

## Sobre la Deriva Genética

La evolución biológica consiste en el cambio de las características hereditarias a través de las generaciones. Los fenómenos que más intervienen en el proceso evolutivo son la selección natural y la deriva genética. En este trabajo vamos a tratar de describir por medio de un sencillo modelo matemático, el fenómeno conocido en genética de poblaciones por *deriva* (o también a veces como *consanguinidad*), Cabrero et AL. (2002), Gale (1990), Drake et al (1998).

Consideremos una población de organismos diploides y cuyo tamaño poblacional, para simplificar, supondremos constante,  $N$ , de generación en generación. También por simplicidad trataremos al tiempo como una variable discreta, cuantificado en término de generaciones. Supondremos apareamiento al azar (*panmixia*) y no consideraremos ni migración, ni mutación, ni selección.

Vamos a estudiar, para un locus autosómico determinado, con dos alelos posibles,  $H$  y  $h$ , como cambian las frecuencias génicas y genotípicas, especialmente la frecuencia de heterocigotos  $Hh$ , de generación en generación.

Sea  $W_j = (X_j, Y_j, Z_j)$  un vector aleatorio que representa el número de individuos con genotipos  $HH$ ,  $Hh$  y  $hh$  respectivamente, en la generación  $j$ . Observemos que  $X_j + Y_j + Z_j = N$  y como el total de alelos para este locus es  $2N$ .

Como consecuencia de la hipótesis de *panmixia*, y en ausencia de mutación, selección o migración, si en la  $j$ -ésima generación  $X_j = x_j$ ,  $Y_j = y_j$  y  $Z_j = z_j$ , al formar un cigoto de la generación  $j + 1$ , tendremos las siguientes probabilidades

$$\begin{aligned} p &= P(HH) = \left(\frac{a_j}{2N}\right)^2, \\ q &= P(Hh) = 2\left(\frac{a_j}{2N}\right)\left(1 - \frac{a_j}{2N}\right), \\ r &= P(hh) = \left(1 - \frac{a_j}{2N}\right)^2 \end{aligned} \quad (1)$$

siendo  $a_j = 2x_j + y_j$ , la frecuencia absoluta del alelo  $H$  en la generación  $j$ .

La distribución de  $W_{j+1} = (X_{j+1}, Y_{j+1}, Z_{j+1})$  condicionada a  $W_j = (X_j, Y_j, Z_j)$  será una trinomial, i.e.:

$$\begin{aligned} P[X_{j+1} = x_{j+1}, Y_{j+1} = y_{j+1}, Z_{j+1} = z_{j+1} | X_j = x_j, Y_j = y_j, Z_j = z_j] &= \\ &= \frac{N!}{x_{j+1}! y_{j+1}! z_{j+1}!} p^{x_{j+1}} q^{y_{j+1}} r^{z_{j+1}} \end{aligned}$$

La función característica de  $W_{j+1} = (X_{j+1}, Y_{j+1}, Z_{j+1})$ , que nos servirá para hallar la esperanza de  $Y_{j+1}$ , la podremos calcular mediante:

$$\begin{aligned} \Phi_{W_{j+1}}(t_1, t_2, t_3) &= E\left(e^{i(t_1 X_{j+1} + t_2 Y_{j+1} + t_3 Z_{j+1})}\right) \\ &= E\left(E\left(e^{i(t_1 X_{j+1} + t_2 Y_{j+1} + t_3 Z_{j+1})} | W_{j+1}\right)\right) \\ &= \sum_{x_j + y_j + z_j = N} P(X_j = x_j, Y_j = y_j, Z_j = z_j) \Phi_{W_{j+1}|W_j}(t_1, t_2, t_3) \\ &= \sum_{x_j + y_j + z_j = N} P(X_j = x_j, Y_j = y_j, Z_j = z_j) (pe^{it_1} + qe^{it_2} + re^{it_3})^N \end{aligned}$$

Por tanto,

$$\frac{\partial \Phi}{\partial t_2} = i \sum_{x_j + y_j + z_j = N} P(X_j = x_j, Y_j = y_j, Z_j = z_j) N(p e^{it_1} + q e^{it_2} + r e^{it_3})^{N-1} q e^{it_2}$$

y como  $iE(Y_{j+1}) = \frac{\partial \Phi}{\partial t_2}(0, 0, 0)$  y  $p + q + r = 1$ , teniendo en cuenta (6), resulta

$$\begin{aligned} E(Y_{j+1}) &= \sum_{x_j + y_j + z_j = N} P(X_j = x_j, Y_j = y_j, Z_j = z_j) a_j \left(1 - \frac{a_j}{2N}\right) \\ &= E(A_j) - \frac{1}{2N} E(A_j^2) \end{aligned}$$

donde  $A_j$  es la variable aleatoria número de alelos  $H$  presentes en la generación  $j$ .

Alternativamente podemos escribir también:

$$E(Y_{j+1}) = E(A_j) - \frac{1}{2N} (\text{var}(A_j) + E^2(A_j)). \quad (2)$$

Vemos pues que  $E(Y_j)$  depende de la esperanza y varianza de  $A_j$ , por lo que procederemos a calcularlas. La variable aleatoria  $A_j$ , el número de alelos  $H$  en la generación  $j$ , puede tomar valores  $0, 1, \dots, 2N$ . Supongamos que en la generación inicial el número de alelos  $H$  que hay en la población es  $c$ , conocido,  $A_0 = c$ . Vamos a calcular los momentos de la variable aleatoria  $A_j$  a partir de la función característica, como hemos hecho anteriormente.

$$\begin{aligned} \Phi_{A_{j+1}}(t) &= E(e^{itA_{j+1}}) = E(E(e^{itA_{j+1}} | A_j)) \\ &= \sum_{k=0}^{2N} P(A_j = k) \Phi_{A_{j+1}|A_j=k}(t) \end{aligned}$$

donde  $\Phi_{A_{j+1}|A_j=k}(t)$  es la función característica de  $A_{j+1}$  condicionada a  $A_j = k$ .

La distribución de la variable aleatoria  $A_{j+1}$  condicionada a  $A_j = k$ , como consecuencia de la hipótesis de *panmixia* y en ausencia de mutación, selección o migración, será una Binomial,  $B(2N, \frac{k}{2N})$ , por tanto

$$\begin{aligned} \Phi_{A_{j+1}}(t) &= \sum_{k=0}^{2N} P(A_j = k) \left(\frac{k}{2N} e^{it} + 1 - \frac{k}{2N}\right)^{2N} \\ \Phi'_{A_{j+1}}(t) &= i \sum_{k=0}^{2N} P(A_j = k) 2N \left(\frac{k}{2N} e^{it} + 1 - \frac{k}{2N}\right)^{2N-1} \frac{k}{2N} e^{it} \end{aligned}$$

y como  $iE(A_{j+1}) = \Phi'_{A_{j+1}}(0)$ , resulta:

$$E(A_{j+1}) = \sum_{k=0}^{2N} P(A_j = k) k = E(A_j)$$

lo que implica, por recurrencia, que

$$E(A_{j+1}) = E(A_j) = \dots = E(A_0) = c.$$

Es decir que el *valor medio* de las frecuencias de los alelos  $H$  permanece constante a lo largo de todas las generaciones.

Para obtener la varianza de  $A_{j+1}$ , calcularemos primero  $E(A_{j+1}^2)$  a través de la derivada segunda de la función característica, ya que  $-E(A_{j+1}^2) = \Phi''_{A_{j+1}}(0)$ .

$$\begin{aligned} \Phi''_{A_{j+1}}(t) &= -\sum_{k=0}^{2N} P(A_j = k) \left( 2N(2N-1) \left( \frac{k}{2N} e^{it} + 1 - \frac{k}{2N} \right)^{2N-2} \frac{k^2}{4N^2} e^{2it} \right. \\ &\quad \left. + 2N \left( \frac{k}{2N} e^{it} + 1 - \frac{k}{2N} \right)^{2N-1} \frac{k}{2N} e^{it} \right) \end{aligned}$$

por lo tanto

$$\begin{aligned} E(A_{j+1}^2) &= \sum_{k=0}^{2N} P(A_j = k) \left( \frac{2N(2N-1)}{4N^2} k^2 + k \right) \\ &= \left( 1 - \frac{1}{2N} \right) E(A_j^2) - E(A_j) \\ &= \left( 1 - \frac{1}{2N} \right) \left( \text{var}(A_j) + E(A_j)^2 \right) - E(A_j) \\ &= \left( 1 - \frac{1}{2N} \right) \left( \text{var}(A_j) + c^2 \right) - c. \end{aligned}$$

y

$$\text{var}(A_{j+1}) = \left( 1 - \frac{1}{2N} \right) \left( \text{var}(A_j) + c^2 \right) + c - c^2 = \left( 1 - \frac{1}{2N} \right) \text{var}(A_j) + c \left( 1 - \frac{c}{2N} \right)$$

Reiterando el proceso, al ser  $A_0$  una constante,  $\text{var}(A_0) = 0$ , tendremos

$$\begin{aligned} \text{var}(A_1) &= c \left( 1 - \frac{c}{2N} \right) \\ \text{var}(A_2) &= c \left( 1 - \frac{c}{2N} \right) \left( 1 + \left( 1 - \frac{1}{2N} \right) \right) \\ \text{var}(A_3) &= c \left( 1 - \frac{c}{2N} \right) \left( 1 + \left( 1 - \frac{1}{2N} \right) + \left( 1 - \frac{1}{2N} \right)^2 \right) \\ &\dots\dots\dots \\ &\dots\dots\dots \\ \text{var}(A_n) &= c \left( 1 - \frac{c}{2N} \right) \left( 1 + \left( 1 - \frac{1}{2N} \right) + \left( 1 - \frac{1}{2N} \right)^2 + \dots + \left( 1 - \frac{1}{2N} \right)^{n-1} \right) \end{aligned}$$

y teniendo en cuenta la fórmula de la suma de los  $n$  primeros términos de una progresión geométrica, tendremos finalmente

$$\text{var}(A_n) = 2Nc \left( 1 - \frac{c}{2N} \right) \left( 1 - \left( 1 - \frac{1}{2N} \right)^n \right)$$

Podemos observar, a partir de la anterior expresión, que, fijado el tamaño poblacional  $N$ , la varianza de la variable aleatoria, número de alelos  $H$ , aumenta de generación en generación. Cuando el número de generaciones tiende a infinito, la varianza será:

$$\lim_{n \rightarrow \infty} \text{var}(A_n) = 2Nb \left( 1 - \frac{b}{2N} \right) = 2N \text{var}(A_1) \quad (3)$$

Volviendo ahora a la expresión (2), tendremos finalmente

$$E(Y_{j+1}) = c \left(1 - \frac{c}{2N}\right) \left(1 - \frac{1}{2N}\right)^{j+1}$$

Si suponemos que la población inicialmente ya estaba en equilibrio Hardy-Weinberg, las frecuencias relativas de los diferentes genotipos al inicio son:

$$f_r(HH) = \left(\frac{c}{2N}\right)^2, \quad f_r(Hh) = 2 \left(\frac{c}{2N}\right) \left(1 - \frac{c}{2N}\right), \quad f_r(hh) = \left(1 - \frac{c}{2N}\right)^2$$

Por lo tanto

$$Y_0 = 2N \left(\frac{c}{2N}\right) \left(1 - \frac{c}{2N}\right)$$

y resulta finalmente a:

$$E(Y_n) = Y_0 \left(1 - \frac{1}{2N}\right)^n$$

lo que significa que la esperanza de la frecuencia del genotipo  $Hh$  en la generación  $n$ , depende de la cantidad inicial de heterocigotos y va disminuyendo a medida que transcurren las generaciones. Si calculamos el límite cuando  $n \rightarrow \infty$  tendremos:

$$\lim_{n \rightarrow \infty} E(Y_n) = \lim_{n \rightarrow \infty} Y_0 \left(1 - \frac{1}{2N}\right)^n = 0$$

Lo que demuestra que, en las condiciones estudiadas, los heterocigotos se extinguen.

A continuación exponemos unas gráficas, en donde se pone de manifiesto los resultados teóricos obtenidos anteriormente. En Figura 1 realizamos simulaciones con 1000 individuos y con 250 alelos  $H$  al inicio. Observamos que en las generaciones, 216, 832, 1748 el alelo  $H$  ha desaparecido y que en la generación 1721, con las mismas condiciones iniciales, todos los alelos de la población, son  $H$ . En la Figura 2, vemos que en las generaciones 299, 678, 1798 todos los alelos son  $H$  y que con las mismas condiciones experimentales, los alelos  $H$  han desaparecido en la generación 1367. En general si el número de alelos inicial  $b$ , está más próximo a cero que a  $2N$ , es más fácil que se alcance el estado estacionario 0 que el  $2N$ . Viceversa, si está más cerca de  $2N$  es más fácil que acaben siendo todos los alelos  $H$ . Además, en general, cuando más cerca esté el número de alelos  $H$  del estado absorbente, 0 ó  $2N$ , o más pequeño sea el tamaño de la población, más rápido se alcanza esta situación. Ver en Figuras 3 que con 250 individuos y con el 50 % de alelos  $H$  al inicio, acaban siendo todos  $H$  en la generación 578 y que con 1000 individuos y con el 50 % de alelos  $H$  al inicio, acaban siendo todos  $H$  en la generación 1620.

Vamos a tratar el tema utilizando una cadena de Markov.

En una población de  $N$  individuos, en las mismas condiciones que las anteriormente descritas, con dos posibles alelos  $H$  y  $h$ , para un locus autosómico determinado, tendremos los  $2N + 1$  estados  $\{A_0, A_1, \dots, A_{2N}\}$ , donde  $A_i$  es el número de alelos  $H$  en el total de la población. En la generación  $k$  tendremos los estados  $\{A_0^k, A_1^k, \dots, A_{2N}^k\}$ .

La probabilidad  $p_{ij}$  de pasar del estado  $A_i$  al estado  $A_j$ , al cambiar de la generación  $k$  a la  $k + 1$  será:

$$p_{ij} = p(A_j^{k+1}/A_i^k) = \frac{(2N)!}{(2N-j)!j!} \left(\frac{i}{2N}\right)^j \left(\frac{2N-i}{2N}\right)^{2N-j}$$

La matriz de transición  $P = (p_{ij})$ ;  $i, j = 0, 1, \dots, 2N$  es de la forma:

$$P = \begin{pmatrix} 1 & \mathbf{0}' & 0 \\ \mathbf{u} & P_1 & \mathbf{v} \\ 0 & \mathbf{0}' & 1 \end{pmatrix} \quad (4)$$

donde:

$\mathbf{0}$  es el vector nulo de orden  $2N - 1$ ,

$\mathbf{u}' = (p_{10}, p_{20}, \dots, p_{2N-10})$ ,  $\mathbf{v}' = (p_{12N}, p_{22N}, \dots, p_{2N-12N})$ , ( $p_{10} = p_{2N-12N}$ ,  $p_{20} = p_{2N-12N}$ ,  $\dots$ ,  $p_{2N-10} = p_{12N}$ ),

y  $P_1$  es una matriz de orden  $(2N - 1, 2N - 1)$  cuyos elementos son mayores que cero e inferiores a la unidad.

Se puede demostrar fácilmente por inducción que

$$P^n = \begin{pmatrix} 1 & \mathbf{0}' & 0 \\ (I - P_1)^n (I - P_1)^{-1} \mathbf{u} & P_1^n & (I - P_1)^n (I - P_1)^{-1} \mathbf{v} \\ 0 & \mathbf{0}' & 1 \end{pmatrix} \quad (5)$$

Cuando transcurre un número suficientemente grande de generaciones, la matriz de transición que da como:

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 1 & \mathbf{0}' & 0 \\ (I - P_1)^{-1} \mathbf{u} & Q(0) & (I - P_1)^{-1} \mathbf{v} \\ 0 & \mathbf{0}' & 1 \end{pmatrix} \quad (6)$$

$Q(0)$  es una matriz  $(2N - 1, 2N - 1)$  cuyos elementos son todos nulos.

El vector columna  $\mathbf{c} = (I - P_1)^{-1} \mathbf{u}$ , tiene por componentes  $c_j = \frac{2N-j}{2N}$ ,  $j = 1, \dots, 2N - 1$  ya que el vector  $\mathbf{u} = (I - P_1) \mathbf{c}$  tiene por componentes:

$$u_j = p_{j0} = \left(\frac{2N-j}{2N}\right)^{2N} \text{ que cumplen}$$

$$u_j = c_j - \sum_{\alpha=1}^{2N-1} p_{j\alpha} c_\alpha$$

En efecto

$$\begin{aligned} \left(\frac{2N-j}{2N}\right)^{2N} &= \frac{2N-j}{2N} - \sum_{\alpha=1}^{2N-1} \frac{2N!}{(2N-\alpha)! \alpha!} \left(\frac{j}{2N}\right)^\alpha \left(\frac{2N-j}{2N}\right)^{2N-\alpha} \left(\frac{2N-\alpha}{2N}\right) = \\ &= \frac{2N-j}{2N} - \left( \left(\frac{2N-\frac{2Nj}{2N}}{2N}\right) \left(\frac{2N-j}{2N}\right)^{2N} - 0 \right) = \end{aligned}$$

$$= \frac{2N-j}{2N} - \left( \frac{2N-j}{2N} - \left( \frac{2N-j}{2N} \right)^{2N} \right) = \left( \frac{2N-j}{2N} \right)^{2N}$$

De manera similar puede demostrarse que el vector columna  $\mathbf{d} = (I - P_1)^{-1}\mathbf{v}$ , tiene por componentes  $d_j = \frac{j}{2N}$ ,  $j = 1, \dots, 2N - 1$

Si en la población inicial tenemos  $q$  alelos  $H$ , el vector estado inicial es  $\mathbf{q} = (0, \dots, 1, \dots, 0)$ , es decir todas las componentes son cero excepto la que ocupa el lugar  $q$ . Transcurrido un número suficientemente grande de generaciones, tendremos un vector estado

$$\mathbf{f} = \left( \frac{2N-q}{2N}, 0, \dots, 0, \frac{q}{2N} \right)$$

es decir estaríamos en el estado  $A_0$  (todos los genotipos de la población serían  $hh$  con probabilidad  $\frac{2N-q}{2N}$  o todos serían  $HH$  con probabilidad  $\frac{q}{2N}$ )

#### Bibliografía

Cabrero, J., Camacho J.P. (2002). Fundamentos de Genética de Poblaciones. En: Soler M. (Ed.) *EVOLUCIÓN. La base de la Biología*, pp. 83-126. Proyecto Sur de Ediciones, S.L.

Drake, J.W., Charlesworth, B., Charlesworth, D., Crow, J.F. (1998). Rates of spontaneous mutation. *Genetics* **148**:1667-1686.

Gale, J.S. (1990) Theoretical population genetics. *Unwin Hyman Ltd. London*.

Martín Ríos.

Departamento de Estadística.

Facultad de Biología.

Universidad de Barcelona.

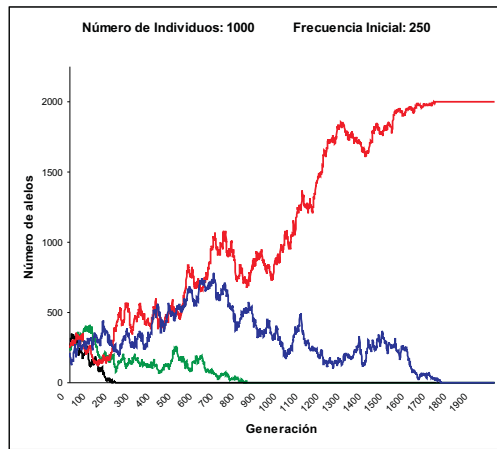


Figura 1:

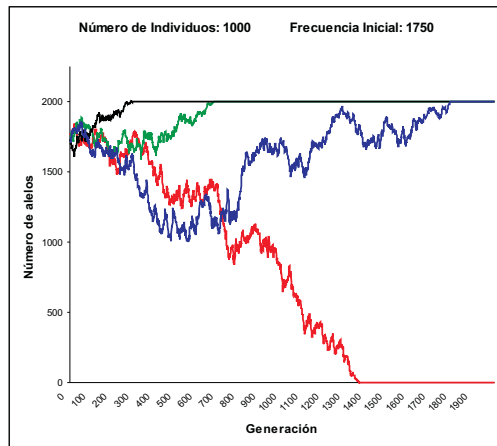


Figura 2:

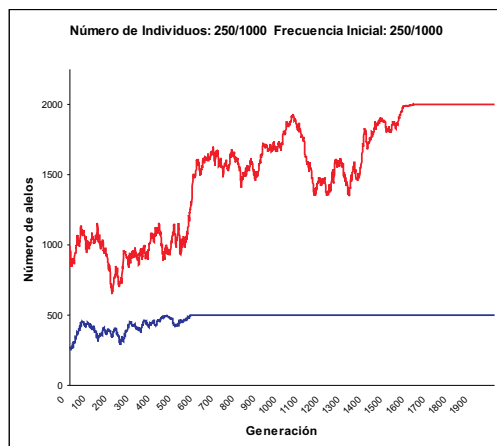


Figura 3: