

Curso de Evolución 2019  
Facultad de Ciencias  
Universidad de la República

Guía de estudios de genética de poblaciones

v. 2009.1

## TABLA DE CONTENIDOS Y MATERIALES (Setiembre de 2019)

### 1. Introducción

Presentación ¿Qué es la microevolución? Variación genética Marco histórico de referencia  
Lecturas recomendadas

### 2. Equilibrio Hardy-Weinberg

El modelo poblacional de Hardy-Weinberg Consecuencias del modelo Hardy-Weinberg  
Homocigosidad y heterocigosidad esperadas Aproximación al equilibrio de Hardy-Weinberg  
Lecturas recomendadas

### 3. Deriva genética

La deriva genética Fluctuaciones de las frecuencias alélicas Formalización del modelo de  
Wright-Fisher La distribución binomial Deriva genética y heterocigosidad "Ritmo" de la deriva  
genética y tamaño poblacional efectivo Lecturas recomendadas

### 4. Mutación

Sesgo mutacional: diferencias entre los sexos Sesgos mutacionales y modelos para  
mutaciones puntuales Variabilidad del genoma mitocondrial Loci hipervariables: los  
microsatélites Mutaciones entre un número discreto de alelos  
Modelo de mutación irreversible  
Modelo de mutación reversible Modelo de alelos infinitos Equilibrio entre deriva  
genética y mutación Lecturas recomendadas

### 5. Selección natural sobre caracteres mendelianos simples

Concepto de selección natural  
Selección sobre caracteres mendelianos simples Un ejemplo: anemia falciforme en  
humanos El alelo C, y el carácter histórico del proceso selectivo  
Principales tipos de selección natural  
Imagen del paisaje adaptativo  
Lecturas recomendadas

### 6. El coalescente

El coalescente: modelo estándar Una aplicación inmediata El coalescente con más de 2  
alelos Tiempos de espera Antigüedad esperada del ancestro común a todos los alelos  
de la muestra Variabilidad genética en un contexto genealógico El coalescente y los  
árboles de genes Divergencia de las poblaciones en ausencia de selección Lecturas  
recomendadas

# Capítulo 1: INTRODUCCIÓN

## Presentación

Esta es la tercera, pero todavía incompleta versión de la guía. Se ha agregado un capítulo sobre el coalescente y las genealogías de alelos. Los estudiantes del curso de Evolución son los destinatarios de esta edición. Espero que les sea de utilidad pese a los capítulos faltantes, a los ocasionales cuadros o figuras prometidos en el texto pero que todavía no existen, y a los inevitables errores que seguramente contiene.

Espero también recibir críticas, sugerencias y comentarios de todo tipo, y muy particularmente indicaciones de errores detectados, de parte de ustedes. Pueden hacerlas llegar por correo electrónico ([lessa@fcien.edu.uy](mailto:lessa@fcien.edu.uy)), o hablando directamente conmigo o con otros docentes del curso. Los estudiantes y docentes de años anteriores identificaron algunos problemas, ahora corregidos, pero seguramente restan otros.

La idea de escribir una guía de estudio de la microevolución, o más modestamente de su núcleo central, constituido por la genética de poblaciones, surgió de la experiencia de dictar el curso de Evolución de la Licenciatura en Biología de la Facultad de Ciencias. No es una idea demasiado original, y de hecho existen excelentes libros de texto, incluyendo algunos de carácter introductorio en la materia, al menos en inglés, a los que un lector interesado en profundizar deberá dirigirse.

El propósito central de esta guía es ofrecer una muy breve introducción, que procura servir para un primer contacto con los conceptos centrales de la genética de poblaciones. He sido muy selectivo con las ecuaciones, y en general las consideradas imprescindibles se presentan, ilustran, y discuten, pero rara vez se demuestran. La guía busca acompañar las clases que sobre el tema incluye el curso de Evolución ya citado, pero puede también utilizarse de modo independiente.

La guía asume un conocimiento básico de genética general, incluyendo nociones de genética mendeliana clásica, de organización de los genes en los cromosomas (o en genomas de organelos como la mitocondria), de estructura y función del ADN, y del código genético que relaciona éste con las secuencias de proteínas. No se asumen conocimientos de genética de poblaciones ni de probabilidad, lo que hará que quienes poseen los fundamentos de estos temas puedan hallar que parte del material es aburrido y detalla trivialidades. En cambio, quienes no poseen nociones suficientes de probabilidad encontrarán descripciones algo informales pero que presumo ayudarán a desarrollar una cierta intuición al respecto. En varias partes del texto, estas descripciones contienen sugerencias de qué aspectos de la teoría de probabilidades son particularmente útiles y merecerían un estudio más serio.

He optado por partir de los problemas genéticos y desde ellos señalar la existencia e importancia de algunas formulaciones de los modelos probabilísticos más sencillos. El camino inverso, es decir comenzar por un estudio de modelos probabilísticos, y construir luego a partir de ellos la genética de poblaciones, es, naturalmente, el que siguen los especialistas y el más eficiente. La única justificación de la opción aquí tomada es que busca facilitar el ingreso al tema desde lo que para muchos estudiantes, en particular de biología, constituye el

primer contacto con el tema.

## ¿Qué es la microevolución?

Existe evidencia de la presencia de seres vivos en la tierra, cuya edad se estima en 4500 millones de años, desde hace unos de 3000 millones de años. Evidentemente, la enorme diversidad de los seres vivos que hoy conocemos no es sino una pequeña parte de la que ha existido en el vasto período de existencia de la vida. También es claro que cualquier especie viviente es producto de esa larga historia, y que dar cuenta de todas sus características requiere el estudio detallado de dicha historia.

Sin embargo, tanto la teoría concebida por Darwin como su heredera moderna, llamada teoría sintética (o, también, neodarwinista) de la evolución, proponen que el principal, aunque no exclusivo, proceso evolutivo transcurre, sin prisa y sin pausa, en el devenir de las variaciones heredables que pasan de una generación a otra de organismos en las poblaciones naturales. En términos modernos y sencillos, la variación genética surge por mutaciones que modifican el material genético pre-existente, y la selección natural dicta el destino de tales variantes. Nótese que el punto de partida en esta proposición es la existencia de seres vivos dotados de un sistema de codificación y transmisión hereditaria basado en los ácidos nucleicos (ADN o ARN). Debió haber una evolución previa que dio lugar a estos seres vivos, pero no nos ocuparemos de ello.

Nótese que esta visión privilegia un proceso (la selección natural) que ocurre en un nivel de organización biológico (organismos en una población), en escalas de tiempo “cortas,” *in extremis* el tiempo que tarda una generación de organismos en dejar lugar a la siguiente.

La microevolución es el estudio de los procesos evolutivos que ocurren en escalas cortas de tiempo. No he definido aún qué constituye una escala corta de tiempo, pero lo importante por el momento es entender que los procesos fundamentales ocurren generación tras generación. La unidad básica de tiempo es por tanto una generación. La selección natural es un proceso importante, aunque no el único, que opera en las poblaciones naturales afectando el destino de las variantes genéticas.

Para algunos biólogos, la evolución se limita, en lo que a procesos se refiere, esencialmente a la microevolución, y la macroevolución es el producto de la microevolución acumulado a lo largo de grandes períodos de tiempo. Para otros, los aspectos “macro” de la evolución involucran otros fenómenos que no pueden reducirse a la microevolución. Como sea, la microevolución en general, y la teoría genético-poblacional en particular, ocupan un lugar central en la teoría evolutiva.

El estudio de la genética de poblaciones suele presentar algunas dificultades, no tanto por las formulaciones matemáticas en sí mismas, sino porque se requiere adquirir el hábito del “pensamiento poblacional,” que al decir de Ernst Mayr constituye uno de los aportes fundamentales de Darwin a la biología. Pero debe notarse que los beneficios y satisfacciones que ofrece la disciplina para quienes asimilan sus fundamentos son múltiples. No es exagerado decir que no hay problema en biología y medicina que no

pueda beneficiarse de un análisis genético-poblacional. Dicho análisis puede revelar por ejemplo qué regiones de un genoma viral han estado sometidos a selección natural, y por ende han sido claves para su supervivencia y éxito, y puede también indicar si una población natural tiene una larga historia de estabilidad demográfica o si, por el contrario, ha sufrido un cambio significativo en su tamaño o distribución.

## Variación genética

El material de base de la genética de poblaciones es la variación genética. Sin ella no hay genética poblacional, o la que existe se reduce a la monótona transmisión de la misma información genética de una generación a la siguiente. Naturalmente, una población uniforme en su composición genética no cambia a lo largo del tiempo, y por ende, estrictamente, no evoluciona. Este estado de cosas no puede, en la práctica, durar demasiado, ya que nuevas mutaciones vendrán a aportar variación más temprano que tarde a nuestra hipotética población carente de variación.

Las mutaciones serán motivo de una sección más adelante. Por el momento, y al solo efecto de convencernos de la inevitable presencia de variación genética, consideremos el genoma humano. Tiene unos 3 mil millones ( $3 \times 10^9$ ) de pares de bases. ¿Es esperable que aparezca alguna mutación nueva en un individuo cualquiera o, por el contrario, cabe esperar que sólo muy raramente cambiará alguno de esos pares de bases en el ADN? Existen tipos muy diferentes de mutaciones, y sus frecuencias también varían. Asimismo, la tasa de mutación varía entre organismos. En el caso de los organismos multicelulares, como los humanos, la división celular incorpora mutaciones con cierta probabilidad, pero interesa también la tasa neta de una generación de humanos a la siguiente (de padres a hijos), mediada por numerosas divisiones celulares (del cigoto al gameto que pasará la información a la siguiente generación). Una estimación de la tasa de mutación en humanos (concretamente aquellas transmitidas por vía paterna, que son más frecuentes) es del orden de  $10^{-8}$  por sitio a los 30 años de edad (la tasa crece con la edad del hombre). Esperamos, por tanto, encontrar en promedio una nueva mutación cada  $10^8$  sitios, o sea unas 30 nuevas mutaciones por genoma. Esta es una expectativa de carácter probabilista, no determinista. Cada par de bases tiene una cierta probabilidad de mutar, y un conteo de grandes números (por ejemplo, un recuento en todo el genoma) debería resultar en un número razonablemente parecido a esta expectativa estadística.

Estas mutaciones nuevas pueden ser difíciles de encontrar. En promedio, ¡deberíamos secuenciar 100 millones de pares de bases para encontrar una nueva mutación! En otras palabras, las mutaciones nuevas son muy numerosas, pero están repartidas en un genoma muy grande. Otra forma de visualizar este resultado es pensar que, si estudiásemos un único par de bases por individuo, encontraríamos en promedio una nueva variante, inexistente en la generación precedente, por cada 10 millones de seres humanos examinados.

Finalmente, una manera también interesante de imaginar el resultado de las mutaciones es pensando en un gen particular. Supongamos que nos interesa el llamado gen de la beta-globina, que codifica para una de las hemoglobinas presentes en los glóbulos rojos. Supongamos también que solamente vamos a considerar las regiones de ese gen que

codifican a los 146 aminoácidos que en la especie humana y en muchas otras constituyen a esa proteína. El código genético es tal que se requieren tres pares de bases (un triplete o codón) para cada aminoácido, de modo que nuestra secuencia de interés está formada por  $146 \times 3 = 438$  pares de bases, que por comodidad redondearemos a 500. Desechemos cualquier diferencia entre esos 500 sitios en el ADN que codifica a la beta-globina, y asumamos también que estos sitios no difieren, en su probabilidad de mutar, del promedio del genoma. Nuestra probabilidad de encontrar un mutante era de una en 10 millones para un sitio cualquiera. Naturalmente, un examen de los 500 pares de bases que constituyen los codones de la beta globina sería 500 veces más fructífero en cuanto a encontrar nuevas mutaciones. En otras palabras, la probabilidad de encontrar una mutación nueva, estimada ahora para esta región, y no para un solo sitio, es aproximadamente  $500 \times 10^{-8} = 5 \times 10^{-6}$ . A nivel de la población, esto significa que, puesto que cualquier población humana tiene millones de individuos, y pasando por alto numerosos detalles, en cada generación hay numerosas nuevas mutaciones, aun restringiendo nuestra atención a esos 500 sitios codificantes de la beta-globina.

En estas conjeturas dirigidas a convencernos que la variación genética debería ser abundante, hemos utilizado las nociones y reglas más elementales y útiles de probabilidad. Ellas son:

- Concepto de probabilidad: Asumimos que cada sitio, cada gen, o cada región del genoma que nos interesa estudiar tiene una cierta probabilidad de ganar una mutación al transmitirse de un individuo a un descendiente. Si pudiésemos repetir infinitamente el ejercicio de formar dicho descendiente, encontraríamos esa frecuencia de mutantes nuevos. No podemos garantizar qué pasará en un individuo concreto, y en cuanto a una muestra de casos, podemos decir que tendremos mayor chance de encontrar una frecuencia parecida a esa probabilidad ideal cuanto mayor sea el tamaño de la muestra. En la práctica, nuestras probabilidades son estimadas a partir de ciertas muestras, y en general no interesa demasiado el valor exacto sino el orden de magnitud (1 en 10.000, 1 en 100.000, etc.). Finalmente, con frecuencia hablaremos de la tasa de mutación, como si fuese única, pero es claro que en estudios detallados esta simplificación deberá remplazarse por nociones más precisas de probabilidades de mutación que pueden variar por región del genoma, según el tipo de mutación que nos interesa considerar, etc.
- Eventos independientes. Para poder estimar el número de mutaciones en un genoma o región del mismo, no basta con asumir, como aproximación, que la tasa de mutación es única para todo el genoma. Es también necesario aceptar que las mutaciones ocurren de modo independiente. Es decir, que el hecho que un sitio haya mutado en nada cambia la probabilidad de que otro sitio cualquiera lo haga.

- Probabilidades de eventos particulares: Estamos con frecuencia interesados en eventos como el observar una mutación, o su complemento, el no observarla. Un caso análogo sería plantearse la probabilidad de observar un determinado alelo, o su complemento, el de observar el caso complementario, es decir cualquiera de los restantes alelos posibles para un locus cualquiera. En éstos y muchos casos, esos eventos tienen probabilidades que suman 1, ya que cubren la totalidad de las opciones.
- “Regla” de la suma: Si la probabilidad de observar uno cualquiera de los eventos en una tirada de un dado es  $1/6$ , la probabilidad de obtener un número par es  $3/6 = 1/2$ .
- “Regla” de la multiplicación: Supongamos que un alelo representa el 10% de los alelos de una población para un gen dado. En otras palabras, la probabilidad elemental de observarlo en un evento simple (tomar un determinado alelo al azar y que resulte de la clase A) es de 10%, o de 0.1. ¿Qué probabilidad tengo de observar dos alelos de la clase A si mi muestra es ahora de dos? Mis chances son de un 10% para el primer alelo, y si se da esa instancia tengo, nuevamente, un 10% para el segundo, que muestreo de modo independiente. 10% de 10% es 1%, o, de manera equivalente,  $p(AA) = 0.1 \times 0.1 = 0.01$ . Debe notarse que para que esto sea cierto, los dos eventos son independientes.

Esta breve e informal introducción a las mutaciones nos indica, en suma, que en cada generación surgen numerosas nuevas variantes genéticas por mutación. Ingenuamente, podríamos suponer que ello implica que la diversidad genética de una población aumentará, gradual pero inexorablemente, generación tras generación. Con igual simplismo cabría pensar que la variación en un sitio, gen, o región cualquiera del genoma refleja sencillamente la tasa de mutación correspondiente. Pero pronto veremos que el destino de las mutaciones involucra otros procesos, por lo que las dos inferencias recién mencionadas son erróneas. La selección natural, por ejemplo, puede favorecer una nueva variante o, por el contrario, actuar en su contra. Una mutación nueva normalmente estará inicialmente representada por una única copia. Si el individuo que la porta no la transmite al menos a un descendiente, esa variante desaparecerá tan rápido como apareció, simplemente por azar. Estos y otros posibles destinos de la variación genética provista por las mutaciones son motivo de estudio de la genética de poblaciones.

### **Marco histórico de referencia**

La genética de poblaciones tiene una historia que, en lo esencial, comienza con el siglo XX. Aunque en esta guía solamente se harán muy breves referencias a esa historia, es importante tener en cuenta el contexto en que los diferentes modelos y trabajos experimentales se cumplieron. Se verá así que la comprensión de la estructura genómica y de la evolución a nivel genético es, en realidad, y pese a sus muchos éxitos, una disciplina con un amplio y abierto futuro, con vastos campos aún inexplorados o apenas atisbados.

La siguiente cronología provee algunos puntos de referencia histórica:

Década	Desarrollo
1850	Darwin y Wallace formulan la teoría de evolución por selección natural.
1860	Mendel establece los fundamentos de la genética.
1900	Las leyes de Mendel son redescubiertas y finalmente valoradas. Hardy y Weinberg establecen los fundamentos de la genética de poblaciones teórica.
1910 1920	Fisher y Wright desarrollan la genética de poblaciones, expresando la teoría darwinista de la selección en modelos poblacionales. Se inicia la formulación de la moderna teoría sintética de la evolución.
19201930	Se consolida la noción de que los genes están localizados en los cromosomas, pese a que se desconoce su naturaleza.
1940	Muller establece que la radiación ultravioleta produce mutaciones. Se establece que el ADN es el material hereditario.
1950	Modelo de doble hélice del ADN. Primeras secuencias de proteínas.
1960	Determinación del código genético vinculando la información del ADN con la secuencia de las proteínas. Se consolida la idea del reloj molecular. La electroforesis de proteínas provee la primera vía sistemática de examinar la variación genética en las poblaciones naturales. Kimura formula su teoría neutralista de la evolución molecular.
1970	Primeras secuencias de ADN.
1980	Estudios de ADN mitocondrial sugieren el origen africano de la especie humana moderna. Mullis inventa la reacción de PCR (polymerase chain reaction). Kingman define el coalescente como modelo de análisis genéticopoblacional.
2000	Se da a conocer la primera versión del genoma humano.

Nótese, por ejemplo, que recién en la década del 1960 se contó con: a) una idea clara de cómo están formados los genes (al menos una clase importante de ellos) y cómo codifican a las proteínas; b) un conjunto mínimo de secuencias de proteínas que podían compararse entre las especies; c) una técnica de estudio empírico y sistemático de variación genética (en sus productos, las proteínas) y de análisis de frecuencias génicas que no requería mutaciones visibles a nivel morfológico.

### Lecturas recomendadas

Gillespie, J. H. 2004. Population genetics—a concise guide. 2ª. Edición. Johns Hopkins University Press, Baltimore. *Como dice el título, una guía compacta de lo esencial de la genética de poblaciones, acompañada de comentarios de un pequeño número de artículos de especial importancia.*

Provine, W. P. 1971. The origins of theoretical population genetics. Univ. of Chicago Press, Chicago. *Una excelente panorámica histórica de la gestación de la genética de poblaciones, desde la época de Darwin y Galton hasta la*



*consolidación de la teoría sintética.*

## Capítulo 2: MODELO HARDY-WEINBERG

En 1908, un matemático inglés (G. H. Hardy) y un fisiólogo alemán (W. Weinberg) formularon independientemente un modelo genético poblacional que lleva sus nombres y ocupa un lugar central en la genética de poblaciones. Como sucede tantas veces con los modelos, el que nos ocupa ha sido injustamente ridiculizado por los naturalistas por su falta de realismo. El eminente evolucionista Ernst Mayr, por ejemplo, no dudó en despreciar, ya no a este modelo solamente, sino a toda la genética de poblaciones como “genética de la bolsa de porotos” (beanbag genetics, en inglés), aludiendo al modo como solían presentarse el modelo en cursos prácticos, donde los alelos están representados por porotos de diferentes colores que se combinan al azar.

Es posible que la incomprensión generalizada del modelo de Hardy-Weinberg se deba a una pobre presentación del mismo en los cursos introductorios. Pero es también probable que el problema sea más profundo, y resida en la débil educación en modelos probabilísticos, que lleva al estudiante a aprender la mecánica de los sencillos cálculos de éste u otros modelos sin llegar por ello a captar su significación. En cuanto a la falta de realismo del modelo, en este caso al menos es correcto decir que su enorme utilidad se debe precisamente, en parte, a esta propiedad. El modelo Hardy-Weinberg, típicamente, se utiliza como hipótesis nula, y el constatar desviaciones significativas de lo esperado según el modelo es indicio de que algo distinto, algo que no se ajusta a lo asumido por el modelo, está sucediendo en la población de estudio.

Como en buena parte de esta guía, en esta sección limitaremos nuestra atención a un único e hipotético gen, que está representado en una población por dos o más alelos (existe variación que merece ser estudiada). No nos interesa por el momento si ese “gen” es del tipo del de la beta-globina, es decir un gen tradicional que codifica una proteína particular, o si se trata de algún otro tipo de región genómica. En principio, podría simplemente tratarse de un único sitio en algún cromosoma, ocupado por un par de bases en la doble cadena del ADN, pero mostrando variación en la población (dicho par es A:T en cierto cromosoma, pero C:G en otro, por ejemplo).

### **El modelo poblacional de Hardy-Weinberg**

El modelo Hardy-Weinberg considera qué sucede en una población ideal con características bien definidas. Para simplificar la discusión posterior, dividiremos arbitrariamente las propiedades de la población en dos clases.

En primer lugar, asumimos ciertas características generales de la población que resultan convenientes, pero destacando que pueden formularse fácilmente variantes del modelo Hardy-Weinberg para acomodar alternativas a estas propiedades. Nuestra población ideal:

- está formada por organismos diploides;
- la reproducción es sexuada;
- las generaciones no se solapan;
- el gen considerado es autosómico;
- no hay diferencias de frecuencias alélicas entre los sexos.

Un segundo grupo de características de la población merece ser destacado separadamente porque en esta guía consideraremos de modo explícito algunos fenómenos que constituyen claros apartamientos del modelo. Estas características del modelo Hardy-Weinberg son:

- los apareamientos ocurren al azar;
- la población es de tamaño infinito;
- no existe migración desde otras poblaciones; no hay mutación;
- la selección natural no opera sobre el gen considerado;

Podemos adelantar la utilidad del modelo al considerar, por ejemplo, el último punto (la ausencia de selección). El modelo Hardy-Weinberg provee predicciones precisas sobre el comportamiento de un gen sobre el que la selección no opera. Si sospechamos que la selección en verdad afecta a un gen determinado, podemos comenzar su estudio demostrando que dicho gen se aparta de modo consistente de las predicciones del modelo. Si rechazamos el modelo (nuestra hipótesis nula), pasaremos a considerar seriamente alternativas. La acción de la selección no es la única fuente de desviaciones de las predicciones del modelo, pero se encuentra entre ellas.

Una importante propiedad de este modelo es la separación de cada generación en dos fases:

- la fase gamética, en la que cada locus está representado por un sólo alelo; las frecuencias alélicas de esta fase se utilizan en el modelo;
- la fase diploide u orgánica, en la que cada locus está representado por dos alelos (uno legado por cada progenitor); las frecuencias genotípicas de esta fase se utilizan en el modelo.

### **Consecuencias del modelo Hardy-Weinberg**

¿Qué podemos esperar del comportamiento de los alelos de un gen en una población con todas las características del modelo Hardy-Weinberg? Supongamos que las poblaciones humanas se ajusten a tal modelo como ejercicio, y consideremos el caso del Factor V (uno de los muchos genes que puede tipificarse a partir de una muestra de sangre). Entre los Yorubas de Nigeria se encuentran dos alelos de este factor. El alelo V tiene una frecuencia de 0.376, mientras que el alelo v se encuentra en una frecuencia de 0.624 (Roychoudhury y Nei, 1988). Llamemos a estas frecuencias p y q, respectivamente:

$$f(V) = p = 0.376 = 37.6\% \quad f(v) = q = 0.624 = 62.4\%$$

Como sabemos, un gen diploide como el considerado, que cuenta con dos variantes alélicas, puede generar solamente tres genotipos diferentes. Así, si un individuo hereda dos copias del alelo V, una de cada progenitor, será homocigota VV. Existirán también homocigotas vv, y heterocigotas Vv.

La primera y más obvia consecuencia del modelo es que las frecuencias alélicas permanecerán estables a lo largo del tiempo. En efecto, no ocurre en el modelo nada que pueda cambiarlas. El tamaño infinito de la población previene fluctuaciones al azar

de las frecuencias (no opera la llamada deriva genética, que consideraremos más adelante), que en una población de tamaño finito ocurrirán inevitablemente. No aparecen alelos nuevos, ni ocurren mutaciones de uno a otro de los alelos presentes. Tampoco existen inmigrantes que aporten nuevos alelos, o simplemente frecuencias distintas de los ya existentes.

La segunda consecuencia del modelo es que existe una correspondencia precisa y estable entre las frecuencias alélicas (también llamadas frecuencias génicas) y las frecuencias genotípicas. Esta estabilidad resulta de agregar, al tamaño infinito de la población y a la ausencia de mutaciones y otras fuentes de variación, el sistema de apareamientos al azar. La población, como suele decirse, es panmíctica.

Consideremos la probabilidad de obtener un individuo homocigota de tipo VV en nuestra población ideal. Este individuo proviene de la combinación al azar de dos alelos, uno legado por el padre (y presente en el espermatozoide) y otro por la madre (en el óvulo), en la formación del cigoto. La probabilidad de que el alelo paterno, que escogemos al azar de entre los disponibles en la población, sea V es

$$f(V) = p = 0.376 = 37.6\%$$

Como el segundo alelo se escoge con independencia del primero, es fácil ver que, habiendo escogido el alelo V en primer lugar con una probabilidad de un 37.6%, tenemos ahora un 37.6% de posibilidades de escoger otra copia de ese mismo alelo para completar el complemento genético de nuestro nuevo individuo. Por lo tanto, deducimos que la frecuencia esperada de homocigotas VV será

$$f(VV) = 0.376 \times 0.376 = 0.141 = 14.1\%$$

En la **Figura 2.1** se representa gráficamente cómo se combinan las frecuencias génicas presentes en los gametos para obtener las frecuencias genotípicas para una población que sigue el modelo Hardy-Weinberg. Razonando brevemente sobre el gráfico se ve que existe una única forma de obtener una combinación homocigota determinada, ya sea VV o vv en nuestro ejemplo, ya que debemos escoger el mismo alelo en ambos gametos. La frecuencia esperada de genotipos homocigotas para un alelo determinado es siempre el cuadrado de la frecuencia de dicho alelo. Así,  $f(VV) = p^2$  y  $f(vv) = q^2$ .

En cambio, hay dos formas de obtener una combinación genotípica heterocigota. En nuestro ejemplo, el alelo V puede provenir del padre y el v de la madre, y esto puede ocurrir con una probabilidad igual al producto de las frecuencias de esos alelos. Es decir, la combinación Vv cuyo alelo V es de origen paterno tiene una frecuencia esperada de pq. Pero el alelo V puede provenir en cambio de la madre, y como no hay diferencias de frecuencias alélicas entre los sexos, el razonamiento es idéntico y esta opción ocurrirá igualmente con una frecuencia pq.

Como resultado, la probabilidad total de observar un heterocigota Vv es  $f(Vv) = 2 pq$ .

Figura 2.1. Representación gráfica del equilibrio Hardy-Weinberg para el caso de dos alelos con frecuencias  $p$  y  $q$ . Las frecuencias alélicas están representadas en los márgenes del cuadrado, cuyos lados se han segmentado en proporción a ellas. Las áreas dentro del cuadrado principal representan las frecuencias esperadas de las diferentes combinaciones.

		Gametos ♀	
		$p$	$q$
Gametos ♂	$p$	$p^2$	$pq$
	$q$	$pq$	$q^2$

Pese a las muchas y flagrantes violaciones de los postulados del modelo Hardy-Weinberg, las poblaciones naturales tienen una sorprendente tendencia a aproximarse, en lo que a frecuencias genotípicas se refiere, a las predicciones del modelo. Ciertamente, ninguna población humana es panmíctica (las probabilidades de apareamiento no son aleatorias, sino que se ven afectadas por factores sociales, geográficos, y muchos otros), ni está completamente libre de inmigración, ni es de tamaño infinito. Las desviaciones resultantes de estos fenómenos son, sin embargo, menores. Para el caso de nuestra población Yoruba y su variación en el Factor V, el equilibrio Hardy-Weinberg permite predecir, resumiendo, las siguientes frecuencias genotípicas aproximadas:

$$f(VV) = 0.141$$

$$f(Vv) = 0.470$$

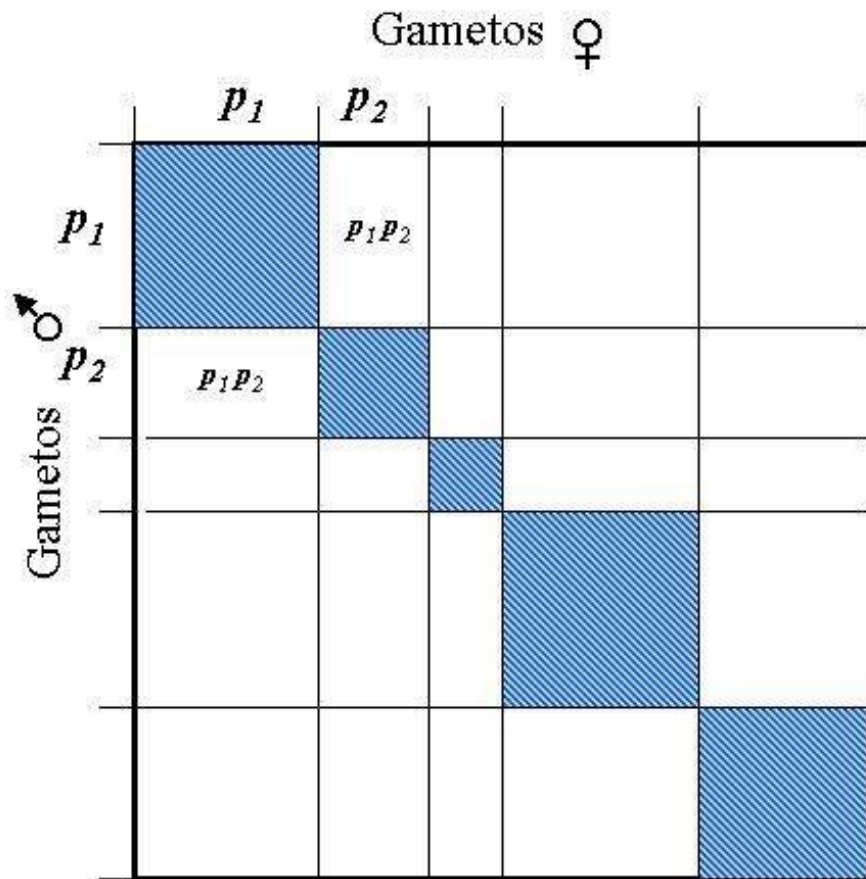
$$f(vv) = 0.389$$

El **Cuadro 2.1** presenta un ejercicio, también basado en un caso de un polimorfismo sanguíneo en una población humana, en el que se introducen algunos elementos adicionales. Por lo pronto, se parte de conteos genotípicos en una muestra experimental. Esta muestra se usa, en primer lugar, para estimar las frecuencias génicas, y luego se usa el modelo Hardy-Weinberg para calcular las frecuencias genotípicas esperadas. Finalmente, se comparan éstas con las observadas.

La representación gráfica que ya usamos en la **Figura 2.1** permite extender el principio del modelo de Hardy-Weinberg a casos más complejos de un modo sencillo e intuitivo. Por ejemplo, como se ilustra en la **Figura 2.2**, si un gen presenta más de dos alelos, basta representar sus frecuencias como segmentos de los lados del cuadrado (que representan los gametos masculinos y femeninos) para obtener las frecuencias genotípicas esperadas en el cuerpo del cuadrado. No hay nada realmente nuevo, desde el punto de vista conceptual, al pasar de dos alelos a tres, cuatro, o más; solamente hay más combinaciones que siguen las mismas reglas.

La elegante uniformidad del equilibrio Hardy-Weinberg se puede apreciar también considerando un problema común en genética, que resulta de las diferentes formas y variables grados de resolución de los análisis genéticos. Por ejemplo, consideremos el gen de la peptidasa A en humanos. Las poblaciones europeas presentan dos alelos  $A_1$  y  $A_2$  con frecuencias aproximadas de 76% y 24%. Estos alelos están definidos por electroforesis de proteínas (Lewontin, 1995). Supongamos ahora que una investigación obtiene, ya no información basada a nivel de proteínas, sino en el ADN del gen en cuestión. Un estudio de esa naturaleza puede revelar variación que no se manifiesta en los análisis proteicos. Por ejemplo, pueden hallarse variaciones en terceras posiciones de algún codón que no resultan en un cambio en el aminoácido correspondiente (los llamamos cambios silenciosos, en contraste con aquellos que resultan en el remplazo de un aminoácido).

Figura 2.2. Representación gráfica del equilibrio Hardy-Weinberg para el caso de varios alelos (solamente  $p_1$  y  $p_2$  están representadas). Las frecuencias esperadas de homocigotas para los diferentes alelos están representadas por las áreas cuadradas sombreadas en la diagonal del cuadrado principal. Al igual que en la Figura 1, las combinaciones heterocigotas están representadas por los rectángulos fuera de la diagonal del cuadrado (solamente las combinaciones de los alelos 1 y 2 están señaladas).



**Cuadro 2.1**<sup>1</sup>. Un estudio de 1000 individuos en Inglaterra observó las frecuencias genotípicas para el gen del grupo sanguíneo NM que se presentan a continuación. Completar las tablas calculando: a) las frecuencias de los alelos N y M; b) las frecuencias absolutas esperadas, según el equilibrio Hardy-Weinberg, para cada genotipo.

Genotipos	Frecuencias genotípicas observadas	Frecuencias genotípicas esperadas
MM	298	
MN	489	
NN	213	
Total	1000	

Frecuencia estimada del alelo M =

Frecuencia estimada del alelo N =

(Nótese que comenzamos por estimar las frecuencias alélicas usando la totalidad de los datos, para luego calcular las frecuencias genotípicas esperadas.) La comparación de las frecuencias esperadas y observadas sugiere que las diferencias entre ellas son pequeñas, y probablemente atribuibles a errores de muestreo. La prueba de chi cuadrado (sobre la que no nos extenderemos) puede usarse para examinar si las desviaciones entre las frecuencias observadas y las esperadas son significativas o si, por el contrario, no hay motivos para rechazar la hipótesis nula de equilibrio Hardy-Weinberg.

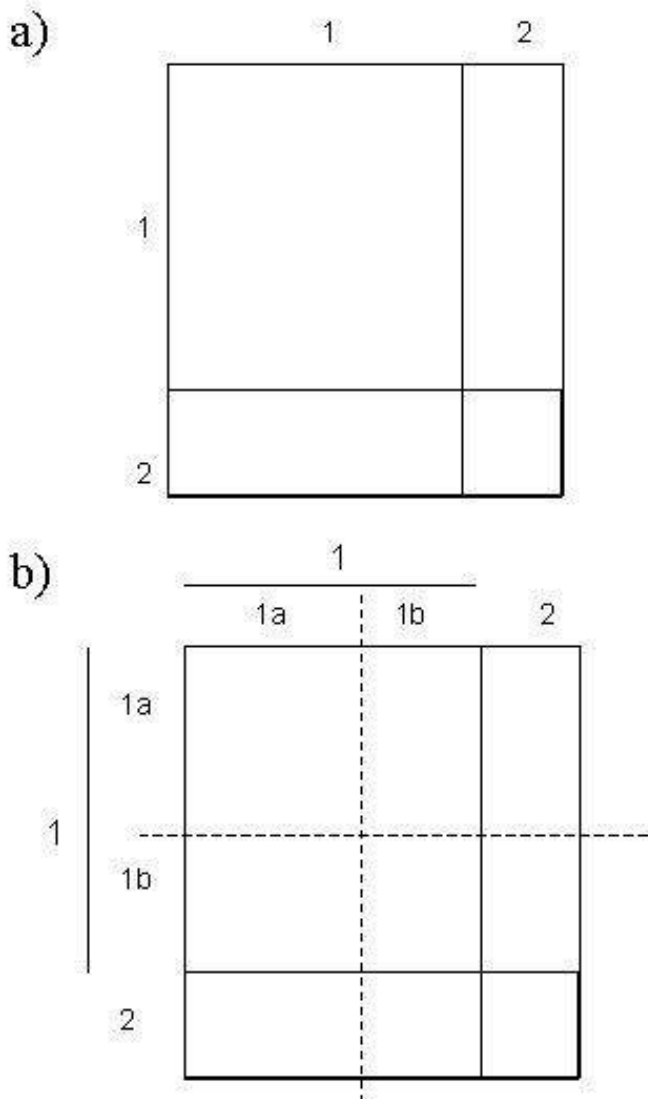
Como ejercicio, calcular el valor del estadístico chi cuadrado e interpretar el resultado teniendo en cuenta que en este caso existe un grado de libertad, y que el valor crítico para  $\alpha = 0.05$  es en este caso  $\chi^2 = 3.84$ .

<sup>1</sup> Ejemplo tomado de Hartl y Clark (1997).



Supongamos que un estudio de este tipo determina que el alelo  $A_1$  clásico es en realidad una mezcla de dos subtipos, indistinguibles a nivel proteico, que llamaremos  $A_{1a}$  y  $A_{1b}$ , de modo tal que constituyen el 60% y el 40%, respectivamente. El examen del equilibrio Hardy-Weinberg con esta nueva información es reconfortante en varios aspectos, como se ve en la **Figura 2.3**.

Figura 2.3. a) Representación gráfica del equilibrio Hardy-Weinberg para dos alelos (1 y 2) de la peptidasa A humana en poblaciones europeas. b) Representación más detallada, en la que se distinguen dos subtipos hipotéticos del alelo 1 (1a y 1b).



Por ejemplo, el cuadrado que representa la frecuencia de homocigotas para el alelo clásico  $A_1$  puede extraerse del cuadro general y subdividirse según las frecuencias de  $A_{1a}$  y  $A_{1b}$ , haciendo caso omiso del resto del diagrama. De hecho, pueden representarse los subtipos en sus diferentes combinaciones, incluyendo las heterocigóticas, del mismo modo. Las predicciones originales del modelo Hardy-Weinberg continúan siendo válidas en cuanto a los alelos clásicos, y al mismo tiempo es fácil ver cómo se incorporan las subdivisiones adicionales provistas por subtipos definidos con técnicas más refinadas.

### Homocigosidad y heterocigosidad esperadas

Ya hemos considerado cómo, a partir de las frecuencias alélicas, el modelo Hardy-Weinberg predice las frecuencias de todas las combinaciones genotípicas posibles. El número de esas combinaciones, por supuesto, aumenta con el número de alelos. Sin embargo, a menudo es útil resumir la información sobre la variabilidad de un locus cualquiera en dos medidas. La primera es la homocigosidad esperada, es decir la frecuencia de homocigotas esperados en equilibrio Hardy-Weinberg. Si llamamos  $p_i$  ( $p_1, p_2, \dots, p_k$ ) a las frecuencias de los distintos alelos (etiquetados de 1 a  $k$ ), entonces la homocigosidad esperada es

$$E(G) = \sum_{i=1}^k p_i^2$$

En la **Figura 2.2**, puede verse claramente que la homocigosidad esperada es la suma de los cuadrados de las frecuencias alélicas, y está representada por la suma de las áreas de los cuadrados en la diagonal del cuadrado principal. Naturalmente, la heterocigosidad esperada está representada en la **Figura 2.2** por el área no cubierta por las diferentes combinaciones homocigotas, es decir:

$$E(H_{HW}) = 1 - E(G) = 1 - \sum_{i=1}^k p_i^2$$

(Nota: usamos en subscrito “HW” para denotar que estamos hablando de la heterocigosidad esperada según el modelo de Hardy-Weinberg. Veremos luego que hay otros sentidos del término en genética de poblaciones.)

Recurriendo a la **Figura 2.1**, puede verse fácilmente que, para el caso de dos alelos, la máxima heterocigosidad esperada es de 50%, y ocurre cuando cada alelo se encuentra a una frecuencia de 50%. El lector puede examinar la **Figura 2.3** y pensar qué combinación de frecuencias alélicas producirá la mayor heterocigosidad esperada posible para tres o más alelos, hasta convencerse de que, cualquiera sea el número de alelos  $k$ , la heterocigosidad máxima se obtiene cuando  $p_1 = p_2 = \dots = p_k = 1/k$ .

### Aproximación al equilibrio Hardy-Weinberg

Hasta ahora hemos descrito el equilibrio Hardy-Weinberg, y también hemos afirmado que las poblaciones naturales se aproximan al mismo con asombrosa determinación pese a que los supuestos del modelo son muy poco realistas. Como ejercicio que puede permitir

atisbar porqué ello ocurre, le proponemos al lector el siguiente caso. Supongamos que partimos de una población inicial con un marcado sesgo en la representación de los alelos en los dos sexos. Por ejemplo, fundamos en el laboratorio una población de *Drosophila* tal que todos los machos son homocigotas para un alelo A, y todas las hembras para un alelo a del mismo gen. ¿En cuántas generaciones se llega al equilibrio Hardy-Weinberg, aceptando naturalmente las reglas del modelo, especialmente la de apareamientos al azar? El lector puede considerar otras situaciones.

### **Lecturas recomendadas**

Gillespie, J. H. 2004. Population genetics—a concise guide. 2a. Edición. Johns Hopkins University Press, Baltimore. *Una excelente introducción a la genética de poblaciones teórica, reducida a su esencia. En particular, el capítulo sobre el modelo Hardy-Weinberg se recomienda para consolidar el material cubierto en este capítulo.*

## Capítulo 3: DERIVA GENÉTICA

Ya hemos adelantado que la genética de poblaciones estudia el destino de la variación genética que surge por mutación en las poblaciones naturales. Un atractivo de la disciplina es que permite comprender cómo suceden cambios en la composición genética de las especies biológicas. Sin embargo, en la sección anterior comenzamos nuestro estudio con un modelo que, curiosamente, implica un equilibrio estable, ya que tanto las frecuencias alélicas como las genotípicas permanecen incambiadas en el modelo de Hardy-Weinberg. Señalamos también el curioso status que goza este modelo, considerado central en la genética de poblaciones a pesar de, o tal vez precisamente debido a la falta de realismo de sus postulados.

En esta sección comenzaremos a remover algunos de esos supuestos y examinar las consecuencias de estas alteraciones. Recordemos los supuestos del modelo Hardy-Weinberg antes de comenzar a cambiarlos. Un primer conjunto de postulados, que continuaremos aceptando a pies juntillas, caracteriza a la población del modo siguiente:

- está formada por organismos diploides;
  - la reproducción es sexuada;
  - las generaciones no se solapan;
  - el gen considerado es autosómico;
  - no hay diferencias de frecuencias alélicas entre los sexos;
- Un segundo grupo de características de la población consiste en que:
    - los apareamientos ocurren al azar;
    - la población es de tamaño infinito;
    - no existe migración desde otras poblaciones;
    - no hay mutación;
    - la selección natural no opera sobre el gen considerado;

Consideraremos ahora poblaciones que ya no serán de tamaño infinito. Curiosamente, y pese a que sabemos que sin mutación no habría variación, y por ende, no tendría objeto estudiar la genética de poblaciones, nuestra presentación considerará primero qué sucede en una población de tamaño finito, todavía libre de mutaciones. El propósito es tomar como punto de partida una cierta variación inicial y examinar su destino sin permitir la aparición de nuevas variantes. La mutación será finalmente admitida en el próximo capítulo.

### **La deriva genética**

La modesta cuota de realismo introducida al pasar de una población de tamaño infinito a una con un número finito de individuos tiene consecuencias de suma importancia. Por lo pronto, y este es uno de los puntos más importantes de esta sección, una población de tamaño finito perderá la monótona estabilidad que caracteriza al modelo poblacional Hardy-Weinberg. Habrá, en otras palabras, evolución.

El proceso evolutivo que pasaremos a considerar, y que resulta del tamaño finito de la población, es peculiar puesto que consiste en evolución sin selección, algo que Darwin jamás consideró seriamente. En ausencia de selección, todos los alelos presentes en una población tienen igual probabilidad de dejar descendientes, pero, en la práctica, podemos observar que algunos dejarán una sola copia de sí mismos en la generación siguiente (este será el caso promedio), mientras que otros tendrán la buena fortuna de legar más de una copia, y otros la mala fortuna de no dejar descendientes. Como veremos, estrictamente una clase alélica que tiene  $n$  copias en una generación puede pasar a tener cualquier valor entre 0 y  $2N$  copias en la generación siguiente... pero, naturalmente no todos esos resultados son igualmente probables. Si una población cuenta con 50 individuos, y por ende 100 alelos, cada uno de ellos verá su suerte librada al azar implícito en la deriva genética.

La deriva genética tiene una curiosa historia. El británico Ronald Fisher, uno de los fundadores de la genética de poblaciones y de ese modo uno de los padres de la moderna teoría sintética de la evolución, fue además un consagrado estadístico. Naturalmente, su formación le colocaba en perfectas condiciones de apreciar los procesos estocásticos como la deriva genética. Sin embargo, Fisher prefirió concentrar sus esfuerzos en la

formalización matemática de la selección natural y en el estudio de su modo de operar y consecuencias. Las razones de esta opción requieren avanzar más en el estudio de la genética de poblaciones. Adelantándonos a dicho estudio, vale la pena consignar que en una población de gran tamaño, como las que Fisher procuró modelar, las presiones selectivas son de gran eficiencia y los efectos del azar pueden, hasta cierto punto, ignorarse, más allá de reconocer que el origen mismo de las nuevas variantes puede ser aleatorio.

Sewall Wright, contemporáneo norteamericano de Fisher, consideró más seriamente los efectos de la deriva genética, en un modelo más complejo de estructura poblacional que también dejaremos para más adelante. En todo caso, en la época de construcción de la teoría sintética (décadas del 20 al 40 del siglo XX), la principal cuestión a resolver era si la recientemente redescubierta genética mendeliana era compatible con la evolución por selección natural (muchos genetistas opinaban lo contrario), y a este propósito tanto Fisher como Wright contribuyeron de modo sustancial. Asimismo, el modelo básico de evolución por deriva genética es el llamado modelo Wright-Fisher, combinando los esfuerzos de ambos investigadores.

Hubo que esperar a la década de los 60 para la deriva genética pasase a ocupar un papel verdaderamente central en las fuertes controversias generadas en el por entonces joven campo de la evolución molecular: Así lo hizo posible el incipiente cúmulo de secuencias de proteínas primero, de frecuencias alélicas luego, y de secuencias de ADN más recientemente. No por casualidad, la audaz tarea de proponer un modelo de evolución de genes y proteínas guiado fundamentalmente por la deriva genética recayó en Motoo Kimura, un discípulo de Sewall Wright, aunque Jukes y Cantor propusieron un modelo también neutral de manera independiente.

Luego de esta disquisición histórica, retomemos la caracterización del problema recordando que nuestra población retiene un número importante de propiedades simples del modelo original de Hardy-Weinberg. Entre ellas, conviene destacar ahora que no hay diferencias de composición genética entre los sexos y que la población es panmíctica, ya que los apareamientos ocurren al azar, sin preferencias o aversiones. La distinción entre los sexos es, en este contexto, innecesaria, ya que basta saber que los organismos diploides producen gametos, que éstos reflejan las proporciones alélicas imperantes en la población, y que se unen al azar para formar los individuos de la siguiente generación. Las sucesivas generaciones no se solapan, lo que simplifica el proceso de evolución de unas a otras.

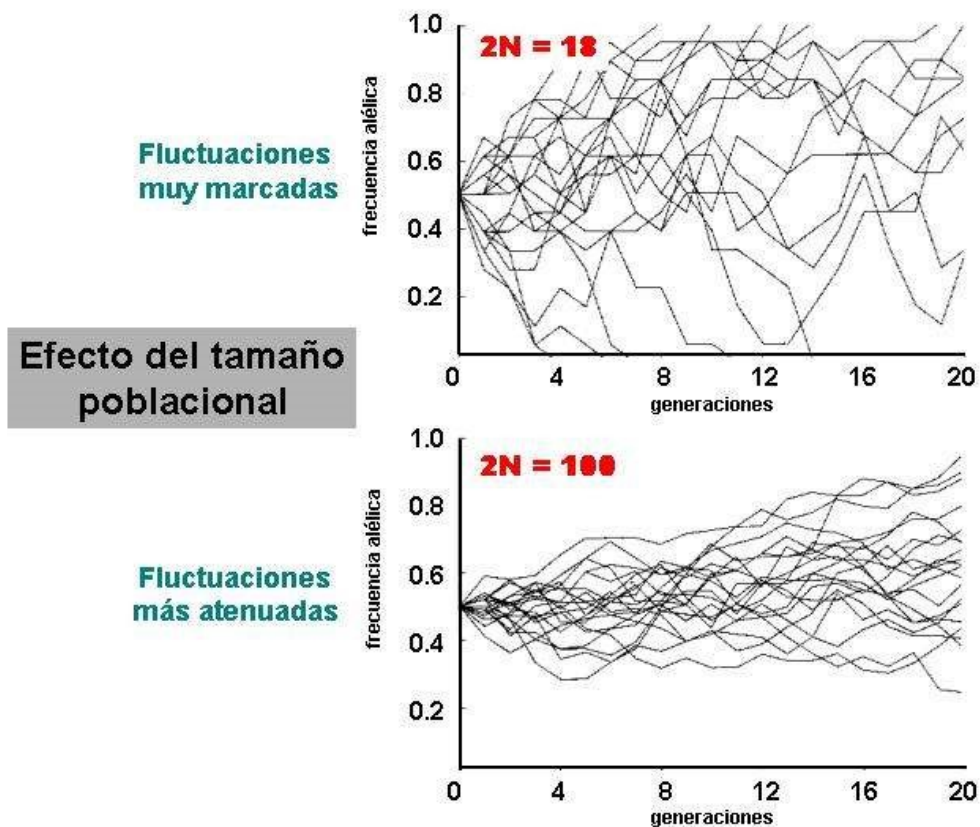
Este modelo, que a primera vista solamente difiere del de Hardy-Weinberg en el tamaño finito de la población, es conocido como modelo de Wright-Fisher, que definieron el sistema demográfico en el que centraremos nuestro análisis. El modelo requiere, además de las frecuencias alélicas iniciales, un parámetro adicional que es, como cabría esperar, el tamaño poblacional, que se define como un número constante de  $N$  individuos. Al estudiar un gen autosómico en un organismo diploide, ello implica que la población tiene  $2N$  alelos. Un detalle evidente, pero importante, es que ese recuento de alelos corresponde a la fase diploide del proceso. Los gametos pueden ser muchísimo más numerosos que  $2N$ , pero reflejarán las frecuencias alélicas de la generación que los produce.

### **Fluctuaciones de las frecuencias alélicas**

La consecuencia más evidente de un tamaño poblacional finito es que ocurrirán fluctuaciones en las frecuencias de los alelos de una generación a la siguiente. Esto resulta del modo como se forma la descendencia (los cigotos derivados de la unión de dos gametos) de una generación cualquiera. Los gametos se combinan al azar para formar cigotos. Más aún, suponemos que el número de gametos es tanto mayor que el número de individuos que la formación de un individuo portador, por ejemplo, de un alelo  $A$  legado por uno de los progenitores, no vuelve ni más ni menos probable que el otro progenitor también aporte ese mismo alelo, ni que el siguiente individuo que vaya a formarse también sea portador de  $A$ . Formalizando, los alelos de una generación se forman muestreando al azar, con reposición, los alelos presentes en la generación precedente.

El muestreo al azar con reposición implicado en la formación de cada generación posibilita que la composición genética de la población cambie a lo largo del tiempo. Así, un alelo puede aumentar o reducir su frecuencia, sin que para ello sea necesario invocar a la selección natural. La **Figura 3.1** ilustra este proceso para series de experimentos (en este caso, simples simulaciones en computadora, que pueden también realizarse con una buena bolsa de porotos) que, aunque tienen en común el comenzar con un alelo representado por el 50% de la población, caen sin embargo en dos clases: en una de ellas se sigue la historia de 9 individuos (18 alelos), mientras que en la otra el tamaño poblacional aumenta a 50 individuos (100 alelos). Las gráficas de la figura trazan, por conveniencia, la frecuencia de un alelo cualquiera, elegido como marcador, a lo largo del tiempo. No hace falta preocuparse, de momento, del resto de los alelos que conforman la población, que naturalmente componen la frecuencia necesaria para completar el 100%.

Figura 3.1



El examen de la Figura 3.1 permite extraer las siguientes conclusiones de carácter cualitativo, que formalizaremos luego:

- 
- 
- En todos los experimentos, se observa que la frecuencia del alelo marcador varía a lo largo del tiempo. En otras palabras, la población inicial, lejos de permanecer estable como en el modelo de Hardy-Weinberg, evoluciona.
- Las fluctuaciones que se observan de una generación a la siguiente no son constantes ni en magnitud ni en signo. En efecto, el alelo marcador aumenta su frecuencia en algunas instancias, la reduce en otras, y aún la retiene en algunos casos. El efecto acumulado es un comportamiento errático de cambios de frecuencias alélicas a lo largo del tiempo.
- Por lo general, las fluctuaciones de frecuencias alélicas son más marcadas en las poblaciones pequeñas (18 alelos) que en las grandes (100 alelos).



Finalmente, y una vez más pese a lo errático del comportamiento de la frecuencia de nuestro alelo marcador, el proceso no continúa indefinidamente, sino que se detiene en una de dos situaciones: o bien cuando el alelo se fija (esto es, llega a ser el 100% de la población) o bien cuando se extingue (deja de existir, o sea llega a una frecuencia de 0).

Estas observaciones contienen, informalmente, lo esencial del proceso que conocemos como deriva genética. Partiendo, al igual que en el modelo Hardy-Weinberg, de una población variable (con alelos de dos o más tipos), la deriva genética produce evolución. Dicho proceso consiste, generación tras generación, en cambios de frecuencias génicas, que constituyen la consecuencia palpable, a corto plazo, de la deriva genética. Aunque no hemos precisado el concepto, sabemos ya que los cambios de frecuencias son aleatorios en magnitud y signo, e intuimos también que aquella es función del tamaño poblacional o, de modo equivalente, del número de alelos. Finalmente, vemos que la consecuencia a largo plazo de la deriva genética es la eliminación de la variación, por fijación de alguno de los alelos que componían la población original, con lo que el proceso se agota a sí mismo, más tarde o más temprano. La evolución, naturalmente, se detendría al llegar a este punto, pero sabemos que esto es un artefacto de nuestra decisión de no permitir por el momento que nuevas mutaciones aporten variación fresca al proceso.

Antes de dar algo de formalidad y precisión al proceso de la deriva genética, observemos nuevamente uno de los paneles de la **Figura 2.1**. Ya hemos visto que las diferentes curvas pueden representar experimentos independientes en que se sigue la evolución de un locus o gen partiendo de las mismas condiciones (definidas por las frecuencias alélicas y el tamaño poblacional) y llevados adelante siguiendo las reglas del modelo de Wright-Fisher. Otra interpretación igualmente válida e instructiva es imaginar que cada curva representa, en realidad, un locus diferente, cuya historia se traza en la misma población. Para que esta interpretación sea válida, debemos asumir que: a) los loci son independientes, es decir no están ligados en el sentido genético del término (por ejemplo, cada locus está en un cromosoma autosómico diferente); y b) para cada locus, partimos de una idéntica frecuencia inicial del alelo que escogemos como marcador.

Nuestras dos interpretaciones se complementan para reflejar propiedades típicas de la deriva genética, a saber que un mismo locus tendría una historia particular cada vez que repitiésemos un experimento siguiendo su evolución, y que en un mismo experimento, cada locus hará lo propio describiendo su propia trayectoria. Ambas cosas son ciertas pese a que en todos los casos el punto de partida es el mismo y las reglas del proceso son idénticas.

### **Formalización del modelo de Wright-Fisher**

En esta sección, procederemos a formalizar algo más el proceso a efectos de consolidar las observaciones cualitativas ya realizadas y reafirmar, al hacerlo, las conclusiones generales sobre el efecto de la deriva genética en las poblaciones. Lejos de agotar las posibilidades en este sentido, nos limitaremos a examinar algunas propiedades del proceso en casos bien simples. Si estos son bien comprendidos, las extensiones a situaciones más complejas quedan al alcance de la mano.

Comencemos con un simple ejemplo de una población de tipo Wright-Fisher con solamente cuatro alelos de dos tipos diferentes<sup>2</sup>, digamos A y a, con frecuencias iniciales respectivas, fijadas arbitrariamente en 0.75 y 0.25. La siguiente generación se obtiene muestreando al azar con reposición a partir de este punto de partida. Escogiendo mentalmente el primer alelo, constatamos que tenemos dos posibles resultados con sus correspondientes probabilidades, que no son más que las frecuencias iniciales:

- Podemos escoger el alelo A, con una probabilidad  $p = 0.75$ .
- Podemos escoger el alelo a, con una probabilidad  $q = 0.25$ .

Es claro que la generación que formaremos, al constar solamente de cuatro alelos, puede conformarse únicamente por las siguientes combinaciones:

- 4A
- 3A y 1 a
- 2A y 2 a
- 1A y 3 a
- 4a

Al pensar sobre la probabilidad de llegar a cada uno de estos destinos, constatamos lo siguiente:

- Existe una única forma de obtener A cuatro veces, ya que ello requiere obtener dicho alelo en cada uno de los cuatro eventos. Como cada evento de muestreo tiene una probabilidad  $p = 3/4 = 0.75$  y es además independiente de los restantes, la probabilidad de obtener 4 A es:

$$p.p.p.p = p^4 = 0.316$$

- En el otro extremo, existe igualmente una única forma de obtener cuatro alelos a, solamente que la probabilidad final es ahora

$$q^4 = 0.004$$

Resolver por razonamiento, y si hace falta enumeración, las restantes posibilidades es tarea sencilla. Comencemos por reconocer que lo que hicimos en los dos casos anteriores fue: 1) multiplicar las probabilidades conocidas de obtener cada alelo de la combinación de interés, tomado de modo aislado; y 2) detenernos, al reconocer que había en cada caso un único modo de llegar a esa combinación. Para resolver los casos restantes, repetiremos el primer paso, y caeremos en la cuenta que requeriremos, además, un segundo paso

---

<sup>2</sup> <sup>3</sup> Seguimos aquí solamente las frecuencias alélicas, pasando por alto algunas complicaciones respecto a la exacta composición de los individuos diploides de ambos sexos. Para un excelente uso de esta opción minimalista, se recomienda estudiar el curioso ejemplo seleccionado por Gillespie (2004), que discute el caso de una población formada por un único individuo hermafrodita que se aparea consigo mismo al azar!

para dar cuenta de los varios modos de llegar a la combinación considerada. Así por ejemplo, para obtener 3 A y 1 a, tenemos que:

- Primer paso: una combinación como AAAa tiene una probabilidad de  $p.p.p.q$ .
- Segundo paso: existe un total de cuatro combinaciones, es decir cuatro formas de llegar a tener 3 A y 1 a, a saber: AAAa AAaA AaAA aAAA.

Cada una de esas cuatro combinaciones es igualmente probable. En suma, la probabilidad acumulada de llegar a tener 3 A y 1 a viene a ser  $4 p^3 q$ . El lector para el que estos razonamientos no resultan totalmente evidentes debería completar los casos restantes, incluyendo los cálculos numéricos, y verificar finalmente que la suma de las frecuencias de las cinco combinaciones posibles es, como corresponde, igual a 1.

Antes de generalizar estos razonamientos, veamos qué significan en la práctica para nuestra población de 4 alelos que comenzó con 3 alelos de una clase y 1 de otra. Observamos que:

- Todas las combinaciones de frecuencias alélicas son en principio factibles de ser obtenidas, en una generación, a partir de la generación inicial. Esto incluye la retención de la frecuencia inicial (en nuestro caso con una probabilidad de 0.422), la fijación de uno cualquiera de los alelos iniciales (el de mayor frecuencia inicial tiene mayor probabilidad de ser el que persista, como es de esperarse), y todas las restantes posibilidades.
- Al igual que en las gráficas de la **Figura 4.1**, la evolución por deriva continuará mientras no ocurra la fijación de uno de los alelos. A diferencia de la simple observación del fenómeno en la **Figura 2.1**, el razonamiento probabilístico nos permite establecer un espectro de posibilidades. De hecho, a partir de las frecuencias iniciales de nuestros dos alelos y del tamaño poblacional o número total de alelos en la población, podemos atribuir probabilidades precisas a cada uno de los posibles resultados del proceso en una generación.

Aunque transitar por los detalles de un ejemplo como el planteado es bueno como ejercicio de tanto en tanto, es también saludable reconocer que el caso particular se inscribe en un modelo general y sencillo, que nos facilitará en mucho el trabajo y nos permitirá examinar con comodidad cualquier caso semejante.

## La distribución binomial

Nuestro problema puede plantearse del siguiente modo. Tenemos un tamaño poblacional fijo de  $2N$  alelos, y los alelos son de dos clases (con frecuencia decimos que hay dos alelos para abreviar, confiando en que ello no damos a entender que  $2N=2$ ). Nuestro modelo de Wright-Fisher compone cada generación mediante el muestreo con reposición a partir de los alelos de la generación precedente. Ello implica, como hemos visto, que la probabilidad de muestrear una clase particular es fija e igual a su frecuencia en la generación precedente. Usamos  $p$  para designar la frecuencia de un alelo cualquiera y  $q$  (que por cierto debe ser igual a  $1-p$ ) para designar la del otro.

Vamos a definir una variable aleatoria, que llamamos binomial, que cuenta el número de

“éxitos”  $i$  como el número de casos en que es muestreado el primer alelo, en una serie de  $n$  ensayos, que en nuestro caso consta de  $n=2N$  alelos. De hecho,  $p$  y  $n$  son los dos parámetros de nuestra distribución, ya que  $q$  queda definido como  $1-p$ .

La distribución de probabilidad discreta de la binomial está dada por:

$$P(i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, \dots, n$$

donde

$$\binom{n}{i} = \frac{n!}{(n-i)! i!}$$

(conocido como el coeficiente binomial) representa las combinaciones de  $n=2N$  elementos tomados de  $a$   $i$ .

El lector puede recurrir al ejemplo más arriba para verificar cómo se obtienen las probabilidades razonadas en el ejemplo con esta ecuación. Además, cada parte de la ecuación describe un aspecto del muestreo:

- la segunda parte  $p^i(1-p)^{n-i}$  usa la “regla” de la multiplicación para calcular la probabilidad de observar una frecuencia cualquiera de cada clase de alelos (así, se reduce a  $p^4$  para el caso de cuatro alelos  $A$ ,  $p^3 \cdot (1-p)$  para tres  $A$  y un  $a$ , etc.).
- la primera parte, es decir la expresión combinatoria (llamada coeficiente binomial), simplemente cuenta de cuantos modos diferentes puede obtenerse un resultado cualquiera (así, hay un solo modo de observar cuatro alelos idénticos, cuatro formas de obtener tres  $A$  y un  $a$ , etc.).

La distribución binomial, dicho sea de paso, es extremadamente útil en muchos contextos. Un ejercicio sencillo que el lector puede hacer es examinar la relación entre frecuencias alélicas y genotípicas del capítulo anterior sobre equilibrio Hardy-Weinberg en términos de esta distribución.

En resumen, mediante la distribución binomial podemos ser más explícitos respecto a las fluctuaciones de frecuencias alélicas que constituyen la deriva genética, y precisar las observaciones cualitativas respecto a la relación entre dichas fluctuaciones y el tamaño poblacional (y el correspondiente número de alelos).

Puede también apreciarse que, pese a la sencillez esencial del modelo binomial, el trabajo de seguir el espectro de posibilidades (las trayectorias evolutivas) de una población cualquiera se vuelve engorroso si el tamaño poblacional es grande y si consideramos varias, y no solamente una generación. Para cada generación debemos calcular la distribución esperada, y luego usar cada uno de los resultados posibles y considerar, a su vez, a qué puede conducir una generación más adelante. Finalmente, deberíamos combinar esos nuevos espectros parciales, y así sucesivamente.

Afortunadamente, ese tedioso proceso no es necesario para comprender cabalmente la deriva genética. Además de las conclusiones cualitativas que ya hemos extraído, los

genetistas de poblaciones recurren a dos simplificaciones que rescatan la esencia del proceso ahorrándonos buena parte de las complicaciones:

En primer lugar, Kimura hizo uso de una ecuación de difusión para aproximar de un modo sencillo el espectro de trayectorias dictado por la distribución binomial. Ya en 1922 Fisher había notado que el efecto de “difusión” que la deriva genética causa en las frecuencias alélicas era análogo al proceso de difusión del calor en un sólido. La solución implica aproximar una distribución discreta (como la binomial, que usamos para conteos de 1, 2,...2N alelos) mediante una continua. Dicha aproximación está fuera de nuestro alcance, pero se comporta muy bien, y tanto mejor cuanto mayor el tamaño poblacional. El libro de Hartl y Clark (1997) compara una solución binomial con la correspondiente aproximación por difusión, y ambas con experimentos de poblaciones de la mosca de la fruta.

En segundo lugar, como ya sugerimos en el capítulo anterior, se puede trazar el curso de un valor global que resume las características de la población, típicamente la heterocigosidad, en lugar de pretender seguir el espectro completo. Examinaremos esta opción muy brevemente a continuación.

### **Deriva genética y heterocigosidad**

Ya hemos insistido hasta el cansancio en la noción de que la deriva genética resulta en (simplemente consiste en) fluctuaciones aleatorias de las frecuencias alélicas. También hemos visto, y conviene destacarlo ahora, que el resultado inexorable del proceso es la fijación o eliminación de uno de los tipos alélicos de la población. Es tan obvio como importante destacar que una vez que se llega a una de estas dos eventualidades el proceso se detiene, ya que el mismo ha agotado la variación sobre la que opera. Sólo la mutación, que no hemos autorizado todavía, puede rescatar al locus que ha llegado a la fijación de un alelo de este estado de inercia evolutiva.

Aprovechemos este momento para insinuar apenas una descripción del modelo de Wright-Fisher como lo que se conoce como cadena de Markov. El tipo de cadena que nos interesa es un proceso que tiene un número discreto de estados (0, 1, 2..., 2N copias del alelo que usamos como marcador); el estado que se alcanza en un tiempo cualquiera, digamos en nuestro caso la décima generación ( $t_{10}$ ), depende únicamente del estado en el tiempo inmediatamente precedente (la generación  $t_9$ ) y de probabilidades de transición de un estado a otro, que son fijas.

En nuestro caso, vemos como efectivamente el espectro de probabilidades para una generación depende solamente del estado en la generación precedente, simplemente caracterizado por las frecuencias alélicas, y de las probabilidades de transición claramente establecidas por la distribución binomial.

Nuestra cadena de Markov tiene dos estados absorbentes, en los que el proceso se detiene por siempre, a saber, precisamente los casos de fijación de uno u otro alelo. Lo importante es que, mientras las frecuencias intermedias, en las que un alelo cualquiera tiene entre 1 y 2N-1 copias, le permiten a la población seguir fluctuando en su composición, los estados absorbentes son trampas de captura definitiva.

Ya hemos visto este fenómeno en la **Figura 3.1**, pero sin enfatizar tanto el fenómeno de absorción. Cuanto más tiempo transcurre, mayores son las probabilidades de observar

dicha absorción en un experimento dado, real o simulado. Y cuanto más experimentos independientes realicemos en idénticas condiciones, mayores chances tendremos entonces de observar algunas fijaciones.

Si nuestra intuición probabilística está ya algo desarrollada, entenderemos que:

- La fijación de algún alelo de los presentes inicialmente en un tiempo dado será tanto más probable cuanto más pequeña sea la población;
- La fijación tenderá también a ocurrir más rápidamente si comenzamos con frecuencias iniciales extremas. Es decir, si nuestro alelo marcador comienza con una frecuencia de 0.99, o de 0.01, está muy cerca ya de un estado absorbente (frecuencia 1 o 0), mientras que si comenzamos con una frecuencia de 0.5, estamos a distancia más saludable de un estado absorbente al que, sin embargo, estamos condenados más tarde o más temprano.

Centrando nuestra atención en la heterocigosidad, como medida sintética de la variación de nuestra población, llegamos a la conclusión que tenderá a reducirse por causa de la deriva. Esta tendencia es, una vez más, probabilística. La heterocigosidad puede de hecho aumentar si la frecuencia de nuestro alelo marcador, por azar, sube de 0.4 a 0.5 en una generación. Pero mientras esta tendencia puede revertirse, la fijación o eliminación son irreversibles. Si seguimos muchos loci independientes, todos ellos autosómicos, a lo largo del tiempo y promediamos la heterocigosidad para seguir su historia, observaremos la tendencia a la reducción. Lo mismo ocurrirá si seguimos un mismo locus a lo largo del tiempo en un gran número de poblaciones, ya que estas dos clases de experimentos son completamente análogas.

Recurriremos al ya recomendado ejemplo de Gillespie (2004) de una población compuesta por un único individuo heterocigota (Aa) en una generación inicial en el tiempo  $t_0$  para ilustrar el ritmo esperado de caída de la heterocigosidad. Este individuo pertenece a una población de tipo Wright-Fisher, o sea que entre otras cosas se aparea al azar consigo mismo, por lo que debe ser hermafrodita. El espectro de trayectorias posibles de unas pocas generaciones está dado por:

Genotipos	generaciones			
	$t_0$	$t_1$	$t_2$	$t_3$
AA	0	1/4		
Aa	1	1/2	1/4	1/8
aa	0	1/4		
E(H)	1	1/2=0.5	0.25	0.125

El lector puede facilmente completar los casilleros vacíos, y constatar que la heterocigosidad esperada E(H) es la expresada en la tabla. El resultado del proceso es bien simple: la heterocigosidad esperada se reduce, en cada generación, a la mitad del valor de la generación precedente. Este resultado drástico se debe, naturalmente, al pequeño tamaño poblacional ( $N=1$ ,  $2N=2$ ) que exagera los efectos de la deriva. Nótese que las frecuencias genotípicas en  $t_1$  son las esperadas en un cruzamiento mendeliano

simple de heterocigotas, como corresponde. Nótese también, luego de completar los casilleros vacíos, como aumentan las frecuencias de los estados absorbentes a expensas del único estado intermedio.

La reducción a la mitad de la heterocigosidad esperada en el ejemplo anterior es un extremo que resulta del pequeño tamaño poblacional. Para generalizar esta observación, comenzaremos por introducir el concepto de endogamia o endocría. Recordemos que nuestro modelo poblacional de Wright-Fisher establece una regla clara (la de muestreo al azar con reposición) para componer a los individuos de una generación a partir de los alelos presentes en la generación precedente.

Consideremos ahora una población típica, con  $2N$  alelos y  $N$  individuos diploides, pero que en un tiempo  $t_0$  la formamos de modo tal que los  $2N$  alelos son todos diferentes. En una población así constituida no existe endogamia, definida ésta como la probabilidad de que dos alelos de un individuo cualquiera sean idénticos por descender de un ancestro común. Es precisamente esta posibilidad la que hemos evitado al construir, artificialmente, una población sin alelos idénticos por descendencia.

Aún en este caso, la deriva genética generará endogamia, o sea individuos compuestos por dos alelos idénticos por descendencia. Consideremos un individuo cualquiera de la generación  $t_1$ , que formamos según la regla del muestreo al azar con reposición a partir de nuestra artificiosa generación inicial. Si elegimos mentalmente un alelo cualquiera de  $t_0$  como el primero de su genotipo, el segundo puede resultar:

- Idéntico al primero con una probabilidad de  $1/2N$ . En este caso tenemos endogamia, representado por el primer individuo homocigota que aparece en la población.
- Diferente al primero en los restantes casos, es decir con una probabilidad de  $1 - (1/2N)$ .

La endogamia, que como se ve resulta en homocigosidad, concepto al que está estrechamente ligada, ha aumentado si caemos en el primer caso, pero no en el segundo. Siguiendo a Sewall Wright, usaremos la letra  $F$  para denotar el coeficiente de endogamia, que es el valor esperado de la endogamia. En nuestro caso, y para la generación en el tiempo  $t_1$ ,  $F$  será el promedio de las expectativas que acabamos de definir. Para calcular el valor esperado, tenemos en cuenta que  $F$  será 1 con una probabilidad de  $1/2N$  y 0 con una probabilidad de  $1 - (1/2N)$ , con lo que se reduce a:

$$F_1 = 1/2N$$

Pero nuestra generación de partida en  $t_0$  fue artificiosa en cuanto no tenía alelos idénticos por descendencia. En la práctica, sabemos que una población siempre tiene alelos de este tipo. De hecho, acabamos de ver que ellos se generarán inexorablemente a un ritmo de  $1/2N$  por generación. Si en lugar de forzar la total ausencia de endogamia en  $t_0$  aceptamos una endogamia inicial  $F_0$ , nuestro valor esperado para  $t_1$  se transforma en:

$$F_1 = 1/2N + (11/2N)F_0$$

La contraparte del aumento de la endogamia es, naturalmente, la reducción de la heterocigosidad esperada, ya que los dos conceptos son complementarios. Podemos por

tanto pasar de uno a otro valor de modo sencillo:

$$1-F_1 = 1 - [1/2N + (1/2N)F_0]$$

$$E(H_1) = (1/2N) H_0$$

El proceso de endogamia y la consiguiente pérdida de heterocigosidad ocurre del mismo modo cada generación, por lo que, generalizando para  $t$  generaciones, tenemos:

$$E(H_{Dt}) = H_{D0} (1 - 1/2N)^t$$

En el caso extremo ya examinado  $N=1$  por lo que la heterocigosidad esperada se reduce a la mitad de una generación a la siguiente. Cuanto mayor sea  $N$ , más tenue será la reducción de la heterocigosidad esperada en función del tiempo.

### **“Ritmo” de la deriva genética y tamaño poblacional efectivo**

En el modelo de Wright-Fisher que hemos considerado hasta ahora, el “ritmo” de la deriva genética está dictado por el tamaño poblacional  $N$ . El modelo en realidad se basa en la regla de muestreo de los alelos de una generación a la siguiente, pudiendo inclusive prescindirse de la fase diploide (los genotipos de los individuos) y centrar la atención exclusivamente en la fase gamética, caracterizada por las frecuencias alélicas. En este sentido, el “número efectivo” de alelos que nos interesa tener en cuenta es  $2N$ .

El concepto de tamaño o número poblacional efectivo no ofrece nada nuevo si nos ceñimos al modelo de Wright-Fisher, pero permite establecer ciertas equivalencias entre éste y otros modelos de los procesos demográficos. Es claro que el modelo puede fácilmente ajustarse para considerar genes ligados al cromosoma X o al cromosoma Y, así como también a otros casos particulares, como el genoma mitocondrial, de transmisión casi exclusivamente materna en muchos animales. A modo de ejemplo, en este último caso el número de individuos se reduce al número de hembras, o sea  $N/2$  si los sexos están igualmente representados en la población, y puesto que cada hembra típicamente tiene una única clase de genoma mitocondrial, el número de alelos también será  $N/2$ . Es decir que en una población como las consideradas hasta ahora, el ADN mitocondrial se comporta como lo predice el modelo de Wright-Fisher para una población de  $N_e = N/2$ . Simplemente el número efectivo de alelos es la cuarta parte del de un locus diploide.

El tamaño poblacional efectivo es, como se ve, un modo conveniente de establecer una equivalencia entre un modelo genético poblacional cualquiera y el modelo de Wright-Fisher. Ahora bien, es importante tener en cuenta que dicha equivalencia no lo abarca todo, sino que se establece respecto a algún aspecto interesante de los modelos. No profundizaremos en la discusión de otros modelos y sus equivalencias, pero estas notas deben servir para advertir al lector de la existencia de estas complicaciones. Nos interesa simplemente señalar algunos efectos de variaciones del modelo demográfico sobre el tamaño efectivo a modo ilustrativo.



Sewall Wright estableció los tamaños poblacionales efectivos de varios modelos que se desvían del considerado hasta aquí. La equivalencia que Wright procuró fue la de calcular un  $N_e$  tal que la población considerada se comportase, en cuanto a su endogamia, como una población Wright-Fisher estándar.

Por ejemplo, puede considerarse cuál será, en este sentido, el tamaño efectivo de una población en la que la proporción de sexos difiere de la igualdad. Los cabañeros, por ejemplo, usan uno o unos pocos machos como reproductores, en combinación con un número generalmente mucho mayor de hembras. En este caso, el tamaño poblacional efectivo es:

$$N_e = \frac{4N_f N_m}{N_f + N_m}$$

donde los subíndices m y h representan a los machos y hembras, respectivamente. Jugando con ejemplos numéricos, el lector puede convencerse de que el tamaño efectivo de un rebaño con gran desbalance entre machos y hembras será notoriamente menor que el número total de individuos reproductores en la población. Esto es sumamente importante para un criador que quiere evitar los efectos negativos de la endogamia (que mencionaremos brevemente cuando hablemos de mutaciones deletéreas).

Otro ejemplo muy nombrado es el de poblaciones que fluctúan en su tamaño, mientras que retienen las restantes características del modelo de Wright-Fisher. Imaginemos a una población que se reproduce dos veces al año en condiciones bien diferentes, de modo tal que el número de individuos reproductores es alto en una estación y bajo en la siguiente, y que esta alternancia se repite año a año. En este caso, considerando una serie de t

$$1/N_e = 1/t (1/N_1 + 1/N_2 + \dots + 1/N_t)$$

El tamaño efectivo en este caso es la media armónica de los tamaños de las generaciones consideradas. Una propiedad de la media armónica es que siempre resulta menor que la media aritmética. De hecho, examinando algunos ejemplos numéricos el lector puede convencerse que el tamaño efectivo se verá más afectado por un ocasional N pequeño que por un ocasional N grande.

Estos ejemplos bastan para notar que, por lo general, el tamaño poblacional efectivo es notoriamente menor que el que un simple conteo de individuos (aún eliminando los que no son adultos reproductores) sugeriría. Y esto tiene sentido, ya que tanto el sesgo de composición sexual como las fluctuaciones demográficas generan endocría, el primero porque unos pocos machos tienen un gran efecto en la composición genética de la población, y las segundas porque las fases de tamaños poblacionales pequeños representan momentos de gran impacto de la deriva genética.

La tendencia de las poblaciones a tener tamaños poblacionales efectivos menores a los que un simple censo de adultos sugeriría es a veces sutil. Para ilustrarlo, recordemos la recomendación de comparar tres casos de deriva ilustrados en Hartl y Clark (1997): una solución basada en la distribución binomial, otra basada en la ecuación de difusión, y otra obtenida experimentalmente con poblaciones de *Drosophila* diseñadas para adherirse al

modelo de Wright-Fisher, incluyendo una proporción de sexos de 1:1. Sin embargo, como notan Hartl y Clark (1997), ha sido posible demostrar que las poblaciones experimentales resultaron en un mayor número de fijaciones (estados absorbentes) a expensas de frecuencias intermedias de los alelos de interés. En otras palabras, ¿el tamaño efectivo de estas poblaciones parece haber sido menor al que se intentó construir! Al parecer, existe suficiente sesgo en el éxito reproductivo de los individuos como para que los apareamientos dejan de ser al azar. Estas variaciones en la reproducción no se vinculan al polimorfismo estudiado, que sigue siendo neutro a todos fines prácticos y se rige por la deriva genética. Pero el tamaño poblacional efectivo es menor al tamaño censal por causa de las quisquillosidades reproductivas de las *Drosophila*. En dichos experimentos, de hecho, la pérdida promedial de heterocigosidad se parece a la que sufriría una población ideal de tipo Wright-Fisher no de 16, sino de 9 individuos (¿4.5 de cada sexo?; no, simplemente 18 en lugar de 32 alelos).

Vamos a concluir estas consideraciones sobre la deriva genética con una cuestión trivial de simbología, y es el constatar que carecemos de un símbolo específico (¿ $d$ ?) para representarla en las ecuaciones de la genética de poblaciones. En esto, la deriva contrasta con la mayoría de los procesos genético-poblacionales respetables, que tienen parámetros asociados (el coeficiente de selección  $s$ , la tasa de migración  $m$ , la tasa de recombinación  $r$ , etc.). La razón es en realidad tan sencilla e importante que sirve para recapitular lo estudiado en este capítulo, y es que la deriva genética está asociada al tamaño poblacional, o más concretamente al tamaño poblacional efectivo, y no requiere de otro parámetro para ser representada. Por lo tanto, la aparición de los símbolos  $N$  y  $N_e$  deben hacernos recordar que la acción de la deriva está siempre presente en la población.

### **Lecturas recomendadas**

Hartl, D. L., y A. G. Clark. 1997. Principles of population genetics. Sinauer Assoc., Sunderland, MA. Un excelente libro de texto. Se recomienda el capítulo 7 sobre deriva genética en general, y las explicaciones y gráficas sobre el proceso de deriva en particular. Este se presenta de tres formas complementarias, que son el desarrollo binomial de un caso simple, su aproximación mediante la ecuación de difusión utilizada por Kimura, y un caso experimental.