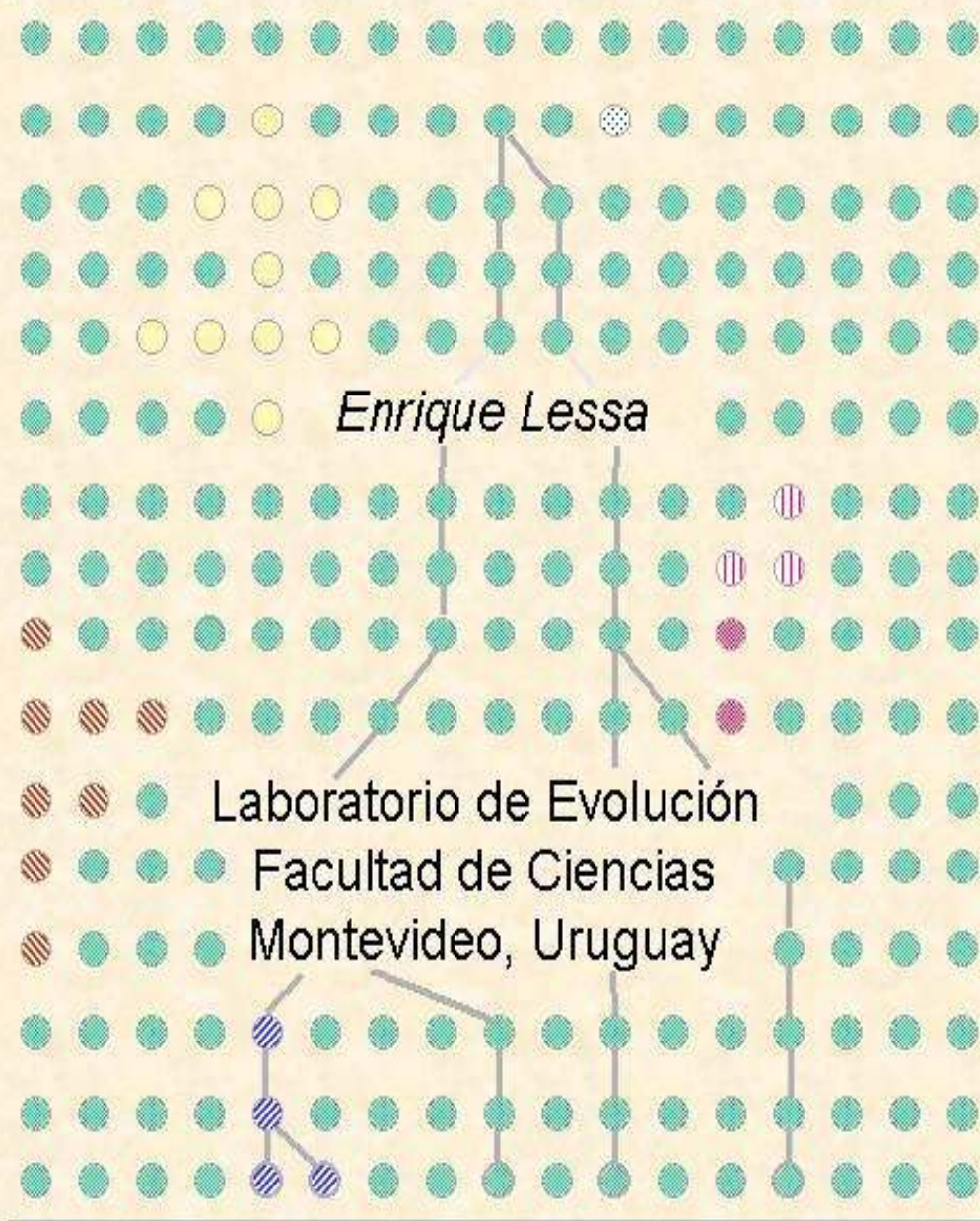


Guía de estudio de genética de poblaciones



lessa@fcien.edu.uy
Casilla de Correos 12106, Montevideo 11300, URUGUAY

TABLA DE CONTENIDOS Y MATERIALES (Agosto 2004)

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
- Homocigosis y heterocigosis esperadas
- Aproximación al equilibrio de Hardy-Weinberg
- Lecturas recomendadas

3. Deriva genética

- La deriva genética
- Fluctuaciones de las frecuencias alélicas
- Formalización del modelo de Fisher-Wright
- La distribución binomial
- Deriva genética y heterocigosis
- “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 2003 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), 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 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 aprehenden 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 en 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 (3×10^9) 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. Pero una estimación conservadora es que la probabilidad de que ocurra una mutación puntual que afecte a un par de bases cualquiera es de uno en un millón (un mutante cada 10^6 pares de bases), o 10^{-6} . Por lo tanto, esperamos que cada nuevo genoma exhiba unas tres mil nuevas mutaciones ($3 \times 10^9 \times 10^{-6} = 3 \times 10^3$). 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 un millón 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 millón 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 betaglobina, 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 betaglobina, 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 un millón 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 $500 \times 10^{-6} = 5 \times 10^{-4}$.

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 encontrar una mutación, o su complemento, el no encontrarla. Un caso análogo sería el encontrar un determinado alelo, o el o 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:* Cuando combinamos distintos eventos, tenemos que calcular probabilidades combinadas simples, que suelen resolverse con la regla de la suma, o la regla de la multiplicación. Hemos aplicado la primera al ver cómo aumenta nuestra probabilidad de observar una mutación al pasar de un único sitio en el ADN a una región de, digamos, 500 sitios. La probabilidad elemental para un sitio se suma a la de los restantes, y como hemos asumido que todos los sitios son equivalentes a estos efectos, ello equivale a multiplicar la probabilidad por sitio por 500. Utilizamos igual razonamiento para calcular cuantas mutaciones serían esperables en el genoma entero. Mentalmente caminamos a lo largo de una cadena de ADN y tomamos nota, para cada sitio, de si aparece mutado o incambiado respecto al ADN progenitor.
- *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) es de 10%, o de 0.1. ¿Qué probabilidad tengo de observar *dos* de estos alelos 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% da una probabilidad de 1%, o sea $0.1 \times 0.1 = 0.01$. Nótese que el mismo caso puede ser examinado utilizando la regla de la suma, pero con una finalidad ligeramente diferente. Si me pregunto qué probabilidad tengo de observar el alelo en cuestión *una vez* en mi muestra de dos, cada una me da una chance, así que debo sumarlas de modo que $0.1 + 0.1 = 0.2$.

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 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.
1920-1930	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 (<i>polymerase chain reaction</i>) Kingman define el coalescente como modelo de análisis genético-poblacional.

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. ¡Ni aún en un período tan reciente era posible secuenciar el ADN!

Lecturas recomendadas

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.*

Hartl, D. L. 2000. A primer of population genetics. Sinauer Assoc., Sunderland, MA. *Un libro excelente en general. En particular, el capítulo 1 comienza con una concisa descripción de la estructura del ADN, el código genético, y de las principales técnicas de estudio de variación genética.*

Avise, J. C. 1994. Molecular markers, natural history and evolution. Chapman & Hall, New York. *Un libro excelente en general, y en particular recomendable por su descripción de las principales técnicas de estudio de la variación genética y por contener una panorámica de las múltiples aplicaciones de la genética de poblaciones experimental, desde el estudio del parentesco hasta el de la evolución.*

Capítulo 2: EQUILIBRIO 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, éste 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 suele 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* 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 propiedades 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 atisbar desde ya la utilidad del modelo al considerar 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:

$$\begin{array}{l} f(V) = p = 0.376 = 37.6\% \\ f(v) = q = 0.624 = 62.4\% \end{array}$$

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 estas frecuencias 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 surge 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

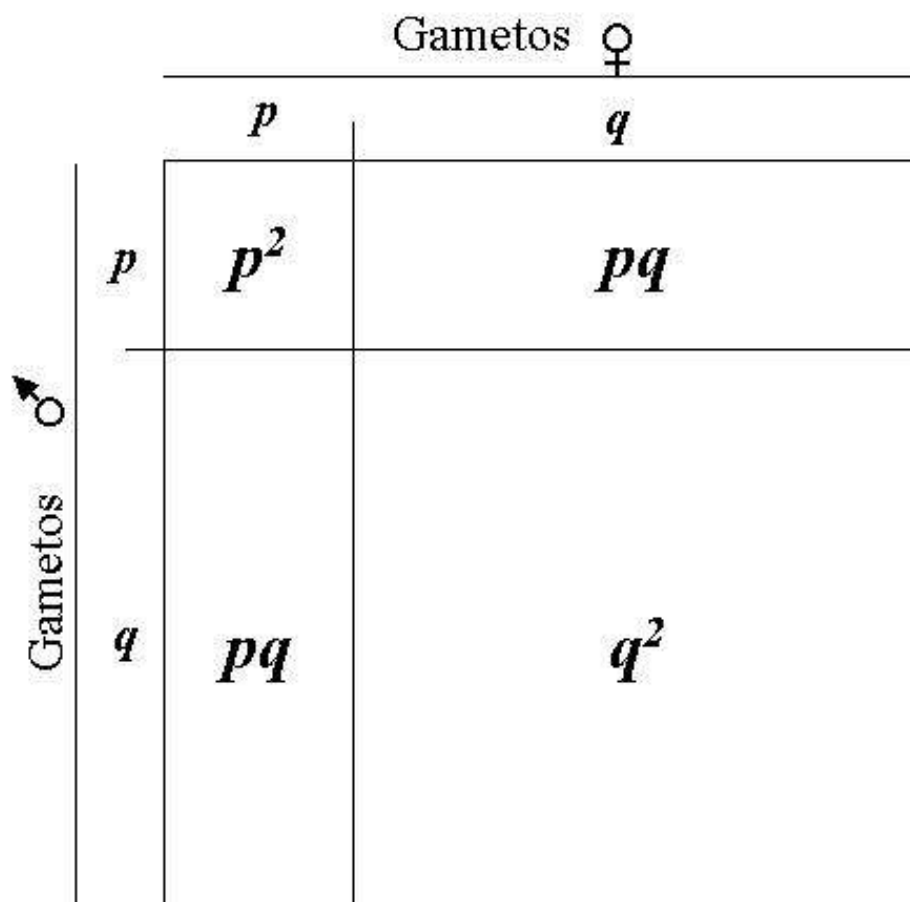
$$f(VV) = 0.376 \times 0.376 = 0.141 = 14.1\%$$

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 .

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.



- 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 $f(Vv) = 2pq$.

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:

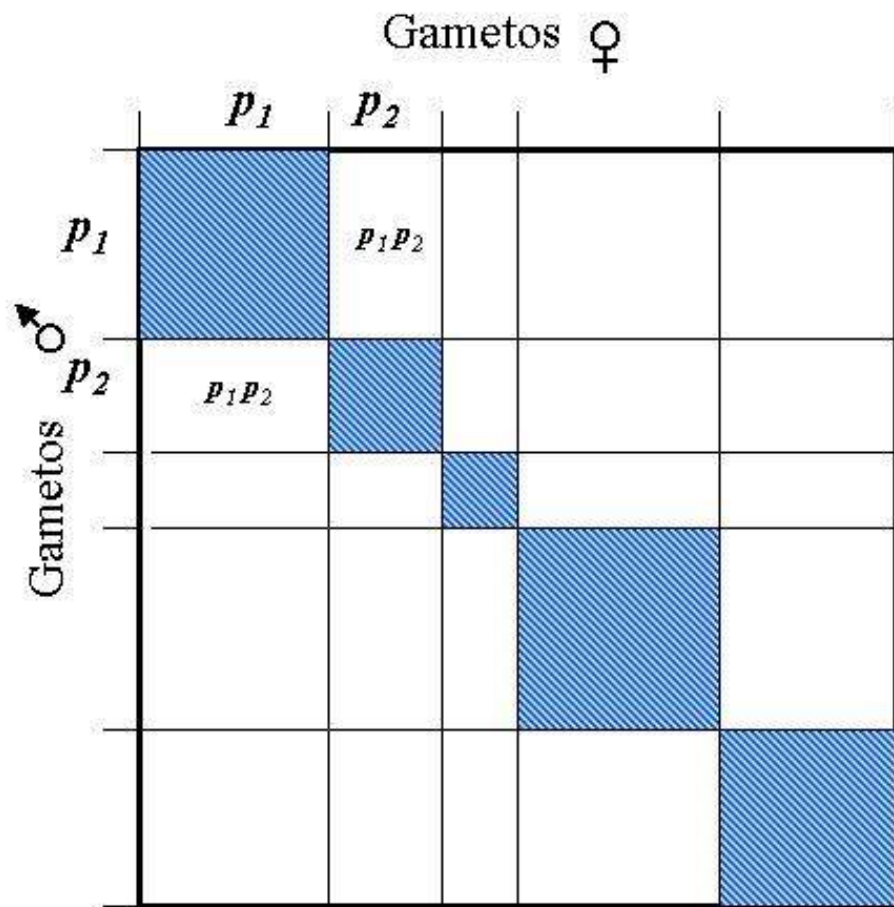
$$\begin{aligned} f(VV) &= 0.141 \\ f(Vv) &= 0.470 \\ f(vv) &= 0.389 \end{aligned}$$

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 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 desplegar sus frecuencias en 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.

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 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

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).



correspondiente (cambios silenciosos, en lugar de cambios que implican el remplazo de un aminoácido).

Cuadro 2.1¹. 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 absolutas observadas	Frecuencias genotípicas absolutas esperadas
MM	298	
MN	489	
NN	213	
Muestra total	1000	1000

Frecuencia estimada del alelo M =
 Frecuencia estimada del alelo N =

La comparación de las frecuencias esperadas (*esp*) y observadas (*obs*) sugiere que las diferencias entre ellas son pequeñas, y probablemente atribuibles a errores de muestreo. La prueba de chi cuadrado (cuyos fundamentos no se detallan) 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. 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 $\chi^2 = 3.84$.

$$\chi^2 = \sum \frac{(obs - esp)^2}{esp}$$

Supongamos que un estudio de este tipo determina que el alelo 1 clásico es en realidad una mezcla de dos subtipos, indistinguibles a nivel proteico, que llamaremos 1a y 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.

¹ Ejemplo tomado de Hartl y Clark (1997).

Por ejemplo, el cuadrado que representa la frecuencia de homocigotas para el alelo clásico 1 puede extraerse del cuadro general y subdividirse según las frecuencias de 1a y 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.

Homocigosis y heterocigosis 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 homocigosis esperada, es decir la frecuencia de homocigotas esperados en equilibrio Hardy-Weinberg. Si llamamos p_i (p_1, p_2 , y así sucesivamente) a las frecuencias de los distintos alelos, entonces la homocigosis esperada es:

$$E(G) = \sum p_i^2$$

En la Figura 2.2 puede verse claramente que la homocigosis 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 heterocigosis esperada está representada en la Figura 2.2 por el área no cubierta por las diferentes combinaciones homocigotas, es decir:

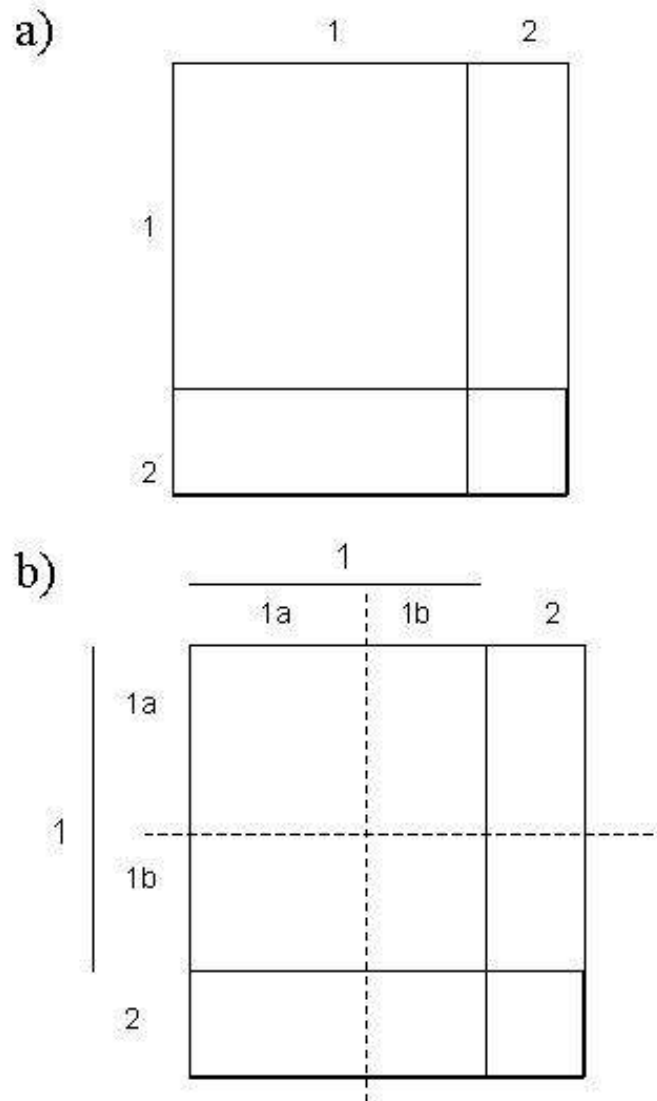
$$E(H) = 1 - E(G) = 1 - \sum p_i^2$$

Recurriendo a la Figura 2.1, puede verse fácilmente que, para el caso de dos alelos, la máxima heterocigosis 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 heterocigosis esperada posible para tres o más alelos.

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.

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 fundamos una población de *Drosophila* con estas características:

Alelo	frecuencia en hembras	frecuencia en machos	frecuencia total
A	1.00	0.00	0.50
a	0.00	1.00	0.50

En otras palabras, cada sexo es homocigota para un alelo. La pregunta es en cuantas 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. 1998. Population genetics—a concise guide. Johns Hopkins Univ. 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 GENETICA

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. 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.

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.

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 alozímicas 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.

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 solamente difiere del de Hardy-Weinberg en el tamaño finito de la población, es conocido como modelo de Fisher-Wright, 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.

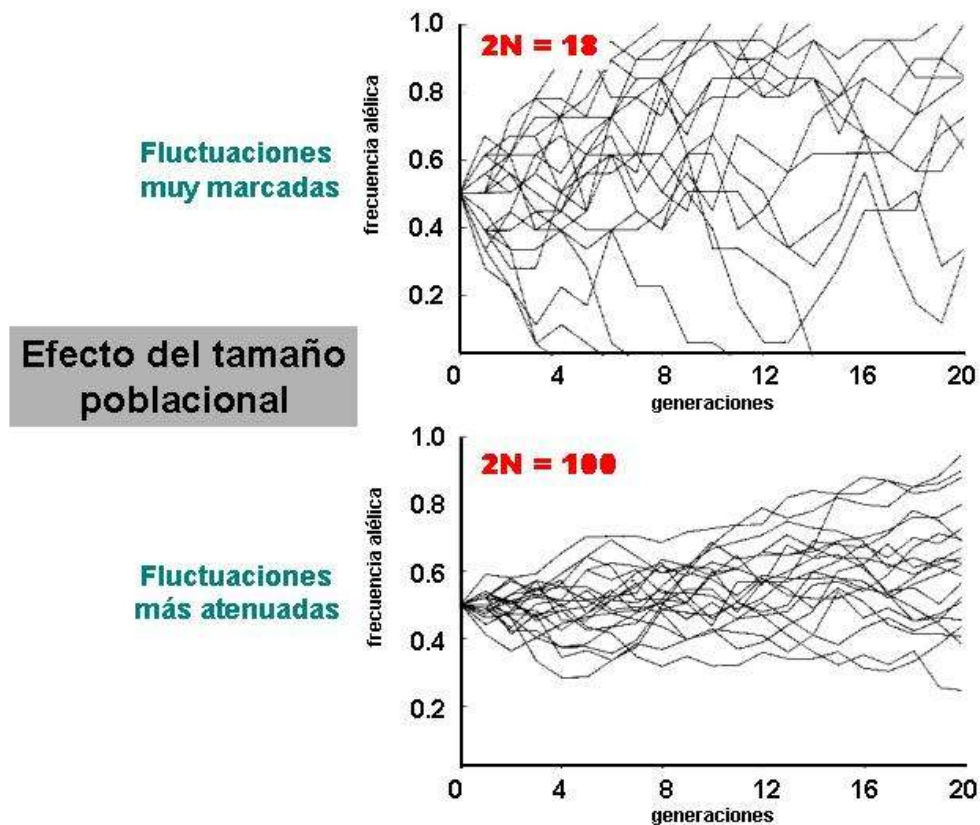
Fluctuaciones de las frecuencias alélicas

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 cigotos 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, y pese a su inconstancia, 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 Fisher-Wright. 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 Fisher-Wright

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 Fisher-Wright con solamente cuatro alelos² de dos tipos diferentes, 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:

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

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:

$$p \times p \times p \times p = p^4 = 0.316$$

- 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

² 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 (1998), que discute el caso de una población formada por un único individuo hermafrodita que se aparee 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^3pq .
- 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, ya que su probabilidad está dada por el producto p^3pq . En suma, la probabilidad acumulada de llegar a tener 3 A y 1 a viene a ser $4p^3q$. 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 indica que $2N=2$). Nuestro modelo de Fisher-Wright 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

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

que 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 , $p^3 \times q$ para tres A y un a , etc.).
- la primera parte, es decir la expresión combinatoria, 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 heterocigosis, en lugar de pretender seguir el espectro completo. Examinaremos esta opción muy brevemente a continuación.

Deriva genética y heterocigosis

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 Fisher-Wright 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 heterocigosis, 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 heterocigosis 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 heterocigosis 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 (1998) de una población compuesta por un único individuo heterocigota (Aa) en una generación inicial en el tiempo t_0 para ilustrar el ritmo de caída de la heterocigosis. Este individuo pertenece a una población de tipo Fisher-Wright, 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 fácilmente completar los casilleros vacíos, y constatar que la heterocigosis esperada $E(H)$ es la expresada en la tabla. El resultado del proceso es bien simple: la heterocigosis 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 exacerba 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 heterocigosis 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 endocria. Recordemos que nuestro modelo poblacional de Fisher-Wright 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 homocigosis, 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:

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

La contraparte del aumento de la endogamia es, naturalmente, la reducción de la heterocigosis esperada, ya que los dos conceptos son complementarios. Podemos por tanto pasar de uno a otro valor de modo sencillo:

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

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

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

$$E(H_t) = H_0 \left(1 - \frac{1}{2N}\right)^t$$

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

“Ritmo” de la deriva genética y tamaño poblacional efectivo

En el modelo de Fisher-Wright 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 Fisher-Wright, 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 Fisher-Wright 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 Fisher-Wright. 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 Fisher-Wright 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:

$$N_e = \frac{4N_m N_h}{N_m + N_h},$$

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 Fisher-Wright. 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 generaciones y comenzando con t_0 el tamaño efectivo cumple la siguiente ecuación:

$$\frac{1}{N_e} = \frac{1}{t} \left(\frac{1}{N_0} + \frac{1}{N_1} + \dots + \frac{1}{N_{t-1}} \right)$$

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 Fisher-Wright, 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 heterocigosis se parece a la que sufriría una población ideal de tipo Fisher-Wright 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.*

Capítulo 4: MUTACION

La mutación es la fuente última de variación y, por lo tanto, de la evolución biológica. En este capítulo vamos a realizar algunas consideraciones sobre la mutación, principalmente sobre el modo como se la incorpora a los modelos genético-poblacionales.

A comienzos del siglo XX, cuando se redescubrieron las leyes de Mendel, la naturaleza discreta de los alelos mendelianos clásicos (las arvejillas eran verdes o amarillas, sin intermedios, y los alelos mantenían su identidad a través de la herencia), los genetistas creyeron que esa discontinuidad entraba en flagrante contradicción con el darwinismo, que insistía en la evolución lenta y gradual. Surgió así el mutacionismo, como una corriente de pensamiento evolucionista que pensaba que el proceso estaba fundamentalmente dictado por la naturaleza de las mutaciones.

Los trabajos de Fisher, Haldane, y Wright, entre otros, se ocuparon de asimilar el mendelismo a la teoría de la evolución de Darwin. Así, el aparente conflicto fue resuelto al reconocer que el proceso evolutivo consta de dos etapas. La *génesis* de la variación reside en el proceso mutacional, cuya diversidad ha sido comprendida de modo creciente desde esa época hasta nuestros días. El *destino* de dicha variación, sin embargo, se define en el juego genético-poblacional que incluye, por lo pronto, al proceso darwinista de la selección natural, pero también otros como la deriva genética, cuyo papel ha sido esbozado en el capítulo anterior. Fisher, particularmente, esbozó de qué modo caracteres fenotípicos continuos, como la altura o la masa corporal, pueden tener sus bases genéticas en alelos mendelianos discretos. La solución, que no consideraremos en detalle en este momento, reside en reconocer que un carácter continuo puede reflejar los efectos acumulativos de variación genética discreta en numerosos genes.

En una celebrada polémica que tuvo lugar en la época de gestación de la teoría sintética, Ronald Fisher y Sewall Wright discreparon sobre la importancia de la deriva genética. Es que Wright, sin desconocer el papel de la selección natural, insistía en que la deriva genética cumplía también un papel importante. Esta visión fue la semilla de la teoría neutralista de la evolución molecular, que Motoo Kimura desarrolló en la década de 1960, y que consideraremos brevemente más adelante.

El reconocimiento de que los procesos mutacionales son muy variables tuvo que esperar, en lo esencial, al advenimiento de la genética molecular, aunque los trabajos de Barbara McClintock sobre elementos transponibles se adelantaron a ella, con métodos clásicos, por décadas. Las mutaciones más sencillas son los cambios puntuales en un nucleótido, que si persisten implicarán también un cambio en la hebra complementaria durante la replicación del ADN. Pero son también mutaciones los cambios cromosómicos, como las inversiones y translocaciones, y las transposiciones como las estudiadas por McClintock. Una importante clase de loci llamados

microsatélites, de amplio uso en análisis de parentesco, sufre mutaciones que resultan en el agregado o la eliminación de motivos de repetición. Así, por ejemplo, un alelo que contiene un motivo GAC repetido 50 veces, puede generar por mutación un alelo con un número mayor o menor de repetidos GAC. Aún el estudio de las mutaciones puntuales, al que ceñiremos la mayor parte de nuestras discusiones, tiene su complejidad, ya que diferentes clases de cambios pueden tener distintas probabilidades de ocurrir.

El presente capítulo está organizado del siguiente modo. En primer lugar, se pasa revista a algunas clases de mutaciones diferentes, a efectos de ilustrar su diversidad sin agotarla. En segundo lugar, se recorren algunos modelos de mutación en una secuencia que es, en lo esencial, histórica y que refleja el nivel de conocimientos y el tipo de datos que, sobre la herencia y sus bases materiales, existía en cada época. Algunos resultados de la genética de poblaciones sobre la mutación y su papel en las poblaciones naturales se presentan en función de esos modelos históricos. Otros resultados importantes, sin embargo, serán presentados en capítulos posteriores.

En este capítulo no consideraremos tampoco la recombinación, que contribuye a la diversidad genética de las poblaciones al producir distintas combinaciones de las variantes generadas por mutación.

Sesgo mutacional: diferencias entre los sexos

En un muy recomendable artículo de revisión, Crow (2000) da cuenta de cómo, ya en la década de 1940, J. B. S. Haldane propuso que el patrón de aparición de la hemofilia, una enfermedad determinada por factores localizados en el cromosoma X, implicaba una tasa mayor de mutación en hombres que en mujeres. Esta temprana observación ha sido confirmada para un número importante de enfermedades.

El patrón de mutación, sin embargo, es complejo y muestra importantes variaciones. La tendencia general es que las mutaciones puntuales parecen estar sesgadas y ocurrir con mucho mayor frecuencia en hombres que en mujeres. Ello está ligado, al menos en varios casos, a la edad, es decir que las mutaciones puntuales son más frecuentes en los hombres a medida que aumenta su edad, lo que no ocurre en las mujeres.

En claro contraste con esta tendencia, las alteraciones cromosómicas (una vez más, al menos algunas bien documentadas), tienen mayor frecuencia, y también aumentan con la edad, en mujeres. El caso mejor conocido es la trisomía del cromosoma 21, que causa el síndrome de Down. Otros tipos de cambios, como las deleciones, no muestran relación con la edad, aunque sí aparecen con mayor frecuencia en mujeres.

Algunos ejemplos citados por Crow (2000) ilustran estas tendencias. En la distrofia muscular de Duchenne, el 93% de las mutaciones puntuales aparecen primero en espermatozoides, mientras que el 87% de las deleciones son de origen materno.

Recordemos aquí las diferencias principales en la línea germinal, que da lugar a los gametos en cada sexo. En el sexo masculino, la espermatogénesis se basa en la multiplicación continua de las células basales en la edad adulta. Esta multiplicación genera, por una parte, gonias que sufren las sucesivas mitosis y las dos divisiones meióticas para producir los gametos, y por otra células basales que persisten como base del ciclo. Como consecuencia, las gonias producidas a edades mayores provienen de células basales que han sufrido un mayor número de divisiones mitóticas. Ello sugiere que cada división ofrece una cierta probabilidad de observar una mutación puntual, y que por lo tanto las mutaciones puntuales aumentan con la edad del hombre.

La producción de gametos en el sexo femenino sigue un curso diferente. Luego de un número limitado de mitosis a partir de las células germinales primordiales, se produce la primera división meiótica de la ovogénesis. En este estado permanecen los oocitos hasta la fecundación, que desencadena la segunda y última división meiótica. Podemos presumir, entonces, que el menor número de divisiones reduce la ocurrencia de mutaciones puntuales, mientras que la detención de la ovogénesis luego de la primera división meiótica podría resultar en una mayor probabilidad de cambio cromosómico.

Para la genética de poblaciones, es importante comprender, entonces, que las tasas de mutación de diferentes clases de cambios serán diferentes. Más aún, cuando intentamos comprender la evolución de un locus cualquiera en términos de mutaciones puntuales, y hablamos de su tasa (única) de mutación, ella será algún tipo de promedio de un gameto típico. De algún modo, esta tasa de mutación promediará las diferencias entre los sexos y los efectos de edad. Claramente, también, será conveniente evitar aglomerar tipos muy diferentes de mutaciones, como los cambios puntuales y las deleciones, bajo una misma tasa de mutación.

Veamos brevemente por qué nos parece aceptable ignorar la diferencia en tasas de mutación entre los sexos, y agruparlas en una única tasa de mutación para los cambios puntuales, y en cambio tenderemos a no aceptar la consideración indistinta de mutaciones puntuales y deleciones. Centraremos nuestra atención en cambios no ligados a los cromosomas sexuales, al igual que lo hemos hecho en toda nuestra discusión, en el entendido que los genes ligados al sexo pueden, con los ajustes debidos, atenderse en modo análogo a los autosómicos.

En el caso de las mutaciones puntuales, ocurrirán, cierto es, con mayor probabilidad en hombres que en mujeres. Pero serán transmitidas igualmente a ambos sexos, de modo que su destino en la población se independizará rápidamente del sexo de sus portadores.

En cambio, no distinguir entre deleciones y mutaciones puntuales implica realmente mezclar dinámicas evolutivas bien diferentes. La probabilidad de una mutación puntual puede depender, por ejemplo, del nucleótido que ocupa un sitio, y, como veremos brevemente más adelante, verse afectada por algún sitio vecino. Todo ello puede tenerse en cuenta en detalle, o promediarse en

una tasa única como aproximación razonable. Pero las deleciones se rigen por otras normas. Tienen, por ejemplo, una tendencia sistemática a involucrar ciertas regiones, en virtud de la estructura secundaria que tiende a adquirir el ADN de hebra simple en el proceso de replicación.

Sesgos mutacionales y modelos para mutaciones puntuales

En esta sección examinaremos muy brevemente algunos modelos más complejos de cambio a nivel nucleotídico. A nivel filogenético, es decir cuando se comparan los cambios acumulados entre las especies a lo largo de, por lo general, escalas de tiempo de millones de años, dichos cambios son mayoritariamente *sustituciones*, término reservado para aquellas mutaciones que se han *fijado* durante la historia de las especies. Varios de los modelos que esbozaremos aquí son precisamente modelos de sustitución, y no, estrictamente, de mutación. Pero para mutaciones no afectadas por la selección natural, existe una estrecha relación, que discutiremos más adelante al hablar de la teoría neutralista, entre sustitución y mutación.

Más allá de estas importantes distinciones, presentaremos algunos de estos modelos en términos mutacionales para dejar sentado que, si bien muchos trabajos genético poblacionales usan modelos muy sencillos de mutación, es posible incorporar al análisis situaciones más complejas.

A nivel de ADN, el modelo más sencillo de mutación, con frecuencia tomado como punto de partida, es llamado modelo de Jukes y Cantor, según el cual todos los tipos de cambio son igualmente probables. Este modelo puede representarse por la siguiente matriz de transición:

Nucleótido original:	Nucleótido nuevo			
	G	A	T	C
G	1-3 μ	μ	μ	μ
A	μ	1-3 μ	μ	μ
T	μ	μ	1-3 μ	μ
C	μ	μ	μ	1-3 μ

En cada generación, un nucleótido cualquiera tiene una probabilidad μ de mutar a cada uno de los restantes tres, y por ende una probabilidad de 1-3 μ de permanecer incambiado. Dicho sea de paso, esta matriz de transición define una cadena de Markov para la evolución de un sitio cualquiera en el ADN. Dicho sitio tiene probabilidades de transición perfectamente definidas de cambiar hacia cualquier otro estado, o de permanecer incambiado.

Las secuencias de regiones no codificantes del genoma nuclear, incluyendo intrones y regiones intergénicas, presentan patrones de sustitución algo diferentes de las implicadas por el modelo de Jukes y Cantor. Si la selección natural no opera, entonces el patrón de sustitución estará determinado por el de mutación. Por eso se procura usar el patrón de sustitución observado en regiones no codificantes para estimar el patrón de mutación. Se presume que

la selección natural en estas regiones no opera (asunción cuestionable), o al menos es mucho más débil que en regiones codificantes (afirmación más aceptable).

En todo caso, un ejemplo de patrones de sustitución del genoma nuclear, basado en seudogenes, es el siguiente (tomado de Graur y Li, 2000):

Nucleótido original:	Nucleótido nuevo			
	G	A	T	C
G	---	20.4	4.4	4.9
A	12.5	---	3.4	4.5
T	3.3	3.3	---	13.8
C	4.6	4.2	20.7	---

Esta tabla contiene frecuencias relativas, y ha sido simplificada excluyendo desvíos estándar y otros detalles. Claramente, existen ciertos tipos de cambio (en gris) que son más frecuentes que otros. En particular, las transiciones (cambios de una a otra purina, es decir de tipo $G \rightleftharpoons A$, o de una a otra pirimidina, o sea $T \rightleftharpoons C$), son más frecuentes que las transversiones. Un avance importante respecto al modelo de Jukes y Cantor, si queremos dar cuenta de esta observación, lo constituye el modelo de dos parámetros de Kimura:

Nucleótido original:	Nucleótido nuevo			
	G	A	T	C
G	$1-\alpha-2\beta$	α	β	β
A	α	$1-\alpha-2\beta$	β	β
T	β	β	$1-\alpha-2\beta$	α
C	β	β	α	$1-\alpha-2\beta$

El modelo de dos parámetros de Kimura reconoce, como se ve, dos tasas diferentes de mutación: una para transiciones y otra para transversiones. La opción entre los dos modelos, o más en general entre modelos de mutación con diferentes grados de complejidad, depende de las circunstancias. En general, una mayor complejidad involucra más parámetros a ser estimados con los mismos datos, y mayor varianza vinculada a las estimaciones. Estos costos se justifican, a veces, si el mayor realismo del modelo realmente es necesario para el problema que queremos resolver.

En muchos casos, en que unas pocas mutaciones puntuales constituyen toda la diversidad de un locus de interés en una población cualquiera, aún la simple distinción de transiciones y transversiones tiene más costos que beneficios. Para estudios a mayores escalas de tiempo, como los de la evolución de un gen a lo largo de millones de años y en linajes diversos, esta y otras distinciones son imprescindibles.

El libro de Graur y Li (2000) contiene una muy buena discusión de éstos y varios otros modelos de sustitución, así como de la evidencia empírica al respecto.

Variación del genoma mitocondrial

El genoma mitocondrial ha sido, en animales, objeto de numerosos estudios evolutivos y genético-poblacionales. Entre sus características principales, al menos para vertebrados, se cuentan las siguientes (hay varias excepciones y algunas dudas que no serán consideradas aquí):

- genoma compacto, de unos 16000 pares de bases, formado en su mayoría por genes sin intrones ni regiones intergénicas;
- transmisión exclusivamente materna, aunque se encuentra en ambos sexos;
- ausencia de recombinación;
- gran variabilidad y tasas de evolución unas diez veces mayores que las típicas del genoma nuclear;
- sesgo composicional, con una marcada deficiencia de G, en especial en terceras posiciones de codón;
- código genético propio, que difiere en el significado de algunos codones del código nuclear.

La variabilidad, herencia uniparental, y ausencia de recombinación, son ventajas que han atraído a muchos al estudio de este genoma. Estas propiedades son comunes en muchos grupos de animales, pero no en plantas. Como contraparte de estas ventajas debe considerarse las complejidades introducidas por el fuerte sesgo composicional y otros detalles estructurales. La ausencia de recombinación simplifica la comprensión de este genoma, pero lleva a que todo este genoma tenga una historia única, y no necesariamente representativa de la del acervo general de la especie, mejor representado en el genoma nuclear. Más allá de estos temas, que serán discutidos más adelante, es claro que el genoma mitocondrial podría merecer, para su estudio evolutivo a largo plazo, un modelo mucho más complejo que los considerados hasta aquí.

Interesa también señalar que la alta tasa de mutación del genoma mitocondrial animal parece estar vinculada, por una parte, a la ausencia de mecanismos sofisticados de reparación, que en el genoma nuclear corrigen buena parte de los errores primarios de incorporación de nucleótidos durante la replicación del ADN. Además, las funciones respiratorias de la mitocondria resultan en un gran número de radicales libres y, en general, un medio en el que el daño oxidativo sobre el ADN propicia las mutaciones. Ello tiene gran importancia biomédica y, a nuestros efectos, da cuenta de las altas tasas de mutación y sustitución observadas en este genoma.

Loci hipervariables: los microsatélites

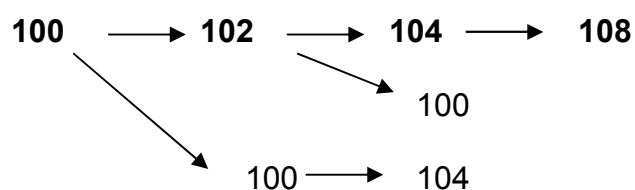
La variación es la base de la dinámica genética poblacional, de modo que, en principio, el análisis de loci muy variables parece ser la panacea para los genetistas de poblaciones. En la década de 1990, se popularizó el uso de un conjunto de loci, llamados microsatélites, que cumplían con la condición de ser hipervariables. Sus características generales son las siguientes:

- Cada locus de este tipo se caracteriza por contener una secuencia repetida un cierto número de veces. Los motivos de repetición pueden ser de un sólo nucleótido (A), dinucleótidos (CT), trinucleótidos, y así sucesivamente.
- Si bien los motivos de repetición pueden ser comunes a muchos loci, éstos se distinguen por las regiones a ambos lados del motivo de repetición, que varían de un locus al siguiente.
- Ello permite amplificar por PCR cada locus por separado, con oligonucleótidos que se asientan a ambos lados del motivo de repetición.
- Los alelos de un locus difieren en el número de copias del repetido. Las mutaciones consisten, por tanto, en el agregado o eliminación de uno o más motivos de repetición durante la duplicación del ADN.
- En general, a mayor tamaño del motivo de repetición, menor variabilidad del locus. Pero todos ellos son mucho más variables que los loci comunes.
- Se trata de una clase de loci ampliamente dispersa en el genoma, en el que se encuentran miles de loci de este tipo.

La alta variabilidad de este tipo de loci se debe a su gran mutabilidad. Las tasas de mutación varían según el motivo de repetición y otros factores, pero en general son del orden de 10^{-4} , o aún 10^{-3} , en contraste con las mutaciones puntuales, que en el genoma nuclear se caracterizan por tasas varios órdenes de magnitud menores.

Para los estudios de genética de poblaciones, la alta variabilidad de los microsatélites tiene connotaciones tanto positivas como negativas. Es positiva en cuanto garantiza la presencia de muchos alelos y, por tanto, un gran poder discriminatorio entre individuos, y un análisis de parentesco poderoso. La gran dispersión y gran número de loci disponibles facilita además los análisis de ligamiento.

Al mismo tiempo, la alta variabilidad y el peculiar tipo de mutación de los microsatélites genera amplio espacio para la aparición recurrente de alelos idénticos por su estado (igual número de repetidos) pero de diferente origen. Ello introduce ruido al igualar variantes de origen dispar. Usando el número de repetidos para identificar a los alelos, es esperable encontrar historias mutacionales como las siguientes en una población:



Vemos en este ejemplo que el alelo "100," por lo pronto, ha surgido al menos tres veces, de modo que su presencia en varios individuos no indica que lo han heredado de un ancestro común en alguna generación precedente. Lo mismo se aplica al alelo "104."

El problema se minimiza en el análisis de parentesco al ampliar el número de loci (para no basar las inferencias en un único locus) y teniendo en cuenta una cierta probabilidad de mutación entre individuos emparentados.

Pero para estudios de la dinámica de la población en general, o de variación geográfica, es decir estudios en los que la escala de tiempo considerada es mayor al simple paso de progenitores a descendientes, el ruido introducido por la mutabilidad y convergencia a alelos de idéntico estado (número de repetidos) se vuelve importante.

Un modelo que intenta abordar este caso, pero que no desarrollaremos aquí, es el llamado modelo de mutación por pasos ("stepwise"). En este modelo, un alelo tiene una probabilidad mayor de mutar a un alelo del que difiere solamente por un repetido. Una versión simple del modelo solamente permite cambios de este tipo. En versiones más flexibles, las probabilidades de transición son función inversa del número de repetidos en que difieren dos alelos.

Claramente, los microsatélites no se ajustarán bien a modelos más sencillos, como el de alelos infinitos, que serán considerados en lo que resta de este capítulo.

Mutaciones entre un número discreto de alelos

Muchos de los caracteres mendelianos estudiados en la primera parte del siglo XX constaban de un número limitado de alelos. La razón es que solamente aquellos alelos responsables de cambios fenotípicos visibles, como el color, eran identificables en muchas especies. De esta época datan dos modelos de mutación tendientes a dar cuenta del comportamiento de estos sistemas:

- Modelo de mutación irreversible, en el que un alelo común o "silvestre" podía mutar a un alelo alternativo con una probabilidad constante, la tasa de mutación μ . Por ejemplo, un alelo "normal" puede mutar a un alelo defectuoso que produce una enfermedad. Hoy sabemos que dicho alelo defectuoso, reconocible a nivel fenotípico, puede en verdad consistir de una colección de diversos mutantes, todos ellos resultantes en una disrupción de la función normal del gen. Pero el modelo puede dar cuenta del comportamiento general del sistema.
- Modelo de mutación reversible, en el que las mutaciones en sentido contrario son también posibles con una tasa ν .

Modelo de mutación irreversible

El efecto de la mutación en este modelo es sencillo si lo examinamos en forma “pura,” es decir sin el efecto simultáneo de la deriva. En este caso, la frecuencia inicial del alelo silvestre cambiará, en una generación, del siguiente modo:

$$p_t = p_{t-1} \times (1-\mu)$$

En palabras, en cada generación la frecuencia del alelo silvestre se verá reducida por una fracción μ debida a la mutación. Al considerar, ya no una, sino varias generaciones, el efecto acumulado tomará la forma de una recursión, en modo análogo a lo sucedido con el efecto de la deriva genética que consideramos en el capítulo anterior:

$$p_t = p_0 \times (1-\mu)^t$$

Unos pocos ejemplos numéricos bastarán para que el lector se convenza de que la erosión de la frecuencia del alelo silvestre causada solamente por mutación será, necesariamente, muy baja para las tasas de mutación usuales. Partiendo de una población pura para el alelo silvestre, ¿cuántas generaciones llevará la reducción de su frecuencia a 0.9 y a 0.5 si la tasa de mutación es de 10^{-6} ? ¿Cuáles serán los valores correspondientes con una tasa de mutación diez veces mayor?

Modelo de mutación reversible

Mientras que es bien probable que un alelo silvestre se vea afectado por mutaciones que le afecten y produzcan un “alelo” (en realidad una colección de alelos dispares) defectuoso con cierta frecuencia, la probabilidad de una mutación que revierta dicha alteración parece muy baja, y en realidad lo es. Un ejemplo más realista de mutaciones reversibles con efectos fenotípicos es el provisto por algunos transposones, cuya inserción cercana a un gen inhibe su expresión. Aquí tenemos un “alelo” silvestre, cuando el transposón está inserto lejos del gen, y uno “mutante” cuando se coloca en sus alrededores. El proceso es reversible, y por ende puede modelarse como si realmente consistiese en mutaciones en dos sentidos entre dos alelos alternativos.

El punto de partida para este modelo de mutaciones reversibles es completamente análogo al ya considerado para las mutaciones irreversibles. El cambio en la frecuencia del alelo silvestre en una generación será:

$$p_t = p_{t-1} \times (1-\mu) + (1-p_{t-1})\nu$$

Esta ecuación tiene, a la derecha de la igualdad, un primer término, idéntico a la antes considerada, que refleja la reducción esperada en la frecuencia del

alelo silvestre por causa de su transformación al mutante. El segundo término da cuenta del aumento de la frecuencia del alelo silvestre por causa de la mutación reversa. Como es de esperarse, los cambios de frecuencias ocasionados solamente por estas mutaciones serán muy pequeños de una generación a otra, tan pequeños como lo dicten las tasas de mutación involucradas.

Igualmente lento será el acercamiento al equilibrio de frecuencias, es decir una frecuencia tal que los cambios introducidos por ambos términos de la ecuación que acabamos de considerar se cancelan. Dejamos al lector la sencilla tarea de demostrar que la frecuencia de equilibrio será:

$$p_{eq} = \frac{\nu}{\mu + \nu}$$

La importancia de la mutación como fuente última de variación es innegable. Su importancia a largo plazo para evitar que la deriva acabe por agotar la variación en una población es igualmente patente. Sin embargo, los modelos mutacionales que acabamos de considerar sugieren, paradójicamente, que la deriva será comúnmente mucho más importante para explicar las fluctuaciones de frecuencias génicas que sufre una población de una generación a la siguiente que las mutaciones. Esta afirmación depende, por supuesto, de los "ritmos" relativos de deriva y mutación. Ya hemos visto en el capítulo anterior que el ritmo de la deriva es $1/2N_e$, donde N_e es el tamaño poblacional efectivo. Por lo tanto, si μ y ν son mucho menores que $1/2N_e$, con frecuencia podemos ignorar esas tasas de mutación en el corto plazo. Una excepción notoria a esta afirmación la constituyen los microsátélites, cuyas tasas de mutación son del orden de 10^{-4} .

Modelo de alelos infinitos

En la década de 1960 se completó el conocimiento del código genético. Además, la electroforesis de proteínas permitió por primera vez evaluar de un modo sistemático la variación en un número interesante de loci en poblaciones de cualquier especie. Al nivel de las proteínas, el número de alelos posibles no es, claramente, infinito. Pero una proteína constituida por una cadena de longitud tan modesta como 100 aminoácidos puede, en principio, dar lugar a un número muy grande de alelos posibles. Cada uno de esos 100 sitios puede estar ocupado por 20 aminoácidos diferentes, de modo que el número teórico de alelos posibles es 20^{100} . Esta cifra es obviamente mucho mayor que el número observado en cualquier población. Más aún, ante estos números, y considerando además que la tasa de mutación de interés (de aquellas mutaciones que resultan en cambios aminoacídicos) es de por sí baja, puede aceptarse como una aproximación razonable que la posibilidad de que coexistan en una población dos mutaciones diferentes que hayan dado lugar a exactamente el mismo alelo es muy cercana a cero.

Este razonamiento llevó a Kimura a pensar que era razonable admitir que si dos alelos eran idénticos en su estado (su composición aminoacídica), debían ser idénticos por descendencia. Ello implicaba que alelos de similar composición eran descendientes de una única mutación que ocurrió en un ancestro común. Dicha mutación, en otras palabras, debió haber ocurrido, a todos fines prácticos, una única vez.

Estos razonamientos y propiedades definen el modelo de alelos infinitos, que ha resultado de gran utilidad teórica y práctica.

Equilibrio entre deriva genética y mutación

Haciendo uso del modelo de alelos infinitos, vamos finalmente a considerar conjuntamente los efectos de la deriva genética y los de la mutación en una población. Lejos de representar un mero ejercicio técnico, vale la pena notar que estamos ahora considerando un modelo de tipo Fisher-Wright, permitiendo (finalmente) la ocurrencia de mutaciones, pero sin admitir todavía la acción de la selección natural. Nuestro modelo implica evolución, que puede sostenerse porque la deriva no agotará definitivamente la variación, que será repuesta por la mutación. Y sin embargo, la ausencia de selección hace que éste sea un modelo neutral de la evolución de una población. Precisamente, el considerar este tipo de modelos fue uno de los aportes sustanciales de Kimura a la teoría moderna de la evolución. Más adelante veremos, sin embargo, que en la visión de Kimura la selección natural no está ausente, sino que cumple un papel diferente al clásico.

Recordemos que la endogamia se define como la probabilidad de que dos alelos muestreados al azar en nuestra población sean idénticos por descendencia. En el capítulo anterior hemos visto que la endogamia en una población aumenta por causa de la deriva. La ecuación

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}$$

da cuenta del proceso como resultado de dos efectos, cada uno representado por un término a la derecha de la igualdad. El primero, $1/2N$ es la probabilidad de que dos alelos deriven de un mismo alelo en la generación precedente, y resulta del sistema de muestreo con reposición impuesto por el modelo de Fisher-Wright. La segunda resulta de combinar la probabilidad de que ello *no* ocurra (el complemento $1-1/2N$), multiplicada por la endogamia que ya poseía la población, y que es retenida. El segundo término nos da, por tanto, la probabilidad de que dos alelos muestreados al azar en una población sean idénticos por descendencia por la endogamia ya existente en la población.

En el modelo de alelos infinitos, toda mutación resulta en un nuevo alelo, previamente inexistente, de modo que el resultado es siempre el aumento de la diversidad genética de la población. Entonces, la endogamia nueva aportada por el primer término de la ecuación, se realizará siempre que ninguna de las

dos copias del alelo ancestral sufra una mutación. Como cada uno de ellos tiene una probabilidad μ de mutar, la probabilidad de que ninguno de ellos lo haga es $(1-\mu)^2$. Del mismo modo, la endogamia que proviene de generaciones precedentes a $t-1$ en la ecuación de más arriba será retenida si ninguno de los dos alelos de la muestra que tomamos para medirla sufrió una mutación. Por lo tanto, este componente de la endogamia también será retenido, por ausencia de nuevas mutaciones, con una probabilidad de $(1-\mu)^2$.

En resumen, podemos superponer el efecto de la mutación al de la deriva genética, modificando la ecuación de más arriba del modo siguiente:

$$F_t = \frac{1}{2N} (1 - \mu)^2 + \left(1 - \frac{1}{2N}\right) F_{t-1} (1 - \mu)^2$$

Los efectos contrapuestos de la deriva genética y la mutación llegarán a un equilibrio, que puede hallarse a partir de esta última ecuación. Despreciando términos del orden de μ^2 , así como aquellos del orden de μ/N , el lector puede llegar a la siguiente ecuación de equilibrio sobre la base de igualar F_t y F_{t-1} :

$$F_{eq} = \frac{1}{1 + 4N\mu} = \frac{1}{1 + \theta}$$

Aparece aquí por primera vez $\theta = 4N\mu$, uno de los parámetros fundamentales de la genética de poblaciones. En el caso de un modelo de Fisher-Wright y para un locus diploide, el valor de θ es el señalado. Aunque no nos interesa aquí detallar formulaciones alternativas, es bueno recordar que θ es 2 veces el número efectivo de alelos de un locus multiplicado por la tasa de mutación. En nuestro caso, el número efectivo de alelos es $2N$, como ya hemos visto. En otras palabras, el tamaño efectivo es igual en nuestro caso al número de alelos. Un modelo Fisher-Wright sin sesgo en la proporción sexual para un gen mitocondrial, por ejemplo, tendrá $N/2$ alelos, o sea que en este caso $\theta = N\mu$.

El parámetro θ tiene que ver con la variabilidad genética de una población, que será función tanto del tamaño poblacional (la deriva genética producirá fluctuaciones menores, en promedio, si el tamaño poblacional es mayor) como la tasa de mutación (que, naturalmente, tendrá una relación positiva con la variabilidad).

Entre los ingredientes de esta última ecuación, algunos pueden estimarse de modo sencillo, y otros son notoriamente difíciles. Aceptando que estamos en equilibrio, F es simplemente la probabilidad de que dos alelos tomados al azar de una población sean idénticos, y por tanto puede estimarse evaluando la homocigosis en una muestra de individuos tomada de la población. En otras palabras, es fácil estimar F , aunque debemos asumir que estamos en equilibrio para utilizar la estimación en la ecuación.

En cambio, el tamaño poblacional efectivo es más difícil de establecer empíricamente en muchas especies. Y la tasa de mutación es suficientemente baja como para que sea inusual la propuesta de examinar numerosos individuos o gametos de ascendencia conocida hasta encontrar un número suficiente de mutaciones como para obtener una estimación razonable.

Más adelante veremos que la teoría neutralista de la evolución molecular provee una solución, ya que permite obtener buenas estimaciones “puras,” es decir separadas del tamaño poblacional, de la tasa de mutación neutral.

Para concluir esta sección, vale la pena centrar nuestra atención en la heterocigosis esperada en una población. El capítulo sobre deriva genética nos permitió ver cómo este proceso lleva a una reducción continua de la heterocigosis. Ahora, y recordando que homocigosis y heterocigosis son valores complementarios, podemos ver que la heterocigosis esperada, en equilibrio con la mutación, será:

$$E(H) = 1 - F_{eq} = \frac{\theta}{1 + \theta}$$

El equilibrio descrito por las ecuaciones de la sección final de este capítulo es bien diferente al del modelo Hardy-Weinberg. Este describe un equilibrio estático de frecuencias de dos alelos y su combinación en las también estáticas frecuencias de los genotipos. Aquí estamos considerando un equilibrio dinámico, en el cual el ingreso en cada generación de nuevos alelos por mutación se compensa por su eliminación por deriva. Otra diferencia importante entre estos equilibrios reside en los tiempos necesarios para arribar a ellos, si partimos de una situación alejada del equilibrio. Vimos en el capítulo 2 que el equilibrio Hardy-Weinberg se alcanza en dos generaciones para un locus autosómico, porque depende de los apareamientos al azar, que rápidamente uniformizan las frecuencias alélicas en los dos sexos. En cambio el equilibrio dinámico entre deriva y mutación depende de dos procesos con tasas característicamente bajas ($1/2N$ y μ). Podemos intuir, por tanto, que este equilibrio tendrá una inercia considerable. Por lo pronto, una población de gran tamaño puede tener menor variación genética que la esperable, si es descendiente de una que, históricamente, fue mucho menor durante un largo tiempo.

Las estimaciones del tamaño efectivo de las poblaciones humanas basadas en tasas mutacionales estimadas por la genética médica y en estas ecuaciones de equilibrio dan valores del orden de $N_e=10^4$. Aunque el tamaño efectivo no incluye individuos que están en edad reproductiva y se ve afectado por otros factores, es evidente que la población humana actual es mucho mayor que ese valor. En otras palabras, la variación genética es menor a la esperable para todos los que somos, ya que la población no ha alcanzado el equilibrio entre deriva y mutación. Debemos por tanto presumir que en esta fase de nuestra historia, y en cada generación, están ingresando más alelos por mutación de los que están desapareciendo por deriva. Naturalmente, esto se aplica

solamente a las mutaciones neutras, ya que seguimos postergando la acción de la selección natural en nuestros razonamientos.

Lecturas recomendadas

- Crow, J. F. 2000. The origins, patterns and implications of human spontaneous mutations. *Nature Reviews/ Genetics* 1:40-47. *Una muy buena revisión de algunos sesgos mutacionales en humanos.*
- Graur, D., y W.-H. Li. 2000. *Fundamentals of Molecular Evolution*. Sinauer Assoc., Sunderland, MA. *Excelente introducción a la evolución molecular. Se recomienda estudiar los capítulos 3 y 4, que contienen descripciones sumarias de varios modelos de mutación y un resumen de los grandes patrones de sustitución conocidos.*
- Hartl, D. L., y A. G. Clark. 1997. *Principles of population genetics*. Sinauer Assoc., Sunderland, MA. *El capítulo 5 contiene una buena descripción de varios modelos mutacionales.*

Capítulo 5: SELECCION NATURAL SOBRE CARACTERES MENDELIANOS SIMPLES

El concepto de selección natural fue formulado de modo independiente por Charles Darwin y por Alfred Wallace. Darwin lo había concebido en 1838, y había comenzado desde entonces la redacción de un largo tratado sobre la evolución. En 1858, Darwin recibió de Wallace una carta en que le solicitaba que comunicase a la Sociedad Linneana un manuscrito adjunto sobre la selección natural. Darwin resolvió presentar, simultáneamente, un apretado resumen de sus propias conclusiones sobre el tema en la misma sesión, y las circunstancias sin duda le impulsaron a publicar, en 1859, un resumen del tratado que venía gestando desde hacía veinte años. Dicha síntesis es *El origen de las especies*, y la versión completa nunca llegó a concretarse, aunque Darwin publicó luego varias obras adicionales sobre diferentes aspectos de la evolución.

Aunque el concepto de selección natural tiene, entonces, 140 años de vida, continúa dando lugar a confusión con sorprendente frecuencia. En este capítulo, comenzaremos por adoptar la más sencilla definición de selección natural. Luego, examinaremos algunas propiedades de los casos más simples de selección sobre caracteres que dependen de un único locus con dos alelos, aunque también consideraremos algunas extensiones. Finalmente, completaremos el capítulo con una breve discusión que procura aclarar algunas de las muchas e innecesarias confusiones que existen sobre el concepto de selección, y esbozar algunas de las muy reales dificultades que se encuentra al intentar aplicarlo.

Concepto de selección natural

Usaremos una definición sencilla de selección natural, que rescata la esencia del concepto. La selección natural es un *proceso* que ocurre si se cumplen las siguientes condiciones, *necesarias* y *suficientes*:

1. Existe *variación fenotípica* (morfológica, fisiológica, etológica, etc.) entre individuos de una población.
2. Dicha variación es, al menos en parte, *heredable*, es decir, tiene un componente genético.
3. Dicha variación afecta el *éxito reproductivo* de los individuos; es decir, existe una correlación entre la variación y el éxito reproductivo, también llamado *eficacia darwiniana* ("*darwinian fitness*").

¿Cómo es posible que un concepto en apariencia tan simple pueda ocasionar confusión y discusiones aparentemente interminables a casi 150 años de su formulación? La respuesta emanará de este y otros capítulos que le siguen, pero vale la pena adelantar algunas características centrales de la selección natural:

- Se trata de un proceso poblacional, no de una "fuerza" que puede medirse en un instante dado. Dicho proceso puede operar en cada generación, pero puede requerir un gran número de ellas para revelar su eficacia. No hay garantías de poder constatar la ocurrencia del proceso en un periodo corto de tiempo. Nótese que varias de estas dificultades son compartidas por todos los procesos poblacionales discutidos hasta ahora (deriva genética y mutación, por ejemplo).
- Involucra, de distinto modo, dos niveles jerárquicos diferentes. La selección opera a través del éxito diferencial de los organismos, en virtud de sus diferencias fenotípicas, pero requiere de variación heredable, cuya fuente última es la mutación, afectando a los fenotipos.
- Se trata de un proceso probabilístico. Una cierta característica fenotípica, heredable al menos en parte, aumenta o reduce la *probabilidad* de contribución reproductiva de los individuos a las siguientes generaciones. Debe existir una relación consistente, en este sentido, entre la variación fenotípica (y, por su correlación con ella, la variación genotípica) y el éxito reproductivo, pero dicha relación no tiene por qué involucrar diferencias dramáticas entre los individuos, o muy fuertes correlaciones entre fenotipo y éxito reproductivo.
- Se trata de un proceso complejo en doble medida. Sabemos que la relación entre genotipo y fenotipo puede revestir muchas formas, desde la sencillez de los caracteres mendelianos hasta los caracteres de base poligénica, desde genes de efectos básicamente independientes hasta aquellos que muestran interacciones epistáticas y efectos pleiotrópicos complejos. A esta complejidad inherente a la genética debe sumarse la complejidad de las relaciones entre los fenotipos y el éxito reproductivo. El concepto de selección no implica, necesariamente, que el mayor éxito reproductivo involucre una mayor producción primaria de huevos o gametos, o una estrategia de inversión permanente en la reproducción. Tampoco implica, necesariamente, que exista un único fenotipo favorable en la población.
- Se trata de un proceso multiforme. Es fácil entender que no tiene sentido plantearse el valor adaptativo de un fenotipo sin tener en cuenta el ambiente, pero hay otras facetas del problema menos obvias, aunque igualmente importantes. Por ejemplo, con frecuencia se ignora que el valor adaptativo de un nuevo carácter depende *también* del contexto genético en que aparece. Una característica fenotípica no tiene un signo adaptativo en sí misma, sino en relación a su impacto en el éxito reproductivo de los individuos, en concierto con sus restantes características. Cuando tratamos el tema directamente a nivel genético, esto se traduce en que el valor adaptativo de un alelo depende de su funcionamiento, tanto en el contexto genómico en que aparece como en el contexto ambiental en que se desempeñan los organismos que lo portan.

Siguiendo a Endler (1986), es instructivo también preguntarse por la relación general entre selección natural y evolución. ¿Puede haber evolución sin selección? Sí, en cuyo caso la evolución ocurre guiada por los procesos de mutación, recombinación, y deriva genética. ¿Puede haber selección sin evolución? Sí, cuando la selección mantiene un equilibrio estable en la

población. En otras palabras, la selección natural es uno de los procesos, sin duda importante, pero no exclusivo, involucrados en la evolución biológica.

Selección sobre caracteres mendelianos simples

Los caracteres mendelianos simples son un buen punto de partida para entender la acción de la selección. Se trata de caracteres que, en condiciones normales, dependen esencialmente de la variación alélica en un locus. Naturalmente, reconocemos que tras la cláusula "en condiciones normales" se esconde una gran complejidad de interacciones necesarias entre el locus de interés, el resto del genoma, y el ambiente.

Comenzaremos con el análisis de un caso, y luego haremos algunas generalizaciones.

Un ejemplo: anemia falciforme en humanos

El caso está bien discutido en Templeton (1982), cuya lectura recomendamos especialmente. Comenzaremos esta discusión con una descripción de los alelos mejor conocidos:

Locus: betaglobina adulta

Dos alelos principales: A (alelo normal), y S (alelo que en homocigosis causa la anaemia falciforme; S viene de *sickle-cell*, glóbulo rojo en forma de hoz).

Diferencia entre los alelos: se trata de una mutación puntual en la segunda posición del sexto codón, que cambia el aminoácido correspondiente de la cadena beta de la hemoglobina de un ácido glutámico (un aminoácido cargado) en el alelo A a una valina (aminoácido neutro) en el alelo S. Esta diferencia en carga neta de la proteína puede revelarse por electroforesis.

Fenotipos:

AA - Individuo normal.

SS - Individuo que padece la anemia falciforme. La cadena de betaglobina mutada altera la organización de la hemoglobina y, a bajas concentraciones de oxígeno, deforma los glóbulos rojos a una forma de hoz característica. Esta enfermedad causa generalmente la muerte prematura de los individuos.

AS - Individuo normal en cuanto a que este genotipo no provoca anemia. Se ha demostrado que estos individuos tienen resistencia a una forma muy difundida de malaria, enfermedad aún hoy muy común en los trópicos.

La descripción anterior es, por supuesto, resultado de numerosos estudios, y de hecho contiene una descripción cualitativa de la acción potencial de la selección natural. Los heterocigotas AS son a la vez resistentes a la malaria y no padecen anemia. Si la malaria es una amenaza importante a la supervivencia y reproducción de los individuos (tenemos amplia evidencia de ello en muchas áreas tropicales), estas características fenotípicas confieren ventajas a los heterocigotas.

Tabla 6.1. Conteos de genotipos de betaglobina, y cálculos derivados de ellas

Genotipos	Frec. observadas	O : frec. relativas observadas	E: frec. relativas esperadas	O/E	w: eficacia relativa	s: coef. de selección
AA	9365	0.756	0.769	0.983	0.879	0.121
AS	2993	0.242	0.216	1.119	1.000	0.000
SS	29	0.002	0.015	0.154	0.138	0.862
Total	12387	1.000	1.000			
Alelos	Frec. alélicas observadas					
A	0.877					
S	0.123					
Total	1.000					

La tabla 6.1 contiene datos de una muestra grande de individuos adultos en una región tropical. La información primaria es, naturalmente, un conteo de genotipos para el gen de la betaglobina en la muestra (frecuencias absolutas observadas). A partir dichos conteos, pueden calcularse (se sugiere verificarlo a medida que se sigue esta explicación) todos los restantes valores. Las frecuencias relativas observadas son simplemente la proporción de cada genotipo en la muestra. Las frecuencias alélicas observadas están implícitas en los datos (verificarlo). De éstas se deducen las frecuencias relativas esperadas de los genotipos, de acuerdo al equilibrio Hardy-Weinberg. La relación O/E sugiere que hay un exceso de heterocigotas, un cierto déficit de homocigotas AA, y un marcado déficit de homocigotas SS respecto a lo esperado. La eficacia darwiniana relativa w estima, precisamente, la magnitud de estas desviaciones de lo esperado para una población en equilibrio Hardy-Weinberg, y el coeficiente de selección s en contra de cada variante es simplemente la diferencia de eficacia relativa de cada variante respecto a una que se toma arbitrariamente como referencia (en la tabla se usa la práctica común de escoger el óptimo como referencia, con lo que todos los otros valores de s son positivos y miden la intensidad de la selección contra los genotipos no óptimos).

La interpretación de este sistema, con las variantes hasta ahora descritas, es la siguiente. Con respecto al óptimo representado por los genotipos AS, tanto AA como SS sufren mortalidad diferencial que reducen su representación entre los adultos de la población. Ello da cuenta de las desviaciones observadas respecto a las frecuencias esperadas de los genotipos, inferidas a partir de las frecuencias alélicas asumiendo un equilibrio de tipo Hardy-Weinberg. Recuérdese que, aún cuando los apareamientos no sean estrictamente al azar, las poblaciones naturales se encuentran en la práctica cerca de dicho equilibrio.

Obviamente, no es imprescindible asumir dicho equilibrio. En el caso de la especie humana, ¿podría el lector proponer una estrategia para estimar el apartamiento de dicho equilibrio debido solamente a la ausencia de panmixia? Dicha posibilidad requiere a) escoger algunos sistemas genéticos no sometidos

a selección, o al menos razonablemente neutros, y b) usarlos para estimar apartamientos de la panmixia. Sobre esa base, podrían ajustarse en algo los cálculos de la tabla 6.1. En la práctica, el caso que nos ocupa involucra presiones selectivas tan fuertes que dichas correcciones no afectarían en forma apreciable los resultados.

Hay otras simplificaciones involucradas en los cálculos, pero una vez más, ellas son menores en este caso. Por ejemplo, la selección puede afectar el éxito reproductivo aún de aquellos individuos que sobreviven a la mortalidad cuyo impacto se estima en la tabla. La mortalidad diferencial es una forma de selección, pero no es la única, ni debe necesariamente formar parte de un régimen selectivo.

El polimorfismo AS de la betaglobina ilustra varias cuestiones importantes del modo en que opera la selección. Vislumbramos cómo fue posible que un alelo S, resultado de una mutación, pudo aumentar su frecuencia en la población, pese a la mortalidad de los homocigotas SS, en virtud de las ventajas de los heterocigotas portadores de dicho alelo. En verdad, un alelo que surge por mutación tiene una frecuencia inicial de $1/2N$, y se encuentra en un único individuo heterocigota. Sabemos además, que mientras su frecuencia permanezca relativamente baja, todos o casi todos los individuos portadores del alelo serán heterocigotas. El lector puede aplicar aquí su conocimiento del equilibrio Hardy-Weinberg para constatar que, por ejemplo, aún una frecuencia relativamente alta, como 10%, resulta en solamente un 1% de individuos homocigotas.

En el caso del alelo S, esto se traduce en que podemos presumir que la selección natural pudo actuar a su favor, a través de las evidentes ventajas de los heterocigotas que ya constatamos, *umentando* su frecuencia hacia valores cada vez más altos. En una primera fase, los homocigotas SS eran prácticamente inexistentes, de modo que la selección no podía actuar en su contra. Habrá llegado un punto, sin embargo, en que la aparición de dichas combinaciones homocigóticas comenzó a redundar en la *eliminación* de alelos S contraponiéndose, de este modo, al aumento de su frecuencia causada por la ventaja de los heterocigotas. Puede vislumbrarse así un *equilibrio*, que dependerá de la eficacia relativa de cada genotipo, de frecuencias de los alelos A y S. Dejamos de lado por ahora la demostración, por cierto sencilla, para enunciar que, si s_1 y s_2 son los coeficientes de selección contra AA y SS, respectivamente, la frecuencia de equilibrio de A, $f(A) = p$ es:

$$p_{eq} = s_2 / (s_1 + s_2)$$

Podemos usar esta ecuación para investigar si las frecuencias de los alelos A y S se encuentran, de hecho, en este equilibrio resultante de la acción de las presiones selectivas antes descritas. De la ecuación de equilibrio surge un valor esperado de $p = 0.877$, que coincide muy bien con la estimación de frecuencias en la tabla 6.1. Esto sugiere que el polimorfismo se encuentra en equilibrio y que, pese a las simplificaciones utilizadas, el análisis de la tabla recoge lo esencial del régimen selectivo.

El caso de este polimorfismo es de libro de texto porque se conocen las causas de la selección, que permiten explicar la presencia, en principio paradójal, y en frecuencias relativamente altas, de un alelo deletéreo en su forma heterocigota. Es éste un ejemplo clásico de un polimorfismo balanceado, es decir mantenido activamente por un tipo particular de selección, conocida como selección equilibradora, en el que la eficacia darwiniana mayor es la de los heterocigotas.

La malaria mata, aún hoy, unas dos millones de personas por año, de modo que el campo ha sido propicio para que, al menos todas las áreas en que esta enfermedad es común, cualquier variante genética que confiera alguna resistencia a esta enfermedad sea favorecida por la selección natural (aunque más abajo veremos que, paradójicamente, la selección puede actuar en contra de algunas de dichas variantes). De hecho, esto da cuenta de importantes enfermedades hereditarias en la especie humana, como las talasemias (anemias causadas por deficiencias de alfa o betaglobinas), y de la presencia de ciertos alelos de la enzima G6PD (glucosa 6 fosfato deshidrogenasa).

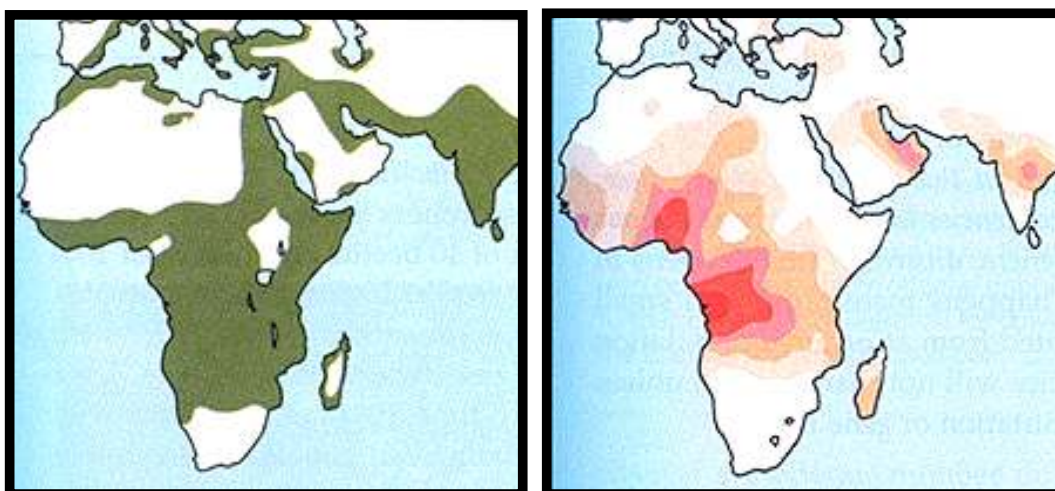


Fig. 5.1. Distribución de la malaria (izquierda) y del alelo S de la betaglobina (derecha) en parte del viejo mundo.

Un corolario de estas observaciones es que, por supuesto, el valor adaptativo de un genotipo es completamente dependiente del contexto. El alelo S es favorecido (en cierta medida) en áreas en que la malaria es importante, pero claramente deletéreo en regiones libres de la enfermedad. De hecho, la frecuencia de este alelo tiene una correlación razonable con la prevalencia histórica de la malaria en distintas regiones del viejo mundo (Fig. 5.1).

El polimorfismo entre los alelos A y S de la betaglobina nos provee un buen ejemplo de selección equilibradora. La magnitud de los coeficientes de selección involucrados es tal que podemos obviar, para entender este polimorfismo, complicaciones debidas a la interacción de los genotipos que nos

ocupan con otros factores genéticos, así como la acción de la selección a través de otros mecanismos menos patentes que la simple mortalidad. De ahí que, en este caso, podemos estimar la selección sobre los genotipos del locus de la betaglobina y hacer de cuenta que esas complejidades no existen. Muchos otros casos de selección han requerido análisis más complejos para su cabal comprensión. Entre ellos se cuenta el ejemplo de texto del llamado melanismo industrial, ilustrado por el polimorfismo de coloración de la polilla *Biston betularia*. Como es sabido, la coloración oscura (melanismo) era una rareza en esa especie, pero pasó a predominar en zonas industrializadas de Inglaterra en el siglo XIX. Sucede que las formas oscuras pasaron a ser menos visibles a los predadores en zonas contaminadas, en las que las cortezas de los árboles perdieron buena parte de la cobertura de líquenes mimetizadas por la coloración estándar, moteada, de esta polilla. Ridely (1996) provee un buen resumen de los estudios de este caso, cuya lectura recomiendo.

El alelo C, y el carácter histórico del proceso selectivo

Siguiendo a Templeton (1982), consideremos ahora el caso de un tercer alelo de la betaglobina, documentado en Africa. Este alelo es recesivo con respecto al alelo común A, pero los homocigotas CC son resistentes a la malaria sin producir la anemia aguda característica de los genotipos SS ya discutidos. ¿Cómo es posible que este alelo, aparentemente óptimo, sea de hecho muy raro en las poblaciones humanas africanas?

Para abordar este problema haremos uso de estimaciones de eficacia darwiniana y de los coeficientes de selección correspondientes, como en la tabla 5.1. Por conveniencia, mantendremos los cálculos usando como referencia a los heterocigotas AS, de modo que los valores de la tabla 5.1 se mantienen, y solamente resta agregar los correspondientes para los genotipos que involucran al alelo C (Tabla 5.2).

Tabla 5.2. Estimaciones de eficacia darwiniana (w) y coeficientes de selección (s), redondeados a una cifra decimal, y características fenotípicas de genotipos de la betaglobina humana.

Genotipos	w : eficacia relativa	s : coef. de selección	fenotipos
AA	0.9	0.1	Susceptible a la malaria
AS	1.0	0.0	Resistente a la malaria
SS	0.1	0.9	Anémico
AC	0.9	0.1	Susceptible a la malaria
SC	0.7	0.3	Anémico
CC	1.3	-0.3	Resistente a la malaria

Está claro en la tabla 5.2 que los genotipos CC son óptimos. Diríamos que tienen todas las ventajas de los heterocigotas AS sin las desventajas de éstos. Y sin embargo el alelo C es raro. La solución proviene de pensar en la

selección natural como un proceso, una historia particular que ocurre una vez que aparecen mutaciones.

Consideremos por ejemplo un punto de partida en el que, a partir del alelo común A, aparecen mutaciones como C y S. Estas mutaciones tendrán frecuencias de $1/2N$ y estarán, como ya vimos, representadas por un individuo heterocigota en cada caso al comienzo del proceso. De hecho, ya discutimos que la aparición de homocigotas es improbable, a no ser que el proceso conduzca a un aumento sustancial de las frecuencias de estos alelos.

El caso es que, mientras los heterocigotas AS cuentan con ventajas con respecto a los homocigotas AA, los heterocigotas AC no tienen ventaja alguna con respecto a éstos. En suma, la selección natural favorece la combinación AS, y con ello tiende a elevar la frecuencia del alelo S, mientras que actúa consistentemente en contra de las combinaciones AC y AA. Pero A es común, mientras que C es raro, de modo que el sistema tiende a un equilibrio exactamente como el descrito más arriba entre A y S. Entretanto, nuestro alelo ideal C es relegado a frecuencias ínfimas, si no eliminado completamente de la población. Templeton (1982) discute otros detalles del caso de buena manera.

En suma, la selección natural actúa contra el alelo ideal puesto que éste no se presenta en su forma homocigótica. La selección lleva a la población a un óptimo relativo, el del equilibrio entre A y S ya descrito, y eso mismo previene el alcance del óptimo absoluto, el de la fijación de C.

Podemos imaginar ahora un punto de partida diferente, en el que solamente la mutación C aparece en la población, sin la molesta presencia del alelo S. En este caso, al comienzo del proceso no hay acción de la selección, puesto que C aparece en individuos heterocigotas, fenotípicamente indistinguibles de los homocigotas AA (Tabla 5.2). En estas condiciones, el alelo C bien puede desaparecer por acción de la deriva genética, en virtud de su baja frecuencia. Si la deriva lo lleva en cambio a una frecuencia más alta, y comienzan a aparecer homocigotas CC, entonces la selección puede llevar rápidamente a la fijación de esta condición.

Esta discusión ilustra algunos detalles importantes del modo como opera la selección que permanecen en la penumbra en una versión trivializada, como la que sugiere que los alelos son siempre “favorables” o “deletéreos”, y que la selección actuará en consecuencia fijando los primeros y eliminando los segundos. Vemos en nuestra discusión que, aún en casos en que los genotipos están asociados a coeficientes de selección tan contundentes que nos eximen de preocuparnos por las interacciones complejas de los genotipos de nuestro locus de interés con el resto del genoma, la selección es un proceso más sutil. Por lo pronto, y siempre:

- la selección es un proceso dependiente completamente del contexto; así, la presencia e importancia de la malaria puede variar en el espacio y en el tiempo, cambiando completamente el régimen selectivo;

- la selección influye en un proceso histórico, de curso cambiante según el punto de partida; así, por ejemplo, la presencia o ausencia de S altera las perspectivas para el alelo C;
- la selección opera simultáneamente con otros procesos poblacionales; así, la deriva puede eliminar al alelo C, o más en general a cualquier alelo nuevo en virtud de su baja frecuencia inicial, independientemente de su valor adaptativo.

Principales tipos de selección natural

Los casos sencillos como los descritos hasta ahora pueden describirse del siguiente modo. En condiciones ambientales estables, y considerando solamente las probabilidades de éxito reproductivo de los genotipos que pueden formarse con un conjunto dado de alelos en un locus (A, S, y C), la población evoluciona según:

- el punto de partida (las frecuencias iniciales de los alelos)
- la siempre presente acción de la deriva genética
- el régimen selectivo, bien caracterizado por los coeficientes de selección (s) o, en modo equivalente, por las eficacias darwinianas (w) de cada genotipo.

Si consideramos una población infinita, podemos ignorar el segundo punto. Sin llegar a dicho extremo, basta asumir que al menos algunos de los coeficientes de selección son bastante mayores que $1/2N$ (o $1/N$, pero no prestaremos atención a esas pequeñeces aquí) para aceptar que la selección guiará el curso del proceso de la población. Recuérdese que $1/2N$ es nuestro informal “coeficiente de deriva” del capítulo 3, y estima por tanto el impacto de dicho proceso. Se vuelve ahora evidente que hemos despreciado la deriva en todos nuestros cálculos en este capítulo, aunque hemos hecho al pasar alguna mención a su posible efecto sobre las mutaciones jóvenes.

En línea con estas simplificaciones, podemos entonces definir distintos regímenes de selección, para configurar los principales tipos de selección, sobre la base de un conjunto de coeficientes de eficacia darwiniana. Vamos a usar w_{11} , w_{12} , y w_{22} , asociados a los genotipos A_1A_1 , A_1A_2 , y A_2A_2 , respectivamente, para un locus con solamente dos alelos:

- Selección direccional, a favor de uno de los homocigotas. Por ejemplo, $w_{11} > w_{12} > w_{22}$, así como también $w_{11} > w_{12} = w_{22}$ (el lector puede pensar a qué corresponden, biológicamente, estos dos ejemplos).
- Selección equilibradora, con $w_{11} < w_{12} > w_{22}$.
- Selección disruptiva, cuando $w_{11} > w_{12} < w_{22}$.

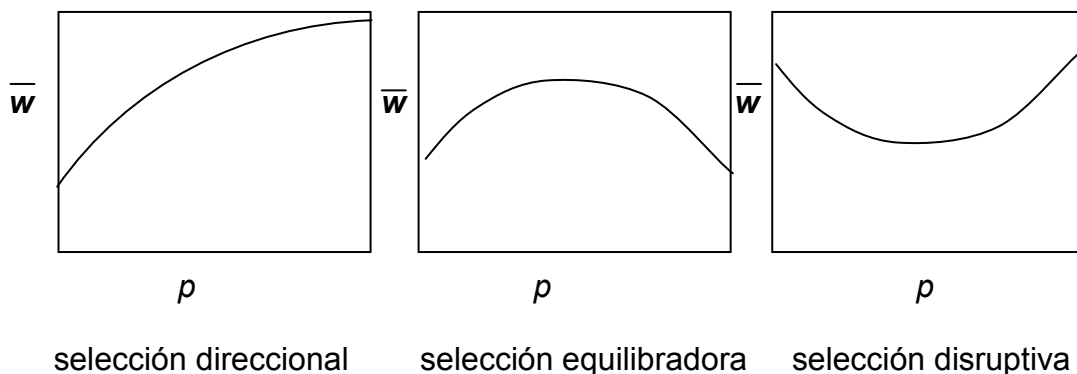
La intuición indica que, eliminada la incertidumbre introducida por la deriva o por la presencia de otros alelos, y en un régimen constante como el descrito, el curso de la evolución será obvio. Ello es cierto en los dos primeros casos. La selección direccional conduce a la fijación del alelo favorable (aunque nos queda pendiente el problema de los tiempos), y la selección equilibradora

conduce a un equilibrio de frecuencias que depende de los valores de nuestros coeficientes.

El caso de la selección disruptiva es diferente. Dependiendo del punto de partida, definido por las frecuencias iniciales de los alelos, el sistema tiende a la fijación de uno u otro alelo. De hecho, hay un equilibrio inestable a una frecuencia intermedia que funciona como un punto de quiebre. La selección favorece a un alelo hacia un lado de dicho punto, tendiendo a su fijación, pero opera en sentido exactamente contrario al otro lado del punto de quiebre. Una vez más, la selección actúa en cada generación según las condiciones dadas. Esto implica, como hemos visto en un contexto algo diferente al considerar el alelo C de la betaglobina, que un alelo joven, óptimo en homocigosis, puede ser eliminado por la selección, puesto que necesariamente estará en heterocigosis en la población.

Imagen del paisaje adaptativo

Podemos traducir estas consideraciones del curso de la evolución de una población como el movimiento en un paisaje adaptativo. El eje horizontal es la frecuencia de uno de nuestros dos alelos, que usamos como referencia, y el eje vertical es la eficacia darwiniana media, \bar{w} , de los individuos de la población. Nótese que este valor es un promedio simple, calculado a partir de los coeficientes asociados a cada genotipo y la frecuencia esperada de dichos genotipos según Hardy-Weinberg. Nuestros tres casos quedan representados del siguiente modo:



Puede decirse que la selección tiende siempre a aumentar la eficacia media de los individuos de la población, puesto que los genotipos más eficaces dejan mayor descendencia, tendiendo a aumentar su representación a expensas de los menos aptos. Esto lleva a la fijación del alelo cuya frecuencia estamos trazando en el primer panel (selección direccional) y a un equilibrio en una frecuencia intermedia en el segundo (selección equilibradora). En estos casos hay un “pico adaptativo” al que tiende la selección. En cambio, la selección disruptiva nos muestra un paisaje con dos picos, que por cierto corresponden a la fijación de uno u otro de los alelos. A cada lado del fondo del valle, la selección moverá a la población cuesta arriba de la ladera correspondiente.

Como toda imagen, puede ayudarnos a conceptualizar el proceso, pero no debe sustituir su estudio directo. Nótese por ejemplo que el paisaje adaptativo cambia de inmediato si cambian las condiciones ambientales. Pero también cambia si surge un nuevo alelo, que agrega otra dimensión al paisaje. Igualmente puede cambiar si no pueden ignorarse las interacciones con el resto del genoma.

No vamos a considerar estos casos. En cambio, vamos a detenernos brevemente para mencionar situaciones en las que, aún para un ambiente estable y un conjunto fijo de alelos, el régimen selectivo varía. Como ejemplo, considérese una especie en la que el color varía como resultado de los genotipos en un locus. Supongamos que los predadores aprenden a identificar a los organismos de esa especie por su color. Un genotipo raro (por ejemplo, el de homocigotas para un alelo joven) será poco predado, ya que el predador aprende a detectar el genotipo más común. Por tanto, el genotipo raro será favorecido por la selección, y aumentará su frecuencia a expensas del corriente. Pero esto conduce a que aquél pierda su rareza, con lo que pasará a ser reconocido por los predadores, perdiendo total o parcialmente su ventaja, o aún entrando en desventaja. Aquí no hay una eficacia darwiniana constante para cada genotipo. Por el contrario, el ajuste de cada genotipo depende de su propia frecuencia. Este régimen, conocido como de selección dependiente de la frecuencia, tiene propiedades bastante diferentes de las de los casos hasta ahora considerados.

Lecturas recomendadas

Endler, J. A. 1986. Natural selection in the wild. Princeton Univ. Press, Princeton. *Libro clásico que pasó revista a la evidencia y criterios para reconocer la acción de la selección natural, e inspiró numerosos trabajos hasta nuestros días.*

Ridley, M. 1996. Evolution. Blackwell Science, New York. *Este excelente libro de texto incluye una discusión detallada del melanismo industrial en la polilla Biston betularia.*

Templeton, A.R. 1982. Adaptation and the integration of evolutionary forces. En Perspectives on Evolution, R. Milkman, Sinauer (ed.), Sinauer Assoc., Sunderland. *¿Es optimizadora la selección natural? En este capítulo, que forma parte del paquete de lecturas del curso de Evolución, Templeton discute el problema sobre la base del clásico caso de la variación en el gen de la beta globina ligada a la anemia falciforme y a la resistencia a la malaria.*

Capítulo 6: EL COALESCENTE

En 1987, Rebeca Cann y colaboradores publicaron un influyente trabajo sobre la evolución humana, basado en el estudio de unos 150 genomas mitocondriales de individuos de muy diversas procedencias geográficas y étnicas. El artículo giraba alrededor de un árbol filogenético que expresaba las relaciones entre esos genomas. Usando estimaciones existentes de la tasa de evolución de esos genomas, estimaron que el ancestro común a éstos debió haber vivido hace unos 150 000 años. Como el genoma mitocondrial se transmite por vía materna, señalaron que ese genoma ancestral debió haber estado presente en una única mujer de aquella época. No hizo falta mucho más para que la prensa la llamase la “Eva mitocondrial.” La filogenia mitocondrial mostraba dos ramas derivadas de dicho ancestro. Una de ellas llevaba exclusivamente a genomas africanos, mientras que la restante conducía a muy diversos genomas, incluyendo aquellos procedentes del resto del mundo pero también, y eso es importante, varios africanos. Sobre estas bases, los autores concluyeron que la especie humana moderna se originó en África hace unos 150 000 años. Finalmente, los autores observaron que el árbol filogenético mostraba una rápida sucesión de ramificaciones cerca de su base, mientras que las ramificaciones más cercanas al presente estaban más espaciadas. Ello les llevó a proponer que la población humana había sufrido una rápida expansión inmediatamente después de su origen y en conjunción con su expansión desde África. En contraste con esos tamaños poblacionales relativamente grandes, propusieron que su origen pudo haber involucrado un “cuello de botella”, es decir un periodo en el que la especie se caracterizó por tamaños poblacionales muy pequeños.

En la Figura 6.1 se representa la filogenia mitocondrial de Cann y colaboradores en forma esquemática. Un examen de ese árbol plantea de inmediato numerosas preguntas. Por lo pronto, estamos acostumbrados a ver filogenias de especies, pero esta es una filogenia de genomas mitocondriales. ¿Debe interpretarse como una filogenia de individuos? Esto tiene poco sentido, ya que no aparece aquí la contribución paterna a la genealogía humana. ¿Refleja entonces la mitad materna de nuestra historia? Y la existencia de un único ancestro común, ¿implica que todos descendemos de una Eva mitocondrial? ¿Porqué se afirma que la forma general de la filogenia apunta a un origen africano? De más está decir que estas y otras preguntas han sido discutidas intensamente desde la aparición del trabajo de Cann y colaboradores. Aunque las interpretaciones originales pecan de cierta ingenuidad, también es cierto que los más finos y completos análisis de los genomas mitocondriales humanos han producido, hasta hoy, resultados asimilables al esquema de la Figura 6.1. En otras palabras, el árbol de Cann y colaboradores ha resistido el análisis crítico. En cambio su interpretación dista mucho de ser evidente. Más importante para nuestros propósitos, las inferencias propuestas no pueden resolverse mirando con detenimiento este árbol filogenético o los datos en que se basó o aplicando las herramientas del análisis filogenético. El mejor abordaje es abandonar por el momento el problema inmediato y recurrir a las herramientas conceptuales de la genética de poblaciones.

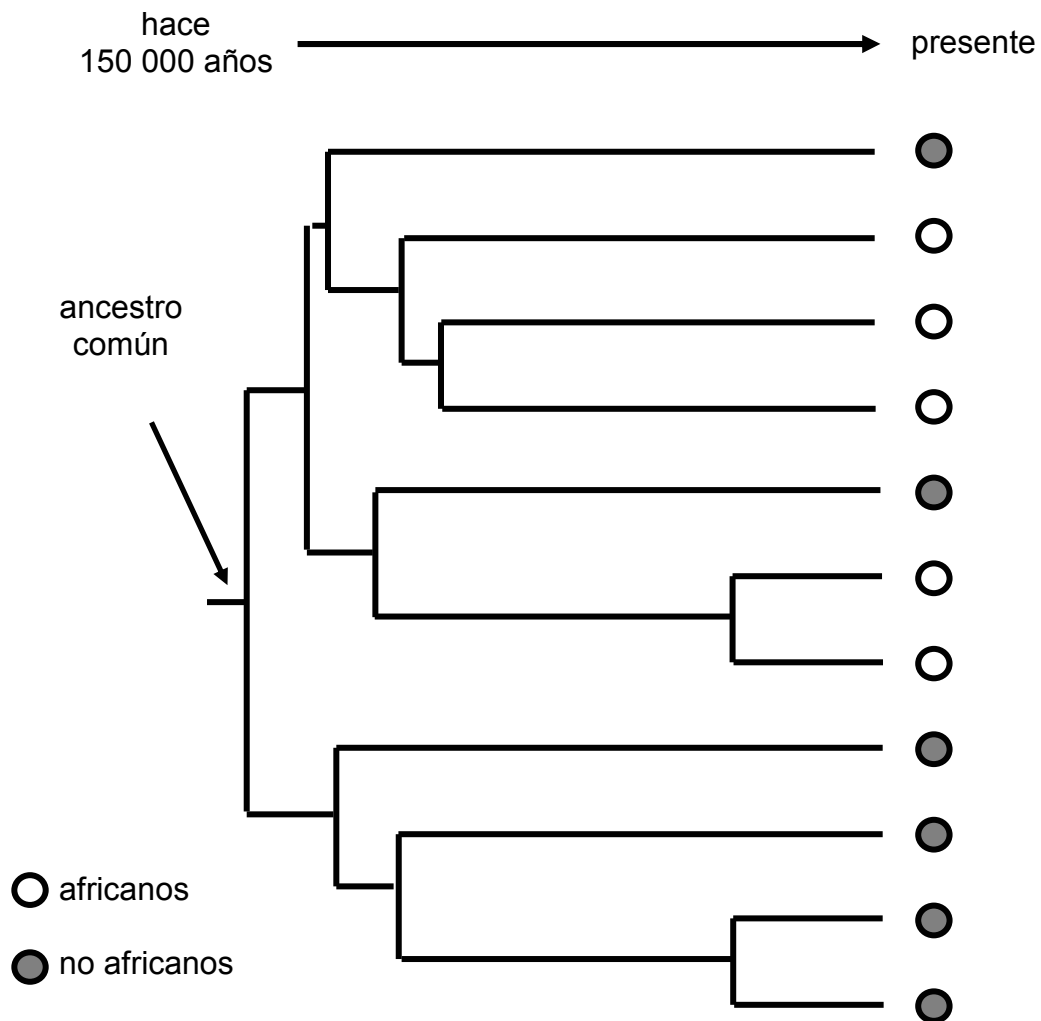


Figura 6.1. Representación esquemática de la filogenia mitocondrial humana de Cann et al. (1987). Cada círculo en las ramas terminales representa el genoma de un individuo. El largo de las ramas es proporcional al tiempo, y el ancestro común tiene unos 150 000 años de antigüedad.

El principal instrumento teórico que usaremos es la teoría de la coalescencia, o simplemente el coalescente. El mismo fue definido por Kingman, un matemático, en 1982. Varios investigadores, entre ellos Tajima (1983) y Hudson (1983), convergieron a los mismos conceptos y realizaron contribuciones a la teoría en forma independiente y esencialmente simultánea.

El coalescente: modelo estándar

Como punto de partida consideraremos una población ideal de tipo Fisher-Wright, tal como la definimos en el Capítulo 2. Recordemos que se trata de una población panmíctica de tamaño constante N , en la que estudiamos el efecto de la deriva genética sobre un locus neutral, diploide, y no ligado a los cromosomas sexuales (autosómico). La población tiene entonces $2N$ alelos en cada generación, que se toman al azar con reposición de la generación precedente. Luego veremos que los ajustes necesarios para considerar un

locus mitocondrial son mínimos. Por conveniencia, y aunque no es imprescindible, consideraremos que no existe recombinación en nuestro locus de interés.

Vamos a introducir dos novedades con respecto al tratamiento que hicimos de la deriva (y más en general de los procesos poblacionales) en los capítulos precedentes:

1. En lugar de ocuparnos del destino de todos los alelos de la población, típicamente agrupados en clases según sus estados, consideraremos las propiedades de una muestra de n alelos tomados en el presente.
2. En lugar de seguir el curso de la población hacia adelante, lo haremos hacia el pasado, partiendo del presente.

Estas dos opciones pueden parecer extrañas a primera vista, y en verdad lo son, pero nos simplificarán enormemente las cosas y han sido esenciales para el éxito del coalescente como aproximación a la genética de poblaciones. La Figura 6.2 presenta una comparación entre la aproximación tradicional (“hacia adelante”) y la coalescente. El código de colores muestra que cada tanto aparecen clases alélicas nuevas, que duran cierto tiempo y, naturalmente, pueden también extinguirse por azar. Pero a partir de una muestra de alelos, muchos de esos eventos quedan fuera de la trama genealógica y son por tanto inaccesibles. Afortunadamente, veremos que también no son necesarios para entender el acervo genético de nuestra población. De algún modo, existe en la genealogía, o mejor dicho en lo que ella ha legado a los alelos de nuestra muestra, información sobre las características más importantes de la población.

Motivados por esa promesa, pensemos ahora en cómo se va formando nuestra genealogía, partiendo del presente. Para una muestra de $n=2$ alelos tomados al azar, ya hemos usado la regla de muestreo al azar con reposición del modelo de Fisher-Wright para aceptar que la probabilidad de que provengan de un ancestro común en la generación precedente es $1/2N$. Llamamos coalescencia al evento de unión de dos alelos a su ancestro común, así que $P(\text{coal} \mid n=2) = 1/2N$.

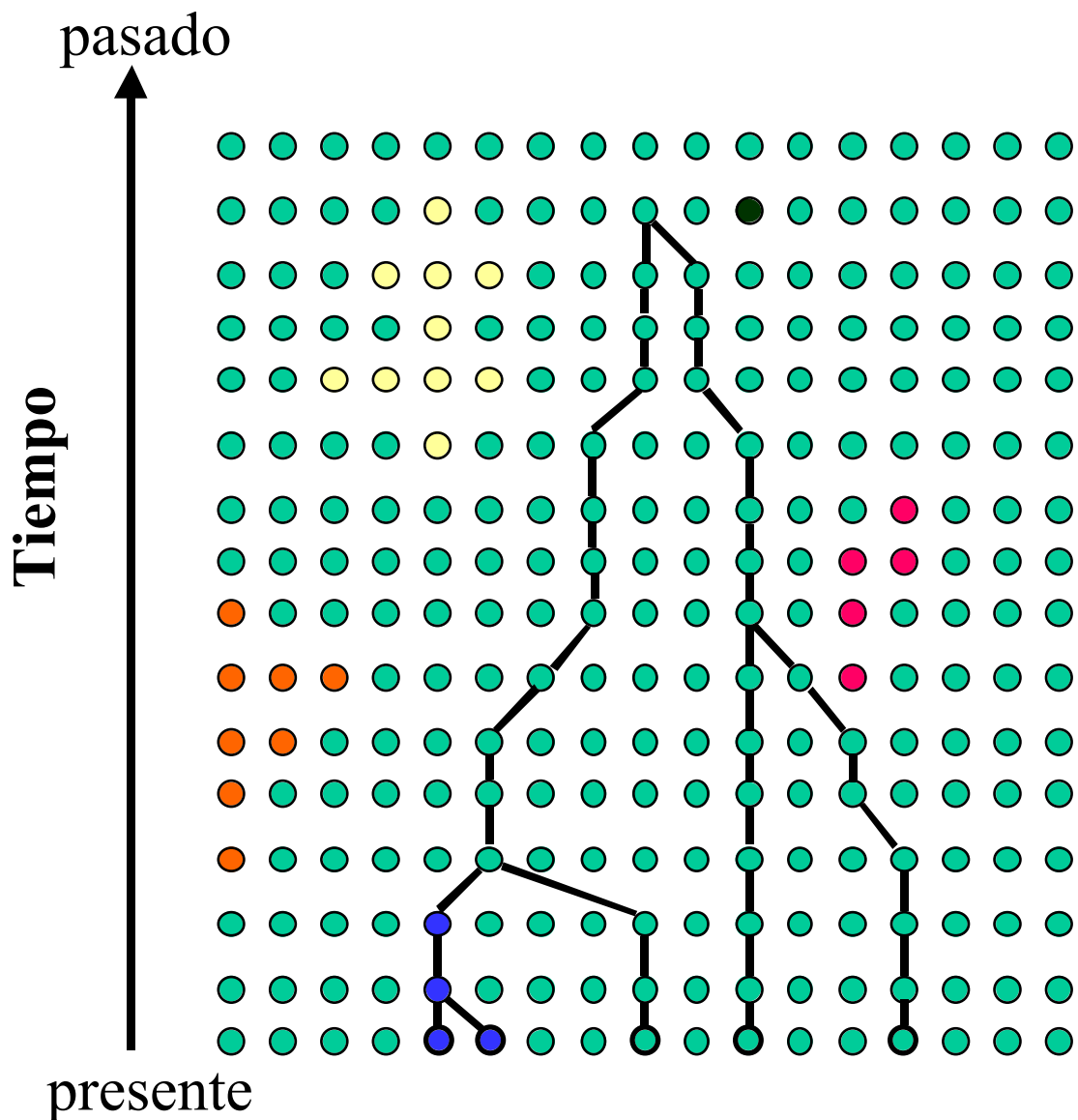


Fig. 6.2. Genealogía de 5 alelos tomados en el presente, trazada hacia sus sucesivos ancestros en las generaciones precedentes. Nótese que solamente los alelos que se encuentran en la línea directa de ascendencia de los incluidos en la muestra están conectados por la genealogía. Los restantes alelos de la población, independientemente de sus estados, no forman parte de la genealogía. Mientras que la genética de poblaciones clásica sigue hacia adelante el curso de toda la población, el coalescente investiga las propiedades de una muestra de alelos. La historia de la población solamente será accesible a través de las propiedades de dicha muestra, y de la genealogía que la vincula a sus ancestros.

Para una muestra de 2 alelos, el proceso se detiene cuando se llega al ancestro común. Si eso no ocurre, tenemos dos alelos ancestrales, o dos linajes, todavía disponibles para coalescer en la generación siguiente hacia el pasado. Pero la probabilidad de coalescencia para esos dos alelos ancestrales es $1/2N$, es decir exactamente la misma que la de nuestros alelos originales.

Una aplicación inmediata

En el Capítulo 3 habíamos definido la heterocigosis como la probabilidad de que dos alelos tomados al azar sean distintos, bajo el modelo poblacional de Fisher-Wright. Usamos además un modelo de alelos infinitos, que involucra una probabilidad de mutación muy baja μ y muchos estados posibles, por lo que cada mutación produce un estado único. Sobre esas bases, calculamos con algo de trabajo y algunas aproximaciones que, en equilibrio entre deriva y mutación, el valor esperado de la heterocigosis $E(H) = \theta/(\theta+1)$, donde $\theta = 4N\mu$.

Bajo esas mismas condiciones, podemos usar el coalescente de $n=2$ alelos para obtener ese resultado en forma inmediata. Para ello basta con razonar que si, en nuestra trayectoria hacia el pasado, ocurre un coalescente antes que cualquiera de los dos linajes haya mutado (o “desmutado”, puesto que vamos hacia atrás), nuestros dos alelos serán idénticos por estado. En cambio, si nos topamos primero con una mutación en cualquiera de los linajes, los dos alelos serán diferentes, y ya no nos importa que, finalmente, lleguen a coalescer.

Nos resta adscribir a cada una de esos resultados una probabilidad por generación. Ya tenemos la probabilidad de que ocurra un coalescente, que es $1/2N$. En cuanto a la probabilidad de mutación, es μ por definición para cada linaje, y por tanto 2μ para que un linaje cualquiera mute. Ello surge de que en cada generación acarreamos dos linajes hasta la precedente. Estamos despreciando la probabilidad de que los dos linajes muten en una misma generación (el lector puede razonar cual es y en qué condiciones es razonable esta aproximación).

El siguiente paso de nuestro razonamiento es tener en cuenta que, para una generación, tanto la probabilidad de coalescencia $1/2N$ como la de mutación 2μ son bajas. Despreciando la probabilidad de que en una misma generación ocurran tanto un coalescente como una mutación (una vez más el lector puede examinar esta posibilidad), llegamos a que nuestro espacio muestral tiene dos eventos excluyentes (la coalescencia o la mutación).

Por lo tanto, la heterocigosis esperada debe ser igual a la probabilidad de que dos alelos hayan sufrido una mutación antes (en reversa) de coalescer, dividida por la probabilidad total, o sea la suma de las probabilidades de coalescencia y mutación que acabamos de deducir. Es decir que $E(H) = 2\mu / [2\mu + (1/2N)]$. Multiplicando numerador y denominador por $2N$ obtenemos $E(H) = \theta/(\theta+1)$.

El coalescente con más de 2 alelos

Para $n=2$ alelos, la probabilidad de coalescencia es $1/2N$ por generación. Para $n=3$ alelos, podemos pensar en la probabilidad de coalescencia como $1/2N$ por cada par disponible para coalescer, es decir $P(\text{coal}|n=3)=3/2N$. Estamos despreciando la probabilidad de que los tres alelos coalescan en un mismo ancestro en una generación, lo cual es razonable si $2N \gg 3$.

Generalizando, si tomo una muestra de n alelos, tendré una probabilidad de coalescencia:

$$P(\text{coal}|n) = \frac{\binom{n}{2}}{2N}$$

Esta es una aproximación que funciona bien si $n \ll 2N$. Notemos que, mientras tengamos n linajes, esta sigue siendo la probabilidad de observar un coalescente en cualquier momento de nuestra incursión hacia el pasado. Más en general, si tenemos i linajes:

$$P(\text{coal}) = \frac{\binom{i}{2}}{2N} \quad i = 2, 3 \dots n$$

El proceso transcurre desde la muestra de n alelos, con $i = n$ en la ecuación, hasta que ocurre el primer coalescente. En esa generación pasa a tener $i = n-1$ linajes, hasta el siguiente coalescente, etc. Finalmente, al coalescer los dos últimos linajes, encontramos el ancestro común a todos los elementos de la muestra.

Una última observación para cerrar esta sección es que el numerador de nuestra última fórmula termina siendo $n(n-1)/2$ (verificarlo).

Tiempos de espera

Aceptemos que el tiempo de espera promedio para un evento infrecuente (como un coalescente) es el inverso de la probabilidad de dicho evento. Llamemos T al tiempo de espera, y coloquémosle un subíndice para indicar cuántos alelos estamos considerando (T_i para i alelos). Se deduce de la sección anterior que los valores esperados son:

$$E(T_2) = 2N$$

$$E(T_3) = 2N/3$$

...

$$E(T_i) = \frac{2N}{\binom{i}{2}} = \frac{4N}{i(i-1)}$$

Por supuesto, i puede simplemente ser n , el número de alelos en nuestra muestra, y el resultado obtenido nos dice que si n es mayor será menor el tiempo de espera (en reversa) para que se produzca el primer coalescente.

Implícita en estos resultados está la sorprendente noción de que son aplicables a un número cualquiera de alelos, en un momento cualquiera de la historia. Es decir que la probabilidad de observar un coalescente para $i = 5$ es la misma si tomamos 5 alelos en el presente o si los consideramos hace 500 generaciones. También es la misma si esos 5 alelos ancestrales son ancestros directos de 5 alelos del presente o si son, en realidad, los 5 antecesores de 10 ancestros del presente, que estamos considerando luego de haberse producido una serie de coalescentes (de $n=10$ a $n-1$, $n-2$... hasta llegar a 5 linajes ancestrales). Es que, en cada generación, el proceso se manifiesta según el número de alelos (una vez más, actuales o ancestrales) de la muestra y el tamaño de la población.

Otro resultado importante es que $E(T_n) < E(T_{n-1}) < E(T_{n-2}) < E(T_2)$. Es decir, los tiempos medios de espera se alargan a medida que el número de linajes se reduce mediante sucesivos coalescentes.

Antigüedad esperada del ancestro común a todos los alelos de la muestra

La antigüedad del ancestro común a todos los alelos de la muestra es fácil de calcular, ya que se trata de sumar los valores esperados de los T_i . Partiendo de la ecuación presentada más arriba, en unos pocos pasos, llegamos a calcular que el valor esperado de la edad del ancestro común, T_{MRCA} (usando la notación del subíndice en inglés, "most recent common ancestor") es:

$$E(T_{MRCA}) = 4N \left(1 - \frac{1}{n} \right)$$

Este resultado es importante, ya que nos dice que, aunque la antigüedad de la genealogía aumenta, como es natural, a medida que agregamos alelos a la muestra, tiende rápidamente a $4N$ generaciones. Este valor debe leerse, en primer lugar, como el doble del número efectivo de genes de la población ($2x$, en nuestro modelo, $2N$ alelos), y esta lectura tiene un valor general. Sustitúyase el $2N$ por el número efectivo de genes de un locus cualquiera, y el resultado es válido.

La segunda e importante lectura de este resultado es que la diversidad genética neutral de una población cualquiera está acotada en promedio, a aquella acumulada en unas $4N$ generaciones. El proceso de homogeneización por deriva genética, que subyace a todos nuestros razonamientos, ocurre a una tasa de $1/2N$ por generación y, en promedio, resulta en que todos los alelos de una población derivan de un único ancestro común en un tiempo de unas $4N$ generaciones.

Nótese también que $E(T_2) = 2N$. Es decir que: a) aún una modesta muestra de dos alelos tomados al azar tiene una historia prolongada subyacente hasta el ancestro común, y b) dicha historia es, en generaciones, aproximadamente la mitad de la historia total, que va desde los alelos de una muestra grande hasta tomada en el presente hasta su ancestro común.

Puesto que no hemos invocado todavía a la mutación, estos resultados genealógicos son la manifestación de la “deriva pura” en las poblaciones, si la variación es neutral. Es natural, por tanto, que dependan del tamaño poblacional, y también que contengan factores de ajuste para el tamaño muestral.

Variabilidad genética en un contexto genealógico

La variación genética neutral, adquirida por mutación, se genera con una probabilidad μ por generación. Puesto que las ramas de nuestra genealogía son, en realidad, generaciones transcurridas a través de sucesivos ancestros, es fácil entender que para calcular la variación genética esperada, basta sumar todas las ramas y multiplicar el resultado por μ . En cada intervalo T_i hay i ramas, de modo que, si llamamos T_{TOT} al tiempo total contenido en la genealogía, la suma de valores esperados es:

$$E(T_{TOT}) = \sum_{i=2}^n iE(T_i)$$

De esta fórmula surgen varios resultados. Reemplazando $E(T_i)$ por $4N/[i(i-1)]$ y sacando $4N$ de la sumatoria como factor común llegamos a:

$$E(T_{TOT}) = 4N \sum_{i=2}^n \frac{1}{(i-1)} = 4N \sum_{i=1}^{n-1} \frac{1}{i}$$

Para establecer la variabilidad esperada en un locus neutral para una población de tipo Fisher-Wright, nos falta únicamente un modelo mutacional apropiado para trabajar con secuencias de alelos. Recurriremos a un modelo emparentado al de alelos infinitos (descrito en el Capítulo 4), conocido como modelo de sitios infinitos. Recordemos que aquel nos permitía, sobre la base de una tasa de mutación baja y un número alto de clases alélicas posibles, aceptar como una aproximación razonable que todos los alelos de igual estado deben ser descendientes de un único alelo (el mutante original). Recordemos también que esta aproximación debe ser razonable para los alelos que coexisten en una población; realmente no necesitamos aceptar que un estado particular apareció una única vez en la historia de la vida.

Nuestro modelo de sitios infinitos tiene asunciones análogas, aplicadas ahora a los distintos sitios de un locus. La idea es que si la tasa de mutación por locus

es baja y el locus tiene una longitud suficiente, será razonable descartar la probabilidad de que un único sitio se vea afectado por más de una mutación. Una vez más, no nos interesa aceptar esta posibilidad para la historia total del locus, sino para los alelos que coexisten en una población y sus ancestros. Nos interesa que las mutaciones involucradas en nuestra genealogía de alelos (ver Figura. 6.2) se repartan entre sitios diferentes.

Con estas condiciones, y usando μ como la tasa de mutación por generación y por locus, podemos calcular $E(S)$, el número esperado de sitios “que segregan” (en el sentido mendeliano) en la muestra de alelos. Los sitios que segregan son los variables o polimórficos, en los que al menos uno de nuestros alelos difiere de los restantes. Para ello no necesitamos acrobacia algebraica alguna, ya que un par de párrafos más arriba calculamos la longitud total esperada de la genealogía, es decir el número total de generaciones que conectan a los alelos con sus sucesivos ancestros. Esas son las generaciones disponibles para que ocurran mutaciones. Por tanto:

$$E(S) = 4N\mu \sum_{i=1}^{n-1} \frac{1}{i} = \theta \sum_{i=1}^{n-1} \frac{1}{i}$$

Esta ecuación refleja algo que ya sabemos: la variación genética de una población de tamaño constante es, para un locus neutral, función del tamaño poblacional y de la tasa de mutación, combinados en θ (la ecuación incluye además un factor de corrección para el tamaño de nuestra muestra). Pero hemos ganado una perspectiva sobre la variación a nivel nucleotídico. En un trabajo que presagió el coalescente, Watterson (1972) llegó a la ecuación de $E(S)$ y propuso que el número de sitios que segregan en una muestra de secuencias de ADN podía servir para estimar θ (despejando este parámetro de la ecuación y remplazando $E(S)$ por el número de sitios observados como variables). Ese estimador, muy utilizado, se conoce en la literatura como θ_W , o también como θ_S . Más importante aún que estos resultados particulares resulta que, como se verá más adelante, tenemos ahora una perspectiva genealógica explícita que, bien explotada, nos permitirá ir mucho más allá de los valores esperados a los que nos hemos visto ceñidos hasta ahora.

Otro estimador de variación muy usado, y vinculado al de Watterson, es el número promedio de diferencias entre los alelos de la muestra, para el que usamos el símbolo π . Su valor esperado es el número esperado de diferencias entre un par cualquiera de alelos. Para averiguarlo basta usar $E(T_2)$, recordar que en ese intervalo tenemos dos ramas de esa longitud (cada una conecta a un alelo de la muestra con el ancestro común), y aplicarles la tasa de mutación μ :

$$E(\pi) = 2E(T_2)\mu = 4N\mu = \theta$$

En la Figura 6.3 se presentan dos genealogías con longitudes de ramas proporcionales a sus valores esperados, y algunas mutaciones adquiridas durante el proceso genealógico.

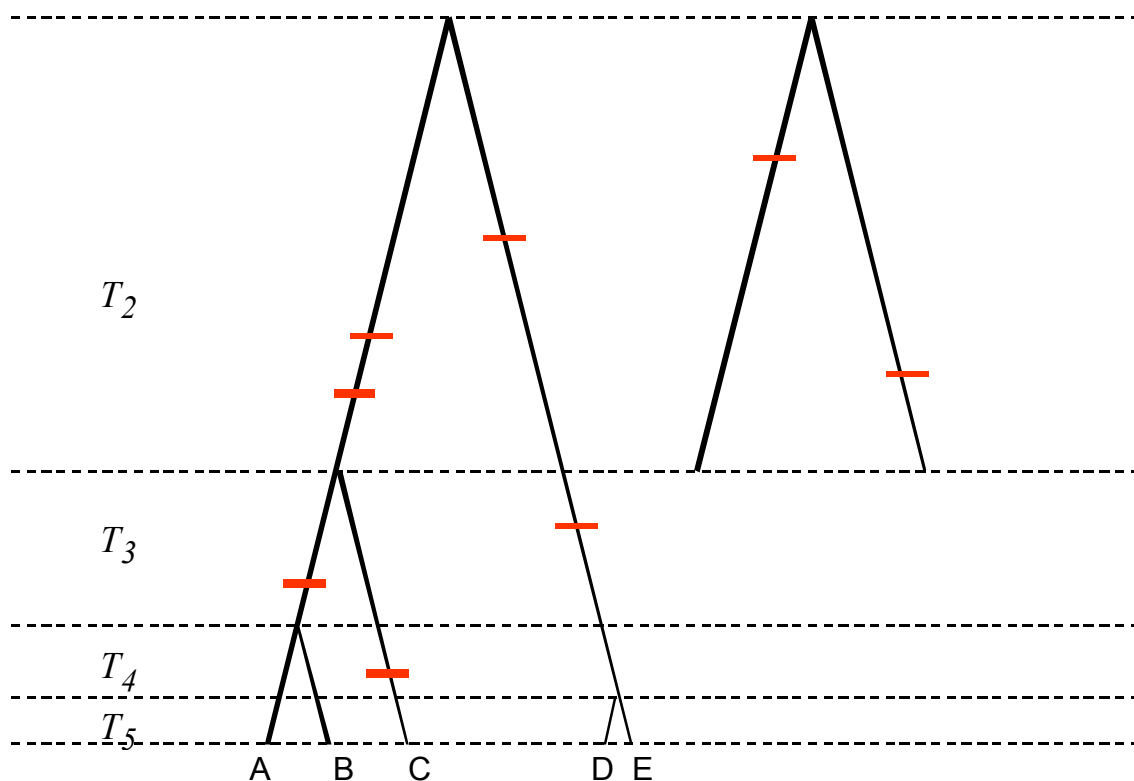


Figura 6.3. A la izquierda, una genealogía de cinco alelos, con intervalos de tiempo proporcionales a sus valores esperados. Las barras horizontales representan algunas mutaciones ocurridas en ciertas generaciones y ramas particulares de la genealogía. A la derecha, lo mismo para una muestra de dos alelos. Nótese que el valor esperado de T_2 es idéntico, ya sea para una muestra de dos alelos en el presente o para los dos linajes ancestrales, luego de pasar, a partir de una muestra de cinco alelos, por T_5 , T_4 y T_3 .

El coalescente y los árboles de genes

Comenzamos este capítulo con un árbol de genes o alelos obtenido para el caso del ADN mitocondrial humano. Estamos ahora en condiciones de abordar algunas de las preguntas que el mismo motivó. Conviene primero aclarar que un árbol de genes no es una genealogía sino una hipótesis filogenética sobre las relaciones entre los alelos. Dicha hipótesis se construye con los métodos del análisis filogenético. En el caso del árbol mitocondrial humano, para el mismo se usó un método de distancias y la asunción del llamado “reloj molecular”, que propone que el cambio molecular se acumula a una tasa estadísticamente constante.

¿Cuál es la relación entre ese árbol de genes y el proceso genealógico descrito por el coalescente? En primer lugar, los alelos de una población están ligados por una trama genealógica subyacente. La misma incluye la topología y la longitud de las ramas en generaciones, o sea los valores de los T_i . Como la genealogía es el resultado de un proceso estocástico, debemos pensar en cada caso particular como una realización de dicho proceso. Los valores

esperados de las ramas son útiles, pero para cada locus, muestra de alelos, y población variarán al azar.

En segundo lugar, nosotros no podemos observar la genealogía. Una realización genealógica acumulará un cierto número de mutaciones, producto de los errores de replicación del ADN de una generación a la siguiente, y éstas se manifestarán en nuestra muestra como sitios variables en el ADN. Por ejemplo, las genealogías de la Figura 6.3 y las mutaciones acumuladas sobre su trayectoria resultarán en seis sitios variables para la muestra de cinco alelos y en dos para la de dos alelos. En el primer caso, no hay mutaciones que distinguan a los alelos A y B, como no las hay para D y E. Un estudio que secuenciase estos cinco alelos revelaría solamente tres secuencias diferentes, es decir tres clases de alelos: (A, B), (C), y (D, E). Como ejercicio, el lector puede inferir de la Fig. 6.3 el número de diferencias que las secuencias revelarían entre cada par de clases.

Recordemos que la Fig. 6.3 solamente muestra los valores esperados de los T_i , pero que las ramas pudieron haber sido más cortas o más largas, y que aún manteniéndolas fijas el proceso mutacional pudo haber producido números diferentes de sitios variables. En otras palabras, el coalescente nos permite entender que una realización cualquiera de la historia de una muestra de alelos tiene dos fuentes de variación: 1) la demografía, que se manifiesta en variaciones en la topología y longitudes de ramas; y 2) la mutación, que en cada generación tiene una probabilidad de aparecer sobre una rama cualquiera de la genealogía. En un sistema neutral como el aquí considerado, la mutación no afecta el éxito de los alelos, por lo que se acumula sobre la genealogía sin afectarla.

Aunque, como es natural, la mutación puede realizarse en el paso de un ancestro a un descendiente con absoluta prescindencia del contexto genealógico más amplio, el patrón de variación total de una muestra de alelos está condicionado por la genealogía. Podemos pensar que la forma y longitud de la genealogía es una realización del proceso demográfico, y que las mutaciones que podremos observar en la muestra están condicionadas por dicha realización. La razón es que solamente veremos las mutaciones acumuladas sobre la genealogía.

En cuanto al árbol de genes, resulta de nuestra capacidad de usar las mutaciones como caracteres para el análisis filogenético. En la Figura 6.4 se muestra en resultado de dicho ejercicio para el caso de la genealogía de cinco alelos de la Fig. 6.3

La Fig. 6.4 permite entender que la filogenia de alelos contiene información parcial sobre la genealogía. Nos es imposible discernir, por ejemplo, que D y E tienen un ancestro común más cercano al presente que A y B. Sin embargo, aunque los datos de este ejemplo simple no permiten establecerlo con certeza, es razonable suponer que la raíz del árbol podría estar en la rama que separa a D y E. Con más información (secuencias más largas, para tener más sitios variables, una muestra más grande, y tal vez un grupo externo) podríamos

reducir en parte estas incertidumbres. Igualmente, los datos de otros loci no ligados permitirían avanzar en la comprensión de las características e historia de la población, ya que dichos loci representarían realizaciones independientes de los procesos mutacionales y de los mismos procesos demográficos.

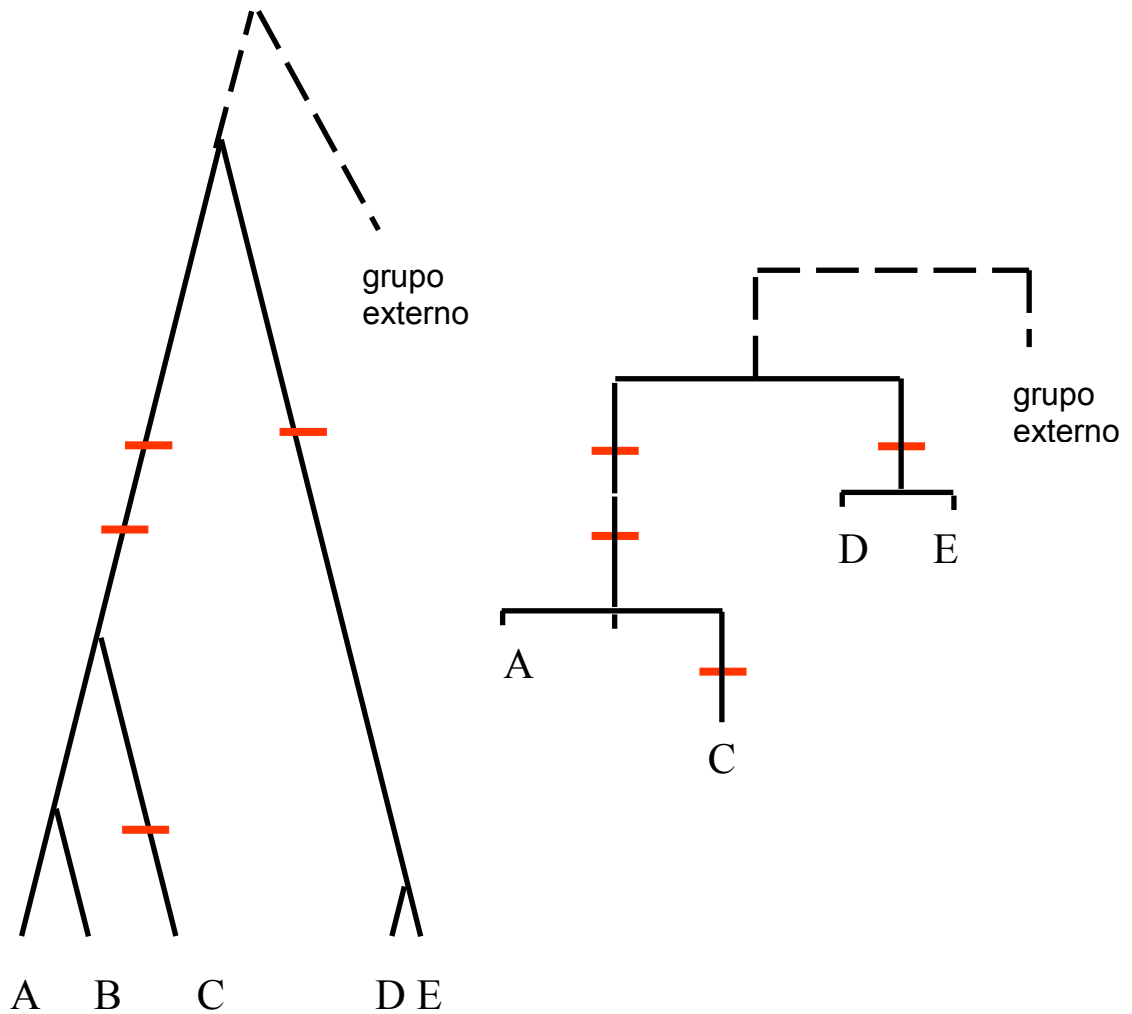


Figura 6.4. A la izquierda, una genealogía de cinco alelos.³ A la derecha, una filogenia inferible a partir de las mutaciones observadas y un grupo externo, con longitudes de ramas proporcionales al número de cambios observados. Los alelos A y B pertenecen a la misma clase alélica, ya que no presentan diferencias. Lo mismo sucede con los alelos D y E.

Pese a todas estas fuentes de incertidumbre, podemos volver ahora a la Figura 6.1 y notar que la filogenia de genomas mitocondriales humanos tiene una característica sobresaliente: las ramas tienden a ser largas cerca del presente y cortas cerca de la raíz del árbol. El esquema de la figura reproduce este patrón, originalmente notado en una filogenia de más de 140 genomas. El gran tamaño de la muestra, y la observación de que la variación en el genoma

mitocondrial humano es muy grande, ofrecen mejores perspectivas de que el árbol de los genes refleje razonablemente las características de la genealogía subyacente. Volviendo a la Fig. 6.3, constatamos que una genealogía representada con longitudes de ramas proporcionales a sus valores esperados tiene ramas cortas cercanas al presente y ramas cada vez más largas a medida que nos acercamos a la raíz. Esto sucede porque los valores esperados de los T_i dependen de una constante ($2N$) común a todos, y del número de pares disponibles para coalescer, que varía en cada intervalo.

Esta observación es cualitativa, pero el problema se puede formalizar y someter a pruebas estadísticas. Los resultados acumulados desde el trabajo original de Cann y colaboradores han sido contundentes en este sentido. El ADN mitocondrial humano no se ajusta bien a las expectativas derivadas del coalescente para nuestro modelo estándar, que asume neutralidad del gen estudiado (en este caso, del genoma mitocondrial completo) y tamaño poblacional constante. Dejaremos para otro capítulo el análisis de cómo la selección natural positiva, que implica apartamientos de la neutralidad de los alelos, afecta a la genealogía. En cambio, nos ocuparemos aquí de entender cómo los cambios demográficos afectan al proceso genealógico.

La respuesta general es sencilla. El proceso genealógico tiene, en cada generación, probabilidades fijas de coalescencia, que solamente dependen del tamaño poblacional y el número de linajes en dicha generación. Es un proceso sin memoria, en el sentido que lo que ha ocurrido en etapas anteriores del proceso no afecta dichas probabilidades. Como consecuencia, si el tamaño poblacional cambia, en lugar de permanecer constante, a lo largo del tiempo, cada parte de la genealogía responderá al $2N$ correspondiente, sin importar cómo ha cambiado este parámetro en otras fases de la historia. La Figura 6.5 muestra el comportamiento promedio del coalescente en una población de tamaño constante y en otra de tamaño variable. Observamos que la genealogía de una población que ha cambiado drásticamente su tamaño está deformada respecto a la estándar. En el caso que se ilustra, una población que fue pequeña en un pasado no muy remoto y ha pasado a un tamaño muy grande en una fase cercana al presente, verá deformadas sus genealogías de alelos como resultado.

Las propiedades de las genealogías nos permiten, entonces, entender cómo se deforman en respuesta al cambio demográfico. A través de los árboles de genes, o mejor aún de un estudio más completo del comportamiento y propiedades estadísticas del coalescente en sus dos aspectos, el demográfico y el mutacional, podemos discriminar una población que ha permanecido constante de otra que ha cambiado en tamaño. Esta conclusión es sorprendente y nos da un arma poderosa para examinar el pasado. En el caso específico del ADN mitocondrial humano, podemos comprender ahora por qué Cann y colaboradores postularon que sus resultados implicaban un crecimiento poblacional rápido a partir de un ancestro.

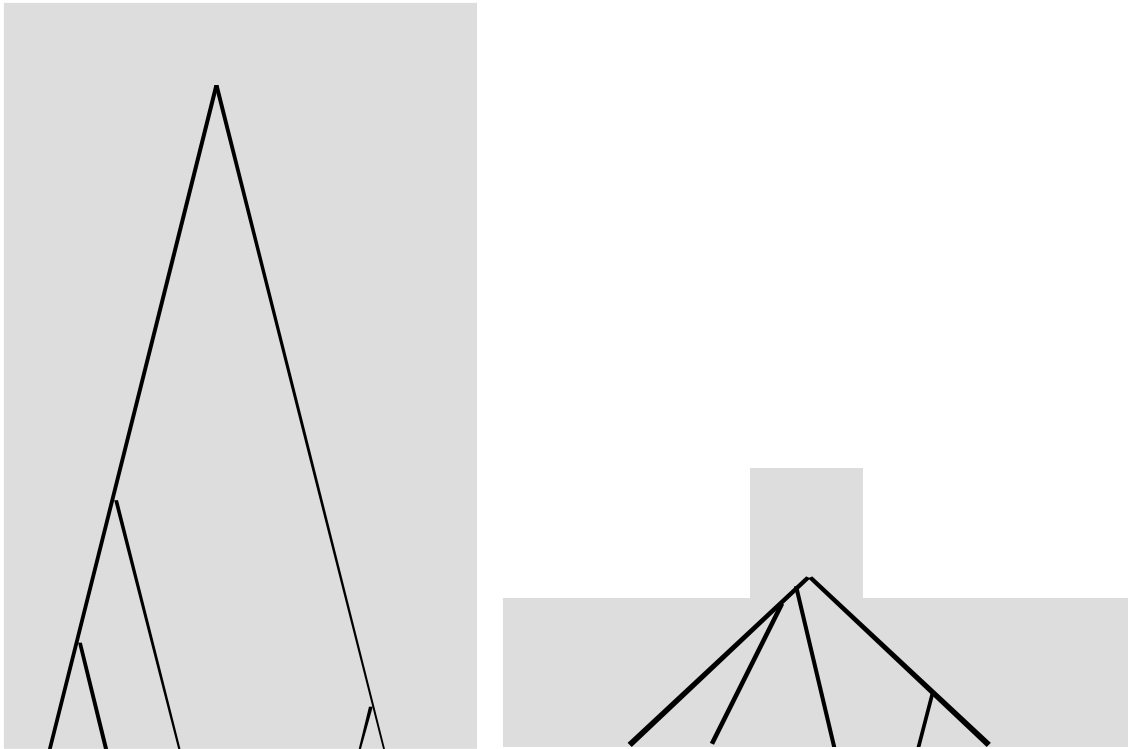


Figura 6.5. El coalescente en dos poblaciones con historias diferentes: a la izquierda, una población de tamaño constante (representado por el ancho de la banda gris de fondo); a la derecha, una de tamaño mayor en el presente, pero mucho menor en el pasado. El coalescente en una población de tamaño constante (izquierda) tiene, en promedio, ramas cortas cerca del presente y ramas largas cerca de su raíz. En cambio, en la población que ha sufrido una expansión significativa, las ramas cercanas al presente son más largas, reflejando su mayor tamaño actual, pero las correspondientes a la época de tamaño muy pequeño son muy cortas.

El siguiente paso del razonamiento de Cann y colaboradores puede entenderse también examinando la Figura 6.5. Vemos que, si el tamaño ancestral de la población fuese muy pequeño, ello precipitaría la coalescencia de los linajes que restasen, al progresar desde el presente hacia el pasado, en un tiempo cercano al momento de cambio del tamaño de la población. Al llegar a un momento en el pasado en que $2N$ es muy pequeño, los tiempos de espera entre eventos de coalescencia se vuelven proporcionalmente más pequeños también. Para llegar a invertir el patrón de longitudes de ramas del coalescente estándar, el tamaño poblacional ancestral tiene que ser realmente chico. Finalmente, si contamos con una calibración de la probabilidad de mutación por generación y una estimación del tiempo generacional, podemos estimar la antigüedad de nuestro ancestro mitocondrial y suponer que vivió cerca del momento de cambio demográfico.

Divergencia de dos poblaciones en ausencia de selección

Más arriba hemos deducido que, para un locus diploide autosómico, el tiempo esperado de fijación de un alelo por deriva genética es $4N$ generaciones.

Conceptualmente, esto es equivalente a nuestro desarrollo de la antigüedad esperada del ancestro común a todos los alelos de la muestra. Si estos derivan de un único ancestro, es porque el mismo se ha fijado como linaje. Esto es interesante, puesto que separa el problema de la fijación de un linaje del problema de la variación que ha acumulado en el camino. Curiosamente, entonces, la deriva genética fija en este sentido linajes sin que necesariamente ello implique que desaparezca la variación genética. Es que no nos interesa tanto ahora la fijación de una única clase alélica, sino el proceso de homogeneidad demográfica producido por la deriva.

Sabemos también que nuestro resultado para el tiempo esperado de monofilia es más general, puesto que resulta del modelo poblacional de Fisher-Wright, en el que los alelos se muestrean al azar en cada generación con prescindencia de la ploidía. El resultado nos dice que el tiempo promedio de fijación es, en generaciones, el doble del número efectivo de alelos en la población. Para un sistema mitocondrial, por tanto, ese tiempo es el doble del número de hembras, o sea unas N generaciones.

Si consideramos ahora dos poblaciones derivadas de una única población ancestral, podemos entender mejor el proceso de divergencia que, inevitablemente, resultará de dicha separación. El proceso transitará por tres fases bien definidas por las relaciones genealógicas entre los alelos de las poblaciones. Al comienzo, las poblaciones serán polifiléticas: ciertos alelos de la población A tendrán alelos más emparentados en la población B que entre los que se encuentran en aquella. Con el transcurso del tiempo, todos los alelos de una de las dos poblaciones, por ejemplo de A, derivarán de un único ancestro exclusivo de esa población. En otras palabras, A será ahora monofilética. Pero sigue siendo posible que en B haya alelos más cercanos a los de A que a sus coterráneos. Es decir que el acervo genético de la población B será parafilético con respecto a A. Finalmente, las poblaciones se volverán monofiléticas recíprocas. Sabemos que, en promedio, este proceso lleva del orden de $4N$ generaciones para un sistema diploide, N generaciones para un sistema mitocondrial, etc. Estos resultados, naturalmente, se aplican en promedio, y en ausencia de selección, y están representados en la Figura 6.6, con la que concluimos este capítulo.

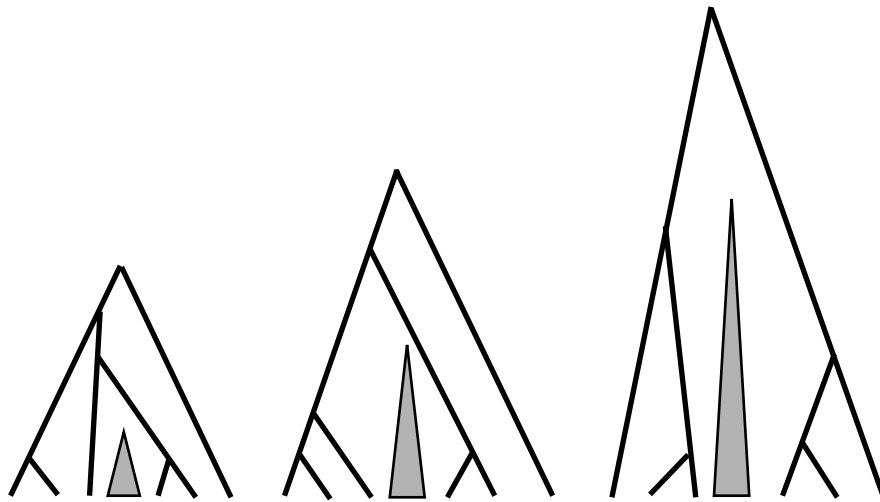


Figura 6.6. Las relaciones genealógicas entre alelos de dos poblaciones progresan desde la polifilia (izquierda) a la monofilia recíproca (derecha), pasando por una fase de parafilia (centro), a medida que transcurre el tiempo desde su separación por una barrera estricta (el triángulo del esquema).

Lecturas recomendadas

Nordborg, M. Coalescent theory. Pp. 179-212 en D. J. Balding, M. Bishop, y C. Cannings (eds.), Handbook of statistical genetics. Wiley, West Sussex, 2001. *Una presentación general del coalescente, bien escrita pero bastante difícil para los estudiantes de biología.*

Pueden consultarse además las breves descripciones de los ya citados libros de Gillespie (1998) y Hartl (2000), aunque ellas no van más allá que este capítulo.