

4. Tamaño poblacional efectivo, deriva y mutación

Enrique Lessa

2024-08-09

Tamaño poblacional efectivo

Para un locus autosómico, nuestro modelo básico de deriva genética, caracterizado en la parte “3. Deriva genética y heterocigosidad”, asume un tamaño poblacional constante N y, para organismos diploides, igual proporción de machos y hembras en la población. En ese contexto, entendemos que el número de alelos a seguir en el proceso de formación de una generación a partir de la precedente varía con el modo de herencia.

En este contexto, podemos decir que en una población diploide, con herencia cromosómica del sexo de tipo XY, y ADN mitocondrial de herencia materna, el número efectivo de alelos es $2N$ alelos para un locus autosómico, $N/2$ alelos para un gen mitocondrial o localizado en el cromosoma Y, y $3N/2$ para un locus del cromosoma X. Sabemos que el “ritmo” o tasa de la deriva genética es función inversa del tamaño poblacional, y aprendimos que dicho “tamaño” no es simplemente el número de adultos, sino que debe ser ajustado para la ploidía y el modo de herencia. Así, el tamaño efectivo de la población para estudiar la evolución de un gen localizado en el cromosoma Y es el número de machos o, si la población tiene una proporción de sexos de 1:1, la mitad de N .

Por convención, para sistemas diploides usamos N como el tamaño efectivo de individuos y tenemos en cuenta la ploidía usando $2N$ de manera sistemática en nuestros cálculos. Si, en cambio, estamos estudiando una población bacteriana, el tamaño efectivo es N , el número de individuos, porque la herencia es haploide. Vamos a definir informalmente un nuevo parámetro “ d ”, la tasa de deriva genética en una población. Así, $d = 1/2N$ y $d = 1/N$ para nuestras poblaciones diploide y haploide, respectivamente.

N_e en una población de tamaño variable en el tiempo

¿Qué sucede si una población de organismos diploide difiere de nuestra población estándar? ¿Por ejemplo, cómo opera la deriva genética en una población que, en lugar de tener un tamaño constante N , varía a lo largo del tiempo? Sabemos bien que esto es más la norma que la excepción, y que las poblaciones naturales fluctúan en tamaño en función de las condiciones ambientales, la carga parasitaria, y muchos otros factores. La respuesta es que la deriva genética operará a una tasa d_e que será el promedio de las tasas en cada generación ($d_1, d_2, \dots, d_t$). En cada caso, $d_i = 1/N_i$. Naturalmente la deriva genética será más intensa cuando d sea grande y menos fuerte cuando d sea pequeña. En promedio:

$$d_e = \frac{1}{t}(d_1 + d_2 + \dots + d_t)$$

Llamamos “tamaño poblacional efectivo” a un valor N_e constante para el cual una población ideal de tipo Wright-Fisher experimenta una tasa de deriva equivalente al de la población de interés, que en este caso tiene N_i variable a lo largo del tiempo.

Reemplazando d_i por $1/N_i$ en la ecuación, obtenemos:

$$\frac{1}{N_e} = \frac{1}{t} \left(\frac{1}{N_1} + \frac{1}{N_2} + \dots \frac{1}{N_t} \right)$$

En palabras, el tamaño efectivo de una población de tamaño variable a lo largo del tiempo es la media armónica de los tamaños poblacionales en cada generación. La media armónica es siempre igual o menor que la media aritmética.

Ejercicio 1. Una población alterna su tamaño entre $N = 100$ y $N = 1000$ individuos.

- a) Calcular el tamaño efectivo de la población y compararlo con la media aritmética.
- b) ¿Cuál es la intensidad de la deriva genética en la población, y cómo se compara con la de una población estable de 550 individuos?

Existen muchos otros factores que afectan el tamaño poblacional efectivo; comunmente, el efecto neto es que $N_e < N$. Siempre que hay variación en el aporte reproductivo de los individuos, el tamaño efectivo se reduce. La razón es que si algunos individuos se reproducen más que otros, aumenta la chance de que varios alelos en la siguiente generación sean copias del mismo alelo ancestral. La contracara de ese resultado es que habrá más alelos de la generación ancestral que no dejarán descendencia, puesto que todos los alelos pugnan por ocupar un número fijo de $2N$ alelos en la siguiente generación.

Un caso de particular interés es el efecto de la proporción de sexos en el tamaño poblacional efectivo. Si dicha proporción difiere de 1:1, entonces uno de los sexos es más abundante que el otro. Pero para un locus autosómico, y con reproducción sexuada, la mitad de los alelos de la siguiente generación proviene de cada sexo.

N_e en función de la proporción de machos y hembras

Para obtener el tamaño efectivo de una población formada por N_m machos y N_f hembras, tenemos que calcular, y luego promediar, el “coeficiente de deriva” según la procedencia de un par de alelos tomados al azar. Tenemos 3 casos posibles para ese par de alelos:

- los dos alelos provienen de hembras de la generación precedente.
- los dos alelos provienen de machos de la generación precedente.
- un alelo proviene de un macho y el otro de una hembra de la generación precedente.

Puesto que la mitad de los alelos de la población proviene de machos y la mitad de hembras, las probabilidades correspondientes son $1/4$, $1/2$, y $1/4$.

A su vez, los “coeficientes de deriva” (probabilidades de coalescencia) correspondientes son $1/2N_f$, 0 , y $1/2N_m$, respectivamente. La siguiente tabla resume las consideraciones anteriores.

Casos	$P(caso)$	$P(coal caso)$
ff	$1/4$	$1/2N_f$
fm	$1/2$	0
mm	$1/4$	$1/2N_m$

En la tabla, ff es el caso en el que los dos alelos tomados al azar provienen de hembras en la generación anterior, fm el caso en que provienen de individuos de distinto sexo, y mm el caso en que provienen de machos.

La probabilidad de coalescencia para esta población es el promedio de cada uno de los tres casos ponderados por su probabilidad de ocurrencia:

$$d = \frac{1}{4} \cdot \frac{1}{2N_f} + \frac{1}{2} \cdot 0 + \frac{1}{4} \cdot \frac{1}{2N_m}$$

Reemplazando d por $1/2N_e$ y despejando, obtenemos:

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

Deriva genética y mutación

Antes de incorporar la mutación, recordemos brevemente las características de nuestros dos modelos:

1. Hardy-Weinberg: no solamente no hay mutación (vemos el comportamiento de las clases alélicas que ya existen en la población), sino que tampoco hay evolución, al no haber tampoco migración, deriva, o selección. En particular, no ocurre deriva porque la población es de tamaño infinito.
2. Wright-Fisher: la población es ahora de un tamaño constante N (en lugar de infinita), lo cual resulta en el proceso de deriva genética. Además, es un modelo haploide porque seguimos el destino de las clases alélicas y sus frecuencias, lo cual es posible para cualquier locus usando la binomial, independientemente de la ploidía y el modo de herencia.

En esta sección, vamos a incorporar la mutación al modelo de Wright-Fisher. A la deriva genética, que a la corta o a la larga solamente genera pérdida de variación, agregamos ahora un proceso que incorpora variación a las poblaciones. El resultado será una población en la cual, de manera aleatoria, ingresan nuevas clases alélicas por mutación y se pierden por deriva, generando un equilibrio dinámico entre los dos procesos.

Una población de tipo Wright-Fisher con mutación tendrá una evolución que depende solamente de dos parámetros: el tamaño poblacional N (más bien, el número de alelos correspondiente), y, en los modelos mutacionales que vamos a considerar, la tasa de mutación μ , que definimos como la probabilidad de mutación por alelo y por generación. Vamos a considerar solamente dos modelos sencillos (y vinculados):

Modelo de alelos infinitos: El número de clases alélicas es tan grande que lo asimilamos a infinito. En la práctica, eso significa que cada clase alélica surge una única vez por mutación. Dos alelos idénticos en estado (con la misma secuencia) necesariamente son, en este contexto, idénticos por descendencia (son descendientes de un mismo alelo ancestral). En términos filogenéticos, no hay homoplasia. Nos interesa que estas consecuencias prácticas sean aplicables a la variación genética de una población.

A esa escala poblacional, el modelo es razonable en buena parte, pero no todos los casos en que μ toma valores bajos. El mismo modelo dejará con frecuencia de tener valor a escala filogenética. Por ejemplo, es frecuente aceptar el modelo para secuencias no repetidas de ADN y de proteínas, pero no para microsatélites u otras regiones hipervariables que, normalmente, requieren otro tipo de modelos.

Ejercicio 2.

- a) ¿Cuántas clases alélicas puede haber en una secuencia de ADN de 100 pares de bases de longitud?
- b) ¿Cuántas clases alélicas puede haber en una secuencia de proteína de 100 aminoácidos de longitud?

Modelo de sitios infinitos: La razón por la cual el modelo de alelos infinitos es razonable es que normalmente los alelos son largas secuencias (de nucleótidos o aminoácidos). En el modelo de alelos infinitos, dos clases alélicas tienen al menos una diferencia en sus secuencias, pero pueden tener más. Para considerar la variación en los sitios de una secuencia larga, pensamos en el modelo de sitios infinitos, donde los sitios son tantos (en el límite infinitos) y la tasa de mutación por sitio es tan baja que, en la práctica, o los sitios son invariantes o han mutado una única vez. Este modelo tampoco tiene homoplasia, y se usa para secuencias no repetidas con tasas de mutación bajas.

Equilibrio entre deriva y mutación

Consideramos una población de tipo Wright-Fisher, y un locus diploide con una probabilidad μ de mutar (por alelo individual y por generación) bajo un modelo de alelos infinitos. Recordemos que estamos considerando un modelo neutral (sin selección), de modo que μ es la tasa de mutación neutral. Vamos a calcular la heterocigosidad esperada para un locus con esas características. La heterocigosidad H es la probabilidad de que dos alelos tomados al azar sean distintos, es decir que pertenecen a dos clases alélicas distintas o, lo que es lo mismo, sus secuencias tienen al menos una diferencia.

Notamos que si tomamos dos alelos en la generación t (por ejemplo, la actual), serán iguales o diferentes según su historia previa. Por causa de la deriva genética, necesariamente tienen un ancestro común en alguna generación en el pasado. Siguiendo la historia de los dos alelos hacia el pasado, llamamos “coalescencia” al momento en que esos dos alelos convergen en su ancestro común más reciente. En cada generación, la probabilidad de coalescencia nos es familiar, aunque sin ese nombre, y es la probabilidad de que dos alelos en una generación t sean copias de un mismo alelo ancestral en la generación inmediatamente precedente $t - 1$. Para nuestro par de alelos, la probabilidad de coalescencia por generación es $1/2N$, nuestro “coeficiente de deriva”.

Dos alelos serán iguales (de la misma clase alélica) si no hubo mutación en ninguna de las dos líneas (o ramas de la filogenia que los une) que los conectan con su ancestro común más reciente. Si ocurrió al menos una mutación en al menos una de las dos líneas, entonces los alelos serán diferentes bajo nuestro modelo mutacional. ¿Cuál es la probabilidad de mutación por generación? Para cada alelo, la probabilidad de mutación en la línea que lo conecta con su ancestro en $t - 1$ es μ . Si μ es pequeño, podemos despreciar que haya dos mutaciones en la misma generación (una en cada línea). Por tanto, la probabilidad de mutación para el par de alelos por generación es, aproximadamente 2μ .

Para calcular la heterocigosidad esperada, consideramos: a) “casos favorables”: los dos alelos son distintos con probabilidad 2μ b) “casos desfavorables”: los dos alelos son iguales con probabilidad $1/2N$ c) el espacio muestral comprende los dos tipos de casos y es la suma de los dos valores anteriores.

Concluimos que la heterocigosidad esperada es:

$$E(H_{DM}) = \frac{2\mu}{2\mu + 1/2N}$$

Multiplicando numerador y denominador por $2N$, obtenemos.

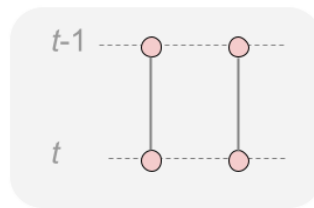
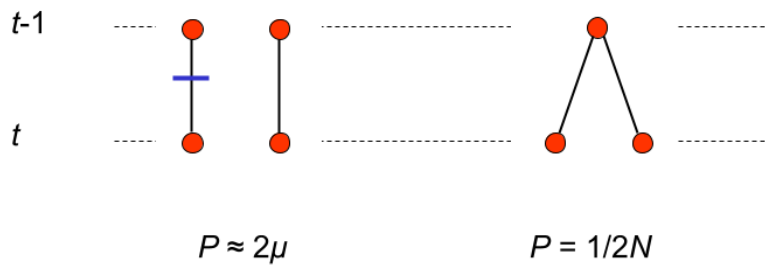
$$E(H_{DM}) = \frac{4N\mu}{4N\mu + 1}$$

El parámetro compuesto $4N\mu$ es tan importante en genética de poblaciones que usamos la letra $\theta = 4N\mu$ para identificarlo. Sustituyendo en la ecuación anterior, obtenemos:

$$E(H_{DM}) = \frac{\theta}{\theta + 1}$$

Observamos que $E(H_{DM})$ crecerá cuando el tamaño poblacional N y/o la tasa de mutación neutral μ aumenten. Intuitivamente, una población grande pierde menos variación por deriva genética que una población pequeña. Y un locus con una tasa de mutación neutral mayor incorpora nuevos alelos con mayor frecuencia que un locus con una tasa de mutación menor. El equilibrio dinámico entre estos dos procesos contrapuestos es una función de la intensidad relativa de los mismos.

La siguiente figura ilustra estos conceptos, y nos permite una aclaración adicional:



En la parte superior, los diagramas ilustran los dos tipos de eventos que hemos considerado (mutación y coalescencia) con sus probabilidades respectivas. Las dos probabilidades relevantes tienen normalmente valores muy bajos (N es grande y μ es pequeño). Por eso despreciamos la posibilidad de que ocurra más de uno de estos eventos por generación.

Ejercicio 3.

Ilustrar al estilo de la figura anterior los siguientes eventos, y calcular sus probabilidades respectivas:

- a) Dos mutaciones en una misma generación.
- b) Un coalescente y una mutación en una misma generación.

La parte inferior de la figura ilustra la circunstancia más común: que al trazar dos alelos hacia la generación inmediatamente precedente no observemos ni una mutación ni una coalescencia ("no pasa nada"). ¿Por qué este evento tan común no interviene en la estimación de $E(H)$? La razón es que "no pasa nada" en una generación seguimos sin saber si en la historia de los dos alelos ocurrió una mutación antes del coalescente o a la inversa. En ese caso, consideramos la probabilidad de que los dos alelos ancestrales (ahora en el tiempo $t - 1$) puedan coalescer o sufrir una mutación cuando los trazamos a $t - 2$... y así sucesivamente a lo largo de las generaciones. Si ocurre un coalescente antes de una mutación, los dos alelos serán iguales. Si ocurre al menos una mutación antes del coalescente, serán diferentes. En suma, solamente importan los valores relativos de $1/2N$ y 2μ .

Ejercicios resueltos

Ejercicio 1. Una población alterna su tamaño entre $N = 100$ y $N = 1000$ individuos.

- a) Calcular el tamaño efectivo de la población y compararlo con el la media aritmética.

El tamaño efectivo es el inverso de

$$\frac{1}{N_e} = \frac{1}{2} \left(\frac{1}{100} + \frac{1}{1000} \right) = 0.055$$

.

$$N_e \approx 182$$

Notamos que el tamaño efectivo está mucho más cerca del mínimo (100) que del máximo (1000).

- b) ¿Cuál es la intensidad de la deriva genética en la población, y cómo se compara con la de una población estable de 550 individuos?

Nuestro informal “coeficiente de deriva”, el inverso del tamaño poblacional, nos da la clave. La intensidad de la deriva en esta población es aproximadamente $1/182$, que es aproximadamente 3 veces más que $1/550$. En comparación con una población estable de 550 individuos, nuestra población, que oscila entre 100 y 1100 individuos pierde, en promedio, heterocigosidad a un ritmo 3 veces mayor.

Notamos, de paso, que tanto en la parte a como en la b las proporciones se mantienen si usamos números de alelos diploides ($2N_e$, 200 y 2000 en los cálculos).

Ejercicio 2.

- a) ¿Cuántas clases alélicas puede haber en una secuencia de ADN de 100 pares de bases de longitud?

Como hay 4 nucleótidos posibles para cada sitio en la secuencia, la respuesta es $4^{100} \approx 1.6 \exp 60$.

- b) ¿Cuántas clases alélicas puede haber en una secuencia de proteína de 100 aminoácidos de longitud?

De manera análoga, en este caso la respuesta es $20^{100} \approx 1.3 \exp 130$.

Estos ejemplos ilustran por qué los modelos de “alelos infinitos” o “sitios infinitos” son aproximaciones razonables para estudiar la variación dentro de las poblaciones.

Ejercicio 3.

Ilustrar al estilo de la figura anterior los siguientes eventos, y calcular sus probabilidades respectivas:

- a) Dos mutaciones en una misma generación.

La probabilidad es μ^2 . Si μ es bajo, μ^2 es despreciable.

b) Un coalescente y una mutación en una misma generación.

Esta probabilidad es aproximadamente $2\mu/2N$, también despreciable si μ es pequeño y N es grande.