

Genética y Probabilidad: pruebas de paternidad y portadores de enfermedades

Rosario Delgado de la Torre

La *genética de poblaciones* estudia la distribución de los genes en las poblaciones, así como los factores que mantienen o cambian la constitución genética de las mismas, de una generación a otra. Una *población* se define como un conjunto de individuos capaces de reproducirse entre sí, que coexisten geográfica y temporalmente, y constituyen una unidad evolutiva. El estudio de la genética de poblaciones es muy importante desde el punto de vista sanitario, ya que la mayoría de las enfermedades tienen una componente genética más o menos importante.

Ya que hablamos de “*genes*”, explicaremos éste y otros términos que aparecerán a continuación:

Los seres humanos, por poner un ejemplo, tenemos en las células 46 cromosomas (22 pares no sexuales y 2 cromosomas sexuales, que en las mujeres son XX y en los hombres XY). Cada cromosoma se compone, básicamente, de ácido desoxirribonucleico (ADN). El segmento de ADN que incluye la información correspondiente a la producción de los aminoácidos de una proteína particular se denomina *gen*.

Se denomina *alelos* a las diferentes formas alternativas de un gen. La localización particular o posición de un gen en un cromosoma se denomina *locus* (en plural, *loci*). Los genes aparecen en parejas en cada locus, y se separan durante la formación de los *gametos* (células sexuales). Cuando dos gametos, uno proveniente de cada progenitor, se unen en la fecundación, la descendencia recibe un gen de cada uno de ellos, en cada locus.

Los individuos con dos alelos diferentes en un locus se llaman *heterocigóticos*, mientras que los que tienen dos copias del mismo alelo se llaman

homocigóticos. Por *genotipo* se entiende la constitución genética de un individuo en uno o varios loci. El término *fenotipo* se refiere a una característica observable de un individuo, asociada con determinado genotipo.

Se dice que un alelo es *recesivo* respecto a una característica o fenotipo, si su efecto sobre dicha característica sólo es visible para los individuos homocigóticos en dicho alelo. Se dice que un alelo es *dominante* en caso contrario, es decir, cuando su efecto sobre la característica es el mismo para los heterocigóticos que para los homocigóticos en este alelo. Se dice que dos alelos son *codominantes* cuando ninguno de los dos es recesivo respecto del otro (es decir, cuando la presencia de cada uno de ellos se deja notar en el fenotipo y no puede quedar oculta por la presencia del otro).

Leyes de Mendel y pruebas de paternidad

Las leyes de Mendel son unos postulados sobre la transmisión de caracteres hereditarios (fenotipos) de una generación a la siguiente. En 1865, el monje agustino austriaco Johann Gregor Mendel (1822-1884), abad del monasterio de Brünn (Chequia), formuló las leyes hereditarias que llevan su nombre, fruto de sus estudios empíricos con algunas especies vegetales (concretamente, guisantes olorosos) de su jardín. El trabajo de Mendel sobre la transmisión de los caracteres de las plantas a través de sucesivas generaciones constituye el fundamento de la genética moderna, aunque no fue reconocido por la comunidad científica hasta su redescubrimiento, en 1900, por parte de Hugo de Vries, Karl Frich Correns y Erich Tschermack.

Supongamos que tenemos una característica o fenotipo controlada por un gen que presenta dos alelos, digamos L (para el dominante) y ℓ para el recesivo (en el caso particular que consideró Mendel, el fenotipo es la rugosidad o no de la piel del guisante, que está controlada por un gen con dos alelos, el dominante L que corresponde a piel lisa y el recesivo ℓ que corresponde a piel rugosa). Es habitual, en el caso de tener un alelo dominante y otro recesivo, el utilizar una letra en mayúscula para denotar el primero (usualmente la inicial) y la misma letra en minúscula para el segundo.

Entonces hay tres tipos de guisante oloroso, atendiendo a este gen (tres genotipos posibles):

LL (liso puro) , $L\ell$ (liso portador de rugoso) y $\ell\ell$ (rugoso)

aunque sólo dos fenotipos: liso (para LL y $L\ell$) y rugoso ($\ell\ell$).

La **primera ley de Mendel** se enuncia así: si se cruzan dos individuos homocigóticos diferentes, esto es, un LL y un $\ell\ell$, entonces todos los descendientes, que forman la primera generación filial, son del mismo tipo, heterocigóticos $L\ell$, y presentan el fenotipo del dominante (en el caso de los guisantes, son lisos) aunque son portadores del recesivo.

Como cada progenitor aporta un alelo a cada hijo, está claro en este caso que uno siempre aporta el alelo L y el otro el ℓ , con lo que todos los hijos han de tener en el locus del gen los dos alelos L y ℓ .

La **segunda ley de Mendel** se refiere a lo que sucede al cruzar dos individuos de la primera generación filial a la que nos referimos en la primera ley, entre ellos. En ese caso, se obtienen descendientes con todas las posibles combinaciones de alelos, en las siguientes proporciones:

LL liso puro ($\frac{1}{4}$), $L\ell$ liso portador de rugoso ($\frac{2}{4} = \frac{1}{2}$) y $\ell\ell$ rugoso ($\frac{1}{4}$)

(en cuanto al fenotipo, hay $1/4$ de rugosos y $3/4$ de lisos).

Aunque Mendel obtuvo esta ley de manera empírica, observando lo que sucedía con miles de plantas de guisante, podemos llegar a ella usando un modelo probabilístico muy sencillo: el modelo consiste en suponer que cada progenitor (que es de la forma $L\ell$), aporta un alelo a cada descendiente, cada uno de ellos con la misma probabilidad. Entonces, las posibles combinaciones de alelos en los descendientes son:

$$LL, \quad L_p\ell_m, \quad \ell_pL_m \quad \text{y} \quad \ell\ell$$

donde hemos tenido en cuenta que para el caso heterocigótico, el alelo L podía provenir del padre y el ℓ de la madre o viceversa, y lo hemos indicado con el subíndice p para el padre y m para la madre (en el caso homocigótico esta distinción es innecesaria).

Como todas las combinaciones son igualmente probables, se han de repartir a partes iguales la probabilidad total (que es 1), correspondiéndoles $1/4$ a cada una. Al ser la segunda y la tercera combinaciones el mismo genotipo, que denotamos genéricamente por $L\ell$, a éste le corresponde una probabilidad de $1/4 + 1/4 = 2/4 = 1/2$.

Hemos visto un modelo probabilístico que se adapta bien a las dos primeras leyes de Mendel. Podemos aplicarlo en otras situaciones y ver qué predice

el modelo. Por ejemplo, si cruzásemos un individuo LL (liso puro, en el caso de los guisantes) con uno $L\ell$ (liso portador de rugoso), ¿qué genotipos podemos esperar obtener en la primera generación filial, y con qué frecuencias o probabilidades? Como uno de los padres siempre aporta el alelo L , mientras que el otro aporta los alelos L y ℓ con la misma probabilidad, la mitad de los hijos tendrá el genotipo LL (liso puro), y la otra mitad $L\ell$ (liso portador de rugoso).

El mismo caso, pero cruzando un individuo $\ell\ell$ (rugoso) con $L\ell$ (liso portador de rugoso) nos da la mitad de los descendientes $L\ell$ (liso portador de rugoso) y la mitad $\ell\ell$ (rugoso).

Estos resultados, predichos por el modelo probabilístico, coinciden con los observados por Mendel sobre sus miles de plantas de guisante. Parece, por tanto, que el modelo probabilístico que hemos considerado para la transmisión de los caracteres hereditarios es adecuado. El modelo se adapta igualmente al caso en que los alelos sean *codominantes*, en vez de uno dominante y el otro recesivo, o a que haya más de dos, e incluso a casos mixtos, como veremos a continuación.

Aplicando la genética mendeliana a los grupos sanguíneos humanos se puede llevar a cabo una **prueba de paternidad**, aunque no siempre será concluyente. Algunas pruebas de paternidad han cobrado cierta notoriedad por haber involucrado a personajes públicos. Una de las más famosas tuvo lugar cuando en 1943 Charles Chaplin se enfrentó a un juicio por paternidad en Hollywood, su abogado recurrió a la genética mendeliana aplicada a los grupos sanguíneos para realizar una prueba de paternidad. Explicaremos ahora en qué consiste:

El grupo sanguíneo humano viene determinado por tres alelos diferentes, conocidos como A , B y O . A y B son dominantes respecto del alelo recesivo O , pero son codominantes entre sí (se manifiestan ambos). Los posibles grupos sanguíneos son (se indican los nombres de los grupos y entre paréntesis los genotipos):

grupo AB (AB), grupo A (AO), grupo B (BO), grupo O (OO)

Entonces, por ejemplo, a partir de un padre de grupo sanguíneo AB (genotipo AB) y de una madre de grupo O (genotipo OO), los posibles genotipos para los hijos son: AO y BO , con la misma probabilidad. Luego los hijos son del grupo sanguíneo A o del grupo B, con probabilidad $1/2$. ¿Es posible que

esta pareja tenga un hijo del grupo O, por ejemplo? La respuesta es que no, salvo mutación genética (muy rara). Luego si se sabe que el hijo es del grupo O y la madre también, el padre biológico no puede ser del grupo AB, salvo mutación (en este caso, aunque no supiésemos el grupo de la madre sabríamos que es imposible que un individuo del grupo O tenga un progenitor del grupo AB). Esta es la base de las llamadas **“pruebas de paternidad”** basadas en el grupo sanguíneo, que permiten descartar, en algunos casos, a un individuo como padre biológico de otro.

Volviendo al “caso Chaplin”, como el grupo sanguíneo de Joan Berry, la madre, era A, y el de la hija era B, el padre debía tener grupo B o AB necesariamente, pero el grupo de Chaplin era O. Tres médicos que declararon en el juicio dieron su conclusión de forma explícita: Chaplin no podía ser el padre de la criatura. Sin embargo, el juez McComb decidió que este argumento biológico era sólo uno más a tener en cuenta, y que había otras pruebas que demostraban lo contrario. En un fallo doblemente ejemplar (por la innovación que supuso el uso de la prueba genética de paternidad en el juicio, y por el grave error judicial que consistió en no darle el valor que ésta merecía), el juez dictaminó la paternidad de Chaplin.

Obviamente, la prueba también podría ser de “maternidad”, pero lo habitual es que la madre biológica sea conocida y que las dudas, si las hay, recaigan sobre la identidad del padre biológico. Hay que decir que actualmente existen pruebas de paternidad basadas en el ADN, mejores que las basadas en el grupo sanguíneo, ya que son siempre concluyentes.

La ley de Hardy-Weinberg de la genética de poblaciones y la estimación del número de portadores de una enfermedad

La ley de Hardy-Weinberg del equilibrio genético proporciona un modelo matemático para el estudio de los cambios evolutivos en la frecuencia de los alelos dentro de una población. En 1908, el matemático G. Hardy y el físico W. Weinberg, de manera independiente, plantearon la hipótesis de que la frecuencia de los alelos y de los genotipos en una población permanecen constantes, de generación en generación, si la población se encuentra estable y en *equilibrio genético*. Esto quiere decir que se cumplan cinco condiciones:

- 1) La población reproductiva es grande.

2) Todos los individuos de la población tienen las mismas probabilidades de reproducirse y el cruce entre dos individuos (progenitores) se produce al azar.

3) No hay mutaciones.

4) No hay migración.

5) No hay selección natural.

Veamos con un ejemplo en qué consiste esta ley:

Supongamos que tenemos una población de 1750 ovejas en equilibrio genético. La hemoglobina de la oveja presenta dos formas diferentes, producidas por dos alelos: A y B , codominantes, de un locus. Con estos dos alelos son posibles tres genotipos:

$$AA, AB \text{ y } BB$$

cada uno de los cuales da un fenotipo diferente (un determinado patrón electroforético), por ser codominantes. Supongamos que se obtienen muestras de sangre de las ovejas y que realizando el análisis electroforético para determinar el tipo de hemoglobina de cada una de ellas se obtiene la siguiente distribución de frecuencias absolutas, o número de ovejas con cada genotipo:

$$AA (910), AB (280) \text{ y } BB (560)$$

Entonces, las frecuencias relativas, llamadas “frecuencias genotípicas” son:

$$AA (910/1750 = 0,52), AB (280/1750 = 0,16) \text{ y } BB (560/1750 = 0,32)$$

y se obtienen dividiendo las frecuencias absolutas para cada tipo de hemoglobina por el número total de individuos.

A partir de las frecuencias genotípicas se pueden determinar las frecuencias alélicas, esto es, las frecuencias de los alelos A y B en la población de ovejas, teniendo en cuenta que la población se encuentra en equilibrio genético y que las leyes de Mendel para el cruce de dos individuos de la población rigen. Como cada individuo AA tiene dos alelos A , cada individuo AB tiene un alelo A y uno B , y cada individuo BB tiene dos alelos B , entonces el número total de alelos de cada tipo en la población es

$$A : 2 \times 910 + 1 \times 280 = 2100 \quad \text{y} \quad B : 2 \times 560 + 1 \times 280 = 1400$$

y las frecuencias (relativas) de los alelos son

$$A : 2100/(2 \times 1750) = 0,6 \quad \text{y} \quad B : 1400/(2 \times 1750) = 0,4$$

En la población de ovejas, las frecuencias alélicas son las mismas en machos y hembras. Entonces, como los cruces se producen al azar (suponemos que la población se encuentra en equilibrio genético), podemos calcular la probabilidad de cada cruce concreto. Por ejemplo, la del cruce $AA \times AB$ es:

$$AA \times AB : 2 \times (0,52 \times 0,16) = 0,1664$$

(se multiplica por dos el producto de las frecuencias relativas de AA y de AB , ya que puede ser el macho AA y la hembra AB o al revés; esto no pasa si los dos progenitores tienen el mismo genotipo). Para los otros cruces tenemos:

$$\begin{aligned} AA \times AA & : 0,52 \times 0,52 = 0,2704 \\ AA \times BB & : 2 \times (0,52 \times 0,32) = 0,3328 \\ AB \times AB & : 0,16 \times 0,16 = 0,0256 \\ AB \times BB & : 2 \times (0,16 \times 0,32) = 0,1024 \\ BB \times BB & : 0,32 \times 0,32 = 0,1024 \end{aligned}$$

(notemos que la suma de las probabilidades de los 6 posibles cruces es 1).

Vamos a calcular las frecuencias genotípicas de los descendientes de cada cruce (siguiendo las leyes de Mendel):

$$\begin{aligned} AA \times AB & \longrightarrow AA (1/2) \quad \text{y} \quad AB (1/2) \\ AA \times AA & \longrightarrow AA (1) \\ AA \times BB & \longrightarrow AB (1) \\ AB \times AB & \longrightarrow AA (1/4), \quad AB (1/2) \quad \text{y} \quad BB (1/4) \\ AB \times BB & \longrightarrow AB (1/2) \quad \text{y} \quad BB (1/2) \\ BB \times BB & \longrightarrow BB (1) \end{aligned}$$

Teniendo ahora en cuenta las frecuencias de cada cruce, tenemos las siguientes frecuencias genotípicas para la primera generación filial (que también suman 1):

$$AA : \frac{1}{2} \times 0,1664 + 1 \times 0,2704 + \frac{1}{4} \times 0,0256 = 0,36$$

$$AB : \frac{1}{2} \times 0,1664 + 1 \times 0,3328 + \frac{1}{2} \times 0,0256 + \frac{1}{2} \times 0,1024 = 0,48$$

$$BB : \frac{1}{4} \times 0,0256 + \frac{1}{2} \times 0,1024 + 1 \times 0,1024 = 0,16$$

Detallemos los cálculos que acabamos de hacer, por ejemplo, para el genotipo AA . Hemos calculado la probabilidad de este genotipo teniendo en cuenta que se puede obtener a partir del cruce $AA \times AB$, con probabilidad $1/2$, del cruce $AA \times AA$, con probabilidad 1 y del cruce $AB \times AB$ con probabilidad $1/4$. Denotemos estas probabilidades por $P(AA / AA \times AB)$, $P(AA / AA \times AA)$ y $P(AA / AB \times AB)$. Hemos usado también las probabilidades de estos cruces, calculadas anteriormente, que son $P(AA \times AB) = 0,1664$, $P(AA \times AA) = 0,2704$ y $P(AB \times AB) = 0,0256$. Para el resto de los cruces, la probabilidad de obtener el genotipo AA a partir de ellos es 0 y por eso no se consideran en el cálculo. Entonces, la probabilidad del genotipo AA es

$$\begin{aligned} P(AA) &= P(AA / AA \times AB) P(AA \times AB) \\ &\quad + P(AA / AA \times AA) P(AA \times AA) \\ &\quad + P(AA / AB \times AB) P(AB \times AB) \\ &= \frac{1}{2} \times 0,1664 + 1 \times 0,2704 + \frac{1}{4} \times 0,0256 = 0,36 \end{aligned}$$

De manera análoga se han calculado las probabilidades de los otros genotipos. La fórmula anterior se conoce como **Fórmula de las Probabilidades Totales** y aparece como denominador en la Fórmula de Bayes (véase *Blade Runner, el "factor humano" y la fórmula de Bayes*, no. 7 del volumen 2006 de *Materials Matemàtics*).

En general, si se conocen las probabilidades de un suceso S (por ejemplo el genotipo AA en la primera generación filial) "condicionadas a diferentes situaciones", S_1, \dots, S_r , así como las probabilidades de estas "situaciones", siempre que éstas cubran todas las posibilidades y sean mutuamente excluyentes, se puede calcular la probabilidad del suceso como suma de los

productos de las probabilidades del suceso condicionadas a cada “situación” por la probabilidad de la “situación”, esto es,

$$P(S) = P(S / S_1) P(S_1) + \cdots + P(S / S_r) P(S_r)$$

En el ejemplo que habíamos considerado, las “situaciones” son los posibles cruces de progenitores, en total $r = 6$, cuyas probabilidades suman 1 y son mutuamente excluyentes, ya que si se da un cruce concreto no se da otro. Como la probabilidad condicionada del suceso AA a algunos de estos cruces es 0, ya que no se pueden obtener hijos con este genotipo a partir de ellos, nos quedamos sólo con los cruces de los que sí puede provenir el genotipo AA en la primera generación filial: $S_1 = AA \times AB$, $S_2 = AA \times AA$ y $S_3 = AB \times AB$ (los que no necesitamos considerar son: $S_4 = AA \times BB$, $S_5 = AB \times BB$ y $S_6 = BB \times BB$).

Consideremos ahora la siguiente cuestión:

¿Qué relación hay entre las frecuencias de los alelos A y B en la generación paterna y las frecuencias genotípicas que acabamos de calcular en la primera generación filial? Si llamamos p a la frecuencia del alelo A y q a la del B , tenemos que

$$p = 0,6 \quad \text{y} \quad q = 0,4, \quad \text{en el ejemplo,}$$

y las frecuencias genotípicas son

$$AA : 0,36 (= p^2), \quad AB : 0,48 (= 2pq) \quad \text{y} \quad BB : 0,16 (= q^2)$$

Su expresión en función de p y q no es casual, siempre se cumple (cuando la población se encuentra en equilibrio genético), y se conoce por **ley de Hardy-Weinberg**. Esta ley dice que, en general, en situación de equilibrio genético, si las frecuencias de los alelos A y B en la generación de los padres son, respectivamente, p y $q = 1 - p$, las frecuencias genotípicas en la primera generación filial son:

$$AA : p^2, \quad AB : 2pq \quad \text{y} \quad BB : q^2$$

(notemos que también $p^2 + 2pq + q^2 = (p + q)^2 = 1^2 = 1$).

Como aplicación de esta ley, veamos con un ejemplo como podemos estimar las frecuencias de los homocigóticos del alelo dominante y de los heterocigóticos, indistinguibles en cuanto al fenotipo, a partir de la observación de la frecuencia de los homocigóticos del alelo recesivo, que sí se distinguen por

el fenotipo. Esto permite estimar, por ejemplo, la proporción de individuos sanos portadores de una enfermedad en una población, a pesar de que no se puedan distinguir en su fenotipo de los sanos (no portadores).

Consideremos una enfermedad o patología (no ligada al sexo) que viene controlada por un gen que tiene dos alelos. Sea B el dominante y b el recesivo. La enfermedad sólo se manifiesta en los homocigóticos bb . En el locus correspondiente al gen hay tres genotipos posibles: BB , Bb y bb , el tercero manifiesta la enfermedad y los dos primeros no, siendo los de genotipo Bb los portadores.

Aunque no es posible conocer el genotipo de todos los individuos simplemente a partir de su fenotipo, es decir, del hecho de que manifiesten o no la enfermedad, y como consecuencia no es posible conocer todas las frecuencias alélicas, si se puede suponer que la población se encuentra en equilibrio genético, se podrá usar la ley de Hardy-Weinberg para estimarlas.

Concretamente, sea p la frecuencia del alelo B y q la del alelo b (con $p + q = 1$). Por tanto, la frecuencia genotípica de los individuos enfermos será q^2 . Si sabemos que el 5 por 1000 de los individuos de la población manifiestan la enfermedad, tenemos que

$$q^2 = \frac{5}{1000} \implies q \cong 0,07071 \quad y \quad p \cong 1 - q = 0,92929$$

Una vez calculadas las frecuencias alélicas, podemos calcular la frecuencia genotípica de los individuos sanos portadores de la enfermedad (los heterocigóticos Bb): es $2pq = 2 \times 0,92929 \times 0,07071 = 0,13142$. También podemos calcular la de los sanos (no portadores, de genotipo BB): $p^2 = 0,92929^2 = 0,86358$. De esta manera, sólo observando la frecuencia de los individuos que manifiestan la enfermedad hemos podido estimar las frecuencias de los sanos y de los sanos portadores de la enfermedad.

También podemos, por ejemplo, estimar cuantos individuos enfermos nacerán aproximadamente, en cada 1000 nacimientos, si los progenitores no manifiestan la enfermedad. Para ello tenemos en cuenta que el fenotipo enfermo sólo se puede observar en la descendencia de individuos que no manifiestan la enfermedad (y con una probabilidad de $1/4$) si ambos padres son portadores. La probabilidad de que se produzca al azar un emparejamiento entre dos progenitores portadores, sabiendo que no manifiestan la enferme-

dad es: $\frac{0,13142^2}{(1 - 0,005)^2}$ ¹. Luego la proporción de individuos enfermos que nacerán de padres que no manifiestan la enfermedad en esta población será de $\frac{1}{4} \times \frac{0,13142^2}{(1 - 0,005)^2} \cong 0,00436$. Si multiplicamos por 1 000 tenemos que de cada 1 000 hijos de padres que no manifiestan la enfermedad, aproximadamente entre 4 y 5 sí la manifestarán.



Rosario Delgado
Dept. de Matemàtiques
Universitat Autònoma de Barcelona
08193 Bellaterra
delgado@mat.uab.cat

Publicat el 18 de setembre de 2006

¹Hemos usado que la probabilidad condicionada es la probabilidad de la intersección, que es la de que los dos sean sanos portadores, dividido por la probabilidad del suceso que condiciona, esto es, que los padres no manifiesten la enfermedad, que es $(1 - 0,005)^2$, pues 0,005 es la probabilidad de estar enfermo.