

Ejercicios a realizar con Familias¹

Indice

Ejercicio S1 Caso de paternidad estandar. Ejemplo basico	3
Ejercicio S2. Caso de paternidad. Mutaciones	4
Ejercicio S3 Articulo CAP (“bone”)	5
Ejercicio S4 Articulo CAP (relacion abuelo-nieto)	7
Ejercicio S5 Caso de paternidad estandar. Indice de Essen-Möller	7
Ejercicio S6. Incesto	8
Ejercicio S7*. Mutaciones.....	10
Ejercicio S8* Mutaciones. Ejercicio teorico (requiere conocimientos de probabilidad)	Error! Bookmark not defined.
Ejercicio S9. Caso de paternidad con mutaciones.....	12
Ejercicio S10. Estudio hermana-hermana	12
Ejercicio S11 Alelos silentes.....	13
Ejercicio S12* Theta-correction. Ejercicio teorico.....	14
Ejercicio S13 Articulo CAP (relacion abuelo-nieto -2-).....	14
Ejercicio S14 Opciones de entrada/salida de datos	14

¹ Thore Egeland, Sept 28, 2012
Norwegian University of Life Sciences
P.O Box 5003
N-1432 Aas Norway
Thore.Egeland@umb.no
<http://arken.umb.no/~theg/alcala>

Ejercicios a realizar con **Familias**

Este documento contiene una serie de ejercicios a realizar utilizando [Familias](http://familias.name/download.html). Utilizaremos la ultima version del programa (Familias 2.0), que puede ser descargada desde <http://familias.name/download.html>. Las soluciones para estos ejercicios aparecen en este [documento](#). Es tambien posible acceder a [videos](#) en los que se muestra como obtener dichas soluciones. Los ficheros de entrada se encuentran [aqui](#). Algunos de los ejercicios que se muestran aqui se solapan con los del [manual](#) del programa. Algunas de las soluciones se muestran con una cantidad inusitada de decimales, simplemente con objeto de comprobar que las soluciones obtenidas son las correctas.²

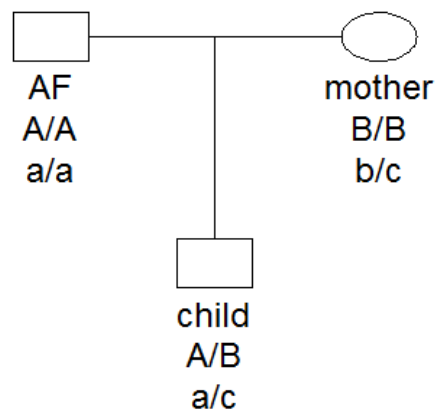
Los ejercicios marcados con un asterisco (*) son de un caracter mas teorico. Pueden saltarse si el objetivo que se persigue es una mejor comprension del programa Familias.

Ejercicio S1 Caso de paternidad estandar. Ejemplo basico

El objetivo de este ejercicio es considerar el caso de paternidad mas simple posible. La Figura 1³ muestra una madre (no se pone en duda), un presunto padre (AF) y un hijo. Se plantean las siguientes hipotesis:

- H_1 : El presunto padre (AF) es el padre biologico. (no se pone en duda la maternidad)
- H_2 : El presunto padre (AF) no tiene relacion alguna (no se pone en duda la maternidad)

Figure1. Standard paternity case.



² Una validacion complete del programa puede encontrarse en “Validation of software for calculating the likelihood ratio for parentage and kinship”, Drábek, FSI: Genetics, Volume 3, Issue 2, Pages 112-118, (2008).

³ Todas las figuras del document han sido obtenidas mediante el paquete de R *paramlink*, disponible en <http://cran.r-project.org/>

Ejercicios a realizar con **Familias**

- a) Considera un unico locus autosomico, S1, con alelos A, B y C (Figura 1). Las frecuencias alelicas son $p_A = p_B = 0.05$ and $p_C = 0.9$. Explica por que la razon de verosimilitudes obtenida es $LR = \frac{1}{p_A} = 20$. Como interpretas esa LR ?
- b) Calcula la LR usando **Familias**.
- c) Tomemos ahora un segundo locus autosomico, S2, con alelos a, b, c y d y frecuencias alelicas 0.1, 0.1, 0.1 y 0.7, respectivamente (Figura 1). Calcula la LR para este marcador a mano. Calcula la razon de verosimilitudes para los dos marcadores usando **Familias**.
- d) Mostrar que la razon de verosimilitudes para los dos marcadores es igual a $1/p_A * 1/p_a$. Verifica el resultado obtenido con **Familias**.
- e) Genera un informe de resultados utilizando la opcion Report. Guarda dicho archivo de resultados. Dicho archivo contiene tanto los parametros y datos de entrada del problema como los resultados finales. Verifica que los resultados son correctos, prestando particular atencion a los valores de LR .
- f) Guardar. Salir de **Familias**.
- g) Abrir **Familias** y leer el proyecto guardado anteriormente.
- h) Consideramos a continuacion una *theta-correction* (θ -correction). Por simplicidad, trabajaremos unicamente con el primer marcador S1. El parametro θ se llama *kinship parameter* (parametro de parentesco o cosanguinidad) en **Familias** y podemos cambiar su valor por defecto en Options. Cambia el valor a $\theta=0,02$. Calcula la LR para el marcador S1. De cara a obtener calculos referentes unicamente a un subgrupo de marcadores, utilizaremos la opcion Systems included. Comprueba que el valor obtenido coincide con el valor teorico dado por

$$LR = \frac{1 + 3\theta}{2\theta + 1 - \theta p_A}$$

- i) Debate acerca de los diferentes parametros que se asumen como ciertos a la hora de resolver el ejercicio.

Ejercicio S2. Caso de paternidad. Mutaciones

Consideramos ahora un caso de paternidad para el cual no se tiene informacion genetica de la madre (Figura 2) con un unico marcador, VWA:

Alelo	14	15	16	17	18	19	20	21
Frec	0,072	0,082	0,212	0,292	0,222	0,097	0,02	0,003

Los datos y las hipotesis son:

- Presunto padre : 14, 15
- Hijo : 16, 17
- H_1 : El presunto padre es el padre biologico.
- H_2 : No existe relacion familiar entre ambos.

Ejercicios a realizar con **Familias**

- Explica por que se obtiene $LR=0$. Confirma el resultado utilizando **Familias**.
- Selecciona el modelo de mutacion Equal probability con tasa de mutacion $R=0.007$ para hombres y mujeres y calcula la LR .
- Sea

$$n = 8 \text{ (numero de alelos),}$$

$$m = R / (n - 1) = 0.007 / 7 = 0.001,$$

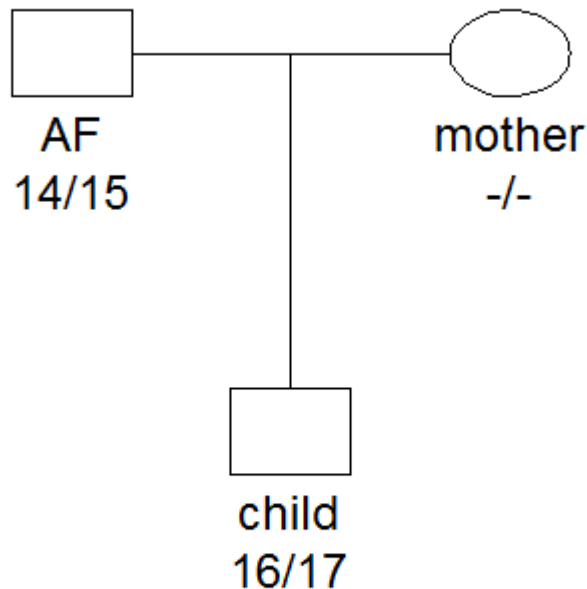
$$p_{16} = 0.212 \text{ y } p_{17} = 0.292.$$

Es posible demostrar (Ejercicio S8) que

$$LR = \frac{m \ p_{16} + p_{17}}{2 \ p_{16} \ p_{17}}.$$

Utiliza dicha formula para comprobar el resultado obtenido con **Familias**.

Figure 2. Paternity case. Mutation



Ejercicio S3 Articulo CAP ("bone")

El College of American Pathologists (CAP) lleva a cabo diferentes tests de cara a evaluar la capacidad de diferentes laboratorios que llevan a cabo pruebas de ADN. A continuacion se indica el test realizado en 2011:

Unos escaladores han encontrado restos oseos en un bosque. Las pruebas parecen dar una idea de quien es el individuo. Se pide que se realice una prueba de cara a determinar si el individuo (bone) esta relacionado con una presunta madre (AM) y la otra hija de dicha madre, la presunta hermana (AS) -Figura 3-. Consideramos las siguientes hipotesis:

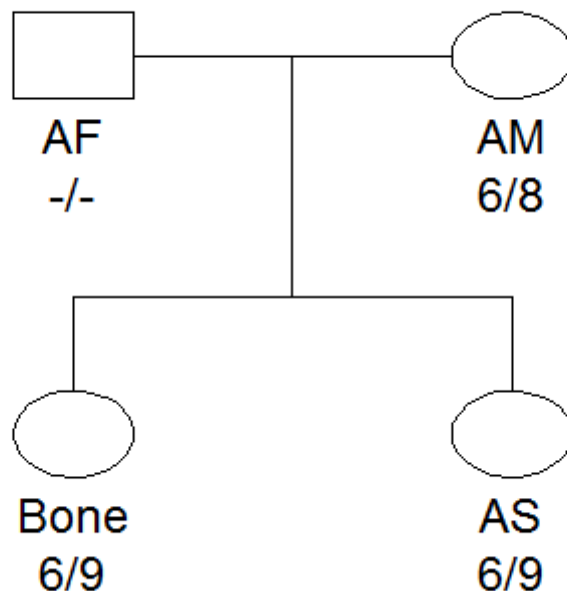
Ejercicios a realizar con **Familias**

- H_1 : El individuo del que proceden los huesos es hija de AM y hermana de AS.
 - H_2 : El individuo del que proceden los huesos no tiene relacion familiar alguna con AM y AS.
- a) Introduce los datos para el marcador que se muestra en la Figura 3 en **Familias** y calcula la LR .
- b) Lee el archivo externo ExS3.txt. Calcula la LR basada en todos los marcadores geneticos.
- c) Usa los resultados obtenidos con **Familias** para encontrar las LR s de cada marcador individual. Comprueba que el valor obtenido con el marcador F13B corresponde con el obtenido en el apartado a). Completa el siguiente cuadro:

LR	$LR(H1/H2)$
D3S1358	
D21S11	
D18S51	
D7S820	
D16S539	
CSF1PO	
F13B	
LPL	
TOTAL	

- d) Uno de los marcadores, el D7S820, da lugar a un LR muy alto, 11189. Cual es la razon de este valor tan alto? Cuanto valdria el LR si eliminamos este marcador?

Figure 3. CAP paper challenge. First marker.
Frequencies: 14:0.122, 15:0.258, 17:0.197, rest:0.423



Ejercicio S4 Artículo CAP (relacion abuelo-nieto)

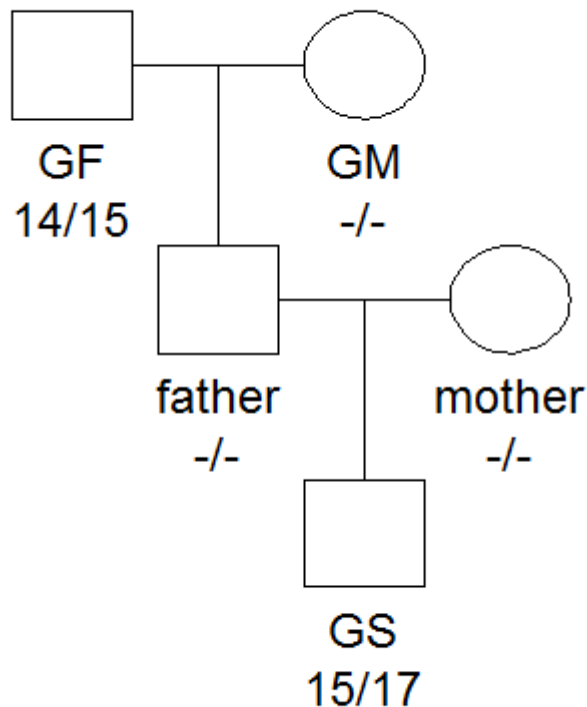
Dos individuos, GF and GS, han sido genotipados. Se consideran las siguientes hipotesis:

- H_1 : GF es el abuelo de GS.
- H_2 : GF y GS no tienen relacion familiar alguna.

La Figura 4 muestra el pedigrí correspondiente a la hipotesis H_1 para el primer marcador genético utilizado (D3S1358).

- Introduce manualmente los datos en **Familias** y calcula la LR para el marcador que se muestra en la Figura 4.
- Calcula la LR basada en todos los marcadores. Lee los datos del archivo ExS4.txt.
- Debate*: En el ejercicio real se afirmó que GF y GS compartían el mismo haplotipo del cromosoma Y, haplotipo con una frecuencia en la población de 0.0025. Puede usarse esta información?

Figure 4. Grand father alternative



Ejercicio S5 Caso de paternidad estandar. Indice de Essen-Möller

(Continuacion del Ejercicio S1). Consideramos las siguientes hipotesis:

- H_1 : El presunto padre (AF) es el padre biológico. (no se pone en duda la maternidad)
- H_2 : El presunto padre (AF) no tiene relacion alguna (no se pone en duda la maternidad)

Ejercicios a realizar con **Familias**

En vez de calcular la LR trabajaremos aquí con el índice de Essen-Möller W , definido matematicamente como la probabilidad de H_1 condicionado a la información genotípica hallada.

Partimos *a priori* de que ambas hipótesis H_1 y H_2 son igualmente probables. En tal caso, puede demostrarse que

$$W = P(H_1 | data) = \frac{LR}{LR + 1}.$$

- Recuerda que $LR=20$ para el marcador S1. Calcula W .
- Recuerda que $LR=200$ para los dos marcadores S1 y S2. Calcula W .
- En **Familias**, W puede obtenerse usando la opción Probabilities. Usa **Familias** para calcular W en ambos casos.
- Debate*: Que valor es preferible, LR o W ?

Ejercicio S6. Incesto

En un caso de paternidad, consideramos las hipótesis siguientes:

- H_1 : AF, el padre de la madre (indiscutible), es también el padre de su hijo.
- H_2 : Otro hombre es el padre de su hijo.

La Figura 5 muestra el pedigrí correspondiente a la hipótesis H_1 . Las frecuencias alélicas son $p_1 = p_2 = p_3 = 0.05$.

- Usa **Familias** para calcular la LR . Comprueba que el resultado coincide con el valor teórico dado por $LR = 1/(2 * p_2)$. Tiene el incesto influencia en el resultado *en este caso particular*?
- La defensa afirma que se deberían tener en cuenta 3 hipótesis, en lugar de las 2 de partida:
 - H_1 : AF, el padre de la madre (indiscutible), es también el padre de su hijo.
 - H_2 : Otro hombre es el padre de su hijo.
 - H_3 : El hermano de la madre es el padre del hijo (ver Figura 6).

La LR puede obtenerse de diferentes formas dependiendo de cuál es la hipótesis que se toma como referencia. Calcula las LR s con **Familias** tomando la segunda hipótesis como referencia. En otras palabras

$$LR_{1,2} = \frac{P(data | H_1)}{P(data | H_2)} \text{ and } LR_{1,3} = \frac{P(data | H_1)}{P(data | H_3)}.$$

Comprueba que $LR_{1,3} = 2 / (1 + p_2)$.

- En aquellos casos en que se tienen más de 2 hipótesis, como ocurre en el apartado anterior, algunos científicos ven preferible calcular probabilidades a posteriori. Calcula las probabilidades a posteriori correspondientes a cada una de estas hipótesis.

$$P(H_1 | data), P(H_2 | data) \text{ and } P(H_3 | data).$$

Figure 5. Incestuous (AF) alternative

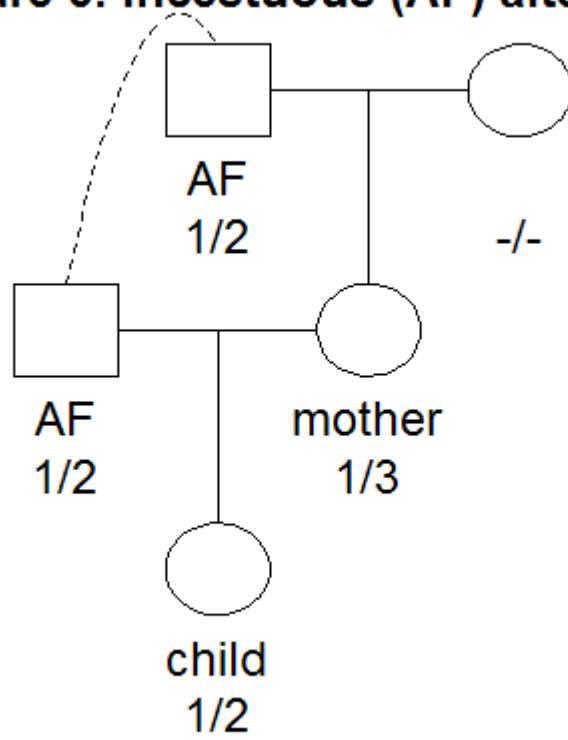
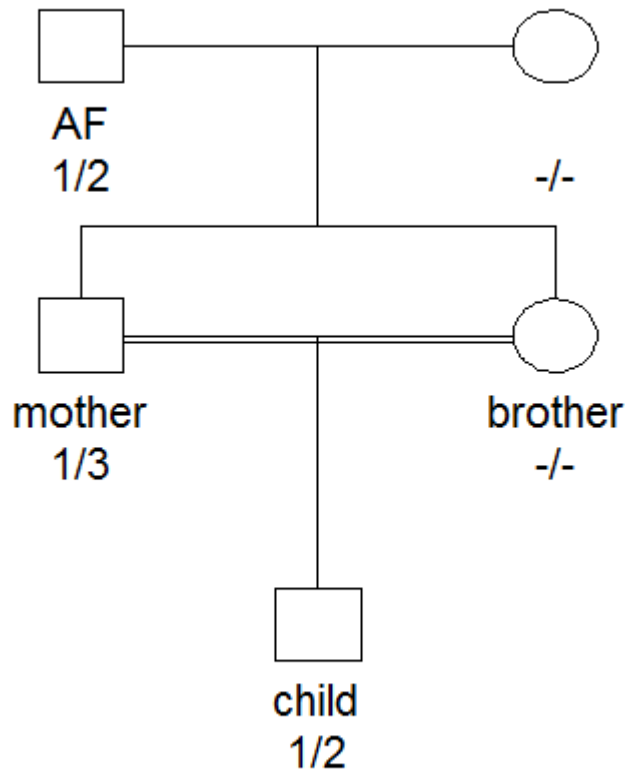


Figure 6. H3: Brother alternative



Ejercicio S7*. Mutaciones

En este ejercicio, que retoma el Ejercicio S2, trabajaremos con los diferentes modelos de mutacion que aparecen en **Familias**. Recordamos las frecuencias alelicas del marcador VWA:

Allele	14	15	16	17	18	19	20	21
Freq	0,072	0,082	0,212	0,292	0,222	0,097	0,02	0,003

Los datos y las hipotesis son:

- Presunto padre : 14, 15
- Hijo : 16, 17
- H_1 : El presunto padre es el padre biologico.
- H_2 : No existe relacion familiar entre ambos.

En **Familias**, se tienen en cuenta cuatro posibles modelos de mutacion. Utilizaremos este ejercicio para probar los cuatro. Trabajaremos con una tasa de mutacion global de $R=0.005$, y el mismo modelo tanto para mujeres como para hombres. Las soluciones se obtendran utilizando el programa, mientras que el ejercicio siguiente es de contenido fundamentalmente teorico.

Ejercicios a realizar con **Familias**

- a) Carga los datos del archivo ExS7.txt en **Familias**.
- b) Utiliza el modelo 4 Equal probability. Calcula la *LR*.
(Resultado; $LR=2.9E-03=0.0029$. Comentario: Para este modelo de mutacion, una mutacion es igualmente probable a cualquiera de los otros alelos).
- c) Utiliza el modelo 3 Probability proportional to frequency. Calcula la *LR*.
(Resultado $LR=6.3E-03=0.0063$. Comentario: Para este modelo es mas probable mutar a un alelo comun –frecuencia alta- que a uno raro –frecuencia baja-).
- d) Utiliza el modelo 2 Probability decreasing with range (equal). Para este modelo son necesarios 2 parametros. La tasa de mutacion no cambia, $R=0.005$. Asigna al segundo parametro, Mutation range, el valor $r=0.5$. Calcula la *LR*.
(Resultado: $LR=4.7E-03=0.0047$. Comentario: En este modelo, la probabilidad de la mutacion depende de la amplitud. Con $r=0.5$, una two step mutation tiene lugar con la mitad de probabilidad que una one-step mutation. Una three step mutation tiene lugar con la mitad de probabilidad que una two step mutation y asi).
- e) Utiliza el modelo 1 Probability decreasing with range (stable) con los mismos parametros que arriba.
(Resultado: $LR=6.4E-03=0.0064$.)
- f) *Comentario:* El modelo 1 (Probability decreasing with range (stable)) y el modelo 3 (Probability proportional to frequency) son *estables* mientras que los modelos 2 y 4 son *inestables*. Si un modelo de mutacion es estable, el hecho de introducir a un nuevo individuo para el que no se dispone de informacion (e.g. el padre del supuesto padre) no da lugar a cambios en la *LR*. Esta es una propiedad deseable, puesto que la introduccion de informacion irrelevante no deberia de producir cambios en los resultados. Por otro lado, los modelos inestables no poseen dicha propiedad y daran lugar a pequeños cambios en tales casos. Verifica que para los modelos 1 y 3 no se producen cambios en situaciones asi, mientras que para los modelos 2 y 4 se muestran ligeros cambios.
- g) *Comentario:* Todos los modelos de mutacion que se utilizan aqui muestran una caracteristica: aunque unicamente 5 alelos (14, 15, 16, 17 y ‘Rest allele’) son necesarios, se incluye informacion para el sistema de 8 alelos dado por VWA. Si introducimos un sistema con 5 alelos, veremos que se producen ligeros cambios en el resultado final.

Ejercicio S8* Mutaciones. Ejercicio teorico (requiere conocimientos de probabilidad)

En este ejercicio, que sirve de confirmacion a los resultados obtenidos en el Ejercicio S7, explicaremos la base matematica en que se apoyan los modelos de mutacion. Los datos e hipotesis de trabajo pueden encontrarse en el Ejercicio S7.

- a) Probar que

$$LR = \frac{P(data | H_1)}{P(data | H_2)} = \frac{p_{16}(m_{14,17} + m_{15,17}) + p_{17}(m_{14,16} + m_{15,16})}{4p_{16}p_{17}} \quad (1)$$

siendo p_{16} la frecuencia alelica del alelo 16 y $m_{14,17}$ la probabilidad de mutar del alelo 14 al 17 (analogo para el resto de notaciones).

Ejercicios a realizar con Familias

- b) Para el modelo de mutacion Equal probability se toma $m_{ij} = m = R / (n - 1)$ donde n es el numero total de alelos. Explica por que la formula del Ejercicio S2 se sigue de la ecuacion (1). Probar que cuando $n=8$ y $R=0.005$, entonces $LR=0.0029$, como en el Ejercicio S7 b).
- c) Consideramos el siguiente modelo proporcional de mutaciones. Por definicion,

$$\begin{aligned} m_{ij} &= kp_j \text{ for } i \neq j, \\ m_{ii} &= 1 - k(1 - p_i). \end{aligned} \tag{1.1}$$

Probar que

$$LR = k = \frac{R}{\sum_{i=1}^n p_i(1 - p_i)} = 0.0063$$

Comentario: La base matematica sobre la que se apoyan los otros dos modelos de mutacion es considerablemente mas tecnica y complicada. Puede encontrarse explicada en la Seccion A.1.4 del [manual](#).

Ejercicio S9. Caso de paternidad con mutaciones

Carga el archivo ExS9.txt. Se consideran las siguientes hipotesis

- H_1 : El supuesto padre es el padre biologico
 - H_2 : El supuesto padre no esta relacionado
- a) Probar que $LR=0$.
- b) Para uno de los marcadores geneticos utilizados, el hijo y su supuesto padre no comparten alelo alguno. Encontrar dicho marcador.
- c) Utilizar el modelo de mutaciones Probability decreasing with range (stable) tanto para hombres como para mujeres, con tasa de mutacion 0.001 y rango de mutacion 0.5 para el marcador PENTA_E y calcular la LR .
- d) Supongamos que se tiene en cuenta una nueva hipotesis H_3 : El hermano del supuesto padre es el padre biologico. Calcular la $LR (H_1/ H_3)$.
- e) *Debate:* Hay un modelo de mutacion que pueda considerarse “mejor” que los demas? Deberia dicho modelo ser usado de forma rutinaria para todos los marcadores?

Ejercicio S10. Estudio hermana-hermana

Se pretende determinar si dos chicas (llamadas sister1 y sister2 en la Figura 7 que se muestra mas abajo) son hermanas (tal y como indica la hipotesis H_1) o si son medio hermanas (tal y como indica la hipotesis H_2 , que se muestra en la parte derecha de la Figura 7). Los alelos del sistema se indican en la Tabla 1 mas abajo. Las frecuencias alelicas son 0.1 para los sistemas S1 y S2, y 0.05 para los sistemas S3-S5.

Figure 7. H1 and H2 (right)

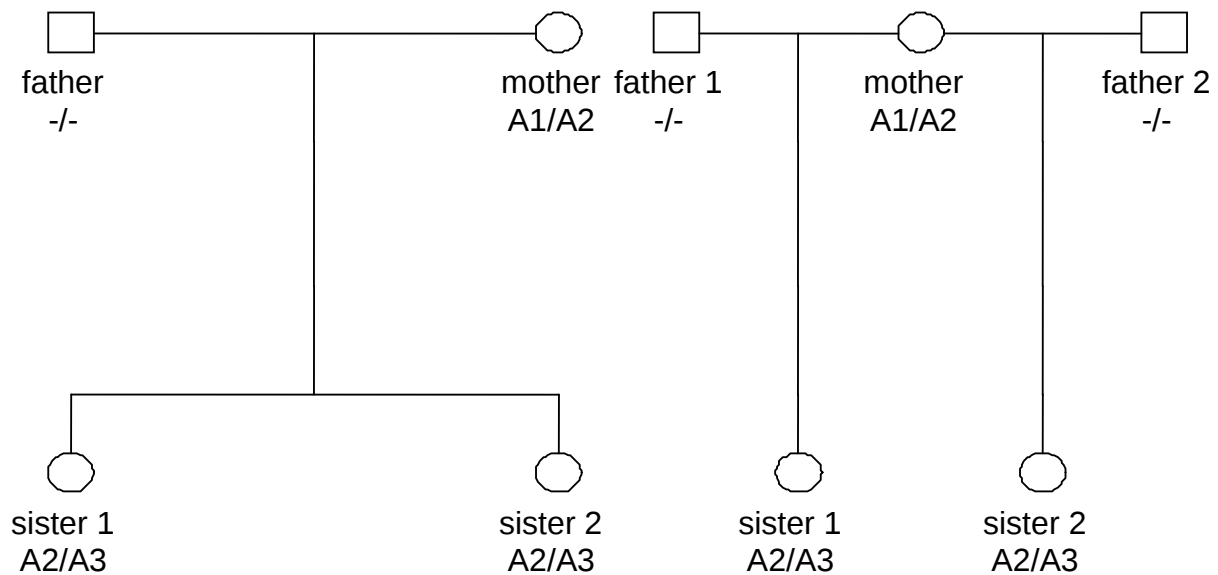


Table 1. Datos genotipicos para el Ejercicio S10.

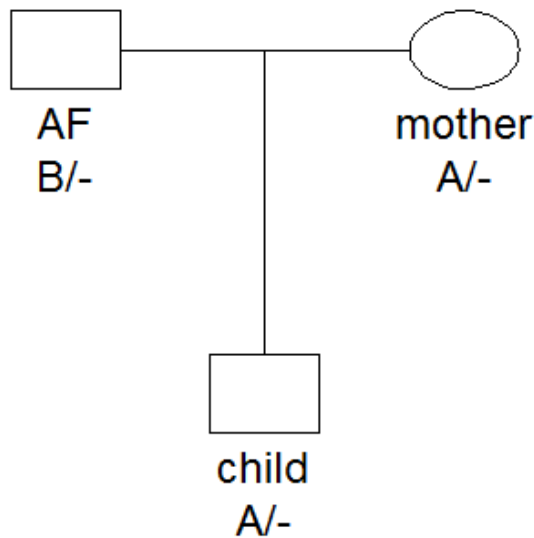
Person	S1	S2	S3	S4	S5
Mother	A_1A_2	A_1A_2	A_2A_3	A_2A_3	A_2A_3
Sister 1	A_2A_3	A_2A_3	A_3A_4	A_3A_4	A_3A_4
Sister 2	A_2A_3	A_2A_3	A_1A_3	A_1A_3	A_1A_3

- Cual es la LR obtenida al comparar la hipotesis de que sister1 y sister2 son hermanas frente a la alternative de que son medio hermanas?
- Debate:* La LR aplicada a este caso no es suficiente para llegar a una conclusion clara. Cuantos marcadores crees que serian necesarios para llegar a una conclusion?

Ejercicio S11 Alelos silentes

Observa la siguiente Figura. Corresponde a un caso de paternidad en el que se sospecha de la existencia de un alelo silente. Incluye en el estudio un alelo silente con frecuencia 0.05 y calcula la LR usando **Familias**. Las frecuencias alelicas de A y B son 0.1

Figure 5. Silent allele?



Ejercicio S12* Theta-correction. Ejercicio teorico

Comprobar teoricamente que se verifica la formula utilizada en el Ejercicio S1 h).

Ejercicio S13 Artículo CAP (relacion abuelo-nieto -2-)

Este ejercicio es una continuacion del Ejercicio S4 introduciendo la *theta correction*: Calcula la *LR* utilizando todos los marcadores. De cara a ahorrar tiempo, lo recomendable es leer los datos del archivo ExS4.txt y cambiar el valor del parametro theta (kinship-parameter) a 0.02. (Ayuda: Esto se lleva a cabo en la ventana Pedigrees, utilizando el boton Options).

Ejercicio S14 Opciones de entrada/salida de datos

El modo mas practico de trabajar con el programa **Familias** es leer los datos de un archivo externo que contenga la informacion necesaria. En algunos casos, puede ser tambien de interes el escribir los resultados en archivos externos. Veremos en este ejercicio como trabajar estas diferentes opciones.

- Lee los datos del archivo externo ExS3.txt.
- Escribe la base de datos en la ventana 'General DNA data'. Guarda el archivo como database.txt.
- Escribe la base de datos de la ventana 'Case Related DNA Data' en un archivo externo llamado casedata.txt.
- Abre un nuevo proyecto en **Familias**.
- Lee database.txt.
- Lee casedata.txt.

Ejercicios a realizar con **Familias**

g) Define the pedigrees, see Exercise S3, and calculate the *LR* once more.