

## EL EFECTO BULMER

A. Blasco

Instituto de Ciencia y Tecnología animal. Universidad Politécnica de Valencia  
 ablasco@dca.upv.es

1. Efecto de la selección sobre la varianza en un modelo con dos loci
  - 1.1. Efecto de la selección sobre la varianza en un locus con dos alelos
  - 1.2. El Desequilibrio de ligamiento
  - 1.3. Efecto de la selección sobre la covarianza entre dos loci
2. Efecto de la selección sobre la varianza en el modelo infinitesimal
  - 2.1. Efecto de la selección sobre la varianza
  - 2.2. Efecto de la selección sobre la suma de las varianzas de cada locus
  - 2.3. El efecto Bulmer

En 1970 James Crow y Motoo Kimura publicaron uno de los libros que más impacto ha tenido en genética cuantitativa, bajo el modesto título de "An introduction to the Genetics of Populations". Allí demostraban que la selección puede modificar la media, pero no la varianza aditiva de un carácter considerada como la suma de las varianzas de todos los loci implicados. Esto parecía conducir a pensar que la varianza aditiva, bajo el modelo infinitesimal, era inagotable. Al año siguiente un matemático, Michael Bulmer, razonaba que la varianza aditiva total debería disminuir con la selección, lo que parecía contradecir el resultado de Crow y Kimura. La única forma de compatibilizar ambos resultados es la existencia de covarianzas negativas entre genes, de forma que la varianza de cada uno de ellos no sea modificada por la selección pero la varianza total disminuya debido a estas covarianzas negativas. Este resultado es contraintuitivo, puesto que parece que si seleccionamos a los mejores individuos, las versiones mejores de los alelos de cada gen deben aumentar de frecuencia y entonces generar covarianzas positivas y no negativas. Abordaremos el tema formal e intuitivamente, y para ello primero examinaremos qué sucede con uno o dos loci antes de pasar al modelo infinitesimal.

### 1. Efecto de la selección sobre la varianza en un modelo con dos loci

#### 1.1. Efecto de la selección sobre la varianza en un locus con dos alelos

Vamos a empezar con un ejemplo simple, un locus aditivo con dos alelos (AA, Aa, aa) con frecuencias génicas (p,q) en una población en equilibrio. Supongamos que el homocigoto AA incrementa el valor del carácter en una cantidad k, el homocigoto aa lo reduce en una cantidad -k y el heterocigoto es intermedio y vale cero. Hacemos selección contra recesivos. Las frecuencias de equilibrio son las que provienen del apareamiento al azar de los gametos.

Genotipo	AA	Aa	aa
Valor	k	0	-k
Frecuencia	$p^2$	$2 p q$	$q^2$
Selección	1	1	$1-S$

Si no hacemos selección, la media del carácter es

$$m = p^2 k + 2pq \cdot 0 + q^2 = (p^2 - q^2) k = (p + q)(p - q) k = (p - q) k$$

La varianza será

$$\sigma^2 = p^2 k^2 + 2pq \cdot 0 + q^2 k^2 - m^2 = (p^2 + q^2) k^2 - (p - q)^2 k^2 = (p^2 + q^2 - p^2 - q^2 + 2pq) k^2 = 2pq k^2$$

que es máxima a frecuencias intermedias ( $p=q=0.5$ ).

Si hacemos selección de forma que eliminamos todos los recesivos ( $S=1$ ), en la generación siguiente la frecuencia del alelo A será

$$p' = \frac{p^2 + pq}{p^2 + 2pq} = \frac{p(p+q)}{p(p+2q)} = \frac{p+q}{p+q+q} = \frac{1}{1+q} \quad q' = 1 - p' = 1 - \frac{1}{1+q} = \frac{1+q-1}{1+q} = \frac{q}{1+q}$$

con lo que la media en la generación siguiente será

$$m' = (p' - q') k = \left( \frac{1}{1+q} - \frac{q}{1+q} \right) k = \frac{1-q}{1+q} k$$

y la varianza

$$(\sigma^2)' = 2p'q'k^2 = 2 \frac{1}{1+q} \cdot \frac{q}{1+q} k^2 = 2 \frac{q}{(1+q)^2} k^2$$

que es diferente de la varianza en la generación anterior. Por ejemplo, si  $p=q=0.5$ , el cambio de varianza es


$$(\sigma^2)' - \sigma^2 = 2 \frac{q}{(1+q)^2} k^2 - 2pq k^2 = 0.44 k^2 - 0.50 k^2 = -0.06 k^2$$

Es decir, que la varianza, que estaba en su máximo cuando  $p=q=0.5$ , se reduce un poco. En el caso de un locus, la varianza se modifica con la selección.

## 1.2. El Desequilibrio de ligamiento

La expresión “desequilibrio de ligamiento” es una de las más desafortunadas expresiones usadas en genética cuantitativa, puesto que puede no haber ligamiento entre genes y existir desequilibrio de ligamiento, o por el contrario genes ligados pueden no presentar este desequilibrio de ligamiento. La forma correcta de expresar la relación entre genes o alelos es decir si hay o no *independencia estadística* entre ellos. La expresión “desequilibrio gamético” es igualmente desafortunada, como veremos más adelante.

Supongamos que un carácter está regulado por dos locus con dos alelos cada uno, (A,a) y (B,b) en cromosomas diferentes. Supongamos que podemos identificar a los nueve tipos de individuos que existirían en nuestra población: AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb, aabb. Si todos los alelos fueran *independientes estadísticamente*, se podría calcular las frecuencias de los genotipos conociendo las frecuencias génicas, y estaríamos en lo que llamamos una situación de equilibrio simultáneo para dos genes. En la tabla siguiente figuran las frecuencias de equilibrio al aparearse al azar espermatozoides y óvulos.

	AB p·r	Ab p·s	aB q·r	ab q·s
AB p·r	AABB p <sup>2</sup> · r <sup>2</sup>	AABb p <sup>2</sup> ·r·s	AaBB p·q·r <sup>2</sup>	AaBb p·q·r·s
Ab p·s	AABb p <sup>2</sup> · r·s	AAbb p <sup>2</sup> ·s <sup>2</sup>	AaBb p·q·r·s	Aabb p·q·s <sup>2</sup>
aB q·r	AaBB p·q·r <sup>2</sup>	AaBb p·q·r·s	aaBB q <sup>2</sup> · r <sup>2</sup>	aaBb q <sup>2</sup> · r·s
ab q·s	AaBb p·q·r·s	Aabb p·q·s <sup>2</sup>	aaBb q <sup>2</sup> · r·s	aabb q <sup>2</sup> · s <sup>2</sup>

Obsérvese que para que haya independencia entre alelos hacen falta ahora DOS condiciones:

1. Apareamientos al azar, como en el caso de un gen, de forma que la unión de gametos sea aleatoria.
2. Que dentro de cada gameto los alelos sean independientes (que haya “equilibrio gamético”).

Esta segunda condición es menos obvia que la primera, y puede cumplirse o no independientemente de que los apareamientos sean al azar. Por ejemplo, dos genes pueden estar ligados y el alelo A puede ir frecuentemente asociado con el B, de forma que la frecuencia de encontrar un gameto AB sea superior a p·r. De hecho no hace falta que estén ligados, pueden estar en cromosomas distintos e ir asociados; por ejemplo, porque estemos seleccionando a favor de un carácter en el que estén implicados ambos genes.

Pongamos un ejemplo; tomemos una población de 200 individuos distribuida como se indica en el cuadro:

	BB	Bb	Bb	
AA	50	42	6	98
Aa	20	44	20	84
aa	2	10	6	18
	72	96	32	200

Las frecuencias génicas son

$$frec(A) = p = \frac{98 + \frac{84}{2}}{200} = 0.70$$

$$frec(a) = q = 1 - p = 0.30$$

$$frec(B) = r = \frac{72 + \frac{96}{2}}{200} = 0.60$$

$$frec(b) = s = 1 - r = 0.40$$

Obsérvese que cuando intentamos calcular las frecuencias genotípicas a partir de las génicas, los resultados son diferentes a los de la tabla. Por ejemplo, si calculamos la frecuencia del genotipo AaBB a partir de las frecuencias génicas, tendríamos<sup>1</sup>

$$frec_{EQ}(AaBB) = 2 \cdot p \cdot q \cdot r^2 = 2 \cdot 0.7 \cdot 0.3 \cdot 0.6^2 = 0.1512$$

mientras que en la tabla, la frecuencia observada del genotipo AaBB es

$$frec(AaBB) = \frac{20}{200} = 0.10$$

No hay *independencia estadística entre alelos*, por tanto están correlacionados y la población no está en “equilibrio de ligamiento”. Obsérvese que los dos genes están en cromosomas diferentes, no pueden estar ligados, pero se usa esta desafortunada frase para describir su dependencia estadística.

Si calculamos la frecuencia observada de los genotipos AA, Aa, aa estas son

$$frec(AA) = \frac{98}{200} = 0.49 \quad frec(Aa) = \frac{84}{200} = 0.42 \quad frec(aa) = \frac{18}{200} = 0.09$$

Que son las que se obtienen al calcularlas a partir de las frecuencias génicas


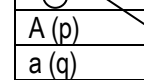
$$frec_{EQ}(AA) = p^2 = 0.7^2 = 0.49 \quad frec_{EQ}(Aa) = 2pq = 2 \cdot 0.7 \cdot 0.3 = 0.42 \quad frec_{EQ}(aa) = q^2 = 0.3^2 = 0.09$$

y lo mismo ocurre con los genotipos (BB, Bb, bb)

$$frec(BB) = \frac{72}{200} = 0.36 \quad frec(Bb) = \frac{96}{200} = 0.48 \quad frec(bb) = \frac{32}{200} = 0.16$$

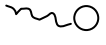
$$frec_{EQ}(BB) = r^2 = 0.6^2 = 0.36 \quad frec_{EQ}(Ba) = 2rs = 2 \cdot 0.6 \cdot 0.4 = 0.48 \quad frec_{EQ}(bb) = s^2 = 0.4^2 = 0.16$$


con lo que llegamos a la aparentemente sorprendente conclusión de que el gen (A,a) está en equilibrio (esto es, en situación de independencia estadística) y el gen (B,b) también, pero juntos no lo están. Esta situación no es sorprendente, puesto para que haya independencia estadística (equilibrio) dentro de un gen sólo hace falta que los apareamientos se produzcan al azar. El que el espermatozoide que lleva el alelo A se una con el óvulo que lleva el alelo A son sucesos independientes y su probabilidad es el producto de las probabilidades de que haya un espermatozoide con el alelo A y un óvulo con el alelo a; esto es,  $p \cdot p = p^2$ , y lo mismo ocurre con los otros posibles sucesos

		
	A (p)	a (q)
A (p)	AA (p <sup>2</sup> )	Aa (p·q)
a (q)	aA (q·p)	aa (q <sup>2</sup> )

para que haya independencia estadística (equilibrio) entre dos genes hacen falta, como hemos dicho, dos condiciones. Una es el apareamiento al azar, pero hace falta además que los alelos que van dentro del espermatozoide (o del óvulo) estén también en estado de independencia estadística; esto es, que reunir juntos un alelo A y un alelo B en un espermatozoide sea el resultado de dos sucesos independientes, aleatorios, con lo que su probabilidad será el producto de las probabilidades  $p \cdot r$  (y lo mismo con el óvulo), lo que se conoce como “equilibrio gamético”. Ambas condiciones son necesarias, y el “equilibrio gamético” no produce independencia estadística si los apareamientos no son al azar.

<sup>1</sup> Llamaré  $frec_{EQ}$  a las frecuencias que corresponden al estado de equilibrio, y a las observadas e el ejemplo les llamaremos  $frec$  simplemente.

	A (p)	a (q)
B (r)	AB (p·r)	aB (q·r)
b (s)	Ab (p·s)	ab (q·s)

	A (p)	a (q)
B (r)	AB (p·r)	aB (q·r)
b (s)	Ab (p·s)	ab (q·s)

Entonces, cuando se aparean al azar el espermatozoide y el óvulo la frecuencia de AaBB será la de la unión de un espermatozoide (AB) con un óvulo (aB), esto es  $(p \cdot r) \cdot (q \cdot r) = p \cdot q \cdot r^2$ , más la de un óvulo (AB) con un espermatozoide (aB), que es también  $p \cdot q \cdot r^2$

Podemos calcular las frecuencias con las que se presentarán los espermatozoides AB, Ab, aB y ab, que son las mismas que las de los correspondientes óvulos. Los espermatozoides (y óvulos) AB son producidos por individuos AABB, AABb, AaBB y AaBb, aunque no todos los producen por igual; los individuos AABB sólo producen gametos AB pero los individuos AABb y AaBB producen la mitad de sus gametos AB, y los AaBb producen sólo la cuarta parte de sus gametos AB. La frecuencia de alelos en los espermatozoides de la siguiente generación será

$$frec(AB) = \frac{50 + \frac{42}{2} + \frac{20}{2} + \frac{44}{4}}{200} = 0.46 \quad frec(Ab) = \frac{\frac{42}{2} + 6 + \frac{44}{4} + \frac{20}{2}}{200} = 0.24$$

$$frec(aB) = \frac{\frac{20}{2} + \frac{44}{4} + 2 + \frac{10}{2}}{200} = 0.14 \quad frec(ab) = \frac{\frac{44}{4} + \frac{20}{2} + \frac{10}{2} + 6}{200} = 0.16$$

y en los óvulos será la misma. Por tanto, las frecuencias observadas y las de equilibrio serán

OBSERVADAS	A	a
B	AB (0.46)	aB (0.14)
b	Ab (0.24)	ab (0.16)

EQUILIBRIO	A (0.7)	a (0.3)
B (0.6)	AB (0.42)	aB (0.18)
b (0.4)	Ab (0.28)	ab (0.12)

**Definición:** Se llama *desequilibrio de ligamiento* a la diferencia entre la frecuencia observada de un gameto (espermatozoide u óvulo) y la que debería tener si estuviera en equilibrio (esto es, en condiciones de independencia estadística).

En nuestro caso, el desequilibrio del gameto AB es

$$D(AB) = frec(AB) - frec_{EQ}(AB) = 0.46 - 0.42 = 0.04$$

$$D(Ab) = frec(Ab) - frec_{EQ}(Ab) = 0.24 - 0.28 = -0.04$$

$$D(aB) = frec(aB) - frec_{EQ}(aB) = 0.14 - 0.18 = -0.04$$

$$D(ab) = frec(ab) - frec_{EQ}(ab) = 0.16 - 0.12 = 0.04$$

Como se ve, coinciden en valor absoluto y cambia el signo. Esto se debe a que las frecuencias de los alelos son las mismas en el equilibrio o no, con lo que

$$frec(A) = frec(AB) + frec(Ab) = frec_{EQ}(AB) + frec_{EQ}(Ab)$$

y por tanto

$$frec(AB) - frec_{EQ}(AB) = -[frec(Ab) - frec_{EQ}(Ab)]$$

o lo que es lo mismo, llamando D a la cantidad en valor absoluto (0.04 en nuestro ejemplo),

$$frec(AB) = frec_{EQ}(AB) + D = p \cdot r + D$$

$$frec(Ab) = frec_{EQ}(Ab) - D = p \cdot s - D$$

$$frec(aB) = frec_{EQ}(aB) - D = q \cdot r - D$$

$$frec(ab) = frec_{EQ}(ab) + D = q \cdot r + D$$

En la siguiente generación, el gameto AB puede venir de un individuo cuyas cromátidas sean  $\frac{A \ B}{x \ y}$  donde 'x' significa alelo 'A' o alelo 'a', e 'y' significa alelo 'B' o alelo 'b'. Si todos los gametos de la generación siguiente vinieran de individuos así, la frecuencia del gameto AB sería la misma en una generación que en la siguiente; llamando  $frec_1$  a las frecuencias en la siguiente generación, tendríamos  $frec_1(AB) = frec(AB)$ .

En la siguiente generación, el gameto AB puede venir también de un individuo cuyas cromátidas sean  $\frac{A \ y}{x \ B}$ . Si todos los gametos de la generación siguiente vinieran de individuos así, la frecuencia del gameto AB sería la frecuencia con que coincide una cromátida A y con una cromátida x B. La frecuencia de las cromátidas A y es p, y la frecuencia de las cromátidas x B es r, la frecuencia conjunta (con apareamiento al azar) es p · r; esto es, la frecuencia de equilibrio,  $frec_1(AB) = p \cdot r$

La frecuencia con que se produce el segundo suceso se llama frecuencia de recombinación 'c', por lo que en la generación siguiente

$$frec_1(AB) = (1 - c) frec(AB) + c \cdot p \cdot r$$

y el desequilibrio de ligamiento será

$$D_1 = frec_1(AB) - frec_{EQ}(AB) = (1 - c) frec(AB) + c \cdot p \cdot r - c \cdot p \cdot r = (1 - c) [frec(AB) - p \cdot r] = (1 - c) \cdot D$$

Esto es, el desequilibrio de ligamiento (i.e.: la dependencia estadística) se reduce una fracción  $1 - c$  por generación. Si los genes están en cromosomas distintos,  $c = 1/2$ , por lo que se reduce a la mitad.

En la figura que sigue (tomada de Falconer, 1960) se puede ver cómo cambia el desequilibrio de ligamiento según lo ligados que estén los genes. La línea con 0.5 se refiere a locus no ligados (en cromosomas separados, por ejemplo). Las leyes de equilibrio que hemos expuesto se generalizan a varios alelos, varios genes, y también a frecuencias dependientes del sexo.

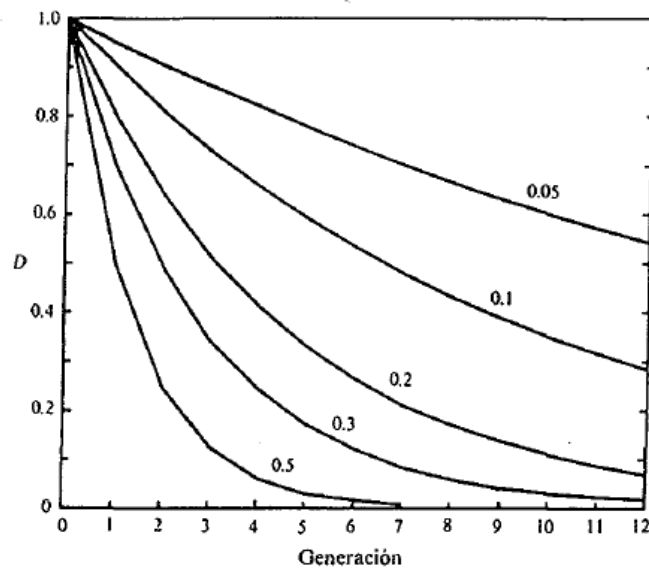


Figura 1. Evolución del desequilibrio de ligamiento D entre dos loci, con el transcurso de las generaciones.

Cuando hay desequilibrio de ligamiento aparecen asociaciones de alelos que luego se irán deshaciendo de generación en generación; por ejemplo, en genómica, los marcadores estarán cada vez menos relacionados con los genes causales y habrá que rehacer la predicción cada tres o cuatro generaciones.

En conclusión:

1. El desequilibrio de ligamiento se puede producir haya o no ligamiento.
2. Dos genes ligados pueden estar en equilibrio de ligamiento
3. Dos genes en cromosomas distintos pueden estar en desequilibrio de ligamiento.
4. El equilibrio gamético puede dar lugar a desequilibrio de ligamiento si no hay apareamientos al azar

El *desequilibrio de ligamiento* es un nombre desafortunado que se utiliza para describir la correlación entre los alelos de uno o más loci. Debería llamarse *Dependencia estadística*.

### 1.3. Efecto de la selección sobre la covarianza entre dos loci

Tomemos el caso de dos genes no ligados (por ejemplo, cada uno en un cromosoma) (AA,Aa,aa) y (BB,Bb,bb) sobre los que actúa la selección favoreciendo a los alelos en mayúscula. Las frecuencias génicas de (A,a) y (B,b) son (p,q) y (r,s) respectivamente. Supongamos que la acción génica es aditiva, y que tener un alelo 'A' o 'B' aporta una cantidad k, y que tener un alelo 'a' o 'b' aporta una cantidad nula (no aporta nada). Esto no supone ninguna pérdida de generalidad porque es un simple cambio de origen de coordenadas, pero facilita los cálculos.

La selección genera desequilibrio de ligamiento (esto es, dependencia estadística). Ya vimos que esto se puede producir porque los apareamientos no sean al azar o porque los gametos no estén en equilibrio de ligamiento (en estado de independencia estadística), por lo que nos limitaremos a ver las consecuencias del desequilibrio de ligamiento (de la dependencia estadística) de los gametos.

Tomemos una variable <sup>(2)</sup> aleatoria 'x' con los valores y frecuencias correspondientes a (A,a), y otra variable aleatoria 'y' con los valores y frecuencias correspondientes a (B,b). En equilibrio (en situación de independencia estadística), tenemos

<sup>2</sup> Formalmente hablando habría que tomar una variable aleatoria X que tuviera las posibilidades X=A y X=a, para las que x=k y x=0 respectivamente, pero no entraré en estos detalles.

A	a	B	b
p	q	r	s
k	0	k	0

	$y_1=k$ $\text{frec}(y_1) = r$	$y_2=0$ $\text{frec}(y_2) = s$
$x_1=k$ $\text{frec}(x_1) = p$	$x_1 \cdot y_1 = k^2$ $\text{frec}(x_1, y_1) = pr$	$x_1 \cdot y_2 = 0$ $\text{frec}(x_1, y_2) = ps$
$x_2=0$ $\text{frec}(x_2) = q$	$x_2 \cdot y_1 = 0$ $\text{frec}(x_2, y_1) = qr$	$x_2 \cdot y_2 = 0$ $\text{frec}(x_2, y_2) = qs$

$$\text{Media } (x) = m_x = p \cdot k + q \cdot 0 = p \cdot k$$

$$\text{Media } (y) = m_y = r \cdot k + s \cdot 0 = r \cdot k$$

$$\text{cov}(x,y) = \sum x_i \cdot y_j \cdot \text{frec}(x_i, y_j) - m_x m_y = k^2 \cdot pr + 0 + 0 + 0 - pk \cdot rk = 0$$

por tanto, en equilibrio (en situación de independencia estadística), la covarianzas entre las variables es cero, como cabía esperar. Ahora supongamos que la selección favorece a los genes A y B, y que tenemos una población seleccionada con las siguientes frecuencias:

$$\text{frec}(A,B) = 0.6$$

$$\text{frec}(A,b) = 0.2$$

$$\text{frec}(a,B) = 0.2$$

$$\text{frec}(a,b) = 0$$

Esta población no está en equilibrio de ligamiento (es situación de independencia estadística). Vamos a verlo. Las frecuencias de los alelos son

$$p = \text{frec}(A) = \text{frec}(A,B) + \text{frec}(A,b) = 0.6 + 0.2 = 0.8$$

$$q = \text{frec}(a) = 1 - \text{frec}(A) = 0.2$$

$$r = \text{frec}(B) = \text{frec}(A,B) + \text{frec}(a,B) = 0.6 + 0.2 = 0.8$$

$$s = \text{frec}(b) = 1 - \text{frec}(B) = 0.2$$

en equilibrio de ligamiento (en situación de independencia estadística),  $\text{frec}(A,B) = p \cdot r = 0.8 \cdot 0.8 = 0.64$ , pero aquí  $\text{frec}(A,B) = 0.60$ , luego no hay equilibrio de ligamiento (no hay independencia estadística) en esta población, debido a que la selección ha favorecido a unos alelos sobre otros. Si calculamos la covarianzas entre genes,

$$\text{Media } (x) = m_x = p \cdot k + q \cdot 0 = p \cdot k = 0.8 k$$

$$\text{Media } (y) = m_y = r \cdot k + s \cdot 0 = r \cdot k = 0.8 k$$

$$\text{cov}(x,y) = \sum x_i \cdot y_j \cdot \text{frec}(x_i, y_j) - m_x m_y = k^2 \cdot \text{frec}(A,B) + 0 + 0 + 0 - pk \cdot rk = 0.60 k^2 - 0.64 k^2 = -0.04 k^2$$

la covarianzas es negativa. Esto va a ocurrir siempre, porque cuando hay selección el gameto AB se ve favorecido, pero también se ven favorecidos los gametos Ab y aB, y como

$$p = \text{frec}(A,B) + \text{frec}(A,b)$$

$$r = \text{frec}(A,B) + \text{frec}(a,B)$$

ese producto es superior a la frecuencia del gameto AB. Lo podemos ver en nuestro ejemplo, en donde llamaremos 'f' a la frecuencia del gameto más seleccionado (A,B), y la frecuencia de los gametos menos seleccionados la expresaremos como una fracción de f; esto es, llamaremos f·c a la frecuencia de los gametos (a,B) y (A,b) donde  $c < 1$ . Finalmente supondremos que el peor de los gametos (a,b) no es seleccionado. Así, las frecuencias serán

$$\text{frec}(A,B) = f$$

$$\text{frec}(A,b) = fc$$

$$\text{frec}(a,B) = fc$$

$$\text{frec}(a,b) = 0$$

y demostraremos que  $p \cdot r > f$



La suma de las frecuencias da la unidad, por tanto

$$f+fc+fc+0=1 \rightarrow f(1+2c)=1 \rightarrow f=\frac{1}{1+2c}$$

$$\begin{aligned} p &= f + fc \\ r &= f + fc \\ pr &= (f + fc)^2 = f^2(1+c)^2 \end{aligned}$$

$$\frac{p \cdot r}{f} = f(1+c)^2 = \frac{(1+c)^2}{1+2c} = \frac{1+2c+c^2}{1+2c} > 1 \rightarrow p \cdot r > f$$

$$\text{cov}(x,y) = k^2 \cdot \text{frec}(A,B) - k^2 \cdot pr = k^2 \cdot [f - pr] < 0$$

Así pues, la selección ha producido covarianzas negativas en el caso de dos loci, para un modelo en el que se favorece la presencia de gametos AB, algo menos los Ab y aB, y menos los ab, lo que es un modelo de selección razonable. Un modelo más general en el que los recesivos no son completamente eliminados da el mismo resultado, porque eliminar completamente los recesivos es un simple cambio de origen de coordenadas.

Obsérvese que la covarianza es proporcional a la cantidad que hemos definido como desequilibrio de ligamiento (dependencia estadística); esto es,

$$\text{cov}(x,y) = \sum x_i \cdot y_j \cdot \text{frec}(x_i, y_j) - m_x m_y = k^2 \cdot \text{frec}(A,B) - k^2 \cdot pr = k^2 \cdot [\text{frec}(A,B) - pr] = k^2 D(AB)$$

Pasaremos a continuación a considerar el modelo infinitesimal.

## 2. Efecto de la selección sobre la varianza en el modelo infinitesimal

El modelo infinitesimal fue propuesto por Mendel en 1865, en su famoso artículo “Experimentos en híbridos de plantas” cuando puso el ejemplo de la existencia de múltiples factores de herencia: “Sería A el color de la flor, compuesto de caracteres independientes  $A_1+A_2+\dots$  (...) Cualquiera que inspeccione los matices de colores que aparecen en las plantas ornamentales difícilmente puede evadir la convicción de que aquí también el desarrollo procede en conformidad con una cierta ley que posiblemente encuentra su expresión *por combinación de varios caracteres de color independientes*” (la cursiva es de Mendel). Posteriormente Yule, en 1902, propuso que la variación continua puede explicarse a través de la suma de muchos factores de herencia, y Pearson, en 1904, examinó matemáticamente el modelo infinitesimal, pero con dominancia completa (heterocigoto igual a uno de los homocigotos), lo que le llevó a rechazarlo por no ajustarse a resultados experimentales, pero Yule (1906) hizo notar que con un modelo aditivo este problema se resolvía. Finalmente Fisher, en un artículo notoriamente difícil <sup>(3)</sup> publicado en 1918, propuso la versión actual del modelo infinitesimal, en el que no sólo consideró la aditividad, como quieren hacer creer los que postulan “nuevos paradigmas”, sino las interacciones dominante y epistática y los apareamientos asociativos.

El modelo infinitesimal sostiene que el valor de un carácter está determinado por los efectos de un conjunto infinito de genes de un efecto infinitesimal cada uno. El modelo infinitesimal es obviamente falso, no existe un conjunto infinito de genes, pero nadie ha pretendido que ni este modelo ni cualquier otro sean “verdaderos”,

---

<sup>3</sup> El artículo es difícil, y no ayuda mucho la habitual mala redacción de Fisher, que no suele exponer con claridad cuáles son sus hipótesis de partida ni los términos que usa. Moran y Smith (1966), en un loable esfuerzo, han comentado el artículo párrafo por párrafo.

sino un aproximación útil a la determinación genética de los caracteres de variación continua <sup>(4)</sup>. No hay efectos infinitesimales de infinitos genes en la naturaleza, pero sí pequeños efectos de muchos genes, por lo que el modelo infinitesimal puede ser en muchos casos una buena aproximación. Mendel propuso en 1865 un modelo para describir la herencia de ciertos caracteres, y nunca habló de los genes sino de “factores de herencia”; no como algo de contenido biológico (en su época la herencia no se explicaba a través de los genes situados en los cromosomas), sino como una teoría explicativa de cierta herencia observada, que obviamente tendría causas biológicas pero eran ignoradas. Fisher estaba completamente de acuerdo con la genética de Mendel cuando propuso el modelo infinitesimal (no así Pearson), e intento *generalizar el modelo de Mendel* a caracteres de variación continua, tal y como el propio Mendel sugería.

Como decía Box (1987), “todos los modelos son falsos, pero algunos son útiles”. Cuando Newton propuso la ley de gravitación universal nunca pretendió que hubiera una fuerza de atracción entre las masas. Esta fuerza plantea muchos problemas: no tiene agente de transmisión, se transmite a través del vacío y atraviesa cualquier obstáculo, y además si aparece una masa nueva o una masa se parte en dos, su efecto de atracción se transmite instantáneamente a todo el universo. Newton siempre insistió en que dos cuerpos se atraían *como si* hubiera una fuerza proporcional a sus masas e inversamente proporcional al cuadrado de su distancia. Planteaba un modelo que era falso pero útil para hacer predicciones, y que ayudaba a comprender la naturaleza del universo. La teoría de la Relatividad, otro modelo falso útil, eliminó la fuerza del modelo evitando algunos de estos inconvenientes, aunque introdujo nuevos problemas como la curvatura del espacio o las paradojas del tiempo. Un conocimiento mayor del genoma y de las rutas metabólicas podrá dar lugar a nuevos modelos, pero el modelo infinitesimal continuará siendo útil en muchos casos, así como la ley de Newton sigue siendo hoy en día habitualmente utilizada en una infinidad de problemas. Ambos modelos permiten hacer predicciones de forma sencilla y ambos aproximan la compleja realidad existente usando unos pocos parámetros con un significado definido y entendible.

## 2.1. Efecto de la selección sobre la varianza

Como hemos dicho, el modelo infinitesimal sostiene que el valor de un carácter está determinado por los efectos de un conjunto infinito de genes de un efecto infinitesimal cada uno. En esas circunstancias, se puede sostener, invocando el Teorema Central del Límite, que la distribución de los efectos de estos genes sigue una ley Normal. La importancia de invocar la Normalidad de las distribuciones genéticas es la gran simplificación de los cálculos que conlleva. Una distribución Normal está definida por dos parámetros claramente interpretables que se pueden estimar con estimadores que tienen buenas propiedades. En el caso multinormal existe una sola moda, una correlación cero implica independencia entre las variables <sup>(5)</sup>, la relación entre las variables es lineal, y tiene otras propiedades interesantes. No siempre es posible invocar el teorema central del límite, pero siempre que sea posible es útil hacerlo, hace mucho frío fuera del mundo de la multinormalidad.

El teorema central del límite dice que la media de un conjunto infinito de variables *independientes*, con la misma distribución, y la misma media <sup>(6)</sup>, y con un efecto infinitesimal cada una de ellas, tiene una distribución

---

<sup>4</sup> Obviamente también en algunos caracteres de variación discontinua, como los caracteres umbral.

<sup>5</sup> Estamos tan acostumbrados a trabajar con la distribución Normal que no reparamos en que en otras distribuciones, por ejemplo en la distribución t, una correlación nula no implica independencia de variables, y una distribución multivariante puede ser multimodal. Otro ejemplo: la media muestral es un estimador suficiente (que recoge toda la información de la muestra) en el caso de la distribución Normal, pero las distribuciones no exponenciales los estimadores suficientes de sus parámetros pueden o no existir o no poder resumir en un solo número la información de toda la muestra.

<sup>6</sup> El teorema es complejo y exige que la media y la varianza existan y sean finitas; en algunos casos estas condiciones no se cumplen, por ejemplo en la distribución de Cauchy (ese esqueleto que todo estadístico guarda en el armario), que tiene un aspecto inocente (parecido a una t) pero que al hacer la integral para calcular la media o la varianza, las integrales no convergen. Versiones menos severas no exigen que las variables tengan la misma distribución, pero obviamente se debe evitar que una variable “destaque”.

Normal. Como aquí cada variable es un locus, la condición de independencia implica que puede haber dominancia, pero no epistasia. Hay versiones menos severas de este teorema que podrían permitir ciertas formas de epistasia, pero no genes mayores ni genes correlacionados que constituyan juntos un efecto grande, ni tampoco formas de epistasia entre genes que den lugar a efectos mayores.

La selección, como hemos visto, produce correlaciones entre loci, pero en el caso del modelo infinitesimal estas correlaciones son lo suficientemente pequeñas como para que se pueda seguir invocando la Normalidad de forma aproximada (Bulmer, 1980, pág. 154, pone un ejemplo convincente de la escasa desviación de la Normal que produce la selección). Aquí supondremos que el valor aditivo de un animal sigue una distribución Normal de media cero (por conveniencia matemática) y varianza  $\sigma_A^2$ . Desde un punto de vista Bayesiano, la distribución del valor aditivo representa el grado de incertidumbre en torno a ese valor, y depende de la información a priori que tenemos sobre la varianza aditiva.

El valor de un individuo está determinado por su genotipo y su ambiente. Usando un modelo simplificado,

$$P = G + E \approx A + E$$

donde ignoramos las interacciones genotipo medio, la dominancia y la epistasia. Supondremos que la varianza ambiental es la misma generación tras generación y examinaremos los cambios en la varianza aditiva. Las distribuciones de A,G,E y P suponemos que son Normales.

Usando la nomenclatura en inglés (C=Child, M=Male, F=Female), el valor aditivo de los hijos es por término medio la media del de los padres, y el de un hijo en concreto

$$A_C = \frac{1}{2}A_M + \frac{1}{2}A_F + A_e$$

donde  $A_e$  es lo que sobra o lo que falta para que las cuentas cuadren. Es un término aleatorio independiente que indica que un mismo padre puede tener hijos con mejor o peor valor aditivo dependiendo de los alelos que aleatoriamente ha transmitido a ese hijo en concreto. Se le suele llamar término Mendeliano.

Si los individuos se aparean al azar (no hay covarianzas entre  $A_M$  y  $A_F$ ), la varianza aditiva de los hijos será

$$\sigma_{A_C}^2 = \frac{1}{4}\sigma_{A_M}^2 + \frac{1}{4}\sigma_{A_F}^2 + \sigma_{A_e}^2$$

y si no hay selección, la varianza de los hijos y de los padres es la misma

$$\sigma_A^2 = \frac{1}{4}\sigma_A^2 + \frac{1}{4}\sigma_A^2 + \sigma_{A_e}^2 \quad \rightarrow \quad \sigma_{A_e}^2 = \frac{1}{2}\sigma_A^2$$

Si hay selección, la varianza aditiva cambia. Haremos un cambio de nomenclatura por comodidad a la hora de escribir, y llamaremos  $V_A$  a la varianza aditiva antes de seleccionar y  $V_A^*$  a la varianza aditiva de los padres seleccionados. Si los individuos se aparean al azar

$$\sigma_{A_C}^2 = \frac{1}{4}\sigma_{A_M}^2 + \frac{1}{4}\sigma_{A_F}^2 + \sigma_{A_e}^2 = \frac{1}{4}V_A^* + \frac{1}{4}V_A^* + \frac{1}{2}V_A = \frac{1}{2}V_A^* + \frac{1}{2}V_A$$

Vamos a calcular la varianza de los padres seleccionados (ver Figura 2).

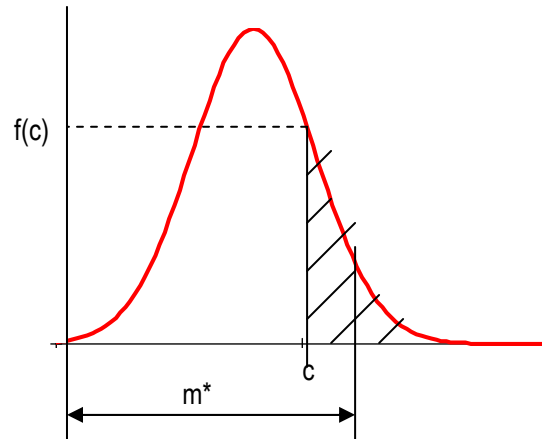


Figura 2. Selección truncada en una variable Normal

En la figura 2 se representa la distribución de los valores aditivos de una población. Hemos supuesto que es Normal con media 0 y varianza  $V_A$ , pero tipificamos la variable y pasa a ser Normal de media cero y varianza

$$V_A = 1$$

porque es más cómodo hacer así los cálculos. Esto no supone ninguna pérdida de generalidad del modelo.

$$f(A) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{A^2}{2}\right)$$

Seleccionamos los mejores valores aditivos a partir de un cierto valor aditivo 'c'. Llamaremos  $m^*$  a la media de los padres seleccionados.

$$m^* = \frac{1}{P} \int_c^{\infty} A \cdot f(A) dA$$

pero obsérvese que

$$\frac{df(A)}{dA} = -A \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{A^2}{2}\right) = -A \cdot f(A) \quad \rightarrow \quad A \cdot f(A) = -df(A)$$

$$m^* = -\frac{1}{P} \int_c^{\infty} df(A) = -\frac{1}{P} [f(A)]_c^{\infty} = -\frac{1}{P} \cdot \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{A^2}{2}\right) \right]_c^{\infty} = \frac{1}{P} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{c^2}{2}\right) = \frac{1}{P} \cdot f(c)$$

La varianza de los padres seleccionados es

$$V_A^* = E(A - m^*)^2 = E(A^2) - (m^*)^2 = \frac{1}{P} \int_c^{\infty} A^2 f(A) dA - (m^*)^2$$

en el apéndice 1 se demuestra que

$$\frac{1}{P} \int_c^{\infty} A^2 f(A) dA = 1 + c \cdot m^*$$

con lo que, sustituyendo,

$$V_A^* = 1 + c \cdot m^* - (m^*)^2 = 1 - m^* (m^* - c) < 1 = V_A$$

es decir, la varianza de los padres seleccionados es menor que la varianza aditiva, y por tanto bajo selección, la varianza de los hijos será menor que la varianza aditiva antes de seleccionar. Por tanto, en conclusión,

La varianza aditiva disminuye con la selección

En su formulación habitual, a la media de los padres seleccionados se le llama “diferencial de selección S” y en el caso de caracteres tipificados “intensidad de selección i”, con lo que en los textos la fórmula aparece como  $V_A^* = 1 - i \cdot (i - c)$ , o destipificando,

$$V_A^* = [1 - i \cdot (i - c)] V_A$$

En la generación t, la varianza sería

$$V_A(t+1) = \frac{1}{2} V_A^*(t) + \frac{1}{2} V_A = \frac{1}{2} [1 - i \cdot (i - c)] V_A(t) + \frac{1}{2} V_A$$

como acabamos de ver,  $1 - i \cdot (i - c) < 1$ , por lo que la varianza se va reduciendo de generación en generación hasta llegar a una varianza de equilibrio. Cuando la selección se deja de hacer selección, el equilibrio de ligamiento se va recuperando hasta recobrar la varianza aditiva del inicio.

En el ejemplo desarrollado hemos supuesto que seleccionábamos sobre la varianza aditiva; esto es, con una heredabilidad  $h^2=1$ . El desarrollo que hemos hecho es el caso en el que más se acusa la reducción de la varianza con la selección, porque cuanto menor sea la heredabilidad, menor será la disminución de la varianza fenotípica, ya que la varianza ambiental es constante y  $V_P=V_A+V_E$ . En la práctica el efecto Bulmer no es excesivamente importante y actúa esencialmente en las dos o tres primeras generaciones. Cuando se relaja la selección la recuperación es rápida y bastan dos o tres generaciones de apareamiento al azar para recuperar la mayor parte de la varianza.

En la tabla siguiente se pone un ejemplo tomado de Bulmer (1980) para una presión de selección alta, del 5%, lo que supone  $i=2.06$ ,  $c=1.65$ , y una heredabilidad alta,  $h^2=0.50$ . Incluso en ese caso la caída de la varianza aditiva no es grande y la recuperación es rápida.

Generación	0	1	2	3	4	5
$V_A(t)$ con selección	100	78	74	74	73	73
Generación		6	7	8	9	10
$V_A(t)$ relajando la selección		87	93	97	98	99

## 2.2. Efecto de la selección sobre la suma de las varianzas de cada locus

Tomaremos primero un locus con dos alelos en equilibrio, sobre el que se realiza selección. Examinaremos el efecto de la selección sobre la varianza del número de descendientes que deja cada genotipo (su aptitud reproductiva). Para examinar el efecto sobre un carácter cuantitativo basta por multiplicar la aptitud

reproductiva por una constante, por lo que las conclusiones serán las mismas. Al final del apartado examinaremos el caso del modelo infinitesimal.

A	a
p	q
$m_A = m + \alpha_1$	$m_a = m + \alpha_2$

AA	Aa	aa
$p^2$	$2pq$	$q^2$
$m_{AA} = m + 2\alpha_1$	$m_{Aa} = m + \alpha_1 + \alpha_2$	$m_{aa} = m + 2\alpha_2$

N: número de individuos de la población

$2n_A$ : número de alelos A de la población.  $2n_A = 2N \cdot p \rightarrow n_A = N \cdot p$

$m_{AA}$ ,  $m_{Aa}$ ,  $m_{aa}$ : aptitudes reproductivas de los individuos AA, Aa, aa respectivamente. Cada aptitud reproductiva es una constante característica de cada genotipo e independiente de las frecuencias génicas. Supondremos aquí que son aditivas, como se expone en el cuadro de arriba. Se puede suponer un modelo más complejo no aditivo donde cada genotipo sea seleccionado de forma diferente, y al final del apartado lo comentaremos.

$m_A$ ,  $m_a$ : aptitudes reproductiva de los gametos con un alelo A, a respectivamente. Son variables dependientes de las frecuencias génicas, como ahora veremos.

m: aptitud reproductiva media. Es una variable dependiente del cambio en las frecuencias génicas.

V: varianza de la aptitud reproductiva en ese locus.

p, q: Son las frecuencias génicas de los alelos A y a. respectivamente. Son variables, cambiando con la selección. Consideraremos generaciones solapadas y que el cambio de frecuencias dp y dq se realiza en un intervalo de tiempo dt.

$$m = p \cdot m_A + q \cdot m_a = p^2 \cdot m_{AA} + 2pq \cdot m_{Aa} + q^2 \cdot m_{aa}$$

$$\begin{aligned} p \cdot m_A &= p^2 \cdot m_{AA} + pq \cdot m_{Aa} \\ q \cdot m_a &= q^2 \cdot m_{aa} + pq \cdot m_{Aa} \end{aligned} \quad \rightarrow \quad \begin{aligned} m_A &= p \cdot m_{AA} + q \cdot m_{Aa} \\ m_a &= q \cdot m_{aa} + p \cdot m_{Aa} \end{aligned}$$

Cuando hay selección, p aumenta y también la aptitud reproductiva media de la población, con lo que el incremento del número de individuos debido a la selección será, en un intervalo dt,

$$dN = N \cdot m \cdot dt \quad \rightarrow \quad \frac{dN}{dt} = N \cdot m$$

y análogamente

$$dn_A = n_A \cdot m_A \cdot dt \quad \rightarrow \quad \frac{dn_A}{dt} = n_A \cdot m_A$$

El incremento de p con la selección será

$$\frac{dp}{dt} = \frac{d}{dt} \frac{n_A}{N} = \frac{1}{N} \frac{dn_A}{dt} + n_A \frac{-1}{N^2} \frac{dN}{dt} = \frac{n_A \cdot m_A}{N} - \frac{n_A \cdot N \cdot m}{N^2} = p \cdot m_A - p \cdot m = p(m_A - m)$$

$$\frac{dq}{dt} = q(m_a - m)$$

$$dp = d(1 - q) = -dq$$

La varianza de la aptitud reproductiva  $V$  se puede ver directamente como la suma de las varianzas de los gametos

$$V = 2 \left[ p \cdot (m_A - m)^2 + q \cdot (m_a - m)^2 \right]$$

o más formalmente

$$\begin{aligned} V &= p^2 (m + 2\alpha_1 - m)^2 + 2pq(m + \alpha_1 + \alpha_2 - m)^2 + q^2 (m + 2\alpha_2 - m)^2 = \\ &= 4p^2\alpha_1^2 + 2pq(\alpha_1^2 + 2\alpha_1\alpha_2 + \alpha_2^2) + 4q^2\alpha_2^2 = \\ &= 2p^2\alpha_1^2 + 2p^2\alpha_1^2 + 2pq(\alpha_1^2 + 2\alpha_1\alpha_2 + \alpha_2^2) + 2q^2\alpha_2^2 + 2q^2\alpha_2^2 = \\ &= 2p^2\alpha_1^2 + 2pq\alpha_1^2 + 2pq\alpha_2^2 + 2q^2\alpha_2^2 + 2(p^2\alpha_1^2 + 2pq\alpha_1\alpha_2 + q^2\alpha_2^2) = \\ &= 2p\alpha_1^2(p + q) + 2q\alpha_2^2(p + q) + 2(p\alpha_1 + q\alpha_2)^2 = \\ &= 2p\alpha_1^2 + 2q\alpha_2^2 + 0 = 2 \left[ p \cdot (m_A - m)^2 + q \cdot (m_a - m)^2 \right] \end{aligned}$$

porque el término

$$p\alpha_1 + q\alpha_2 = p(m_A - m) + q(m_a - m) = p m_A + q m_a - m(p + q) = m - m = 0$$

como es de esperar, ya que  $\alpha_1$  y  $\alpha_2$  están referidas a la media.

El cambio en la varianza debido a la selección será

$$\frac{dV}{dt} = 2 \left[ (m_A - m)^2 \frac{dp}{dt} + (m_a - m)^2 \frac{dq}{dt} + p \cdot \frac{d}{dt} (m_A - m)^2 + q \cdot \frac{d}{dt} (m_a - m)^2 \right]$$

En el apéndice 2 se demuestra que los dos últimos sumandos son cero, por lo que

$$\begin{aligned} \frac{dV}{dt} &= 2 \left[ (m_A - m)^2 \frac{dp}{dt} + (m_a - m)^2 \frac{dq}{dt} \right] = 2 \left[ (m_A - m)^2 \cdot p \cdot (m_A - m) + (m_a - m)^2 \cdot q \cdot (m_a - m) \right] = \\ &= 2 \left[ (m_A - m)^3 \cdot p + (m_a - m)^3 \cdot q \right] \end{aligned}$$

Cuando sumamos todos los genes, tenemos que

$$\frac{d \sum V}{dt} = 2 \sum_i p_i \cdot (m_i - m)^3$$

esta expresión es el coeficiente de simetría de una distribución. Apelando al teorema central del límite, cuando sumamos infinitos genes de efecto infinitesimal, la distribución resultante es una Normal, y la normal tiene un coeficiente de simetría nulo, con lo que

$$\frac{d \sum V}{dt} = 0 \quad \boxed{\text{No hay cambio en la suma de las varianzas aditivas de los genes debido a la selección.}}$$

En el caso de que haya dominancia aparece un término no nulo, pero al sumar para todos los genes, la dominancia es a veces positiva y a veces negativa, y su suma total es cero. El mismo argumento se puede utilizar para la epistasia.

### 2.3. El efecto Bulmer

Si tenemos un modelo infinitesimal en el que

$$A = A_1 + A_2 + \dots$$

$$\text{Var}(A) = \text{Var}(A_1) + \text{Var}(A_2) + \dots + 2\text{cov}(A_1, A_2) + 2\text{cov}(A_1, A_3) + \dots$$

Repasemos las conclusiones del apartado 2.1 y del 2.2:

1. La varianza aditiva disminuye con la selección
2. No hay cambio en la suma de las varianzas aditivas de los genes debido a la selección

en conclusión,

3. La selección genera covarianzas negativas entre loci.

La aparición de estas covarianzas negativas debidas a la selección indica que la selección genera desequilibrio de ligamiento (esto es, dependencia estadística), lo que es perfectamente lógico ya que cuando hay selección los alelos no se eligen al azar, por tanto habrá una proporción mayor de alelos favorables al carácter seleccionado.

A este fenómeno se le denomina “efecto Bulmer” por haber sido propuesto por el estadístico Michael Bulmer en 1971. Siguiendo la conocida “Ley de Stigler”, ninguna ley o teorema que lleva un nombre propio ha sido propuesto por primera vez por esa persona (<sup>7</sup>), y este es también el caso del “efecto Bulmer”, que se encuentra sin demostrar y expuesto de forma intuitiva en la tercera edición (no en las anteriores) del famoso libro de Lush “Animal breeding plans”.

El efecto Bulmer se puede también comprender de forma intuitiva.

1. *La selección no afecta a la suma de las varianzas aditivas de cada gen*

Supongamos un modelo con N genes con dos alelos ( $A_i, a_i$ ) de frecuencias (p,q) y de efecto infinitesimal respectivamente (k, -k).

A	a
p	q
k	-k

La selección produce un cambio infinitesimal en las frecuencias génicas debido a que el valor de los alelos es infinitesimal. Llamando  $m^*$  y  $p^*$  a la media y a la frecuencia del alelo A tras la selección,

La media en la población será

$$m = N(p \cdot k - q \cdot k) = N \cdot k [p - (1-p)] = N \cdot k (2p - 1)$$

y el cambio en la media debido a la selección

$$m^* - m = N \cdot k (2p^* - 1 - 2p + 1) = 2 N k (p^* - p)$$

---

<sup>7</sup> La ley de Stigler no es de Stigler, naturalmente. La propuso por primera vez el sociólogo de la ciencia Robert K. Merton, profesor de la universidad de Columbia. Stephen Stigler, profesor de la universidad de Chicago y famoso historiador de la Estadística, lo reconoció así al promulgarla (Stigler, 1980).



Obsérvese que  $N$  tiende a infinito y  $k$  tiende a ser infinitamente pequeño, y aunque  $p$  y  $p^*$  no son infinitesimos, también el incremento en la frecuencia ( $p^*-p$ ) del alelo  $A$  es infinitamente pequeño, pero sin embargo la media aumenta con la selección en un valor que no es infinitesimal. Necesitamos un modelo que permita el cambio de la media con la selección, a pesar de estar sumando infinitos infinitesimos. Para ello sostenemos que  $k$  y ( $p^*-p$ ) son infinitesimos de orden  $N^{-1/2}$

$$k \propto \frac{1}{\sqrt{N}} \qquad p^* - p \propto \frac{1}{\sqrt{N}}$$

con lo que el producto  $Nk(p^*-p)$  ya no es un infinitesimo y la media puede aumentar con la selección en cantidades apreciables.

La varianza aditiva en un gen es  $2pqk^2$ , y la suma de las varianzas de todos los genes es

$$V = N \cdot 2pq \cdot k^2 = 2Nk^2(p-p^2)$$

El incremento e la suma de varianzas será

$$V^* - V = 2Nk^2(p^* - p^{*2} - p + p^2) = 2Nk^2[p^* - p - (p^{*2} - p^2)] = 2Nk^2[p^* - p - (p^* + p)(p^* - p)]$$

donde  $p^*+p$  es una cantidad no infinitesimal, pero el producto  $(p^*+p)(p^*-p)$  es un infinitesimo del mismo orden que  $p^*-p$ , con lo que la cantidad entre corchetes es un infinitesimo del mismo orden que  $p^*-p$ .

$$[p^* - p - (p^* + p)(p^* - p)] \propto \frac{1}{\sqrt{N}}$$

Ahora bien,  $k^2$  es un infinitesimo de orden  $1/N$

$$k^2 \propto \frac{1}{N}$$

con lo que finalmente el cambio en la suma de varianzas de cada locus es infinitesimal.

$$V^* - V \propto \frac{1}{\sqrt{N}}$$

y podemos decir que la suma de las varianzas no aumenta con la selección.

## 2. La selección disminuye la varianza aditiva de los genes en su conjunto

Parece obvio que si selecciono a los mejores padres, sus valores aditivos serán más parecidos que si los tomo al azar en la población, por lo que su variabilidad será menor y su varianza menor. Si el valor aditivo se transmite a los hijos, estos tendrán una varianza menor respecto a si sus padres hubieran sido elegidos al azar. Por tanto, la varianza aditiva debe disminuir con la selección.

## 3. Como consecuencia de 1 y 2, la selección genera covarianzas negativas entre genes

Sabemos que la varianza total es la suma de las varianzas de cada uno de los genes, más las covarianzas entre ellos. Sabemos que la varianza de cada gen no se ve alterada por la selección, por lo que la suma de las varianzas tampoco se verá alterada. Sabemos que la varianza total disminuye con la

selección. Necesariamente las covarianzas tienen que ser negativas. Este resultado es fuertemente contraintuitivo, porque uno espera encontrar al alelo A junto al B, al C, etc., pero la selección actúa sobre todo el genoma, y los gametos AB son seleccionados junto a los Ab y a los aB como hemos visto antes. Esto se generaliza a ABCD... con AbCd..., aBcD..., etc. y en conjunto la proporción de asociaciones de alelo recesivo junto alelo dominante es mayor que la de alelos dominantes, lo que da lugar a covarianzas negativas, como vimos en el apartado 1.3.

## Referencias

- Box, G. E. P., Draper, N. R. 1987. *Empirical Model Building and Response Surfaces*, John Wiley & Sons
- Bulmer M. G. 1971. The effect of selection on Genetic variability. *Amer. Nat.* 105: 201-211.
- Bulmer M. G. 1980. *The mathematical theory of quantitative genetics*. Clarendon press. Oxford.
- Crow J., Kimura M. 1970. *An Introduction to Population Genetics theory*. Harper & Row. New York
- Falconer 1960. *Introduction to Quantitative Genetics*. Ronald press. New York
- Fisher R. 1918. The correlation between relatives on the supposition of Mendelian Inheritance. *Trans. Royal Soc. Edinburgh.* 52:339-433.
- Mendel G. 1865. Experimentos en híbridos de plantas. En C.Stern, E.R. Sherwood. *El origen de la genética*. Alhambra.
- Moran P.A.P., Smith C.A.B. 1966. Comentario on R. Fisher's paper on the correlation between relatives on the supposition of Mendelian Inheritance. Cambridge Univ. press. Cambridge.
- Pearson K. 1904. On a generalizad theory of alternative inheritance, with special referente to Mendel's laws. *Phil Trans. Royal Soc. A.* 203:53-86.
- Stigler S. M. 1980. Stigler's law of eponymy. *Transactions of the New York Academy of Sciences* 39: 147–157.
- Yule G. U. 1902. Mendel's laws an their probable relation to intra-racial heredity. *New. Phytol.* 1:192-207, 222-238.
- Yule G. U. 1906. On the theory od inheritance of quantitative compound characters on the basis of Mendel's Laws –apreliminary note. Report 3<sup>rd</sup> Int. Conf. Genetics pp. 140-142.

## Apéndice 1

Necesitamos unos resultados intermedios. Si  $f(x)$  es  $N(0,1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\frac{d}{dx} f(x) = (-x) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = -x \cdot f(x)$$

$$\frac{d}{dx} [x \cdot f(x)] = f(x) + x \frac{df(x)}{dx} = f(x) - x^2 f(x) \quad \rightarrow \quad x^2 f(x) = f(x) - \frac{d}{dx} [x \cdot f(x)]$$

$$x^2 f(x) dx = f(x) dx - d[x \cdot f(x)]$$

$$\frac{1}{P} \int_c^\infty x^2 f(x) dx = \frac{1}{P} \int_c^\infty f(x) dx - \frac{1}{P} \int_c^\infty d[x \cdot f(x)] = \frac{1}{P} \cdot P + \frac{c \cdot f(c)}{P} = 1 + c \cdot m^*$$

## Apéndice 2

Demostraremos que

$$p \cdot \frac{d}{dt} (m_A - m)^2 + q \cdot \frac{d}{dt} (m_a - m)^2 = 0$$

para lo que hacen falta dos resultados intermedios:

$$\frac{dm_A}{dt} = \frac{d}{dt} (p \cdot m_{AA} + q \cdot m_{Aa}) = m_{AA} \frac{dp}{dt} + m_{Aa} \frac{dq}{dt}$$

y hemos visto en la página 14 que

$$\frac{dp}{dt} = p(m_A - m) \quad \frac{dq}{dt} = q(m_a - m)$$

con lo que procedemos ya a hacer las derivadas.

$$\begin{aligned} p \cdot \frac{d}{dt} (m_A - m)^2 &= 2p(m_A - m) \frac{dm_A}{dt} - 2p(m_A - m) \frac{dm}{dt} = \\ &= 2 \left[ \frac{dp}{dt} \cdot \frac{dm_A}{dt} - \frac{dp}{dt} \cdot \frac{dm}{dt} \right] = \\ &= 2 \left[ \left( \frac{dp}{dt} \right)^2 m_{AA} + m_{Aa} \frac{dp}{dt} \cdot \frac{dq}{dt} - \frac{dp}{dt} \cdot \frac{dm}{dt} \right] \end{aligned}$$

análogamente

$$q \cdot \frac{d}{dt} (m_a - m)^2 = 2 \left[ \left( \frac{dq}{dt} \right)^2 m_{aa} + m_{Aa} \frac{dp}{dt} \cdot \frac{dq}{dt} - \frac{dp}{dt} \cdot \frac{dm}{dt} \right]$$

sumando ambos términos,

$$p \cdot \frac{d}{dt} (m_A - m)^2 + q \cdot \frac{d}{dt} (m_a - m)^2 = 2 \left[ \left( \frac{dp}{dt} \right)^2 m_{AA} + 2 m_{Aa} \frac{dp}{dt} \cdot \frac{dq}{dt} + \left( \frac{dq}{dt} \right)^2 m_{aa} - \frac{dm}{dt} \cdot \left( \frac{dp}{dt} + \frac{dq}{dt} \right) \right]$$

pondremos ahora todo en función de dp

$$dq = d(1-p) = -dp \quad \rightarrow \quad \frac{dp}{dt} \cdot \frac{dq}{dt} = - \left( \frac{dp}{dt} \right)^2 \quad ; \quad \left( \frac{dq}{dt} \right)^2 = \left( \frac{dp}{dt} \right)^2 \quad ; \quad \frac{dp}{dt} + \frac{dq}{dt} = 0$$

sustituyendo  $m_{AA}$ ,  $m_{Aa}$ ,  $m_{aa}$  por sus valores en función de los efectos de los alelos, tenemos

$$p \cdot \frac{d}{dt} (m_A - m)^2 + q \cdot \frac{d}{dt} (m_a - m)^2 = 2 \left[ (m + 2\alpha_1) \left( \frac{dp}{dt} \right)^2 - 2(m + \alpha_1 + \alpha_2) \left( \frac{dp}{dt} \right)^2 + (m + 2\alpha_2) \left( \frac{dp}{dt} \right)^2 \right] = 0$$