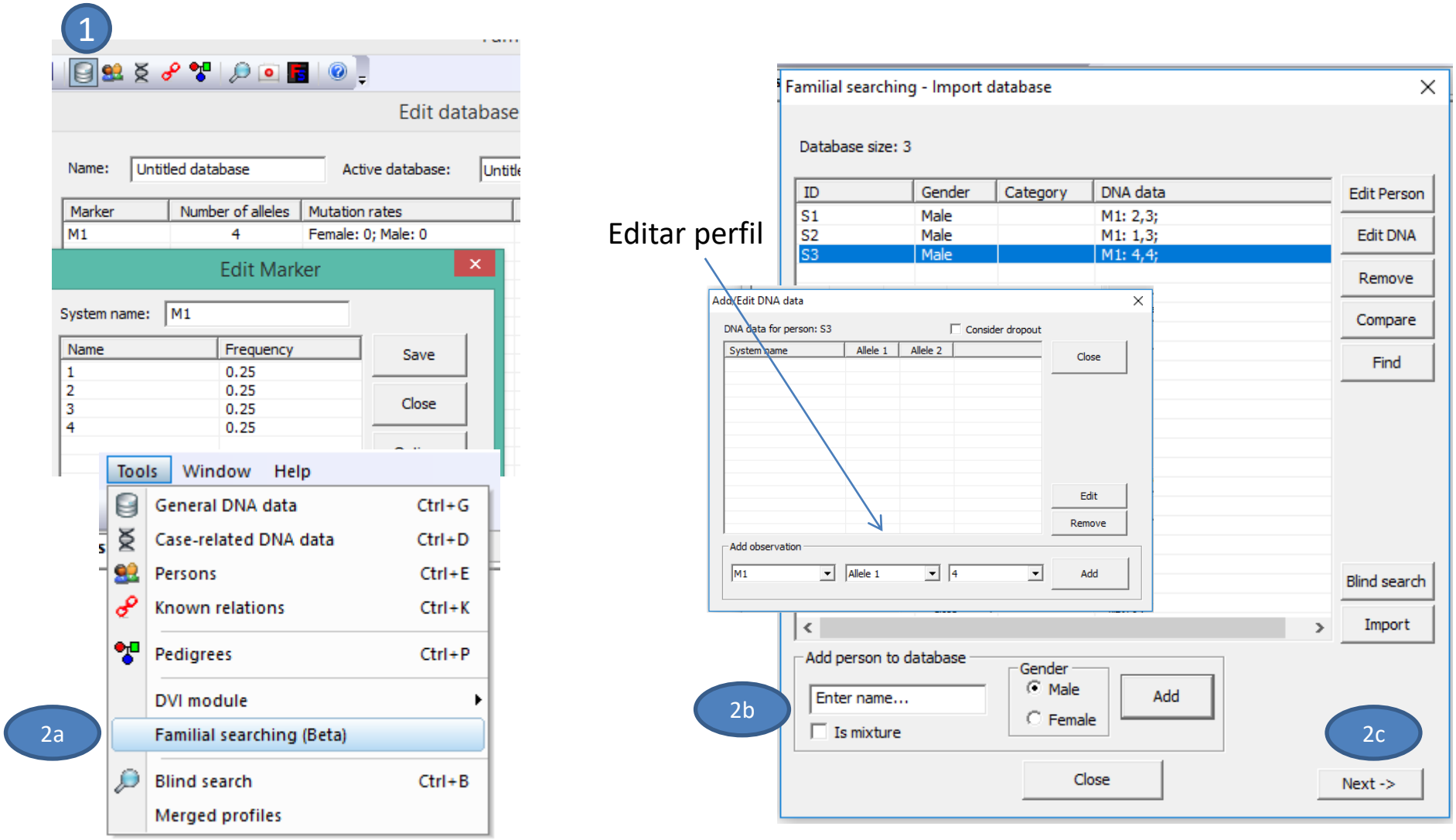
Seleccione el primer sospechoso/traza y clic en «Edit DNA» para introducir el perfil
En el módulo Familial Searching , **los perfiles se introducen de manera diferente** (ya que puede ser necesario añadir más de dos alelos en el caso de mezclas de ADN)
Para introducir S1=2-3, tiene que:

- Seleccionar «Allele 1» en el desplegable A y después seleccionar el valor B de este primer alelo (2, en este ejemplo). **No presione Add aún**
- Select «Allele 2» en el desplegable y después seleccionar el valor de este segundo alelo (3, en este ejemplo). Ahora, presionar Add para obtener C .





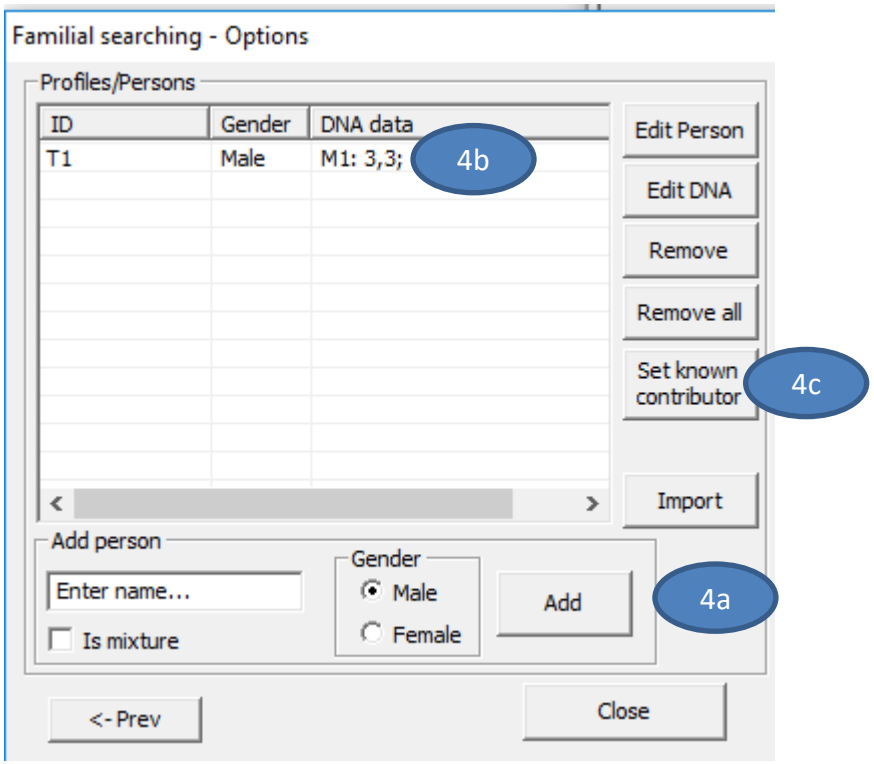
Definiendo la base de datos de sospechosos



Definir el marcador 1, entrar en el módulo Familial searching 2a

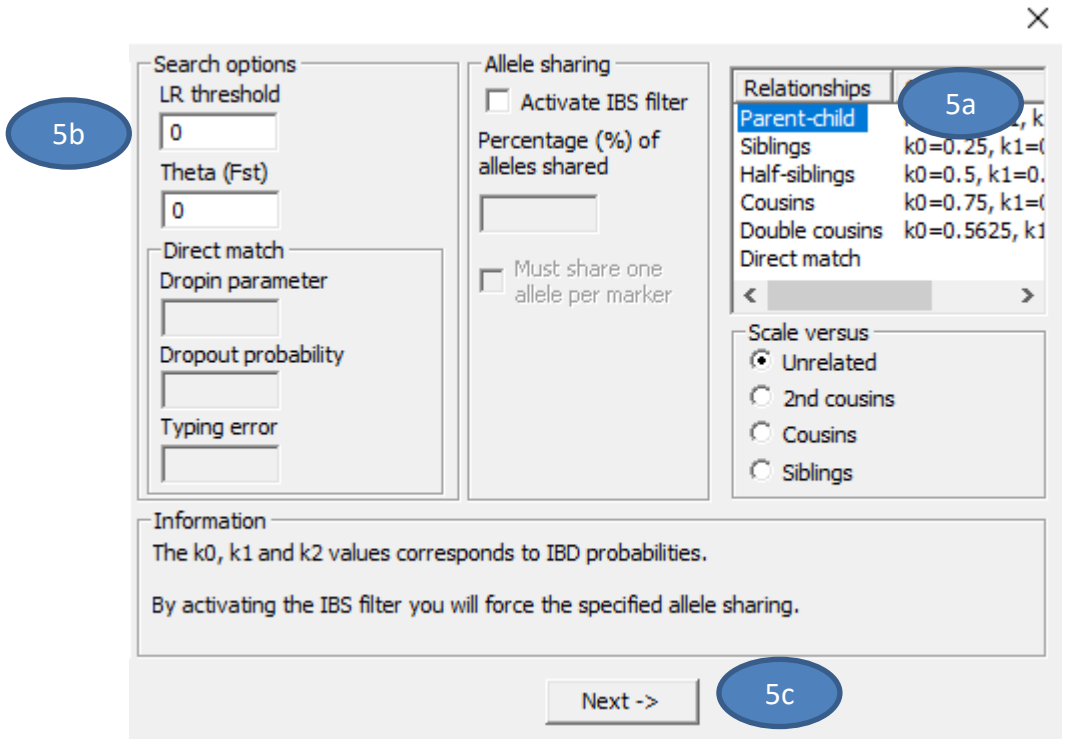
Añadir personas y perfiles genéticos 2b Presionar Next 2c

Definiendo los perfiles a comparar (*target*(*s*))



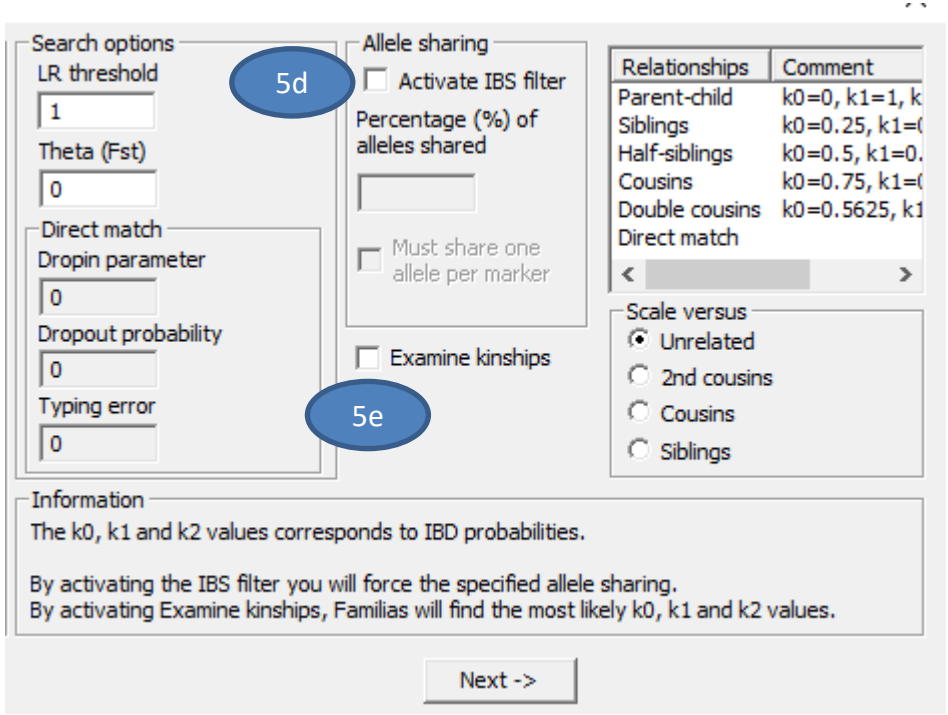
- Defina el target, en este caso una muestra dubitada. 4a
  Presionar 'Add'
- Edit los datos de ADN como se indica 4b
- Podemos fijar contribuyentes conocidos (ej. Víctimas en mezclas) 4c

Definiendo las opciones de búsqueda



- Especifique Parent-child 5a como la relación que queremos incluir en la búsqueda.
- Fije el umbral de LR a 0 (cero) 5b y deje el resto de parámetros con sus valores por defecto.
- Presione 'Next' 5c

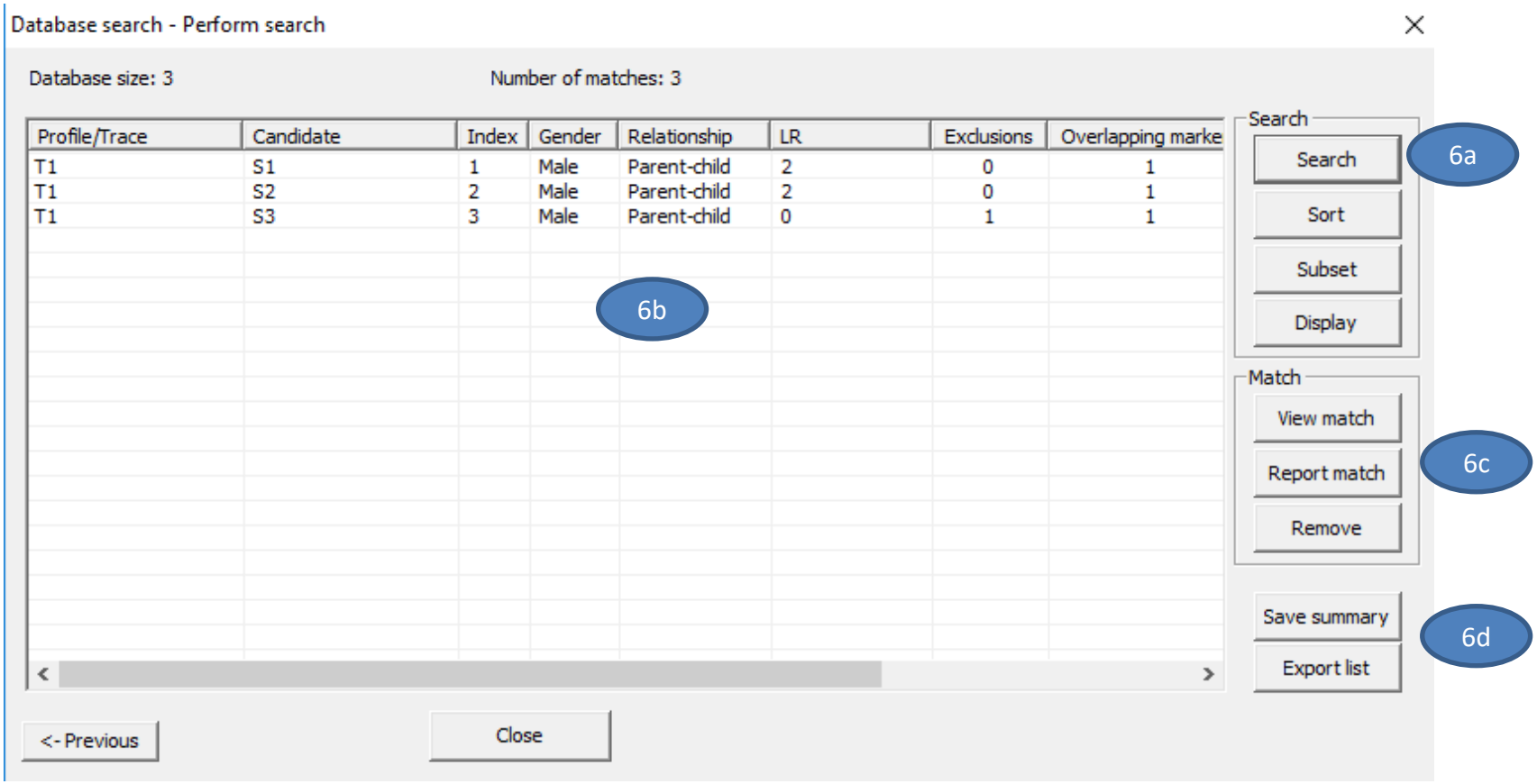
Opciones avanzadas



Se puede usar un filtro IBS 5d para eliminar *matches* con bajo núm. de alelos IBS compartidos

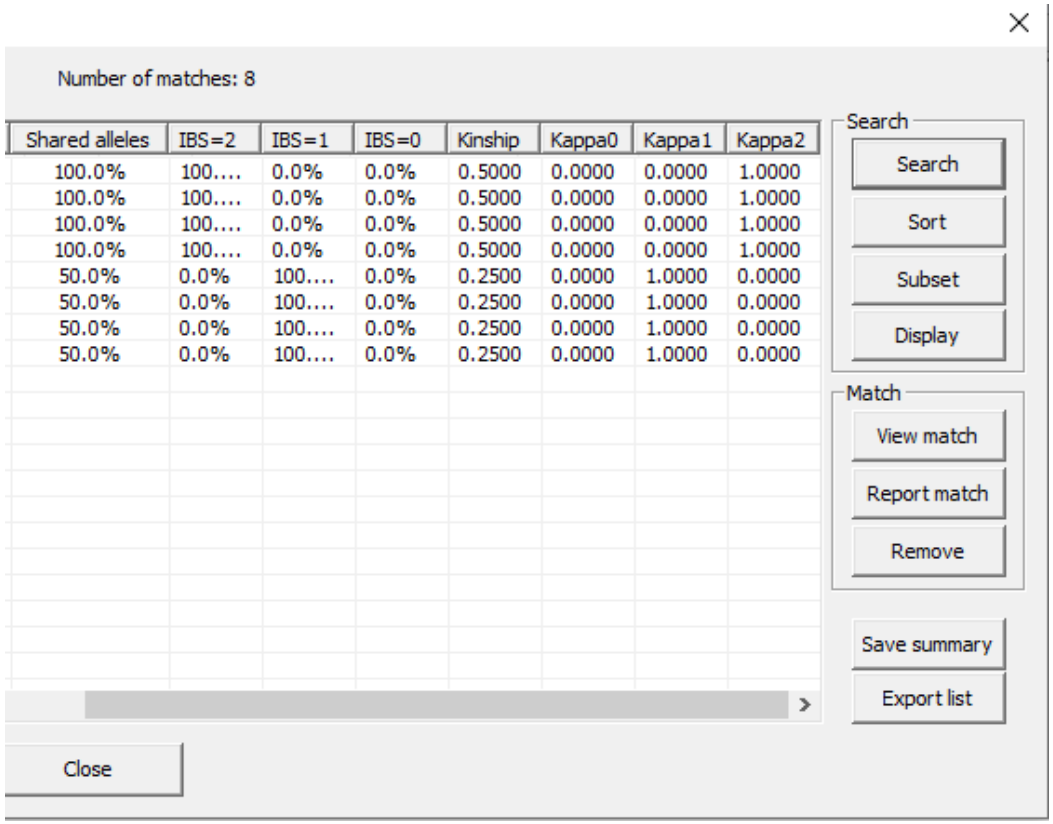
Se puede seleccionar «Examine kinships» 5e para que Familias realice una búsqueda informada entre los valores más probables de k0, k1 y k2. Esta búsqueda es computacionalmente intensiva y sólo se realiza para las relaciones que excedan el umbral de LR.

Búsqueda familiar – Resultados básicos



- Presione 'Search' 6a
- Los resultados se muestran como *matches* entre un perfil/traza (target) y un candidato (persona de la base de datos). Los resultados se ordenan de acuerdo al valor del LR. Además se muestran otros datos 6b
- Puede investigar los resultados con más detalle 6c y puede crear informes. 6d

Búsqueda familiar – Resultados avanzados



Shared alleles	IBS=2	IBS=1	IBS=0	Kinship	Kappa0	Kappa1	Kappa2
100.0%	100....	0.0%	0.0%	0.5000	0.0000	0.0000	1.0000
100.0%	100....	0.0%	0.0%	0.5000	0.0000	0.0000	1.0000
100.0%	100....	0.0%	0.0%	0.5000	0.0000	0.0000	1.0000
100.0%	100....	0.0%	0.0%	0.5000	0.0000	0.0000	1.0000
50.0%	0.0%	100....	0.0%	0.2500	0.0000	1.0000	0.0000
50.0%	0.0%	100....	0.0%	0.2500	0.0000	1.0000	0.0000
50.0%	0.0%	100....	0.0%	0.2500	0.0000	1.0000	0.0000
50.0%	0.0%	100....	0.0%	0.2500	0.0000	1.0000	0.0000

- Se pueden ver otros estadísticos de la búsqueda desplazándose hacia la derecha.
- Alelos compartidos (total), así como la fracción de marcadores con 0, 1 o 2 alelos IBS
- Si «Examine kinships» se ha activado en las opciones de búsqueda, se realiza una búsqueda dirigida que encuentra los valores IBD más probables así como el coeficiente de parentesco.

Familial Search importando desde archivos

- Abra Familias e importe la base de datos de frecuencias:
  - **Tools > General DNA data** formado del archivo: http://familias.name/tutorial/database-dvi-tutorial.txt
- Importe la base de datos de sospechosos
  - **Tools > Familial searching > Import** formato del archivo: http://familias.name/tutorial/offenders-fs-tutorial.txt
- Importe el perfil target
  - **Tools > Familial searching > Next > Import**
    formato del archivo: http://familias.name/tutorial/target-fs-tutorial.txt
- La búsqueda ya puede hacerse, obteniéndose los resultados anteriores. Se proporcionan más detalles sobre la importación de archivos en la sección 8 de este manual. Familias puede manejar archivos .xml de CODIS.

Recuerde: puede copiar y pegar los links en su buscador de internet y después guardar los archivos usando Ctrl+s en su teclado

8. Archivos para importar (formatos)

8.1 Archivo con la base de datos de frecuencias

Estándar (datos separados por tabulaciones)

Todas las versiones de Familias reconocen en formato estándar. Este formato se describe en la figura y se guarda como archivo de texto.

TPOX	
8	0.2
9	0.3
9.3	0.5
D12S391	
20	0.1
21	0.3
22	0.4
23	0.2

Con formato de Matriz

El formato matriz es el que comunmente se encuentra en las publicaciones de frecuencias alélicas. Es muy importante el identificador de alelo o alelos, arriba a la izquierda.

Alleles	TPOX	D12S391
8	0.2	
9	0.3	
9.3	0.5	
20		0.1
21		0.3
22		0.4
23		0.2

8.2 Archivo con los datos del caso (*Case data input, samples*)

Datos separados por tabulador 1

Este tipo de archivo se puede crear fácilmente en Excel (o herramienta similar) y se debe guardar como archivo de texto separados por tabulaciones. Abajo puede ver un ejemplo del formato.

Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
Person 1	8	9.3	20	21
Person 2	9.3	9.3	22	23

Datos separados por tabulador 2

Familias también permite archivos de texto separado por tabuladores pero con una sola columna por marcador.
Puede ver el formato a continuación.

Sample	TPOX	D12S391
Person 1	8,9.3	20,21
Person 2	9.3,9.3	22,23

GeneMapper

Familias puede importar datos de perfiles desde GeneMapper (o datos con formato similar al de GeneMapper). El formato exacto se describe en la figura.

Sample	Marker name	Allele 1	Allele 2
Person 1	TPOX	8	9.3
Person 1	D12S391	20	21
Person 2	TPOX	9.3	
Person 2	D12S391	22	23

Formato XML

Familias reconoce el formato xml de CODIS. A la derecha se muestra un ejemplo de este formato, en el cual puede ver las etiquetas (tags) necesarias. No son necesarias otras etiquetas (normalmente añadidas por CODIS).

Puede encontrar un ejemplo en http://familias.name/Files/example.xml

```xml
<?xml version="1.0" encoding="utf-8"?>
<CODISImportFile xmlns="urn:CODISImportFile-schema">
  <SPECIMEN SOURCEID="N/A">
    <SPECIMENID>Person 1</SPECIMENID>
    <SPECIMENCATEGORY>Father</SPECIMENCATEGORY>
    <LOCUS>
      <LOCUSNAME>TPOX</LOCUSNAME>
      <ALLELE>
        <ALLELEVALUE>8</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>9.3</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS>
      <LOCUSNAME>D12S391</LOCUSNAME>
      <ALLELE>
        <ALLELEVALUE>20</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>21</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS>
  </SPECIMEN>
</CODISImportFile>
```

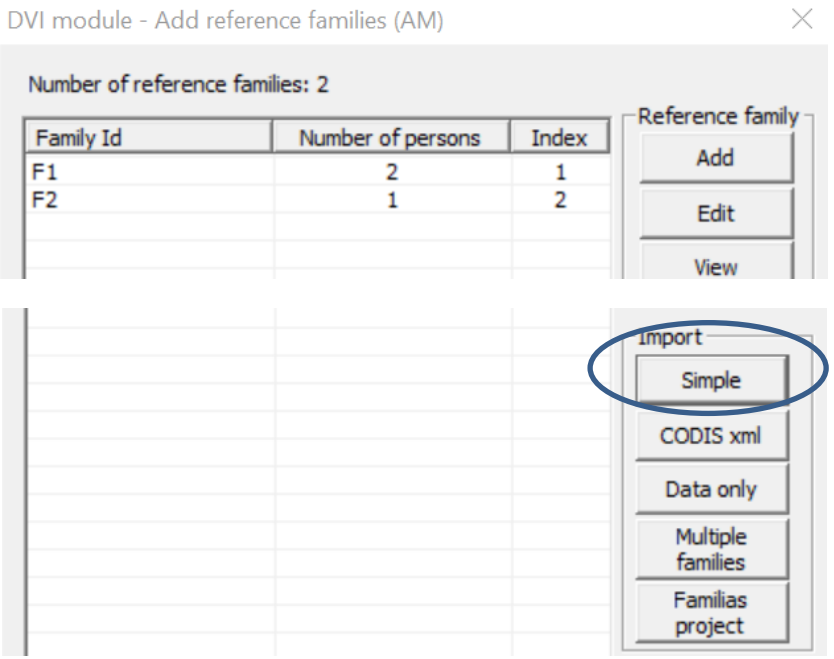
8.3 Datos de las Familias de referencia (DVI)

Relaciones familiares

Familias reconoce algunas relaciones de parentesco estándar que se pueden definir en los archivos que se van a importar. Abajo puede ver el listado exhaustivo. Por favor, tras la importación, compruebe todos los datos para asegurarse de que Familias los ha interpretado correctamente. Puede ver un ejemplo en la página 83.

[Father], [Mother], [Parent], [Son], [Daughter], [Child], [Sister], [Brother], [Sibling], [Half Sister], [Half Brother], [Half Sibling], [Direct], [Identity], [Paternal Half Sister], [Maternal Half Sister], [Maternal Half Brother], [Paternal Half Brother], [Grandmother], [Grandfather], [Paternal Grandmother], [Maternal Grandmother], [Paternal Grandfather], [Maternal Grandfather], [Grandson], [Granddaughter], [Grandchild], [Uncle], [Aunt], [Niece], [Nephew], [Wife], [Husband], [Spouse]

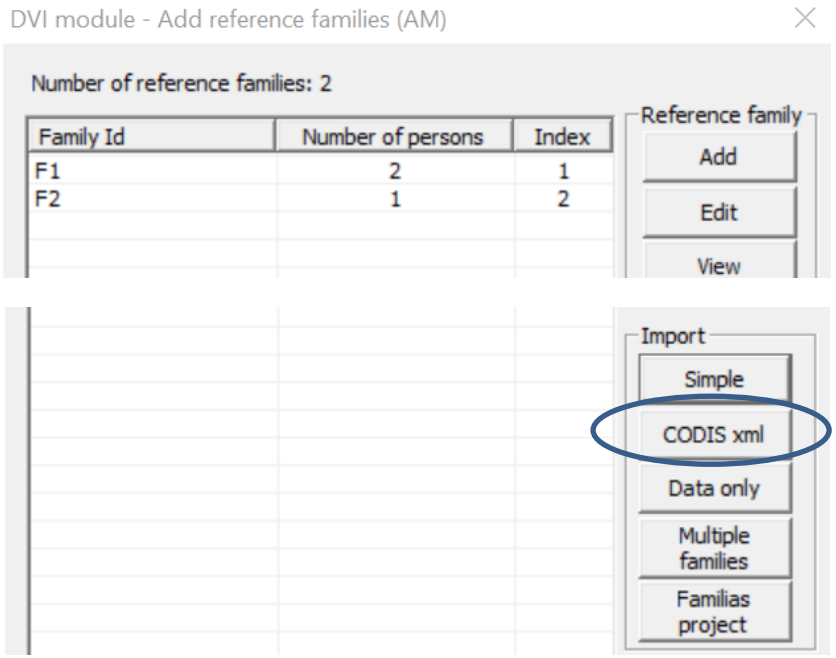
Importación Simple



Similar a los archivos separados por tabulación descritos anteriormente para la entrada estándar de datos de ADN. Se pueden incluir varios individuos en el mismo archivo. Nota, no se puede incluir información sobre la familia, cada individuo se coloca en una familia. Se puede incluir una línea especificando la relación de parentesco con la persona desaparecida, como puede verse abajo.

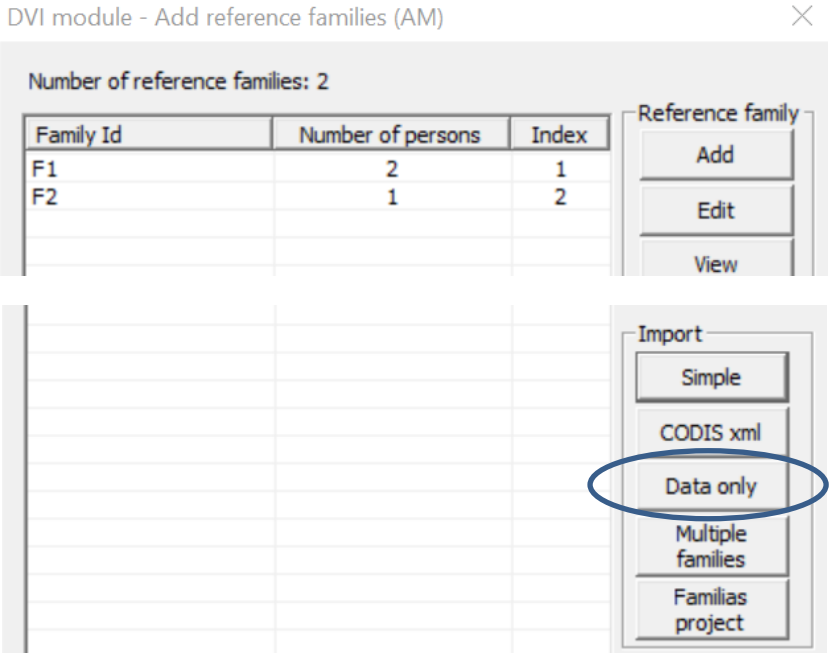
Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
[Brother]				
Person 1	8	9.3	20	21

CODIS xml



El formato se ha descrito anteriormente para la entrada estándar de datos de ADN. Esta versión reconocerá las categorías de CODIS y creará las relaciones familiares basándose en la información de CODIS.

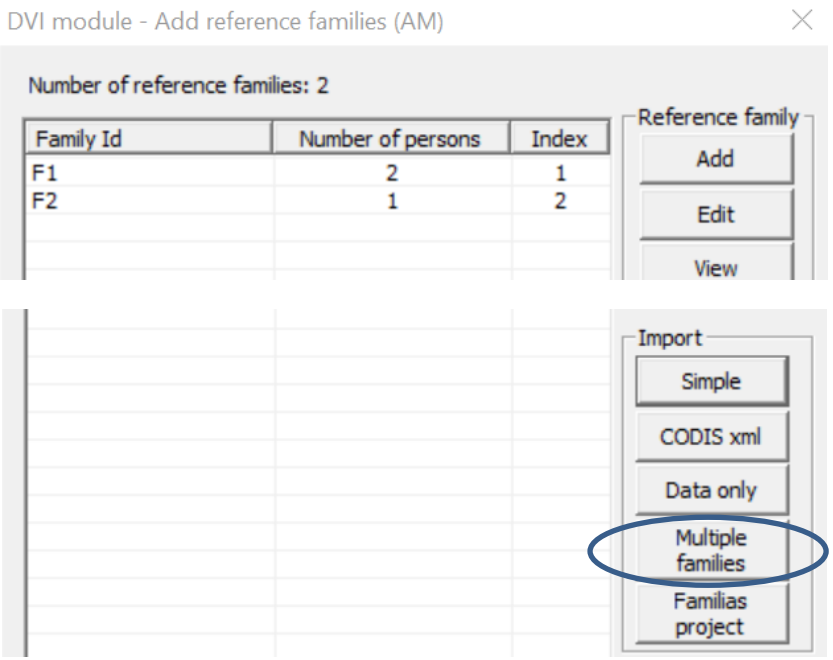
Sólo Datos



Este formato se parece al formato Simple, pero añade una columna para designar el nombre de la familia, ver abajo.

Family	Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
Family 1	Person 1	8	9.3	20	21

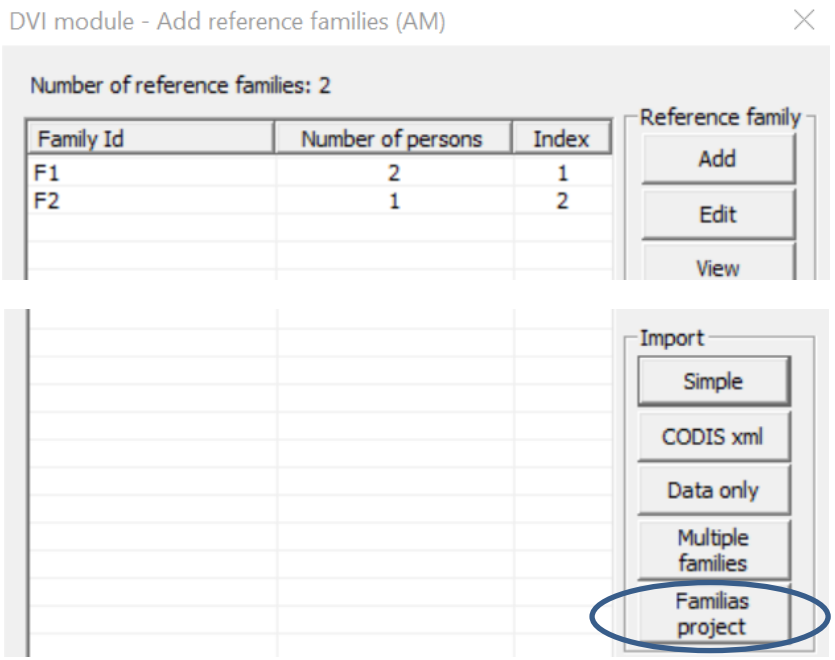
Varias familias



Este formato es una extensión del formato «sólo datos», pero incluye una columna adicional para especificar el parentesco con la persona desaparecida dentro de cada familia, ver figura abajo.

Family	Relationship	Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
Family 1	[Brother]	Person 1	8	9.3	20	21

Proyectos de Familias



Esta opción le permite importar un proyecto de Familias estándar con los datos de ADN y los pedigríes. Familias creará automáticamente las familias (una por cada archivo de Familias). La persona desaparecida debe denominarse MISSING PERSON.

9. Archivos exportados e informes

9.1 Informes generados por Familias

Formato rtf

Es el formato estándar de Familias, es decir, el formato en el que Familias genera los informes. Se puede generar para que incluya diferente tipo de información: (i) simple, que contiene sólo información sobre los LRs y los pedigríes y (ii) extenso, que contiene toda la información necesaria para reproducir los resultados.

Con la extensión rtf, lo que se genera realmente es un archivo .txt y por lo tanto no se trata de un archivo con el texto formateado.

Formato csv

Este formato genera un archivo separado por comas (csv). Se muestra un ejemplo en la figura.

```
Output generated by Familias, version 3.2.2
Ped 1 vs Ped 2;
Marker;LR
SE33;8.22655
Total LR;8.226551889
Database;Norwegian
```

Formato tab

Este formato genera un archivo separado por tabulaciones. Se muestra un ejemplo en la figura.

```
Output generated by Familias, version 3.2.2
Marker  LR
SE33    8.22655
Total LR    8.226551889
Database    Norwegian
```

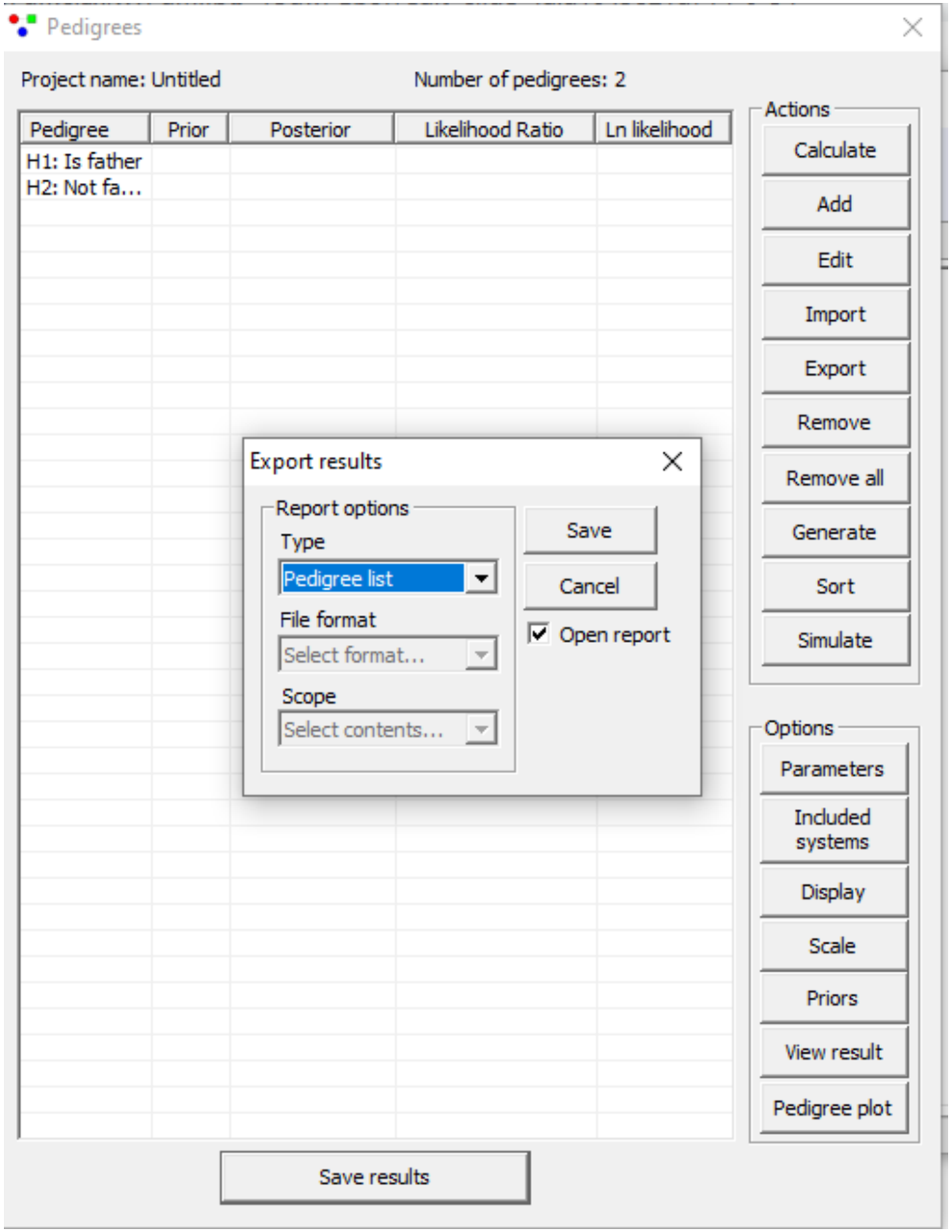
Formato xml

Este formato genera un archivo xml de contenido integral. El archivo se puede abrir también como un proyecto de Familias. Se muestra un extracto en la figura.


```xml
<?xml version="1.0" encoding="utf-8"?>
<FAMILIAS>
    <INFORMATION>
        <TITLE>Familias xml file</TITLE>
        <VERSION>3.2.2</VERSION>
        <TIMESTAMP>
        Mon Dec 11 09:53:31 2017
        </TIMESTAMP>
        <PROJECTNAME>testing_opt</PROJECTNAME>
        <DATABASEUSED>Norwegian</DATABASEUSED>
    </INFORMATION>
    <FREQUENCYDATABASES>
        <FREQUENCYDATABASE>
            <THETA>0</THETA>
            <DATABASENAME>Norwegian</DATABASENAME>
            <MARKERS>
                <MARKER>
                    <MARKERNAME>SE33</MARKERNAME>
                    <ISEXCLUDED>1</ISEXCLUDED>
                    <MARKERMODEL>
                        <FEMALEMUTATIONMODEL>Equal probability (Simple)</FEMALEMUTATIONMODEL>
                        <FEMALEMUTATIONRATE>0</FEMALEMUTATIONRATE>
                        <FEMALEMUTATIONRANGE>0</FEMALEMUTATIONRANGE>
                        <FEMALEMUTATIONRATE2>2.71134e-309</FEMALEMUTATIONRATE2>
                        <MALEMUTATIONMODEL>Equal probability (Simple)</MALEMUTATIONMODEL>
                        <MALEMUTATIONRATE>0</MALEMUTATIONRATE>
                        <MALEMUTATIONRANGE>0</MALEMUTATIONRANGE>
```

Listado de Pedigríes

Este formato genera archivos de texto separado por tabuladores con el listado de todos los pedigríes, así como de los likelihoods, los LRs y los posteriors (si están disponibles)



9.2 Otros

Archivos Generales

Hay varios diálogos/ventanas que permiten exportar la tabla de resultados completa, por ejemplo los resultados de una búsqueda en el módulo DVI. Este tipo de exportación genera simplemente un archivo .txt separado por tabulaciones que puede procesarse después en Excel o una herramienta similar.

Estas exportaciones se explican por sí mismas.

Informes resumen

Hay varios diálogos/ventanas que permiten exportar resúmenes, por ejemplo, el módulo Blind Search permite al usuario generar un resumen de los resultados de una búsqueda, como también lo hace el módulo DVI en las búsquedas.

Exportación de matriz

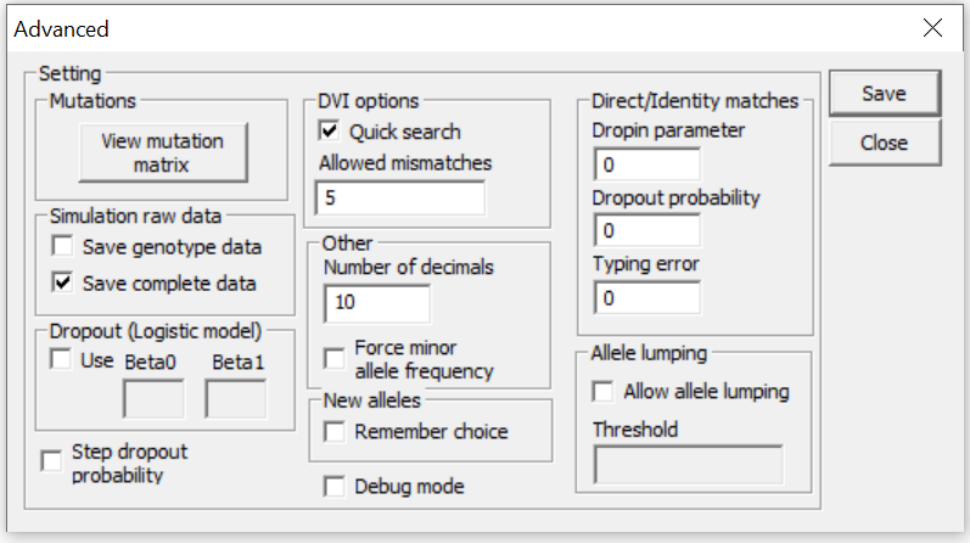
El módulo *Blind Search* permite la exportación de una matriz de distancia. La matriz será de n x n elementos, siendo n el número de individuos en la búsqueda. La matriz compara todos los individuos y calcula, o bien,

1. La distancia IBS entre dos genotipos
2. El coeficiente de parentesco, estimado usando una aproximación de máxima probabilidad
3. Las estimas kappa2 y kappa0, usando el mismo método que en 2.

10. Avanzado

Descripción

En Familias hay algunas opciones/configuraciones avanzadas, a las que puede acceder vía
File -> Advanced, vea la figura de la página siguiente. Algunas configuraciones están destinadas a usuarios expertos.



View mutation matrix – para ver y/o guardar la matriz de mutación de los marcadores definidos.
Save genotype data – Decide si se guardan o no los genotipos de las simulaciones
Save complete data – Decide si se guarda o no todo el dato bruto (genotipos y probabilidades) de una simulación
Dropout – Habilita un modelo logístico de dropout en el que se necesitan las alturas de pico (se puede importar usando sólo el formato Genemapper). Sólo recomendado para usos avanzados.
Step dropout probability – Si se habilita junto con el dropout en un perfil, Familias calculará el LR para un rango de probabilidades de dropout (similar a un análisis de sensibilidad, sensitivity analysis).
DVI options – Por defecto se habilita una función de búsqueda rápida en la que Familias primero calcula los LRs descartando mutaciones y después teniendo en cuenta la mutación en los marcadores en los que la probabilidad es cero. Aquí se configura también el número permitido de mismatches para que aún así Familias lleve a cabo el cálculo. Si deshabilita esta característica el tiempo de computación se puede extender.
Number of decimals – Configura el número de decimales que se muestran y guardan cuando los números son relevantes.
Force minor allele frequency – Si se habilita, Familias forzará el uso de frecuencias mínimas (definidas para cada marcador) cada vez que una frecuencia alélica está por debajo de este valor. Advertencia! Puede causar que la suma de las frecuencias alélicas sea mayor que 1.
New Alleles – Recuerda la elección cada vez que se detecta un alelo nuevo. Por ejemplo, la misma frecuencia para todos los alelos nuevos
Debug mode - Esto activará algunas opciones de depuración, principalmente mediante el volcado de archivos en el directorio de instalación de Familias.
Direct/Identity matches – Valores usados en la comparación directa. También se pueden cambiar en la ventana de diálogo de Blind Search.
Allele lumping – No implementado actualmente. Si se habilita causará el agrupamiento de alelos, lo cual producirá tiempos de computación más cortos. El agrupamiento es una aproximación por la cual todas las transiciones/mutaciones con una probabilidad menor que un umbral se agrupan.

11. Miscelánea

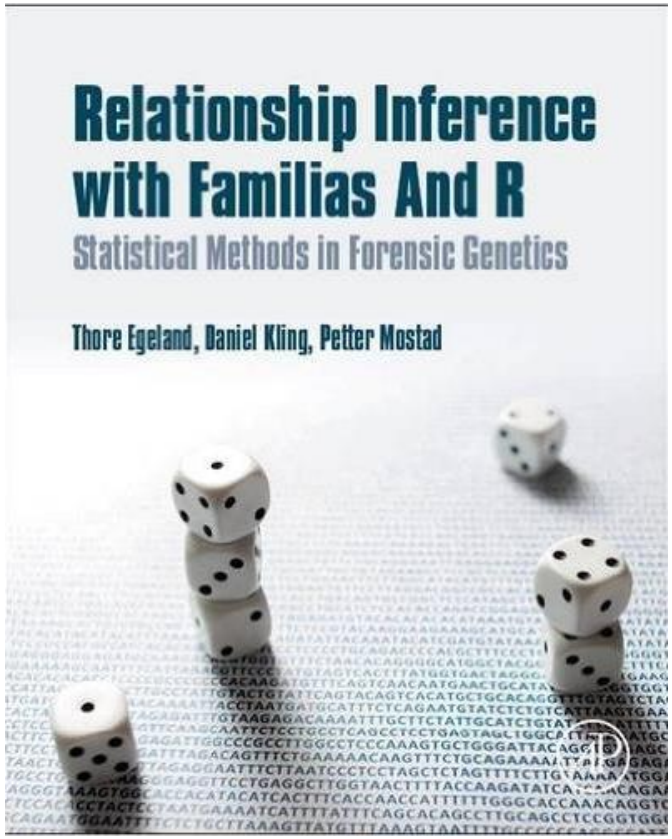
Crear una base de datos de frecuencias

- **File > Create database** se puede usar para
  - Crear una base de datos de frecuencias alélicas
  - Proporcionar parámetros estadísticos poblacionales, como heterocigosidad
  - Realizar búsqueda ciega con los perfiles del estudio poblacional
- Ejemplo simple (file) basado en una muestra de 10 individuos analizados para dos marcadores.

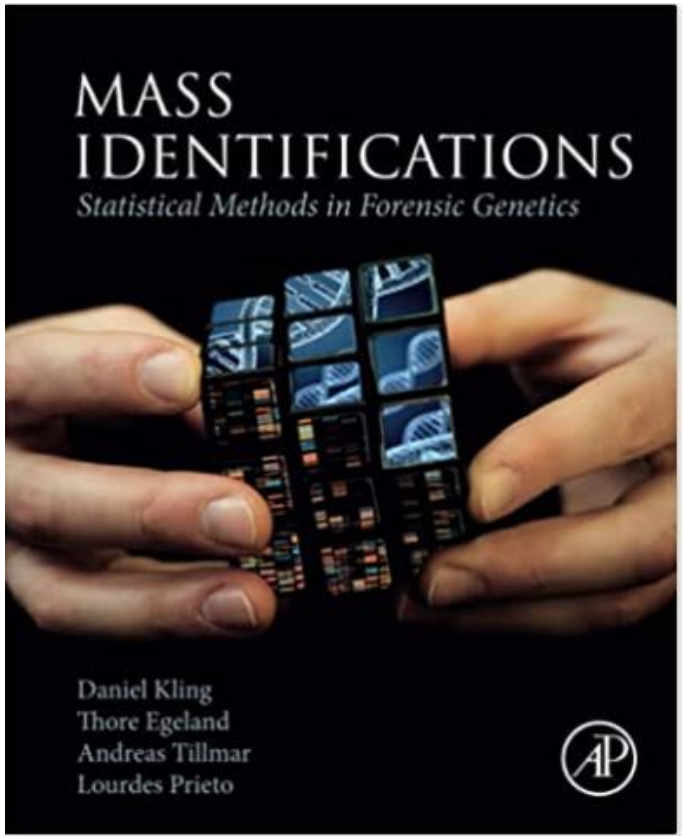
SampleID	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2
ID1	13	14	30	30
...	...	...	...	...
ID10	11	13	30	31.2

- Importar: **File > Create database > Import**
- Crear un archivo de Familias: **File > Create database > Create**
- Parámetros estadísticos: **File > Create database > Statistics**
  Tel archivo contiene info sobre parámetros estadísticos abajo del todo y se puede importar a Familias
- Realizar blind search: **File > Create database > Check data**

Más información en: http://familias.no, http://familias.name y en los libros:



Relationship inference



Mass Identifications

Links útiles

Bases de datos (listas para usar):
- Noruega
- Somalí: Noreste de África (Somalia, Etiopía, Eritrea)
- Asia central (Pakistán/Afganistán)
- Sureste de Asia (Tailandia, Vietnam)
- Española

Elias Hernandis ha diseñado una herramienta muy útil para convertir y producir archivos de datos para Familias (y otros programas). Ver https://leapdna.org/.

Artículos que hacen referencia a Familias. Validación.

Puede ser útil consultar artículos que mencionan Familias

Se puede obtener un listado actualizado aunque no exhaustivo en:

Referencias al artículo de Egeland et al. (2000)

Referencias al artículo de Kling et al. (2014)

Puede ver la validación de software para parentesco, Familias incluido, en:

Drabek (2008)