Tamaño poblacional efectivo (N_e)

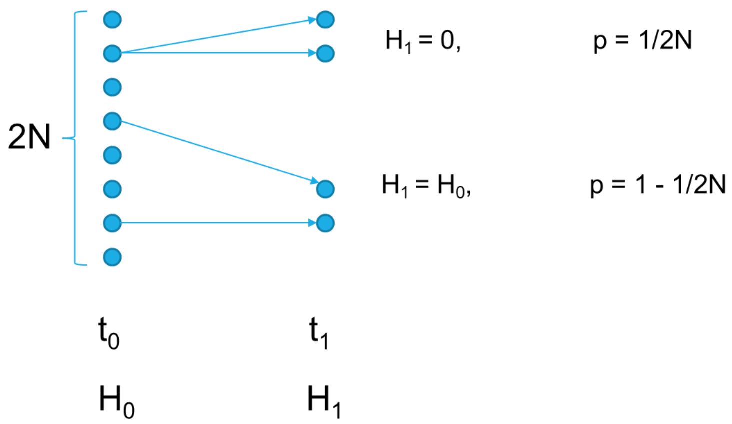
Curso de Evolución 2021

Contexto: pérdida de heterocigosidad por deriva genética en el modelo Wright-Fisher (WF)

Definimos la heterocigosidad *H* como la probabilidad de que dos alelos tomados al azar sean diferentes (i.e., pertenezcan a clases alélicas distintas, es decir que tengan al menos una diferencia en sus secuencias).

En un modelo WF sin mutación, consideramos la esperanza de la heterocigosidad en el tiempo t_1 $E(H_1)$, dado un cierto valor H_0 en t_0. Para eso, consideramos dos eventos mutuamente excluyentes que pueden resultar de muestrear dos alelos al azar en t_1:

- Los dos alelos son copias de un mismo alelo ancestral en t_0:
  - Este evento ocurre con probabilidad 1/2N
  - En este caso el valor de $H_1 = 0$ (no hay mutación, de modo que los dos alelos son copias idénticas de un mismo alelo ancestral).
- Los dos alelos son copias de alelos distintos en t_0:
  - Este evento es el complementario del antes considerado, de modo que ocurre con probabilidad 1-1/2N.
  - En este caso, el valor de H_1 es H_0: puesto que los dos alelos en H_1 son copias idénticas de dos alelos en t_0 mantienen sus propiedades, incluyendo la probabilidad de que sean diferentes.



La esperanza de una variable es la suma de los valores que toma dicha variable, cada uno ponderado por su probabilidad. Como H_1 toma solamente dos valores (0 y H_0), usamos, por comodidad, la fórmula de la esperanza para variables discretas:

$$E\left(H_1\right) = \sum_{i=1}^{3} H_i p_i$$

Así obtenemos:

$$E\left(H_1\right) = 0 \cdot \frac{1}{2N} + H_0\left(1 - \frac{1}{2N}\right) = H_0\left(1 - \frac{1}{2N}\right)$$

En palabras, la heterocigosidad esperada en una generación es igual a la heterocigosidad en la generación precedente menos una quita proporcional de una fracción 1/2N de la heterocigosidad en la generación precedente. Podemos definir informalmente una "tasa de deriva" "d" = 1/2N, artilugio que usaremos más abajo para definir esta fracción para loci que no son autosómicos.

Observaciones:

- Esta es una esperanza. En una trayectoria particular para un locus, H puede fluctuar en cualquier dirección, siguiendo las normas de la binomial, siempre que no se pierda la variación alélica, en cuyo caso H=0 (estado absorbente).
- Si observamos múltiples trayectorias con el mismo punto de partida, el promedio de los valores de H en cada generación se aproximará a la esperanza.

Para obtener la esperanza de H_2, usamos el mismo razonamiento, por lo cual:

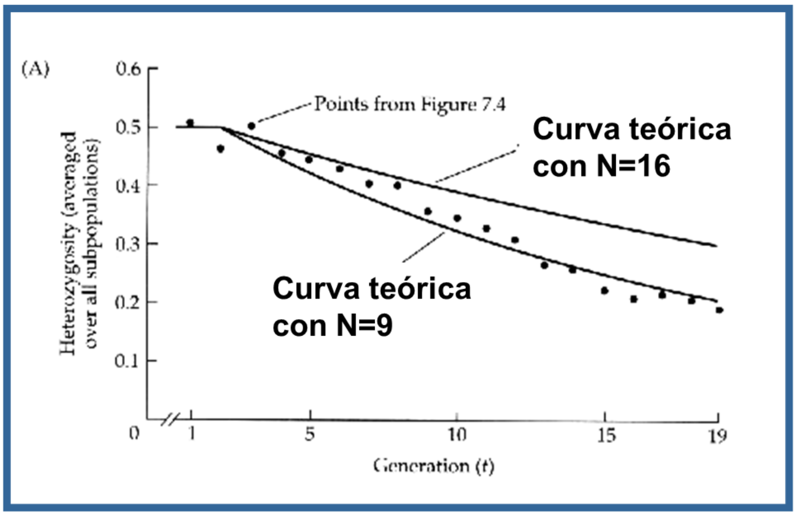
$$E\left(H_2\right) = H_1\left(1 - \frac{1}{2N}\right) = H_0\left(1 - \frac{1}{2N}\right)^2$$

Generalizando:

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

Tamaño efectivo estimado empíricamente

Recordemos la gráfica que traza la heterocigosidad observada media en función del tiempo, en un experimento con numerosas cajas de poblaciones de la mosca de la fruta con N = 16 (8 machos y 8 hembras).



Si la población se comportase como predice el modelo WF, deberíamos observar una pérdida media de heterocigosidad como la indicada en la figura. En cambio, observamos una pérdida de heterocigosidad que ajusta mucho mejor al mismo modelo con N=9. Las causas pueden ser variadas y complejas (la reproducción está distribuida de manera desigual entre los machos, por ejemplo)… pero podemos seguir usando tranquilamente el modelo WF si, en lugar de N (el número de individuos adultos) usamos un valor corregido, que llamamos N_e.

Tamaño efectivo para loci que no son diploides autosómicos

Recordemos que el modelo de WF es haploide, en el sentido de que el muestreo de una generación a partir de la precedente se hace sobre los alelos individuales de ésta. En otras palabras, el número efectivo de alelos para un sistema diploide autosómico es 2N. Por tanto, el ritmo de pérdida de heterocigosidad 1/2N es el inverso del número de alelos (2N porque hemos razonado siempre sobre loci diploides autosómicos).

De manera análoga, podemos entender cuál es el número efectivo de alelos, en el modelo WF, para otros tipos de loci.

Locus	*Número efectivo de alelos*	*"d"*
Diploide autosómico	2N	1/2N
Mitocondrial	$N_f = N/2$	$1/2N_f = 1/4N$
Cromosoma Y	$N_m = N/2$	$1/2N_m = 1/4N$
Cromosoma X	¾ . 2N = 3N/2	2/3N

Observamos que el ritmo de la deriva genética será mayor en genes mitocondriales y ligados al cromosoma Y que en loci autosómicos porque en estos dos casos el número efectivo de alelos será N_h y N_m, respectivamente. Si la proporción de sexos es 1:1, en ambos casos el tamaño efectivo es ¼ de N.

En el caso de loci en el cromosoma X, y asumiendo una proporción sexual de 1:1, observamos que hay 3 alelos ligados al X por cada 4 alelos autosómicos, y hacemos las cuentas en función de ello.

Proporción sexual y tamaño efectivo

Vamos a considerar ahora nuestro locus diploide autosómico para una población de tamaño N compuesta por dos subpoblaciones: N_f hembras y N_m machos. Queremos obtener el ritmo de pérdida esperada de heterocigosidad "d". Como tenemos dos subpoblaciones y tomamos pares de alelos al azar, tenemos tres casos a considerar, y para cada uno tenemos "d" y una probabilidad asociada:

- Los dos alelos provienen de las hembras: este evento ocurre con $p_1 = 1/4$ (notar que, independientemente de la razón de los sexos, la mitad de los alelos de la generación 1 proviene de las hembras de la generación precedente), y

$$d_1 = d_f = 1/2N_f$$

- Los dos alelos provienen de los machos: en forma análoga, $p_2 = 1/4$, y

$$d_2 = d_m = 1/2N_m$$

- Los dos alelos provienen de sexos diferentes: $p_3 = ½$. En este caso, los alelos son necesariamente descendientes de distintos alelos en t_0, así que

$$d_3 = 0$$

Combinando los resultados anteriores, tenemos que:

$$E(d) = \sum_{i=1}^{3} d_i p_i$$

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

$$\frac{1}{N_e} = \frac{1}{4N_f} + \frac{1}{4N_m}$$

Por tanto:

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

Observamos que N_e es el doble de la media armónica de N_f y N_m. Observamos también que si $N_f = N_m = N/2$, entonces $N_e = N$.

Esto es importante porque nos indica que el uso del modelo haploide WF para estudiar la pérdida de heterocigosidad por deriva genética es correcto.

Por otra parte, observamos que un sesgo en la proporción de sexos, que implica que la población tiene números desiguales de machos y hembras, tiene un tamaño efectivo menor al número de adultos. Algunos ejemplos:

N_f	N_m	N	N_e
50	50	100	100
70	30	100	84
90	10	100	40

Sabemos que la media armónica es siempre igual o menor que la media aritmética. En términos biológicos, como la mitad de los alelos de una generación proviene de cada sexo, la intensidad de la deriva está marcada principalmente por el número de individuos del sexo de menor frecuencia. En el último ejemplo, el tamaño efectivo es 40 (el doble de la media armónica entre 90 y 10), mientras que el doble de la media aritmética es, naturalmente, 100.

Tamaño poblacional variable

Utilizando una lógica análoga de obtener el tamaño efectivo como un promedio ponderado de ritmos de deriva en subpoblaciones (de machos y hembras, por ejemplo), es fácil entender que el tamaño poblacional efectivo de una población de tamaño poblacional variable N_1, N_2, … N_t es la media armónica de los tamaños poblacionales:

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