

CAPÍTULO 5

BASES ESTADÍSTICAS DE LA GENÉTICA FORENSE

METODOLOGÍA ESTADÍSTICA EN LAS PRUEBAS DE ADN EN LA CIENCIA FORENSE ANIMAL. IDENTIFICACIÓN INDIVIDUAL Y FILIACIONES

Juan Pedro Lirón, María Elena Fernández, Guillermo Giovambattista

5.1. Introducción

La utilización de la tipificación a nivel ADN en las ciencias forenses se ha convertido en una herramienta extremadamente útil para diferenciar un individuo de otro o establecer relaciones de parentesco, y se fundamenta en principios científicos y técnicos que son aceptados universalmente. Estas nuevas técnicas moleculares permiten el estudio de la diversidad biológica en el nivel más básico: el propio material genético, el ADN. Los principios básicos y las metodologías convencionales de la genética de poblaciones y la estadística pueden ser utilizados para interpretar los resultados de análisis de ADN en casos forenses. Debido a la aparición relativamente nueva de las técnicas de ADN, los tribunales han sometido sus fundamentos y métodos de aplicación a un minucioso estudio y control de calidad. En los comienzos de la utilización de los perfiles de ADN en la corte, se consideraba que la interpretación de los resultados poseía un cierto y aparente grado de complejidad, lo que impedía un pleno uso de sus invaluables cualidades. Sin embargo, con el paso de los años, las

metodologías experimentales y estadísticas han evolucionado notablemente, constituyendo hoy en día uno de los campos más sólidos y reconocidos en la corte como elemento de prueba.

El objetivo principal de este capítulo consiste en introducir las bases estadísticas de las pruebas de ADN y explicar cómo sus resultados pueden utilizarse para la determinación de relaciones de parentesco y la resolución de casos forenses en la ciencia animal.

Desde la introducción de las técnicas de ADN en las ciencias forenses en la década de los 80, ha habido un gran avance en las metodologías y herramientas de estudio de la variabilidad genética a nivel molecular, incrementado dramáticamente la sensibilidad y la discriminación de las técnicas de análisis. Actualmente, las técnicas moleculares más utilizadas en las ciencias forenses se basan en la aplicación de la técnica de PCR y la utilización de perfiles genéticos basados en marcadores moleculares nucleares como los microsatélites (STRs) y más recientemente en los SNPs y los marcadores uniparentales (ADNmt y cromosoma Y). Si bien los principios matemáticos y de genética de poblaciones pueden ser aplicados con ciertas modificaciones a marcadores genéticos que poseen distintos tipos de herencia y ubicación cromosómica, los modelos estadísticos desarrollados en este capítulo se aplican a loci de herencia codominante (siempre podemos inferir los dos alelos que posee un individuo para un genotipo determinado) y ubicados en cromosomas no sexuales o autosómicos (ver *Notación*).

Notación

Supongamos que un locus o gen A puede tomar varias formas alternativas o alelos, A_i . Un individuo posee dos alelos, cada uno heredado de un progenitor. Cuando los dos alelos de un individuo son iguales, por ejemplo A_iA_i decimos que el genotipo del individuo es homocigota. En cambio, si el individuo recibió diferentes alelos de sus progenitores para este locus, por ejemplo A_iA_j , decimos que es de genotipo heterocigota.

En un determinado cruzamiento, un individuo transmite a su descendencia uno de los dos alelos que posee para cada gen. La elección del locus particular que transmite es considerada aleatoria en el sentido que cada alelo posee la misma probabilidad de ser heredado. Este modelo de herencia, propuesto por Mendel en 1865, es la base genética de las pruebas de paternidad. Si una madre posee los alelos A_1A_2 y su cría posee el genotipo A_2A_3 , resulta claro que el alelo A_2 ha sido heredado de su madre y el A_3 de su padre. Entonces, todos los machos que no posean el alelo A_3 , por ejemplo padres con los genotipos A_1A_2 , A_2A_2 o A_1A_4 , pueden ser excluidos como posibles padres de la cría.

5.2. Interpretación de la evidencia

5.2.1 identificación individual: casos forenses

Los marcadores genéticos se han aplicado a la identificación individual desde el descubrimiento de los grupos sanguíneos, jugando un rol fundamental en las disputas de paternidad. Simultáneamente a la aparición de estas técnicas bioquímicas, hubo una evolución concomitante de la estadística aplicada a las ciencias genéticas forenses. A final de los años 80, los perfiles de ADN comenzaron a utilizarse para excluir una persona específica como origen de cierta evidencia biológica encontrada en la escena del delito. En estos casos,

es necesario establecer la probabilidad de hallar el perfil genético del acusado en caso de no ser éste el aportante de la evidencia. Con el paso del tiempo se ha incrementado el número de marcadores genéticos utilizados, disminuyendo notablemente la probabilidad de que dos individuos tomados al azar de la población coincidan con un mismo perfil genético y por ende justificando la hipótesis que el sospechoso es el verdadero aportante de la evidencia genética. Un ejemplo común de la aplicación de las ciencias forenses es la resolución de casos de robo de ganado, o abigeato. En estos casos, los perfiles de ADN se obtienen a partir de muestras biológicas halladas en la escena del robo o evidencia (pelos, sangre, restos de carne y otros tejidos) y de esta forma pueden ser comparados con los perfiles de ADN de muestras biológicas secuestradas a los sospechosos del delito o referencia (carne, manchas de sangre en la ropa o cuchillos). La comparación de los perfiles genéticos de la evidencia y la referencia puede arrojar tres resultados posibles:

1. Si los perfiles de ADN de la evidencia coinciden con los de la referencia en todos los loci estudiados, este hallazgo se interpreta como una "*inclusión*" o una "imposibilidad de excluir".
2. Si los dos perfiles no son consistentes en alguno de los loci, el hallazgo se puede interpretar como una no coincidencia o "*exclusión*".
3. Si no hay datos suficientes para apoyar una conclusión, el hallazgo se interpreta como "*no concluyente*".

Si ha ocurrido una coincidencia en el perfil genético de la muestra de evidencia y referencia obteniéndose una *inclusión*, este resultado puede tener dos interpretaciones posibles:

1. Los perfiles de ADN de la evidencia coinciden con los de la referencia, dado que las dos muestras *proviene del mismo individuo* (por ejemplo, los restos de tejidos secuestrados en el lugar del robo pertenecen al mismo animal que las muestras de carne secuestradas

al sospechado de cometer el ilícito). Esta es la hipótesis de la fiscalía o acusador.

2. Las muestras de la evidencia *no provienen del mismo individuo* al cual pertenecen las muestras de la referencia y por ende los perfiles genéticos de la evidencia coinciden con otro individuo de la población. Esta es la hipótesis de la defensa.

Desde sus comienzos, la identificación genética se ha basado en principios científicos, evitando que en las conclusiones hubiera opiniones de expertos basadas en la experiencia y no en la objetividad. Uno de las principales características de la presentación de la evidencia genética en la corte es que los informes no son de tipo determinista, es decir que eliminan las afirmaciones categóricas de identificación o exclusión que existen en otras disciplinas. En su lugar, los resultados de las pericias se presentan en forma de razones probabilísticas y se expresan como relaciones de verosimilitud dentro de un contexto matemático de tipo bayesiano.

Índice forense “likelihood ratio” (LR). El índice forense es una forma de cuantificar el peso de la evidencia genética. La verosimilitud LR de una hipótesis H dada la evidencia E puede expresarse como $LR(H \setminus E)$. La verosimilitud de una hipótesis (H_1) se evalúa siempre en forma relativa a otra hipótesis (H_2). Esta comparación se denomina razón de verosimilitud:

$$LR_{H_1, H_2 \setminus E} = \frac{\Pr(E \setminus H_1)}{\Pr(E \setminus H_2)} \quad (1)$$

donde $\Pr(E \setminus H_1)$ es la probabilidad condicional de la evidencia E bajo la hipótesis H_1 y $\Pr(E \setminus H_2)$ la probabilidad condicional de la evidencia E bajo la hipótesis H_2 .

En los casos de identificación individual se compara el patrón genético de los restos orgánicos de los animales faenados ilegalmente

(referencia) con las muestras biológicas secuestradas a los imputados (evidencia). Por lo tanto, en el marco de las probabilidades condicionales, el término E corresponde a los genotipos de las dos muestras biológicas analizadas (evidencia y referencia). En el caso que estos dos perfiles genéticos sean iguales, lo que corresponde a una inclusión, las hipótesis alternativas a contrastar son las siguientes:

H_f, hipótesis de la fiscalía o del acusador: la muestra secuestrada a los imputados (evidencia) perteneció al mismo individuo que los restos orgánicos correspondientes a los animales faenados ilegalmente (referencia).

H_d, hipótesis de la defensa: la muestra secuestrada a los imputados no se corresponde con el animal faenado y por ende corresponde a otro animal de la población.

A partir de aquí, introduciremos alguna terminología adicional y denotaremos al genotipo de la evidencia como **G_e** y al genotipo de la muestra de referencia como **G_r**. Nótese que en el caso que estamos tratando $G_e = G_r$. Además, designaremos como **I** a cualquier otro tipo de información o evidencia no genética que eventualmente posea la corte en relación a las hipótesis de la defensa y la acusación. Previo a la tipificación de ADN, la probabilidad de H_f estaba condicionada por I: $\Pr(H_f/I)$. Luego de la evidencia genética, la probabilidad H_f pasa a estar condicionada además por G_e y G_r: $\Pr(H_f/G_e, G_r, I)$. De la misma manera, aplicamos el mismo razonamiento con H_d: $\Pr(H_d/G_e, G_r, I)$. Claramente, H_d y H_f son sucesos exhaustivos y mutuamente excluyentes. Esta propiedad establece uno de los principios fundamentales de la interpretación de la evidencia, que expresa que para poder establecer la certeza de una proposición es necesario considerar al menos una proposición alternativa. En principio podemos considerar sólo la información de origen no genético y expresar la relación entre las dos hipótesis como una razón de probabilidades (*odds* en inglés) *a priori* en favor de la fiscalía:

$$\frac{\text{Pr } H_f / I}{\text{Pr } H_d / I} \quad (2)$$

La forma de incorporar la razón de probabilidades *a priori* a favor de la fiscalía es ordenando la ecuación con la hipótesis H_f en el numerador. De este modo expresamos cuántas veces es más probable que el acusado sea culpable que inocente. Invirtiendo la fórmula (LR en favor de la defensa) expresaríamos cuánto más probable es que el acusado sea inocente que culpable.

A continuación, incorporamos la evidencia genética y lo expresamos como una razón de probabilidades *a posteriori* en favor de la fiscalía:

$$\frac{\text{Pr } H_f / G_r, G_e, I}{\text{Pr } H_d / G_r, G_e, I} \quad (3)$$

Aplicando el teorema de Bayes a la expresión anterior y reemplazando (G_e, G_r) por E , llegamos a la siguiente ecuación:

$$\frac{\text{Pr } H_f / E, I}{\text{Pr } H_d / E, I} = \frac{\text{Pr } E / H_f, I}{\text{Pr } E / H_d, I} * \frac{\text{Pr } H_f / I}{\text{Pr } H_d / I} \quad (4)$$

$$\text{razón a posteriori} = \text{LR} * \text{razón a priori}$$

Esta ecuación es de central importancia en la interpretación de genética forense ya que permite distinguir el rol de la corte – establecido en la parte izquierda de la igualdad – del rol del científico forense, que corresponde al primer término del lado derecho de la igualdad. Mientras que la corte evalúa la probabilidad de culpabilidad o inocencia del acusado en relación con las evidencias, el científico aborda otro tipo de problemática diferente: “cuál es la probabilidad de la evidencia si la hipótesis de la fiscalía fuese cierta” y “cuál es la probabilidad de la evidencia si la hipótesis de la defensa fuese cierta”.

La ecuación 4 muestra que la probabilidad *a posteriori* viene determinada por la razón *a priori* multiplicada por un cociente de dos probabilidades o razón de verosimilitud LR (*likelihood ratio*):

$$LR = \frac{\Pr G_e, G_r / H_f, I}{\Pr G_e, G_r / H_d, I} \quad (5)$$

expandiendo la ecuación anterior:

$$LR = \frac{\Pr G_e / G_r, H_f, I}{\Pr G_e / G_r, H_d, I} * \frac{\Pr G_r / H_f, I}{\Pr G_r / H_d, I} \quad (6)$$

Los términos $\Pr G_r / H_f, I$ y $\Pr G_r / H_d, I$ denotan la probabilidad de la muestra de referencia secuestrada al sospechoso de hurto dado que pertenezca o no, respectivamente, al mismo animal que la evidencia tomada en el lugar del robo. Es importante observar que estos términos son equivalentes y se simplifican en la ecuación porque la probabilidad del genotipo G_r no varía entre las dos hipótesis H_d y H_f , sea la muestra de referencia y evidencia originadas del mismo animal o no.

El índice forense (Evetts y Weir, 1998) se estima entonces como:

$$LR = \frac{\Pr G_e \setminus G_r, H_f, I}{\Pr G_e \setminus G_r, H_d, I} \quad (7)$$

si se considera primero el numerador, en donde las muestras de referencia y evidencia pertenecen al mismo individuo y no existen errores en la tipificación de los marcadores genéticos. La probabilidad según la hipótesis H_f de que G_e sea igual a G_r , dado que pertenecen al mismo individuo, es igual a 1. Recordemos que estamos considerando un caso de inclusión genética. Entonces, el índice LR se transforma en:

$$LR = \frac{1}{Pr G_e | G_r, H_d} \quad (8)$$

Nos queda por último calcular la probabilidad del denominador. Para asignar dicho valor, se debe responder la siguiente pregunta: ¿Cuál es la probabilidad de observar el genotipo de la evidencia G_e igual al de la referencia G_r , si la muestra de la evidencia pertenece a otro individuo diferente del que aportó la muestra de referencia? Si se considera a “otro individuo” como un animal tomado al azar de la población de origen de la muestra hallada en el lugar del hecho, la probabilidad de que la evidencia y la referencia posean idéntico genotipo depende de las frecuencias génicas de dicha población (ver Anexos 5.5.1 y 5.5.2). De esta forma, el índice LR se calcula como:

$$LR = \frac{1}{P} \quad (9)$$

donde P es la frecuencia esperada del genotipo en la población y se calcula a partir de bases de datos de individuos tipificados para los mismos marcadores moleculares que los utilizados para construir los perfiles genéticos del caso. Supongamos que podemos calcular la frecuencia P del genotipo G en la población de referencia, a la cual pertenece la muestra secuestrada al imputado, considerando que no pertenece al mismo animal que se posee como evidencia $Pr G_r | G_e, H_d$. Asignamos de esta manera la probabilidad P al denominador de la ecuación 9. Por ejemplo, si la frecuencia de P en la población es 0,0001, el LR se calcula como $LR = 1/0,0001 = 100.000$. Esta evidencia puede ser presentada de la siguiente forma: “La evidencia es diez mil veces más probable si la muestra de referencia perteneció al mismo individuo faenado en el lugar del hurto que a algún otro animal desconocido”.

5.2.2 Pruebas de parentesco. Índice de paternidad

5.2.2.1 Paternidad en tríos: madre, cría y padre alegado

Para evaluar la evidencia en casos de paternidades disputada, se utiliza un marco teórico similar al empleado en los casos de identificación individual. El índice de paternidad (IP) es un término utilizado en las pruebas de parentesco para referirse a la razón de verosimilitud (LR). En un típico caso de paternidad discutida donde se posee el perfil genético de la madre, la cría y el padre putativo o alegado, utilizaremos la siguiente nomenclatura:

M: madre de la cría y G_M su genotipo.

C: cría de la madre *M* y G_C su genotipo.

P: padre alegado y G_P su genotipo.

Los genotipos G_M , G_C y G_P constituyen la evidencia *E*. A su vez, denominaremos a las dos hipótesis de contraste en el cálculo de LR como:

H_p : el padre alegado es el padre de la cría.

H_d : algún otro individuo y no el padre alegado, es el padre de la cría.

Siguiendo el marco bayesiano utilizado en los casos de identificación individual, podemos interpretar la evidencia de la siguiente manera:

$$\frac{\Pr H_p / E, I}{\Pr H_d / E, I} = \frac{\Pr E / H_p, I}{\Pr E / H_d, I} * \frac{\Pr H_p / I}{\Pr H_d / I} \quad (10)$$

razón a posteriori = LR * razón a priori

De forma análoga a la ecuación 6 de la sección anterior, supondremos que no existe otra información *I* que la evidencia genética *E* y expandiendo LR obtenemos:

$$LR = \frac{\Pr G_C, G_M, G_P / H_p}{\Pr G_C, G_M, G_P / H_d}$$

$$LR = \frac{\Pr G_c / G_m, G_p, H_p}{\Pr G_c / G_m, G_p, H_d} * \frac{\Pr G_m, G_p / H_p}{\Pr G_m, G_p / H_d} \quad (11)$$

Tanto H_p como H_d no cambian el valor de G_m y G_p del segundo término, por lo tanto este cociente es igual a 1. Entonces:

$$LR = \frac{\Pr G_c / G_m, G_p, H_p}{\Pr G_c / G_m, G_p, H_d} \quad (12)$$

El cálculo del índice de paternidad proviene de evaluar el numerador y denominador de esta ecuación para las diferentes combinaciones de alelos que puedan tener los genotipos del trío: madre, hijo y padre alegado. El numerador expresa la probabilidad del genotipo de la cría dado los genotipos de la madre y padre alegado, y puede ser evaluado mediante un cuadrado de Punnett. Como poseemos el genotipo de la madre, es posible determinar el alelo materno A_m y el paterno A_p que posee la cría. Incorporando esto en el denominador de la ecuación anterior, tenemos:

$$\frac{\Pr G_c / G_m, G_p, H_d}{\Pr A_m, A_p / G_m, G_p, H_d} = \frac{\Pr G_c / G_m, G_p, H_d}{\Pr A_m / G_m, G_p, H_d * \Pr A_p / G_m, G_p, H_d} \quad (13)$$

Asumiendo que $\Pr(A_m/G_m, G_p, H_p) = \Pr(A_m/G_m)$, el denominador de LR se convierte en:

$$\Pr A_m / G_m * \Pr A_p / G_m, G_p, H_d \quad (14)$$

Es fácil notar que la probabilidad $\Pr(A_m/G_m)$ de que la cría reciba un alelo materno es 1 o $\frac{1}{2}$ si la madre es homocigota o heterocigota, respectivamente. Generalmente esta probabilidad se denomina factor mendeliano materno, M_M . La segunda parte del denominador considera

la probabilidad del alelo paterno A_P considerando que no es heredado del padre alegado, sino de otro padre. Este término de la ecuación se denomina factor mendeliano paterno, M_P . Para calcularlo, necesitamos conocer las frecuencias génicas de este alelo en la población de los posibles padres de la cría. A continuación ejemplificaremos el cálculo del índice de paternidad para algunas combinaciones de genotipos de los padres y la cría:

Ejemplo 1.

Consideremos los siguientes genotipos del trío: $G_M=cd$, $G_P=ab$ y $G_C=ac$. Para el cálculo del numerador la ecuación 12, $Pr(G_C/G_M, G_P, H_P)$, tenemos que los cuatro posibles genotipos de la cría vienen del siguiente cuadro.

		Alelos del padre	
		a	b
Alelos de la madre	c	ac	bc
	d	ad	bd

Los cuatro genotipos posibles de la cría son equiprobables y por lo tanto la probabilidad de que la cría tenga el genotipo ac es de $\frac{1}{4}$. Ahora resta calcular el denominador. Como la madre es heterocigota, la primera parte del denominador o factor mendeliano materno, $Pr(A_m/G_m)$, es $\frac{1}{2}$. El alelo paterno es $A_p = a$, entonces $Pr(A_p/G_M, G_P, H_d)$ es la frecuencia del alelo a en la población, que denominaremos p_a . De esta forma, obtenemos el índice de paternidad para este ejemplo:

$$IP = \frac{\frac{1}{4}}{\frac{1}{2} p_a} = \frac{1}{2 p_a}$$

Ejemplo 2.

Consideremos ahora los siguientes genotipos del trío: $G_M=cc$, $G_P=ab$ y $G_C=ac$. Para el cálculo del numerador $Pr(G_C/G_M, G_P, H_P)$, tenemos dos posibles genotipos de la cría que provienen del siguiente cuadro.

		Alelos del padre	
		a	B
Alelos de la madre	c	ac	Bc
	c	ac	Bc

Los 2 genotipos posibles de la cría son equiprobables y por lo tanto la probabilidad del genotipo ac es $\frac{1}{2}$. Como la madre es homocigota, el factor mendeliano materno, $\Pr(A_m/G_m)$, es 1. El alelo paterno A_p es a, entonces $\Pr(A_p/G_M, G_P, H_d)$ es la frecuencia del alelo a en la población, p_a . El índice de paternidad para este ejemplo viene expresado entonces como:

$$IP = \frac{\frac{1}{2}}{1 p_a} = \frac{1}{2 p_a}$$

Ejemplo 3:

Consideremos ahora los siguientes genotipos del trío: $G_M=ab$, $G_P=bc$ y $G_C=ab$. Para el cálculo del numerador, los cuatro genotipos posibles de la cría se obtienen a partir del siguiente cuadro.

		Genes del padre		
		b	C	
Genes de la madre	a	ab	Ac	
	b	bb	Bc	

Los cuatro genotipos posibles de la cría son equiprobables y por lo tanto la probabilidad de tener el genotipo ac es de $\frac{1}{4}$. Como la madre es ab y el hijo ab, puede haberle pasado cualquiera de sus dos alelos con probabilidad de $\frac{1}{2}$. Dependiendo de qué alelo pasó la madre, el alelo paterno A_p podrá ser a con probabilidad p_a o b con probabilidad p_b . De este modo, el denominador quedará como $(\frac{1}{2} p_a + \frac{1}{2} p_b)$. Finalmente, el índice de paternidad para este ejemplo equivale a:

$$IP = \frac{\frac{1}{4}}{\frac{1}{2} p_a + \frac{1}{2} p_b} = \frac{1}{2(p_a + p_b)}$$

Como se ha mencionado anteriormente, el índice de paternidad puede variar según los alelos del trío madre, cría y padre alegado. Habitualmente, en una prueba de parentesco se tipifican decenas de marcadores genéticos y por ende se realiza un igual número de cálculos de IP. Es conveniente en estos casos calcular el IP para todas las combinaciones de genotipos de la cría, madre y padre alegado, como se muestra en la Tabla 5.1.

Genotipo materno	Genotipo de la cría	Genotipo del padre alegado	IP
Aa	Aa		
Ab		aa	$\frac{1}{p_a}$
Bb	Ab		
Bc			
Aa			
Ab	aa		
Ac		ab	$\frac{1}{2p_a}$
Bb	Ab		
Bc			
Bc			
Cc	ac		
Cd			
		aa	$\frac{1}{p_a}$
Ab	ab	ab	$\frac{p_a p_b}{2(p_a + p_b)}$
		ac	$\frac{1}{2(p_a + p_b)}$

Tabla 5.1. Cálculo del índice de paternidad para las combinaciones posibles de tipo inclusión de genotipos maternos, paternos y de la cría.

5.2.2.2 Paternidad en dúos: cría y padre alegado

En la práctica veterinaria, en los casos de paternidades discutidas, es muy común no poseer la información genética de la madre. Por ejemplo, en la comparación de un toro posible padre de una o varias crías. En este caso, la evidencia consta sólo del genotipo del padre putativo G_p y el genotipo de la cría G_c . Denominaremos a las dos hipótesis de contraste en el cálculo de IP como:

H_p : el padre alegado es el padre de la cría.

H_d : algún otro individuo y no el padre alegado, es el padre de la cría.

Siguiendo el marco bayesiano utilizado anteriormente, podemos interpretar la evidencia de la siguiente manera:

$$LR = \frac{\Pr G_c / G_p, H_p}{\Pr G_c / G_p, H_d} \quad (15)$$

La probabilidad del genotipo de la cría en el denominador es independiente del genotipo del padre alegado, dado que según H_d éste último no es el verdadero padre. Por lo tanto, la ecuación se reduce a:

$$LR = \frac{\Pr G_c / G_p, H_p}{\Pr G_c / H_d} \quad (16)$$

A continuación ejemplificamos el cálculo del índice de paternidad para distintas combinaciones de genotipos del padre putativo y la cría.

Ejemplo 1.

Consideremos los siguientes genotipos del dúo: $G_p=ac$ y $G_c=ab$. Para calcular el numerador $\Pr(G_c=ab/G_p=ac, H_p)$ asumimos que el padre alegado es el verdadero padre de la cría. Como no poseemos la información de la madre, no podemos calcularlo mediante un cuadro de Punnett, como hicimos para calcular el IP en los tríos. En nuestro ejemplo el alelo a de la cría proviene del padre con probabilidad $\frac{1}{2}$, ya que el mismo es heterocigota. El alelo b es heredado de la madre y, como no poseemos la información del genotipo materno, suponemos que deriva de una madre desconocida de la población. De esta forma, le asignamos a este alelo b una probabilidad equivalente a su frecuencia génica poblacional p_b , siendo el numerador $\frac{1}{2} * p_b$. Para calcular el denominador $\Pr(G_c=ab/ H_d)$, debemos determinar la probabilidad de que el genotipo de la cría $G_c=ab$ provenga de la población. Considerando

que conocemos las frecuencias génicas poblacionales, de donde proviene el padre verdadero según H_d y asumiendo que existe equilibrio de Hardy-Weinberg (ver Anexo 5.5.3), calculamos el denominador $\Pr(G_c=ab/ H_d) = 2 \cdot p_a \cdot p_b$. Finalmente, el índice de paternidad equivale a:

$$IP = \frac{\frac{1}{2} p_b}{2 p_a \cdot p_b} = \frac{1}{4 p_a}$$

En la Tabla 5.2 se calcula el IP para todas las combinaciones posibles de genotipos de la cría y padre alegado.

Genotipo materno	Genotipo de la cría	IP
aa	Aa	$\frac{1}{p_a}$
aa	Ab	$\frac{1}{2 p_a}$
ab	Aa	$\frac{p_a \cdot p_b}{4 p_a p_b}$
ab	Ab	$\frac{1}{4 p_a}$

Tabla 5.2. Cálculo del índice de paternidad para las combinaciones posibles de tipo inclusión de genotipos paternos y de la cría.

Cuadro 5.1. Ejemplo del cálculo de las frecuencias genotípicas según el principio de Hardy-Weinberg (H-W).

Proporciones genotípicas esperadas según el principio de Hardy-Weinberg para un locus con dos alelos		Frecuencias alélicas para el esperma	
		A ₁ (p ₁)	A ₂ (p ₂)
Frecuencias alélicas para los óvulos	A ₁ (p ₁)	A ₁ A ₁ (p ₁ p ₁)	A ₁ A ₂ (p ₁ p ₂)
	A ₂ (p ₂)	A ₁ A ₂ (p ₂ p ₁)	A ₂ A ₂ (p ₂ p ₂)

Las proporciones genotípicas que se originan a partir de las frecuencias génicas de la población según el principio de H-W pueden expresarse de forma simbólica como:

Homocigotas: A_iA_i; P_{ii} = p_i²,

Heterocigotas: A_iA_j; P_{ij} = 2p_ip_j, i ≠ j.

En el caso de que una población no posea sus proporciones genotípicas de acuerdo al principio de H-W, sólo basta una generación de apareamiento al azar para alcanzarlo, pues en cada generación las proporciones genotípicas sólo dependen de las frecuencias génicas de la población. Como mencionamos anteriormente, las ecuaciones utilizadas a lo largo de este capítulo asumen que las poblaciones se encuentran en equilibrio de H-W. Sin embargo, existen modificaciones de estas fórmulas generales para ajustar los valores en situaciones donde existen desviaciones de los valores esperados para tal equilibrio. Uno de los ajustes más utilizados en las ciencias forenses es el que contempla la existencia de sub-estructuración poblacional.

5.3 PODER DE EXCLUSIÓN

La disponibilidad de varias técnicas de análisis de la variabilidad genética como los polimorfismos bioquímicos, los grupos sanguíneos y especialmente los marcadores de ADN, ha permitido su aplicación en la identificación individual y en el control de filiación. Una forma de cuantificar la eficacia de un conjunto particular de marcadores genéticos es calculando la probabilidad de exclusión (PE) *a priori*, la que se define como la probabilidad de que un sistema genético específico confiera evidencias que conduzcan a la exclusión de un determinada relación genética (ejemplo, la paternidad discutida de un padre, la correspondencia entre la evidencia y la referencia en casos forenses, etc.). Este índice se evalúa en el contexto de diferentes escenarios que se presentan habitualmente en los casos de paternidad e identificación individual. En las situaciones analizadas en este capítulo se asume que los marcadores genéticos utilizados poseen herencia de tipo codominante y están en equilibrio de ligamiento y Hardy-Weinberg (ver Anexos 5.5.3 y 5.5.4).

5.3.1 Poder de exclusión en casos de paternidad

En la sección anterior se ha definido que si un padre alegado es compatible con todos los marcadores genéticos de la cría, obtenemos una *inclusión genética*. Además, notamos que este resultado puede tener dos interpretaciones posibles: i) H_p : el padre alegado es el padre de la cría. ii) H_d : algún otro individuo de la población y no el padre alegado, es el padre de la cría. De esto surge que la efectividad de un marcador genético para resolver disputas de paternidad puede ser determinada calculando su habilidad de excluir padres falsos. Cuanto más pequeña es la probabilidad del denominador de la ecuación de verosimilitud, que en los casos de parentesco denominamos Índice de Paternidad (IP), mayor peso poseerá la evidencia.

5.3.1.1 Poder de exclusión de paternidades en tríos: $PE(p/m,c)$

Podemos calcular el poder de exclusión en los casos de paternidad considerando todas las combinaciones de genotipos maternos y de la cría y verificando en cada caso qué genotipos paternos pueden excluir la relación de paternidad. De esta manera, la sumatoria del valor de probabilidad de todas las posibles exclusiones constituye el poder de exclusión de este sistema genético. Como ejemplo, consideremos un marcador bialélico donde una madre posee el genotipo A_1A_1 y su cría A_1A_2 . Con esta información podemos determinar que el alelo materno de la cría es A_1 y el paterno A_2 . Por ende, todos los padres que no posean A_2 quedan excluidos de la paternidad. Como se trata de un locus con dos alelos, el único padre excluido puede ser el de genotipo A_1A_1 . La probabilidad de exclusión de este caso particular se calcula como:

$$\Pr(G_m=A_1A_1) * \Pr(G_c=A_1A_2/G_m=A_1A_1) * \Pr(G_p=A_1A_1) \\ p_1^2 * p_2 * p_1^2 = p_1^4 * p_2$$

En la Tabla 5.3 se muestran todas las combinaciones posibles madre, cría y padre excluido para un locus multialélico.

Madre		Cría		Padre excluido		
Tipo	Pr G_m	tipo	Pr G_c/G_m	Tipo		Pr G_p
A_iA_i	p_i^2	$A_i A_i$	p_i	$A_w A_x,$	$w, x \neq i$	$(1-p_i)^2$
		$A_i A_j$	p_j	$A_w A_x,$	$w, x \neq j$	$(1-p_j)^2$
A_iA_j $j \neq i$	$2p_i p_j$	$A_i A_i$	$p_i/2$	$A_w A_x,$	$w, x \neq i$	$(1-p_i)^2$
		$A_j A_j$	$p_j/2$	$A_w A_x,$	$w, x \neq j$	$(1-p_j)^2$
		$A_i A_j$	$(p_i + p_j)/2$	$A_w A_x,$	$w, x \neq i, j$	$(1-p_i - p_j)^2$
		$A_i A_k$	$p_k / 2$	$A_w A_x,$	$w, x \neq k$	$(1-p_k)^2$
		$A_i A_i$	$p_k/2$	$A_w A_x,$	$w, x \neq k$	$(1-p_k)^2$

Tabla 5.3. Exclusiones de paternidad para un locus multialélico.

El poder de exclusión de este marcador genético consiste en la sumatoria del producto de las tres columnas de cada fila de la tabla. Este cálculo del poder de exclusión individual PE_i puede expresarse mediante la siguiente ecuación (Jamieson y Taylor, 1994):

$$PE_i(p/m,c) = 1 - 2 \sum_{i=1}^k p_i^2 \sum_{i=1}^k p_i^3 - 2 \sum_{i=1}^k p_i^4 - 3 \sum_{i=1}^k p_i^5 - 2 \left(\sum_{i=1}^k p_i^2 \right)^2 - 3 \sum_{i=1}^k p_i^2 \sum_{i=1}^k p_i^3 \quad (17)$$

donde p_i corresponde a la frecuencia génica del i -th alelo y p_j la frecuencia génica del j -th alelo, siendo i diferente a j .

5.3.1.2 Poder de exclusión en dúos: $PE(p/c)$

El poder de exclusión en casos de paternidad cuando no se conoce el genotipo de uno de los progenitores (generalmente el padre) se calcula de forma similar al caso de paternidad en tríos. En este caso no existe el genotipo materno y la probabilidad de la cría, por ende, no es condicional sobre la madre. Como ejemplo, si consideramos un marcador bialélico donde la cría posee el genotipo A_1A_1 , los únicos padres excluidos en este son de genotipo A_2A_2 . La probabilidad de exclusión de este caso particular se calcula como:

$$\Pr(G_c=A_1A_1) * \Pr(G_p=A_2A_2) \\ p_1^2 * p_2^2 = p_1^2 * p_2^2$$

La probabilidad de exclusión de un marcador genético para todas las combinaciones de crías y padres excluidos puede determinarse con la siguiente ecuación (Jamieson y Taylor, 1997):

$$PE_i(p/c) = 1 - 4 \sum_{i=1}^k p_i^2 \sum_{i=1}^k 2p_i^2 - 4 \sum_{i=1}^k p_i^3 - 3 \sum_{i=1}^k p_i^4 \quad (19)$$

donde p_i corresponde a la frecuencia génica del i -th alelo y k es el número de alelos

5.3.1.3 Poder de exclusión en casos de cambio de progenie: PE (c/p,m)

En este caso se desea determina el poder de exclusión de una relación de filiación entre una cría y sus dos padres putativos (por ejemplo, sustitución de progenie). En esta situación la probabilidad de exclusión (Jamieson y Taylor, 1997) se estima con la siguiente fórmula:

$$PE(c / m, p) = 1 - \frac{4 \prod_{i=1}^k p_i^2}{3 \prod_{i=1}^k p_i^2 + 2 \prod_{i=1}^k p_i^2} \quad (20)$$

donde p_i corresponde a la frecuencia génica del i -th alelo y k es el número de alelos.

5.3.2 Poder de exclusión en casos de identificación individual: PE (e/r)

El poder de exclusión de un locus cuando se comparan dos genotipos (evidencia y referencia), también denominado poder de discriminación, se obtiene calculando el complemento de la probabilidad de que dos individuos tomados al azar de la población presenten genotipos idénticos (*matching*). El PE en este caso se estima mediante la siguiente fórmula:

$$PE(e/r) = 1 - \frac{\sum_{i=1}^k p_i^4 + 4 \sum_{j=1}^k p_j^2 \sum_{i=1}^k p_i^2}{4 \sum_{i=1}^k p_i^2 \sum_{j=1}^k p_j^2} \quad (21)$$

donde p_i y p_j corresponden a la frecuencia génica de los i -th y j -th alelos y k al número total de alelos. El primer término corresponde a todos los posibles homocigotos, y el segundo a las diferentes combinaciones posibles de genotipos heterocigotas (sin tener en cuenta el orden).

5.3.3 Poder de exclusión combinado simple y doble

La probabilidad de exclusión depende de las frecuencias génicas y del número de alelos. Si PE_i es la probabilidad de exclusión de un locus i , la probabilidad de exclusión combinada simple PEC_s de varios loci se determina calculando la probabilidad de que al menos un locus del sistema conduzca a la exclusión. O, en otras palabras, PEC_s es igual al complemento de la probabilidad de que todos los marcadores genéticos incluyan. De esta manera, calculamos el PEC_s como (Boyd, 1954; Weir, 1990):

$$PEC_s = 1 - \prod_{i=1}^k PE_i \quad (22)$$

donde PE_i corresponde a la probabilidad de exclusión del i -th locus y k al número de loci.

Para reducir el efecto de los errores de tipificación genética, de las mutaciones y de la presencia de alelos nulos se ha sugerido la aplicación del criterio de exclusión en más de un locus (Chakraborty *et al.*, 1974, Pemberton *et al.*, 1995). La probabilidad de exclusión combinada de dos o más loci (PEC_d) puede ser calculada como el complemento de la probabilidad que ocurra ninguna o solo una exclusión (Chakraborty *et al.*, 1974):

$$PEC_d = 1 - \prod_{i=1}^k (1 - PE_i) \prod_{j=1}^k (1 - PE_{ij}) \quad (23)$$

donde PE_i corresponde a la probabilidad de exclusión del i -th locus, k al número de loci, y PE_{ij} al poder de exclusión individual de los j -th loci diferentes del i -th locus considerado en cada término del producto.

5.3.4 Poder de exclusión para marcadores uniparentales

Las ecuaciones descritas en los apartados anteriores se refieren a marcadores autosómicos que poseen una herencia mendeliana

codominante. Como se mencionó anteriormente, cada individuo posee dos alelos para cada locus, uno heredado de cada progenitor. Sin embargo, en genética forense es un hecho común el uso de marcadores uniparentales, como el ADNmt (linaje materno) y los marcadores del cromosoma Y (linaje paterno). Estos loci, a diferencia de los anteriores, se caracterizan por poseer una sola copia o haplotipo por individuo heredada de uno de los progenitores. Es por esta razón que en los casos de identificación genética mediante marcadores uniparentales, el poder de exclusión se calcula como:

$$PE(e/r) = 1 - \prod_{i=1}^k p_i^2 \quad (24)$$

donde p_i es la frecuencia del haplotipo mitocondrial o del cromosoma Y en la población, siendo el término $\prod_{i=1}^k p_i^2$ la probabilidad de coincidencia de dos muestras tomadas al azar de la población (*random match probability*; Angleby y Savolainen, 2005).

5.5. ANEXOS

Anexo 5.1. Principios de genética de poblaciones

En este capítulo se introducen las bases estadísticas de los cálculos, tanto de los índices forenses como del poder de exclusión en los casos más frecuentes de paternidades e identificación individual. En las ecuaciones descritas para los cálculos forenses, considerábamos la hipótesis de que la evidencia y la referencia, a pesar de poseer idéntico genotipo, no eran parte del mismo animal y por lo tanto la evidencia provenía de otro individuo de la población. Asimismo, en los cálculos de los índices de paternidad, la hipótesis alternativa consideraba que el alelo paterno de la cría no provenía del padre dubitado sino de otro individuo de la población. En todos estos casos, es necesario entonces

estimar de alguna manera la probabilidad de que estos alelos o genotipos pertenezcan a la población general. La estimación de estas probabilidades constituye una parte muy importante de la genética forense y se funda en principios establecidos dentro del marco conceptual de la Genética de Poblaciones.

Para obtener la probabilidad de que un individuo seleccionado al azar de la población posea el mismo perfil genético que la evidencia de ADN, necesitamos conocer la frecuencia de dicho genotipo en la población. Estos valores generalmente se obtienen a partir de una muestra de la población total, que suele denominarse base de datos de ADN. En la práctica de rutina de los laboratorios forenses, para determinar el perfil genético de las evidencias se utilizan decenas de marcadores moleculares. Por lo tanto, para una determinada base de datos, existen miles de millones de genotipos posibles para este conjunto de marcadores utilizados, y debido a esta limitación debemos utilizar forzosamente las frecuencias génicas de cada alelo para estimar la frecuencia esperada de un cierto perfil de ADN. A consecuencia de lo expuesto, para realizar dichas inferencias estadísticas, necesitamos establecer ciertas asunciones acerca de la estructura genética de la población, y ahí es donde la Genética de Poblaciones juega un papel clave.

Anexo 5.5.2. Frecuencias alélicas y genotípicas

Es usual en genética designar a un gen con una letra y a cada alelo con un subíndice numérico. Por ejemplo A_6 se refiere al sexto alelo del locus A, y B_{12} al doceavo alelo del locus B. Cuando queremos aplicar alguna regla general o fórmula a un locus determinado, generalmente utilizamos letras como subíndices, por ejemplo i o j . Además, al referirnos a la frecuencia o proporción de un alelo determinado utilizamos la letra p con un subíndice que indica el nombre del alelo. Por ejemplo, denotamos particularmente que el alelo A_5 se encuentra en la población en una proporción del 20% como $p_5=0,2$ y generalmente que

cierto locus A_i posee una frecuencia p_i . Es importante notar que la sumatoria de las frecuencias alélicas para todos los alelos i de un locus es igual a uno, es decir .

Anexo 5.5.3. Apareamiento al azar y proporciones en equilibrio de Hardy-Weinberg

En general y de manera simplificada, en los cálculos forenses se suele asumir que los individuos de una población se reproducen en forma aleatoria. Sin embargo, es evidente que los individuos de una especie determinada no se aparean aleatoriamente. Por ejemplo, una vaca de la raza Holstein en Gran Bretaña no tiene la misma oportunidad de aparearse con un macho de la raza Holando de la provincia de Santa Fé, Argentina, que con un toro de su mismo establecimiento lechero. Sin embargo, los apareamientos no se realizan de acuerdo a los genotipos que poseen los individuos para los marcadores utilizados en los análisis forenses. En este caso, y a pesar de la separación geográfica de estos animales, si las frecuencias alélicas locales no se alejan de las halladas para la raza Holstein a nivel mundial, las proporciones genotípicas de ambas poblaciones pueden ser similares.

Las frecuencias genotípicas esperadas en una población con apareamiento al azar se denominan proporciones de Hardy-Weinberg, y surgen intuitivamente de suponer que estas proporciones son el resultado de la combinación aleatoria de las gametas que provienen de sus progenitores. Por ejemplo, pensemos en un locus que posee dos alelos A_1 y A_2 con frecuencias p_1 y p_2 , respectivamente. Los posibles genotipos para este locus surgen de la combinación aleatoria de sus alelos: A_1A_1 con frecuencia esperada $p_1 \cdot p_1 = p_1^2$; A_2A_2 con frecuencia esperada $p_2 \cdot p_2 = p_2^2$ y A_1A_2 con frecuencia esperada $p_1 \cdot p_2 + p_2 \cdot p_1 = 2 \cdot p_1 \cdot p_2$.

Anexo 5.5.4. Múltiples loci y equilibrio de ligamiento

Una población con apareamiento al azar (y en ausencia de otras fuerzas evolutivas como selección y migración) durante un cierto número de generaciones consigue un estado de equilibrio, en el cual la frecuencia de un perfil genético para varios loci (genotipo multiloci) es equivalente al producto de las frecuencias genotípicas de sus loci individuales. Cuando la población arriba a dicho estado, decimos que se encuentra en Equilibrio de Ligamiento (LE). El concepto de equilibrio de ligamiento es utilizado frecuentemente en la práctica cuando se calcula el índice forense de un conjunto de loci. El procedimiento de cálculo de este índice se conoce como ley del producto, pues consiste simplemente en multiplicar el valor del índice forense obtenido para cada loci individual.

Una diferencia importante entre el equilibrio Hardy-Weinberg y el de ligamiento es que este último no se alcanza en una sola generación de apareamiento al azar, sino en forma gradual. Para un par de loci no ligados (que no se encuentran en el mismo par de cromosomas homólogos) el desequilibrio de ligamiento (LD) se reduce a la mitad en sucesivas generaciones. Es importante notar que, aunque de una forma más gradual, los loci que se encuentran en un mismo cromosoma pueden alcanzar el estado de equilibrio. Sin embargo, la mayor parte de las aplicaciones forenses utilizan marcadores genéticos que se encuentran en distintos cromosomas homólogos y que se encuentran en LE. Otro punto a tener en cuenta es que la subestructuración poblacional actúa aumentando el LD entre algunos loci y de forma inversa reduciendo el desequilibrio entre otros marcadores. De esta forma, la ley del producto puede considerarse como una forma conservativa de calcular el índice forense para un conjunto de loci.

5.6. Referencias Bibliográficas

1. Chakraborty R., Shaw M., Schull W.J. (1974) Exclusion of paternity: the current state of the art. *Am J Hum Genet* 26:477–488
2. Evett I.W., Weir B.S. (1998) *Interpreting DNA Evidence* Sinauer Associates. ISBN 0-87983-155-4
3. Jamieson A. (1994) The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. In: *Festschrift in Honour of Dr Clyde J. Stormont*. *Animal Genetics* 25 (Supplement 1), 37–44.
4. Jamieson A. & Taylor S.C.S. (1997) Comparison of three probability formulae for parentage exclusion. *Animal Genetics* 28, 397–400.
5. Pemberton J.M., Slate J., Bancroft D.R., Barrett J.A. (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol Ecol* 4:249–252
6. Weir B.S. (1990) *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.
7. Angleby H, Savolainen P. Forensic informativity of domestic dog mtDNA control region sequences. *Forensic Sci Int* 2005;154:99–110.