

2. MÉTODOS CLÁSICOS DE ANÁLISIS DE CARACTERES CUANTITATIVOS

Bernardo Ordás, Rosa Ana Malvar

Grupo de genética y mejora de maíz

Misión Biológica de Galicia (CSIC), apartado 28 36080 Pontevedra

e-mails: bordas@mbg.csic.es; rmalvar@mbg.csic.es

2.1. Varianza genética y heredabilidad

2.1.1. Introducción

2.1.2. Medias de un carácter cuantitativo. Efectos genéticos

2.1.2.1. Un gen con efectos aditivos y dominantes

2.1.2.2. Dos genes con efectos aditivos y dominantes

2.1.2.3. Múltiples genes con efectos aditivos y dominantes

2.1.2.4. ¿Es el modelo aditivo-dominante adecuado?

2.1.2.5. Estimación de los parámetros genéticos

2.1.2.6. Interpretación de los parámetros

2.1.2.7. Efectos genéticos. Estimaciones experimentales

2.1.3. Varianzas genéticas

2.1.3.1. Estimación de varianzas: evaluación de distintos genotipos

2.1.3.2. Estimación de varianzas genéticas. Diseños de cruzamientos

2.1.3.2.1. Estimación de varianzas genéticas: hermanos completos

2.1.3.2.2. Estimación de varianzas. Diseño I de Carolina del Norte

2.1.3.2.3. Estimación de varianzas. Diseño II de Carolina del Norte

2.1.3.2.4. Estimación de varianzas. Diseño III de Carolina del Norte

2.1.3.2.5. Estimación de varianzas. Dialelos

2.1.3.2.6. Estimación de varianzas: Distintas generaciones

2.1.3.2.7. Métodos de estimación de los componentes de varianza

2.1.3.2.8. Varianzas genéticas. Estimaciones experimentales

2.2. Selección

2.2.1. Introducción

2.2.2. Respuesta a la selección a corto plazo

2.2.2.1. Especies autóгамas.

2.2.2.2. Especies alógamas

2.2.3. Respuesta a la selección a largo plazo

2.3. Referencias

Caracteres cuantitativos son aquellos controlados por múltiples genes de forma que su herencia conjunta no es mendeliana, aunque cada gen individual sí se comporta como mendeliano. El estudio de la herencia de los caracteres cuantitativos es de gran importancia en mejora genética de plantas porque muchos de los caracteres de interés son de este tipo, por ejemplo, el rendimiento. Si bien las modernas técnicas moleculares permiten soñar con un futuro en el que sea posible el estudio individual de genes cuantitativos, a día de hoy la detección y análisis de éstos ha resultado en su mayoría infructuosa. Esto es debido a que los caracteres más complejos como el rendimiento están controlados por cientos o miles de genes de efecto muy pequeño lo que dificulta su estudio. El análisis genético de los caracteres cuantitativos, por tanto, se ha basado hasta el momento en gran medida en parámetros estadísticos tales como las varianzas genéticas o la heredabilidad. El conocimiento de estos parámetros sirve al mejorador a la hora de diseñar el programa de mejora más adecuado y para realizar la selección de forma más precisa y eficaz. En este capítulo se revisarán los diferentes diseños que permiten estudiar los efectos y las varianzas genéticas, así como los factores a considerar (número efectivo, intensidad de selección, etc) cuando se realiza selección artificial de un carácter cuantitativo.

2.1. Varianza genética y heredabilidad

2.1.1. Introducción

Un carácter es cualitativo si se debe a la acción de uno o dos genes y las diferencias alélicas producen fenotipos claramente diferentes con poca influencia ambiental. Por contra, un carácter es cuantitativo cuando la variación observada se debe a la segregación de muchos genes polimórficos y los efectos de las diferencias alélicas sobre el fenotipo son generalmente pequeñas en comparación con los efectos del ambiente. Cuando un mejorador estudia un carácter cuantitativo mide el valor fenotípico de ese carácter. El fenotipo (F) lo podemos dividir en los valores atribuibles al genotipo (G) y al ambiente (M).

$$F = G + M$$

Esta suma de efectos indica que lo que se analiza no es sólo el valor del genotipo sino la influencia del ambiente sobre los distintos genotipos que puede variar de unos ambientes a otros. Algunos factores macroambientales pueden ser controlados por el mejorador como por ejemplo el riego y los fertilizantes, pero otros son totalmente incontrolables. En cualquier caso la separación de los factores ambientales de los genéticos es de primordial importancia para el estudio de cualquier carácter cuantitativo.

La variación de los fenotipos, medida como la varianza fenotípica (VF) se puede descomponer en la varianza de los efectos genéticos (VG) y ambientales (VM).

$$VF = VG + VM$$

Hay que tener en cuenta que los valores genotípicos y las desviaciones ambientales pueden estar correlacionados, en cuyo caso hay que añadir un término más que consiste en el doble de la covarianza de G y M. Además, puede haber interacción entre

genotipos y ambientes en cuyo caso habrá un componente de la varianza atribuible a esta interacción.

En principio para saber si un carácter se puede mejorar es interesante conocer qué proporción de la varianza fenotípica es varianza genética (heredabilidad en sentido amplio). Desde los trabajos de Fisher (1918) y Wright (1935) se utiliza ampliamente un modelo en el que la varianza genética puede descomponerse en tres componentes:

- Componente debido al efecto aditivo de los genes (varianza aditiva, VA)
- Componente debido a las desviaciones del modelo aditivo por interacciones entre los alelos del mismo locus (varianza dominante, VD)
- Componentes debido a las desviaciones del modelo aditivo por interacciones entre alelos de distintos loci (varianza epistática, VE)

Por tanto:

$$VG = VA + VD + VE$$

Si un gen determinado presenta efecto aditivo, el valor genotípico del carácter analizado debe variar cuando se sustituye un alelo del gen por otro. Para clarificar estos conceptos nos referiremos a un carácter determinado por un solo gen (*A*) con dos alelos *A1* y *A2*. Consideremos que se está evaluando el daño producido por un insecto en función de la longitud del daño y el valor genotípico de los individuos $A1A1 = 5$ cm y el de $A1A2 = 6$ cm; esto significa que la presencia del alelo *A2* produce un cambio de 1 cm, por lo que el genotipo $A2A2 = 7$ cm. No obstante, si hay efecto dominante existen interacciones entre alelos del mismo locus y por lo tanto, el valor genotípico de $A1A2$ no coincidirá con la suma de los efectos aditivos del alelo *A1* y *A2*. Además, también puede existir efecto epistático debido a la interacción de alelos de loci distintos.

La varianza aditiva es el componente más importante puesto que es la causa fundamental del parecido entre parientes y, por lo tanto, la determinante de las propiedades genéticas observables de la población y la respuesta de ésta a la selección. En la práctica lo interesante es conocer la proporción de la varianza genética aditiva con respecto a la varianza fenotípica. Este cociente es lo que se denomina heredabilidad en sentido estricto (Falconer y Mackay, 1996).

Durante los últimos 70 años se han descrito varios métodos para analizar la herencia de los caracteres cuantitativos. Hallauer y Miranda (1988) y Mather y Jinks (1982) describen ampliamente los distintos métodos que pueden utilizarse. Se pueden distinguir dos tipos de métodos: los que se basan en la media de un carácter cuantitativo y por tanto permiten estimar los efectos genéticos y los que se basan en la varianza del carácter pudiendo estimar las varianzas genéticas y por ende la heredabilidad.

2.1.2. Medias de un carácter cuantitativo. Efectos genéticos

La media de los fenotipos de los individuos pertenecientes a una determinada familia y la variación entre ellos se debe a la acción conjunta de los genes, del ambiente y de la interacción. Diferentes familias tendrán diferentes medias y diferentes varianzas porque contienen diferentes genotipos.

El caso más sencillo:

Familias que se obtienen del cruzamiento de dos líneas puras. Se suele trabajar con 6 generaciones básicas: P_1 y P_2 = parentales; F_1 = cruzamiento entre las líneas parentales; F_2 =autofecundación de la F_1 ; Bc_1 y Bc_2 = retrocruzamientos de la F_1 con cada uno de los padres.

Analizaremos varias situaciones:

2.1.2.1. Un gen con efectos aditivos y dominantes

Partimos de dos líneas puras que son idénticas para todos sus genes excepto para el gen *A*. Este gen tiene dos alelos A^+ y A^- que son responsables del valor fenotípico más alto y más bajo respectivamente del carácter de interés. Los genotipos de distintas generaciones son:

$$P_1 = A^+A^+ \quad P_2 = A^-A^- \quad F_1 = A^+A^-$$

Si: $m = (P_1 + P_2)/2$ media de los fenotipos de los dos padres; a = desviación de m debido al efecto aditivo del gen; y d = desviación de m debido al efecto dominante.

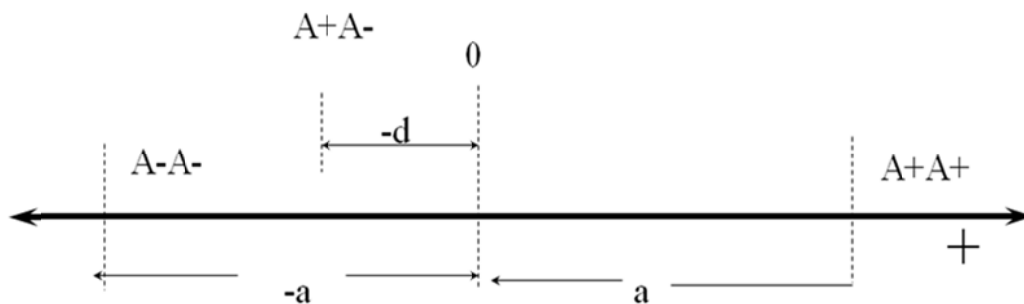
Entonces, obtenemos los siguientes valores genéticos

$$P_1 (A^+A^+) = m + a \quad m = (P_1 + P_2)/2$$

$$P_2 (A^-A^-) = m - a \quad \text{y} \quad a = (P_1 - P_2)/2$$

$$F_1 (A^+A^-) = m + d \quad d = F_1 - (P_1 + P_2)/2$$

Representado gráficamente el valor fenotípico en una escala +/-



Si $|d| > a$ sobredominancia; si $|d| = a$ dominancia completa; si $|d| < a$ dominancia parcial.

Si producimos la F₂ autofecundado la F₁ o cruzando individuos de la F₁, y los retrocruzamientos cruzando la F₁ con P₁ para obtener el Bc₁ y la F₁ con P₂ para obtener el Bc₂ podemos obtener las medias fenotípicas de estas generaciones (Tabla 2.1)

Tabla 2.1. Valor genotípico y media fenotípica de las generaciones F₂, Bc₁ y Bc₂

	F ₂			Bc ₁		Bc ₂	
Genotipo	A+A+	A+A-	A-A-	A+A+	A+A-	A-A-	A+A-
Frecuencia	1/4	1/2	1/4	1/2	1/2	1/2	1/2
Valor genético	m + a	m + d	m - a	m + a	m + d	m - a	m + d
Media fenotípica	1/4(m + a) + 1/2 (m+d) + 1/4 (m-a) = m + 1/2 d			1/2(m + a) + 1/2 (m+d) = m + 1/2 a + 1/2 d		1/2(m - a) + 1/2(m+d) = m - 1/2 a + 1/2 d	

Por tanto conocemos la relación que existe entre la media de los fenotipos y los valores genéticos de las seis generaciones básicas.

2.1.2.2. Dos genes con efectos aditivos y dominantes

Cuando los padres difieren en dos genes se puede hacer extensivo el modelo de un gen pero hay que tener en cuenta que la distribución de los alelos en los padres afecta a las estimaciones de los parentales y de los retrocruzamientos. Los alelos + y - pueden estar en los parentales asociados o dispersos:

Asociados P₁ = A+A+B+B+ y P₂ = A-A-B-B- ; dispersos P₁ = A+A+B-B- y P₂ = A-A-B+B+

Por definición a_A y a_B son positivos, por lo tanto en valor del componente aditivo será mayor si los genes están en asociación (a = a_A + a_B) que si están en disociación (a = a_A - a_B). El padre medio no depende de la distribución de los alelos. Ni el componente dominante porque es d = d_A + d_B en ambas situaciones. Lo que si hay que tener en cuenta es que los valores de d pueden ser positivos o negativos. Cuando la dominancia es negativa en un gen y positiva en otro se denomina dominancia ambidireccional mientras que la si la dominancia es sólo positiva o negativa se denomina dominancia unidireccional.

2.1.2.3. Múltiples genes con efectos aditivos y dominantes

Cuando el carácter está regulado por muchos genes es difícil distinguir los efectos individuales de cada gen. Esto es especialmente cierto para los efectos genéticos aditivos ya que depende de la distribución de los alelos en los padres.

Si k genes están en asociación y k' están en dispersión, el valor de los padres es:

$$P_1 = m + \left(\sum_1^{k-k'} a_i - \sum_1^{k'} a_i \right) \quad P_1 = m + \left(\sum_1^k a_i - 2 \sum_1^{k'} a_i \right)$$

$$P_2 = m - \left(\sum_1^{k-k'} a_i - \sum_1^{k'} a_i \right) \quad P_2 = m - \left(\sum_1^k a_i - 2 \sum_1^{k'} a_i \right)$$

Si definimos un coeficiente de asociación/dispersión r_a como

$$r_a = \frac{\left(\sum_1^k a_i - 2 \sum_1^{k'} a_i \right)}{\sum_1^k a_i}$$

$r_a = 1$ significa que hay asociación completa; si $r_a = 0$ la dispersión es completa, mientras que si $0 < r_a < 1$ la asociación es parcial. Ahora podemos definir

$$[a] = r_a \sum_1^k a_i$$

donde $[a]$ = nos indica los efectos aditivos netos después de tener en cuenta los alelos dispersos en los padres. Por lo tanto, los valores genéticos de los padres son: $P_1 = m + [a]$ y $P_2 = m - [a]$.

El genotipo de F_1 independientemente de los genes que están en asociación o en dispersión será $A+A-B+B-.....K+K-$ y por tanto $F_1 = m + d_A + d_B + + d_K = m + [d]$

representa los efectos dominantes netos e indica la dirección de la dominancia en la mayoría de los k genes. Los valores genéticos de distintas generaciones se presentan en la Tabla 2.2.

Tabla 2.2. Valores genéticos de distintas generaciones y los valores de la media de los padres (m), el efecto aditivo neto ([a]) y el efecto dominante neto ([d]).

Valores genéticos	Generación	m	[a]	[d]
$P_1 = m + [a]$	P_1	1	1	0
$P_2 = m - [a]$	P_2	1	-1	0
$F_1 = m + [d]$	F_1	1	0	1
$F_2 = m + \frac{1}{2} [d]$	F_2	1	0	0,50
$Bc_1 = m + \frac{1}{2}[a] + \frac{1}{2} [d]$	F_3	1	0	0,25
$Bc_2 = m - \frac{1}{2}[a] + \frac{1}{2} [d]$	F_4	1	0	0,125
	F_∞	1	0	0
$F_n = m + \frac{1}{2^n} [d]$	Bc_{11}	1	0,50	0,50
$Bc_{1n} = m + (1 - \frac{1}{2^n})[a] + \frac{1}{2^n} [d]$	Bc_{21}	1	0,75	0,25
$Bc_{2n} = m - (1 - \frac{1}{2^n})[a] + \frac{1}{2^n} [d]$	Bc_{12}	1	-0,50	0,50
	Bc_{22}	1	-0,75	0,25

De los coeficientes de las tablas está claro que si hay dominancia direccional la media de F_n cambiará con cada generación de autofecundación tendiendo a alcanzar el valor de la media de los padres. Si un cruzamiento presenta vigor híbrido o heterosis, es decir la F_1 supera la media de los padres, la disminución del valor con cada generación de autofecundación se conoce como depresión consanguínea.

2.1.2.4. ¿Es el modelo aditivo-dominante adecuado?

Conocemos por tanto el valor esperado de las medias de distintas generaciones si los únicos efectos genéticos son los aditivos y los dominantes. Existe una manera de comprobar que los datos de un carácter se adecúan a un modelo aditivo-dominante. Si

sólo existe aditividad y dominancia: $B_{C1} = \frac{1}{2} P_1 + \frac{1}{2} F_1$ o $2B_{C1} - P_1 - F_1 = 0$; si existe algún factor más, como por ejemplo la epistasia, entonces $2B_{C1} - P_1 - F_1 \neq 0$ es decir $2B_{C1} - P_1 - F_1 = A$, siendo A una cantidad significativamente distinta de 0. Sabemos que si $A = \sum k_i y_i$, es decir una suma de funciones lineales, la varianza (s^2) de A es $s^2_A = \sum k_i^2 y_i^2$ por lo tanto $s^2_A = 4 s^2_{B_{C1}} + s^2_{P_1} + s^2_{F_1}$. Conociendo la varianza de A podemos calcular si A difiere significativamente de cero. Para ello calculamos $t(g_l) = \frac{A}{\sqrt{s^2_A}}$ siendo g_l la suma de los grados de libertad de las varianzas B_{C1} , P_1 y F_1 . Si $A = 0$ el modelo aditivo-dominante explica la herencia del carácter que estamos estudiando. Hay dos relaciones más: $B = 2B_{C2} - P_2 - F_1$ y $C = 4F_2 - 2F_1 - P_1 - P_2$. Para asegurarnos que el modelo aditivo-dominante es correcto A , B y C tienen que ser igual a cero. Esta prueba fue propuesta por Mather en 1949 y le denominó 'scaling test'.

2.1.2.5. Estimación de los parámetros genéticos

Pretendemos estimar los parámetros genéticos m , $[a]$ y $[d]$ utilizando la media fenotípica de las 6 generaciones básicas. Disponemos de 6 ecuaciones para resolver 3 incógnitas podemos estimar m , $[a]$ y $[d]$ utilizando los métodos de la regresión. Como simplificación consideremos que queremos estimar sólo m y $[a]$. Se pueden escribir las medias de las generaciones (y_i) como: $Y_i = c + b x_i$ donde $c = m$ y $x_i =$ coeficientes de $[a]$ en la generación i , por lo que $b = [a]$. Estadísticamente estas estimaciones son las que minimizan las desviaciones entre los valores observados y esperados. Podemos testar la significación de cada estima utilizando una t de Student. También podemos utilizar un análisis de varianza y comprobar si el cuadrado medio residual es significativo. Si el residuo es significativo entonces existen otros factores además de m y $[a]$ que deben incluirse en el modelo. Ahora se trata de una regresión múltiple $Y_i = c + b_1 x_i + b_2 x_i + \dots$ donde $b_1 = [a]$, $b_2 = [d]$...

Hemos asumido que las medias de todas las generaciones tienen el mismo error. Es decir que las varianzas de todas las medias de las generaciones son iguales (s_i^2/n_i). Pero

esto no tiene por qué ser cierto. La variabilidad esperada en una línea pura es menor que una F_2 por lo que es habitual utilizar distinto número de individuos para caracterizar distintas generaciones, por lo que es conveniente ponderar las medias de las generaciones para realizar el análisis de regresión. Se utiliza como ponderación el recíproco de la varianza. Como tenemos N medias generacionales podemos estimar N parámetros, como uno de ellos es m podemos estimar $N-1$ efectos. Nosotros pretendemos explicar la variación entre las medias observadas y estimadas con el mínimo de parámetros posibles. Si solo estimamos m entonces validaremos el modelo mediante una χ^2 con $N-1$ grados de libertad, si incluimos un parámetro más en el modelo la comprobación se realiza con un grado de libertad menos. Esta prueba es conocida como 'joint scaling test'. Si el modelo aditivo – dominante no es suficiente habrá que buscar otros efectos como epistasia o efectos maternos.

2.1.2.6. Interpretación de los parámetros

Si $[a] = 0$ sólo podemos afirmar que no hay variación aditiva en el cruzamiento si el carácter se debe a un único gen es decir cuando $[a] = a$. Si $[d] = 0$ sólo no hay dominancia si $[d] = d$ ya que se podrían producir compensaciones si hay dominancias de sentido contrario.

Para un gen definimos la ratio de dominancia como d/a y podemos decir que:

- Si $d/a = +1$, el alelo $A+$ es completamente dominante sobre el alelo $A-$
- Si $d/a = -1$, el alelo $A-$ es completamente dominante sobre el alelo $A+$
- Si $0 < d/a < +1$, el alelo $A+$ es parcialmente dominante sobre el alelo $A-$
- Si $-1 < d/a < 0$, el alelo $A-$ es parcialmente dominante sobre el alelo $A+$
- Si $d/a = 0$ no hay dominancia
- Si $d/a > +1$ el alelo $A+$ es sobredominante sobre el alelo $A-$
- Si $d/a < -1$ el alelo $A-$ es sobredominante sobre el alelo $A+$

$$[d]/[a] = \sum_{i=1}^k \frac{d_i}{r} \sum_{a=1}^k a_i$$

Con dos o más genes: El numerador podría ser cero si hay dominancia bidireccional y el denominador podría ser cero debido a la dispersión génica. Por estos problemas la ratio se considera como 'ratio potencial' porque realmente nos indica cuál de los padres tiene más alelos dominantes. Por otra parte, cuanto mayor sea el número de generaciones estudiadas tendremos más capacidad para estimar un mayor número de efectos, por ejemplo se podrían estimar efectos epistáticos.

2.1.2.7. Efectos genéticos. Estimaciones experimentales.

El diseño de medias generacionales, debido a su sencillez, ha sido y es ampliamente utilizado para estimar los efectos genéticos y por ende conocer la herencia de múltiples caracteres cuantitativos, especialmente en plantas autóгамas, pero también en plantas alógamas que admiten la autofecundación, como el maíz, o en las que se pueden obtener líneas puras por distintos métodos.

Dada la profusión de bibliografía (en una búsqueda realizada en abril del 2012 en la Web of Knowledge utilizando como criterios de búsqueda 'análisis de medias generacionales' y 'aditividad y dominancia' hemos encontrado 248 registros) sólo comentaremos algunos estudios recientes. Las medias generacionales actualmente se utilizan ampliamente en el estudio de la herencia de la resistencia a enfermedades. Así Lyimo et al. (2011) estudiaron la resistencia a la ceporiosis del maíz y sus resultados indican que la magnitud de los efectos aditivos, dominantes y epistáticos dependen del cruzamiento analizado y de los caracteres estudiados. Sin embargo, los efectos dominantes y los epistáticos dominante × dominante son importantes en la resistencia parcial a la fusariosis de la espiga de trigo (*Fusarium graminearum*) medida como porcentaje de espiguillas infectadas, mientras que los efectos aditivos y aditivo × aditivo son importantes para los caracteres relacionados con el desarrollo de la

enfermedad (Fakhfakh et al., 2011). El modo de herencia de la resistencia a *Septoria tritici* en trigo varía con la agresividad del aislado utilizado en la inoculación (Benjdi et al., 2011). Así cuando el inóculo es poco agresivo la herencia sigue un modelo aditivo y dominante, cuando la agresividad aumenta se ajusta a un modelo con efectos epistáticos, y si los niveles de agresividad son altos entonces el modelo epistático con dos genes no es suficiente para explicar la variación observada en el estudio de medias generacionales. Las especies *Pythium* causan pudrición de la semilla, ahogamiento, y pudrición de la raíz del cártamo (*Carthamus tinctorius* L.) y la resistencia a esta enfermedad está bajo control genético con efectos aditivos, dominantes y epistáticos significativos (Ghaderi et al., 2011). La resistencia al nematodo *Rotylenchulus reniformis* en la accesión de algodón (*Gossypium barbadense*) GB713 se debe a uno o más genes (Gutierrez et al., 2011).

Además de los caracteres de resistencia las medias generacionales han sido útiles para determinar la herencia de otro tipo de caracteres, por ejemplo el contenido en azúcar de la caña de sorgo para utilizarlo en la producción de etanol (Audilakshmi et al., 2010); tolerancia a la salinidad en trigo (Dashti et al., 2010); la transición de fases en maíz (Ordás et al., 2009) o la producción y otros caracteres agronómicos en cebada (Eshghi y Akhundowa, 2009).

2.1.3. Varianzas genéticas

Wright (1935) fue el primero en proponer métodos para la estimación de las varianzas genéticas. En la actualidad existe una amplia bibliografía sobre los diseños adecuados para los cálculos de estos parámetros. En general, para llevar a cabo estos diseños primero es necesario realizar una serie de cruzamientos y posteriormente ensayar las progenies procedentes de estos cruzamientos con un diseño adecuado. El análisis de varianza apropiado al modelo permite estimar una serie de componentes de varianza basándose en los cuadrados medios esperados. Estos componentes se pueden expresar

como covarianzas entre los descendientes que a su vez son funciones lineales de los distintos componentes de la varianza.

La población que queremos estudiar debe de cumplir las siguientes condiciones: herencia diploide, que no existan correlaciones entre el ambiente y los descendientes, que no exista herencia materna, que no exista alelismo múltiple, los genes que determinan los caracteres que queremos analizar tiene que ser independientes o estar en equilibrio de ligamiento y los miembros elegidos para los cruzamientos tienen que ser una muestra aleatoria de la población.

2.1.3.1. Estimación de varianzas: evaluación de distintos genotipos

Uno de los diseños más sencillo para estimar varianzas genéticas es evaluar diferentes genotipos en un ambiente utilizando un diseño de bloques al azar con repeticiones. Para ello, los distintos genotipos tienen que ser una muestra al azar de la población que queremos estudiar. Por tanto los genotipos se consideran factores aleatorios. Con este diseño podemos estimar la varianza genética pero no la podemos descomponer en sus componentes. Así mismo solo podemos calcular la heredabilidad en sentido amplio. Cada observación tendrá el siguiente valor:

$$Y_{ij} = \mu + R_j + G_i + e_{ij}$$

Y_{ij} = valor de la observación j^{th} del genotipo i^{th} ; μ = media del experimento; R_j = efecto de la repetición j^{th} ; G_i = efecto del genotipo i^{th} ; e_{ij} = error asociado con la ij^{th} . El análisis de varianza apropiado al modelo se presenta en la Tabla 2.3. y los parámetros que se pueden estimar en la Tabla 2.4.

Tabla 2.3. Análisis de varianza de genotipos

Fuentes de variación	gl	CM	CM esperados	
			CM fijos	CM aleatorios

Repeticiones	r-1	CMr	$\sigma^2 + g\sigma^2r$	$\sigma^2 + g\sigma^2r$
Genotipos	g-1	CMg	$\sigma^2 + rk^2g$	$\sigma^2 + r\sigma^2g$
Error	(r-1)(g-1)	CMe	σ^2	σ^2
Total	rg-1			

r = repeticiones, g = genotipos, e = error σ^2 = varianza, CM = cuadrado medio, gl = grados de libertad

Tabla 2.4. Estimaciones de parámetros basados en el análisis de varianza

Parámetros	Estimaciones
La media de un genotipo	$\mu + G_i + \Sigma R_j/r$
Error estándar de la media de genotipos	$\sqrt{CMe/r}$
Mínima diferencia significativa MDS0,05	$t(gle, 0,05) \sqrt{2CMe/r}$
Varianza fenotípica total	$\sigma^2 + \sigma^2g$
Varianza fenotípica de las medias de fenotipos	$\sigma^2 /r + \sigma^2g$
Varianza genotípica	$(CMg - CMe)/r$
Heredabilidad basada en las unidades experimentales	$\sigma^2g/ (\sigma^2 + \sigma^2g)$
Heredabilidad basada en las medias	$\sigma^2g/ (\sigma^2 /r + \sigma^2g)$
Var(var)	$2/c^2 \sum (CM_i^2 /gl_i+2)$
Var(VG)	$2/r^2 ((CMg^2 /((g-1)+2) + (CMe^2 /((g-1)(r-1)+2))$

r = repeticiones, g = genotipos, e = error, σ^2 = varianza, CM = cuadrado medio, μ = media del experimento, R_j = efecto de la repetición j^{th} ; G_i = efecto del genotipo i^{th} , gl = grados de libertad

También es común repetir la evaluación en varios ambientes. Se suele utilizar un diseño en bloques al azar con r repeticiones en m ambientes. Cada observación tendrá el siguiente valor:

$$Y_{ijk} = \mu + M_i + R(M)_{ij} + G_k + GM_{ik} + e_{ijk}$$

Y_{ijk} = valor de la observación j^{th} del genotipo k^{th} en el ambiente i^{th} ; M_i = efecto del macroambiente i^{th} ; $R(M)_{ij}$ = efecto de la repetición j^{th} dentro del ambiente i^{th} ; G_k = efecto

del genotipo k^{th} ; GM_{ik} = efecto de la interacción del ambiente i^{th} con el genotipo k^{th} ; e_{ijk} = efecto del error asociado con la observación del genotipo k^{th} en la repetición j^{th} y en el ambiente i^{th} . En las tablas 2.5 y 2.6 se muestran el análisis de varianza y los parámetros estimados respectivamente, consideramos que todos los factores son aleatorios.

Tabla 2.5. Análisis de varianza de genotipos evaluados en varios ambientes

Fuente de variación	gl	CM	CM esperados, factores aleatorios
Ambiente	m-1	CMm	$\sigma^2 + g\sigma^2r(m) + r\sigma^2gm + mr\sigma^2g + rg\sigma^2m$
Repeticiones(ambientes)	(r-1)e	CMr(m)	$\sigma^2 + g\sigma^2r(m)$
Genotipos	g-1	CMg	$\sigma^2 + r\sigma^2gm + mr\sigma^2g$
Genotipo \times ambiente	(g-1)(e-1)	CMg \times m	$\sigma^2 + r\sigma^2gm$
Error	(r-1)(g-1)e	CMe	σ^2
Total	rge-1		

r = repeticiones, g = genotipos, m = ambientes, e = error, σ^2 = varianza, CM = cuadrado medio, gl = grados de libertad

Tabla 2.6. Estimaciones de parámetros basados en el análisis de varianza

Parámetros	Estimaciones
Error estándar de la media de genotipos	$\sqrt{CMg \times m / mr}$
Mínima diferencia significativa MDS _{0,05}	$t_{(g gm, 0,05)} \sqrt{2CMg \times m / mr}$
Varianza fenotípica total	$\sigma^2 + \sigma^2gm + \sigma^2g$
Varianza fenotípica de las medias de fenotipos	$\sigma^2 / rm + \sigma^2gm / r + \sigma^2g$
Varianza genotípica	$(CMg - CMg \times m) / rm$
Heredabilidad basada en las unidades experimentales	$\sigma^2g / (\sigma^2 + \sigma^2gm + \sigma^2g)$
Heredabilidad basada en las medias	$\sigma^2g / (\sigma^2 / rm + \sigma^2gm / r + \sigma^2g)$
Var(var)	$2/c^2 \sum (CM_i^2 / gl_i + 2)$
Var(VG)	$2/r^2 ((CMg^2 / ((g-1)+2) + (CMg \times m^2 / ((g-1)(m-1)+2)))$

r = repeticiones, g = genotipos, m = ambientes, e = error, σ^2 = varianza, CM = cuadrado medio, gl = grados de libertad

2.1.3.2. Estimación de varianzas genéticas. Diseños de cruzamientos

2.1.3.2.1. Estimación de varianzas genéticas: hermanos completos

Diseñado por Mather (1949) consiste en cruzar por parejas n individuos de la F_2 o de una población, obteniendo $n/2$ familias. Se suelen denominar familias de hermanos completos (full sib) o familias biparentales. Si evaluamos las familias en un diseño con r repeticiones y dentro de cada parcela experimental evaluamos k individuos de cada familia, el análisis de varianza sería (Tabla 2.7):

Tabla 2.7. Análisis de varianza de familias de hermanos completos

Fuente de variación	gl	CM	CM esperados
Repeticiones	$r-1$		
Entre familias	$n/2-1$	CMc	$\sigma^2_w + k \sigma^2 + rk\sigma^2_c$
Error	$(r-1)(n/2 -1)$	CMe	$\sigma^2_w + k \sigma^2$
Total	$rn/2 - 1$		
Dentro de familias	$rn/2(k-1)$	CMw	σ^2_w

r = repeticiones, n = individuos de la población, k = individuos dentro de cada familia, CM = cuadrado medio, c = cruzamientos (entre familias), w = dentro de familias

Podemos comprobar, mediante el test de la F , si la diferencia entre cruzamientos es más grande que la diferencia dentro de cruzamientos. Tiene la ventaja de que es un diseño muy simple pero la información que obtenemos no es suficiente para diseñar un programa de selección a largo plazo y solamente sabemos si hay suficiente variabilidad para iniciar el programa de selección.

2.1.3.2.2. Estimación de varianzas. Diseño I de Carolina del Norte

Estas estimaciones además de poderse aplicarse a una F_2 también se pueden aplicar a una población en general. Fue diseñado por Comstock y Robinson (1948). Para llevar a cabo este diseño un número m de individuos, que actúan como machos, se cruzan con f hembras cada uno. Por lo tanto obtendremos $m \times f$ familias de hermanos completos y

m familias de medios hermanos. Cada familia de hermanos completos estará formada k individuos. La evaluación se llevará a cabo en un ambiente con r repeticiones. El análisis de varianza se expone en la Tabla 2.8 y las estimaciones de los parámetros en la Tabla 2.9.

Tabla 2.8. Análisis de varianza del diseño I de Carolina del Norte

Fuentes de variación	gl	CM	CM esperados
Repeticiones	r-1		
Machos	m-1	CMm	$\sigma^2 + r\sigma^2f/m + fr\sigma^2m$
Hembras (machos)	m(f-1)	CMf/m	$\sigma^2 + r\sigma^2f/m$
Error	(r-1)(mf -1)	CMe	σ^2
Total	nmr - 1		
Dentro de familias	rmf(k-1)	CMw	σ^2w

CM = cuadrado medio, gl = grados de libertad, r = repeticiones, m = machos, f = hembras, k = individuos dentro de familias, e = error, w = dentro de familias, σ^2 = varianza

Tabla 2.9. Estimaciones de parámetros utilizando el diseño I de Carolina del Norte

Párametros	Estimaciones	
σ^2m	$(CMm - CMf/m) / fr$	$\frac{1}{4} VA$
σ^2f/m	$(CMf/m - CMe)/r$	$\frac{1}{4} VA + \frac{1}{4} VD$
VA	$4 \sigma^2m$	
VD	$4(\sigma^2f/m - \sigma^2m)$	
h^2 basada en medias	$4 \sigma^2m / (\sigma^2/r + 4\sigma^2f/m)$	

CM = cuadrado medio, VA = varianza aditiva, VD = varianza dominante, r = repeticiones, m = machos, f = hembras, k = individuos dentro de familias, e = error, w = dentro de familias, σ^2 = varianza

En este caso la varianza entre familias es independiente de la variación ambiental y así la variación entre machos representa únicamente a la varianza aditiva. Además como ventajas tenemos que estimamos la VA, VD y h^2 . La principal desventaja es que la VD tiene asociados errores importantes. Si lo evaluamos en varios ambientes, entonces el

análisis de varianza y el método de estimación serán los expuestos en las Tablas 2.10 y 2.11.

Tabla 2.10. Análisis de varianza del diseño I de Carolina del Norte evaluado en varios ambientes

Fuentes de variación	gl	CM	CM esperados
Ambientes	e-1	CME	
Repeticiones (amb)	e(r-1)	CMr(E)	
Machos	m-1	CMm	$\sigma^2 + r\sigma^2f/me + rfr\sigma^2me + re\sigma^2f/m + fre\sigma^2m$
Hembras (machos)	m(f-1)	CMf/m	$\sigma^2 + r\sigma^2f/me + re\sigma^2f/m$
Machos × amb	m-1	CMmE	$\sigma^2 + r\sigma^2f/me + r\sigma^2me$
Hembras(machos)× amb	m(f-1)	CMf/mE	$\sigma^2 + r\sigma^2f/me$
Error	(r-1)(mf-1)	CMe	σ^2
Total	nmr - 1		
Dentro de familias	rmf(k-1)	CMw	σ^2w

CM = cuadrado medio, gl = grados de libertad, r = repeticiones, e = ambientes, m = machos, f = hembras, k = individuos dentro de familias, CMe = cuadrado medio del error, w = dentro de familias, σ^2 = varianza

Tabla 2.11. Estimaciones de parámetros utilizando el diseño I de Carolina del Norte cuando las familias se evalúan en varios ambientes

Párametros	Estimaciones	
σ^2m	$(CMm + CMf/mE - CMf/m - CMmE)/rfe$	$\frac{1}{4} VA$
σ^2f/m	$(CMf/m - CMf/mE)/re$	$\frac{1}{4} VA + \frac{1}{4} VD$
σ^2me	$(CMmE - CMf/mE)/rf$	$\frac{1}{4} VAE$
σ^2f/me	$(CMf/mE - CMe)/re$	$\frac{1}{4} VAE + \frac{1}{4} VDE$
VA	$4 \sigma^2m$	
VD	$4(\sigma^2f/m - \sigma^2m)$	
VAE	$4 \sigma^2me$	
VDE	$4(\sigma^2f/me - \sigma^2me)$	
h^2 basada en medias de re	$4 \sigma^2m / (\sigma^2/re + 4\sigma^2f/me + 4\sigma^2f/m)$	

CM = cuadrado medio, CMe = cuadrado medio del error, VA = varianza aditiva, VD = varianza dominante, VAE = varianza aditiva×ambiente, VDE =varianza dominante × ambiente, r = repeticiones, m = machos, f = hembras, E y e = ambiente, σ^2 = varianza

Es necesario evaluar un gran número de familias de hermanos completos por lo que muchas veces las familias se organizan en grupos de tal forma que se reduzca el error experimental.

2.1.3.2.3. Estimación de varianzas. Diseño II de Carolina del Norte

Este diseño, también propuesto por Comstock y Robison (1948), es similar al diseño I salvo que ahora todos los machos se cruzan con todas las hembras creando así grupos de medios hermanos de hembras y grupos de medios hermanos de machos. Consideremos que m machos se cruzan con f hembras, que los evaluamos en un ambiente con r repeticiones y que de cada familia evaluamos k plantas por repetición. El ANOVA se indica en la Tabla 2.12 y las estimaciones en la 2.13

Tabla 2.12. Análisis de varianza del diseño II de Carolina del Norte

Fuentes de variación	gl	CM	CM esperados
Repeticiones	$r-1$		
Entre MH machos	$m-1$	CM _m	$\sigma^2 + r\sigma^2mf + r\sigma^2m$
Entre MH hembras	$f-1$	CM _f	$\sigma^2 + r\sigma^2mf + r\sigma^2f$
Machos × hembras	$(m-1)(f-1)$	CM _{mf}	$\sigma^2 + r\sigma^2mf$
Error	$(r-1)(mf-1)$	CME	σ^2
Total	$rmf - 1$		
Dentro de HC	$rmf (k-1)$	CM _w	σ^2w

CM = cuadrado medio, gl = grados de libertad, MH = medios hermanos, HC = hermanos completos, r = repeticiones, m = machos, k = individuos dentro de familias, CME = cuadrado medio del error, w = dentro de familias.

Tabla 2.13. Estimaciones de parámetros utilizando el diseño II de Carolina del Norte

Párametros	Estimaciones
------------	--------------

σ^2m	$(CMm - CMfm) / rf$	$\frac{1}{4} VA$
σ^2f	$(CMf - CMfm) / rm$	$\frac{1}{4} VA$
σ^2fm	$(CMfm - CMe) / r$	$\frac{1}{4} VD$
VA	$4 \sigma^2m \text{ ó } 4 \sigma^2f$	
VD	$4\sigma^2fm$	
h^2 basada en medias	$4 \sigma^2m / (\sigma^2/r + 4\sigma^2fm + 4 \sigma^2m)$	En función de m
h^2 basada en medias	$4 \sigma^2f / (\sigma^2/r + 4\sigma^2fm + 4 \sigma^2f)$	En función de f

CM = cuadrado medio, VA = varianza aditiva, VD = varianza dominante, r = repeticiones, m = machos, f = hembras, k = individuos dentro de familias, e = error, w = dentro de familias, σ^2 = varianza

El diseño II tiene la ventaja de que las estimaciones de la VA, VD y h^2 son más precisas que en el diseño I ya que combinando machos y hembras aumenta la precisión de la VA y, que además, la VD se puede testar directamente. No obstante tiene la desventaja que se necesitan muchos cruzamientos. Si lo evaluamos en varios ambientes, el análisis de varianza será el expuesto en la Tabla 2.14 y las estimaciones en la Tabla 2.15:

Tabla 2.14. Análisis de varianza del diseño II de Carolina del Norte evaluado en varios ambientes

Fuentes de variación	gl	CM	CM esperados
Ambientes	e-1	CME	
Repeticiones (amb)	e(r-1)	CMr(E)	
Machos	m-1	CMm	$\sigma^2 + r\sigma^2mfe + r\sigma^2me + r\sigma^2mf + r\sigma^2m$
Hembras	f-1	CMf	$\sigma^2 + r\sigma^2mfe + r\sigma^2fe + r\sigma^2mf + r\sigma^2f$
Machos × hembras	(m-1)(f-1)	CMmf	$\sigma^2 + r\sigma^2mfe + r\sigma^2mf$
Machos × amb	(m-1)(e-1)	CMmE	$\sigma^2 + r\sigma^2mfe + r\sigma^2me$
Hembras × amb	(f-1)(e-1)	CMfE	$\sigma^2 + r\sigma^2mfe + r\sigma^2fe$
Machos × hembras × amb	(m-1)(f-1)(e-1)	CMmfe	$\sigma^2 + r\sigma^2mfe$
Error	(r-1)(mfe-1)	CMe	σ^2
Total	rmf - 1		

CM = cuadrado medio, gl = grados de libertad, r = repeticiones, E y e = ambientes, m = machos, f = hembras, CMe = cuadrado medio del error, w = dentro de familias, σ^2 = varianza

Tabla 2.15. Estimaciones de parámetros utilizando el diseño II de Carolina del Norte

Párametros	Estimaciones	
σ^2m	$(CMm+CMmfE-CMfm-CMmE)/rfe$	$\frac{1}{4}$ VA
σ^2f	$(CMf+CMmfE-CMfm-CMfE)/rme$	$\frac{1}{4}$ VA
σ^2fm	$(CMfm - CMfmE)/re$	$\frac{1}{4}$ VD
σ^2me	$(CMmE - CMfmE) /rf$	$\frac{1}{4}$ VAE
σ^2fe	$(CMfE - CMfmE)/rm$	$\frac{1}{4}$ VAE
σ^2fme	$(CMfmE - CMe)/r$	$\frac{1}{4}$ VDE
VA	$4 \sigma^2m$ ó $4 \sigma^2f$	
VD	$4\sigma^2fm$	
VAE	$4 \sigma^2me$ ó $4 \sigma^2fe$	
VDE	$4\sigma^2fme$	
h^2 basada en medias	$4 \sigma^2m /(\sigma^2/re + 4\sigma^2fme/e + 4 \sigma^2me/e + + 4\sigma^2fm + 4 \sigma^2m)$	función de m
h^2 basada en medias	$4 \sigma^2f /(\sigma^2/re + 4\sigma^2fme/e + 4 \sigma^2fe/e + + 4\sigma^2fm + 4 \sigma^2f)$	función de f

CM = cuadrado medio, CMe = cuadrado medio del error, VA = varianza aditiva, VD = varianza dominante, VAE = varianza aditiva×ambiente, VDE = varianza dominante × ambiente, r = repeticiones, m = machos, f = hembras, E y e = ambiente, σ^2 = varianza

2.1.3.2.4. Estimación de varianzas. Diseño III de Carolina del Norte

Para llevar a cabo este diseño individuos de una población F_2 se retrocruzan con las líneas parentales de la F_2 . Los probadores no son elegidos al azar, si no que son las líneas parentales del cruzamiento, por eso se consideran factores fijos en el análisis de varianza. Consideremos n individuos que se cruzan con las líneas parentales. Se evalúan en un diseño de bloques al azar con r repeticiones. En cada parcela experimental se evalúan k individuos. El análisis de varianza se muestra en la Tabla 2.16 y las estimaciones en la Tabla 2.17.

Tabla 2.16. Análisis de varianza del diseño III de Carolina del Norte

Fuentes de variación	gl	CM	CM esperados
Repeticiones	r-1		
Probador	1	CMt	$\sigma^2 + r\sigma^2tF_2 + mk^2t$
F ₂	n-1	CMF ₂	$\sigma^2 + 2r\sigma^2F_2$
Probador × F ₂	n-1	CMtF ₂	$\sigma^2 + r\sigma^2tF_2$
Error	(r-1)(2n -1)	CMe	σ^2
Total	2nr-1		
Dentro de probador × F ₂	2nr(k-1)	CMw	σ^2w

CM = cuadrado medio, t = probadores, n = individuos, CMe = Cuadrado medio del error, w = dentro de familias, k = individuos dentro de familias

Tabla 2.17. Estimaciones de parámetros utilizando el diseño III de Carolina del Norte

Párametros	Estimaciones	
σ^2F_2	$(CMF_2 - CMe)/2r$	$\frac{1}{4} VA$
σ^2tF_2	$(CMtF_2 - CMe)/r$	VD
VA	$4 \sigma^2F_2$	
VD	σ^2tF_2	
h ² basada en medias	$4 \sigma^2F_2 / (\sigma^2/r + \sigma^2tF_2 + 4 \sigma^2F_2)$	

CM = cuadrado medio, CMe = cuadrado medio del error, VA = varianza aditiva, VD = varianza dominante, r = repeticiones, t = probadores, σ^2 = varianza

Este diseño estima con igual precisión la VA y VD e incluso permite conocer si existe sobredominancia mediante $F = CMtF_2/CMF_2$. Sin embargo, no permite estimar interacciones intergénicas, pero esto sería posible si incluimos como tercer probador la F₁. Este diseño se denomina 'triple test cross' y fue propuesto por Kersey y Jinks (1968). Al igual que los diseños I y II, el diseño III se suele evaluar en varios ambientes (Tablas, 2.18 y 2.19).

Tabla 2.18. Análisis de varianza del diseño III de Carolina del Norte evaluado en varios ambientes

Fuentes de variación	gl	CM	CM esperados
Ambientes	e-1		
Repeticiones(amb)	(r-1)e		
Probador	1	CMt	
Probador × amb	1(e-1)	CMtE	
F ₂	n-1	CMF ₂	$\sigma^2 + 2r\sigma^2F_2e + 2re\sigma^2F_2$
Probador × F ₂	n-1	CMtF ₂	$\sigma^2 + r\sigma^2tF_2e + re\sigma^2tF_2$
F ₂ × amb	(n-1)(e-1)	CMF ₂ E	$\sigma^2 + 2r\sigma^2F_2e$
Probador × F ₂ × amb	(n-1)(e-1)	CMtF ₂ E	$\sigma^2 + r\sigma^2tF_2e$
Error	(r-1)(2ne - 1)	CMe	σ^2
Total	2ner-1		

CM = cuadrado medio, CMe = cuadrado medio del error, r = repeticiones, t = probadores, E y e = ambientes, σ^2 = varianza, n = individuos

Tabla 2.19. Estimaciones de parámetros utilizando el diseño III de Carolina del Norte evaluado en varios ambientes

Párametros	Estimaciones	
σ^2F_2	$(CMF_2 - CMF_2E)/2re$	¼ VA
σ^2tF_2	$(CMtF_2 - CMtF_2E)/re$	VD
σ^2F_2e	$(CMF_2E - CMe)/2r$	¼ VAE
σ^2tF_2e	$(CMtF_2E - CMe)/r$	VDE
VA	$4 \sigma^2F_2$	
VD	$4 \sigma^2tF_2$	
VAE	$4 \sigma^2F_2e$	
VDE	$4 \sigma^2tF_2e$	
h ² basada en medias	$4 \sigma^2F_2 / (\sigma^2/re + \sigma^2tF_2e/e + \sigma^2tF_2 + 4 \sigma^2F_2e + 4 \sigma^2F_2)$	

CM = cuadrado medio, CMe = cuadrado medio del error, VA = varianza aditiva, VD = varianza dominante, VAE = varianza aditiva×ambiente, VDE = varianza dominante × ambiente, r = repeticiones, t = probadores, E y e = ambientes, σ^2 = varianza

2.1.3.2.5. Estimación de varianzas. Dialelos

Sin duda alguna es el diseño más utilizado por los mejoradores aunque en muchos casos no se utiliza para estimar componentes de las varianzas. Para obtener el dialelo completo es necesario que n genotipos se crucen en todas las combinaciones posibles incluyendo recíprocos. Así se obtendrán n^2 familias de hermanos completos. Evidentemente el poder obtener un dialelo completo depende mucho del tipo de organismo con el que trabajemos. Por ejemplo, si la especie con la que trabajamos es maíz podemos obtener n líneas puras procedentes de una población F_2 . Las líneas puras las podemos cruzar entre sí obteniendo los cruzamientos en los dos sentidos y autofecundando las líneas para obtener lo que equivale al cruzamiento de una línea consigo misma. Es decir se pueden obtener las n^2 familias.

También se puede trabajar con dialelos parciales. Griffing (1956) describió cuatro métodos experimentales posibles dependiendo del número y tipo de familias que se evalúen. Así si partimos de p líneas puras tenemos los métodos:

- . Método 1. Padres, F_{1S} y recíprocos. p^2 combinaciones
- . Método 2. Padres y F_{1S} , recíprocos no. $\frac{1}{2} p (p+1)$ combinaciones
- . Método 3. F_{1S} y recíprocos pero no padres. $p (p-1)$ combinaciones
- . Método 4. F_{1S} ni padres ni recíprocos. $\frac{1}{2} p (p-1)$ combinaciones

Asimismo tenemos dos modelos:

- Modelo fijo. Los padres del dialelo son factores fijos y se estiman aptitudes combinatorias generales y específicas. La aptitud combinatoria general (ACG) se define como el comportamiento medio de una línea en sus cruzamientos con otras líneas y la aptitud combinatoria específica (ACE) como la desviación del valor de un cruzamiento de la media de la ACG de sus padres.
- Modelo aleatorio. Los padres son un conjunto de líneas derivadas al azar de una población. Se estiman los componentes de la varianza de la población de referencia

Para una mejor comprensión estudiaremos el método 4. Partimos de p padres y tendremos $p(p-1)/2$ cruzamientos posibles. Si estos cruzamientos los evaluamos en un solo ambiente con r repeticiones utilizaremos el análisis de varianza de la Tabla 2.20.

Tabla 2.20. Análisis de varianza de un dialelo

Fuentes de variación	Gl	CM	CM esperados
Repeticiones	$r-1$		
Cruzamientos	$p(p-1)/2 - 1$	CMc	$\sigma^2 + r\sigma^2c$
Error	$(r-1)(p(p-1)/2 - 1)$	CMe	σ^2

CM = cuadrado medio, gl = grados de libertad, r = repeticiones, p = padres, c = cruzamientos, e = error

Podemos dividir ortogonalmente la suma de cuadrados de los cruzamientos. Por sencillez en la Tabla 2.21 consideramos 5 padres.

Tabla 2.21. Ejemplo de los cruzamientos posibles utilizando 5 padres

Padres	Padres					medias
	1	2	3	4	5	
1		X12	X13	X14	X15	X1.
2			X23	X24	X25	X2.
3				X34	X35	X3.
4					X45	X4.
5						X5.

Si consideramos 5 padres tenemos 9 grados de libertad en la fuente de variación de cruzamientos. Los podemos dividir en 4 grados de libertad para las medias $(p-1)$ y 5 grados de libertad para la variación entre celdas dentro de medias, ya que si conocemos las medias y 3 celdas del padre 1 y dos celdas del padre 2 podríamos

estimar el resto. Otra forma de explicarlo sería: si ponemos las restricciones de que las desviaciones de cada padre tienen que sumar 0 y por cada restricción restamos un grado de libertad, obtenemos que 10 celdas – 5 padres = 5 grados de libertad. Así el nuevo ANOVA se muestra en la Tabla 2.22 para el modelo aleatorio y en la Tabla 2.23 para el modelo fijo.

Tabla 2.22. Análisis de varianza de un dialelo. Modelo aleatorio

Fuentes de variación	gl	CM	CM esperados
Repeticiones	r-1		
Cruces	$P(p-1)/2 - 1$	CMc	$\sigma^2 + r\sigma^2c$
Entre media	p - 1	CMacg	$\sigma^2 + r\sigma^2s + r(p-2)\sigma^2acg$
Entre celdas/m	$p(p-3)/2$	CMace	$\sigma^2 + r\sigma^2ace$
Error	$(r-1)(p(p-1)/2 - 1)$	CMe	σ^2

CM = cuadrado medio, gl = grados de libertad, r = repeticiones, p = padres, c = cruzamientos, acg = aptitud combinatoria general, ace = aptitud combinatoria específica

Tabla 2.23. Análisis de varianza de un dialelo. Modelo fijo

Fuente de variación	gd	CM	CM esperados
Repeticiones	r-1		
Cruces	$p(p-1)/2 - 1$	CMc	$\sigma^2 + rk^2c$
GCA Entre media	p - 1	CMg	$\sigma^2 + (r(p-2)/(p-1))k^2g$
SCA Entre celdas/m	$p(p-3)/2$	CMs	$\sigma^2 + (2r/p(p-3))k^2s$
Error	$(r-1)(p(p-1)/2 - 1)$	CMe	σ^2

CM = cuadrado medio, gl = grados de libertad, r = repeticiones, p = padres, c = cruzamientos, acg = aptitud combinatoria general, ace = aptitud combinatoria específica

Tanto en el modelo aleatorio como en el modelo fijo podemos saber si la ACG y la ACS son significativas utilizando el test de la F, y podemos estimar los efectos para el modelo fijo. Así: $X_{ijk} = \mu + r_k + g_i + g_j + s_{ij} + e_{ijk}$; donde μ = media, r_k = efecto de la repetición, g_i y g_j = ACG, s_{ij} = ACE, y e_{ijk} = error experimental. Podemos calcular las Estimaciones que se presentan en la Tabla 2.24.

Tabla 2.24. Estimaciones de parámetros utilizando un dialelo

Parámetros	Estimaciones	
g_i	$(1/p(p-2))(pX_{i.} - 2X_{..})$	M fijo
s_{ij}	$X_{ij} - (1/(p-2))(X_{i.} + X_{j.}) + (2/(p-1)(p-2))X_{..}$	M fijo
Varianza σ^2	CMe	
Varianza de un cruce	σ^2/r	
Varianza cruces con un padre común	$\sigma^2/(r(p-1))$	
Varianza de g_i	$(p-1)/(p(p-2))\sigma^2$	
Varianza de s_{ij}	$((n-3)/(n-1))\sigma^2$	
VA si F = 0	$4\sigma^2acg$	M aleatorio
VD si F = 0	$4\sigma^2ace$	M aleatorio
VA si F = 1	$2\sigma^2acg$	M aleatorio
VD si F = 1	$2\sigma^2ace$	M aleatorio

g_i o acg = aptitud combinatoria general, s_{ij} o ace = aptitud combinatoria específica, σ^2 = varianza, p = padres, CMe = cuadrado medio del error

Los otros métodos de dialelo se explican con detalle en Griffing (1956). Cabe destacar que Gardner y Eberhart (1966) diseñan un dialelo especialmente útil para analizar poblaciones y sus cruzamientos.

2.1.3.2.6. Estimación de varianzas: Distintas generaciones

En las generaciones no segregantes P_1 , P_2 y F_1 los individuos son genéticamente idénticos, por lo tanto toda la variabilidad que podamos encontrar es ambiental. Las varianzas de estas generaciones nos permiten estimar estimaciones independientes de la varianza ambiental (VM). Si las varianzas son homogéneas y calculadas sobre el

mismo número de individuos podemos estimar VM como el promedio de las tres varianzas. En el caso en que el número de individuos en cada generación fuese diferente se debe de sumar la suma de cuadrados debida a cada generación y dividirla por la suma de los grados de libertad.

En la F₂ la VG = $\{ \frac{1}{4} (m + a)^2 + \frac{1}{2} (m + d)^2 + \frac{1}{4} (m - a)^2 \} - \frac{1}{2} (m + \frac{1}{2} d)^2 = \frac{1}{2} a^2 + \frac{1}{4} d^2$. Si los alelos tienen la misma frecuencia, VG la podemos dividir en el componente genético aditivo (VA*) y en el componente dominante (VD*) (* significa que es la VA y VD en el caso en que los alelos tienen la misma frecuencia). Para un gen, VA* = $\frac{1}{2} a^2$ y VD* = $\frac{1}{4} d^2$. En el caso de muchos genes en los que no hay epistasia ni ligamiento VA* = $\frac{1}{2} \Sigma a^2$ y VD* = $\frac{1}{4} \Sigma d^2$.

En cuanto a Bc₁, para un gen la VG = $\frac{1}{4} a^2 + \frac{1}{4} d^2 - \frac{1}{2} ad$; Para k genes VG = $\frac{1}{2} VA^* + VD^* - VAD$ siendo VAD = $\frac{1}{2} (\delta_{1a_1d_1} + \delta_{2a_2d_2} + \dots + \delta_{ka_kd_k})$ donde $\delta_i = +1$ si el alelo está en P₁ y $\delta_i = -1$ si no lo está. Por lo tanto VAD es el producto de los efectos aditivos y dominantes que están segregando en el cruce. Por ende, en Bc₂, VG = $\frac{1}{2} VA^* + VD^* + VAD$.

Por definición VA, VD y VM no pueden ser negativas ya que son varianzas, se pueden obtener valores negativos por errores de muestreo pero no deberían diferir significativamente de cero. VAD es una covarianza y su signo dependerá de la dirección de la varianza.

El mejorador está interesado en conocer si el cruzamiento muestra varianza genética significativa y que proporción es heredable. Esto lo podemos saber si comparamos las varianzas de las generaciones segregantes y las que no los son. Si la varianza

combinada de las generaciones P₁, P₂ y F₁ es significativamente menor que las varianzas de la F₂ y Bc indicarían que la varianza genética es significativa para el carácter que estamos estudiando en este cruzamiento. Para comparar las varianzas calculamos las F VF₂/VM, Bc₁/VM y Bc₂/VM. Sí la varianza genética es significativa entonces podemos calcular: VA* = 2VF₂- VBc₁ -VBc₂; VD* = VBc₁ – VBc₂ – VF₂ – VM y VAD= ½ (VBc₁ –VBc₂).

Al igual que ocurre en el diseño de medias generacionales, podemos buscar el modelo que mejor se ajusta. En la Tabla 2.25. Se indican las varianzas de las generaciones en función de VM, VA*, VD* y VAD.

Tabla 2.25 Varianza de las distintas generaciones

Generación s ²	Parámetros			
	VM	VA*	VD*	VAD
P ₁	1	0	0	0
P ₂	1	0	0	0
F ₁	1	0	0	0
F ₂	1	1	1	0
Bc ₁	1	½	1	-1
Bc ₂	1	½	1	1

VM = varianza ambiental, VA = varianza aditiva, VD = varianza dominante, VAD = varianza aditiva× dominante

Si las varianzas VM no son iguales en las generaciones no segregantes entonces tenemos que estimar más varianzas. VA* no se ve afectada por la dispersión genes. Por lo tanto no hay una relación directa entre el efecto aditivo ([a]) y VA*, por ejemplo [a] puede ser muy pequeño y VA* grande debido a la dispersión génica. Por otra parte [a] puede ser grande y VA* pequeño cuando en el carácter está regido por muchos genes con efectos pequeños. Este resultado ocurre cuando trabajamos con un solo

cruzamiento sin embargo [a] y VA tendrán valores similares (en la misma dirección) cuando se interpretan simultáneamente varios cruzamientos. Por otra parte, si tenemos un efecto dominante ([d]) significativo así como VD* significativa sabemos que el carácter tiene dominancia direccional (para los alelos + si [d] es positivo y para los alelos - si [d] es negativo). Por el contrario no hay dominancia cuando tanto el efecto como la varianza no son significativos. Por último, si [d] es pequeño pero significativo y la varianza no es significativa se puede decir que hay poca dominancia.

Además de estimar las varianzas utilizando las seis generaciones de un diseño de medias generacionales también podemos utilizar las familias F₃. Si se autofecunda una muestra al azar de n individuos F₂ obtenemos n familias F₃ si evaluamos para un carácter k individuos de cada familia F₃, usando un diseño completamente aleatorizado. Podríamos llevar a cabo el análisis de varianza de la Tabla 2.26 y estimar los parámetros como se indica en la Tabla 2.27.

Tabla 2.26. Análisis de varianza de un diseño de familias F₃

Fuente de variación	gl	CM	CM esperados
Entre familias	n-1	CMe	$\sigma^2e + r\sigma^2d$
Dentro de familias	n(k-1)	CMd	σ^2d
Total	nk -1		

gl = grados de libertad, CM = cuadrado medios, n = familias F₃, d = dentro de familias, e = entre familias

Tabla 2.27. Estimaciones de parámetros utilizando familias F₃

Parámetros	Estimaciones	
σ^2d	CMd	$\frac{1}{2} VA^* + \frac{1}{2} VD^* + VM$
σ^2e	$(CMe - CMd)/r$	$VA^* + \frac{1}{4} VD$
VA*	σ^2e	Si $VD^* = 0$
VM	$\sigma^2d - \frac{1}{2} VA^*$	Si $VD^* = 0$

CM = cuadrado medio, d = dentro de familias, e = entre familias, VA = varianza aditiva, VD = varianza dominante, r = repeticiones, VM = varianza ambiental, σ^2 = varianza

Si se siguen avanzando en generaciones de autofecundación de tal forma que consideramos n familias F₂ de las cuales se han autofecundado n' plantas F₃ de cada familia F₂ para obtener nn' familias F₄, si evaluamos k individuos de cada familia F₄ entonces tendríamos el análisis de varianza expuesto en la Tabla 2.28 y podríamos estimar los parámetros como se indica en la Tabla 2.29.

Tabla 2.28. Análisis de varianza de la evaluación de F₂, F₃ y F₄

Fuente de variación	gl	CM	CM esperados
Entre F ₂	n-1	CM1	$\sigma^2_3 + r\sigma^2_2 + n'r\sigma^2_1$
Entre F ₃ dentro F ₂	n(n'-1)	CM2	$\sigma^2_3 + r\sigma^2_2$
Entre F ₄ dentro de F ₃	nn'(k-1)	CM3	σ^2_3
Total	nn'k - 1		

n = familias F₃, n' = familias F₄ de cada F₃, k = individuos de cada familias F₄, CM = cuadrado medio, σ^2 = varianza

Tabla 2.29. Estimaciones de parámetros utilizando familias F₃

Parámetros	Estimaciones	
σ^2_1	$(CM1-CM2)/n'r$	$VA^* + 1/16 VD^*$
σ^2_2	$(CM2-CM3)/r$	$1/2 VA^* + 1/8 VD^*$
σ^2_3	CM3	$1/4 VA^* + 1/4 VD^* + VM$

CM = cuadrado medio, d = dentro de familias, e = entre familias, VA = varianza aditiva, VD = varianza dominante, r = repeticiones, VM = varianza ambiental, σ^2 = varianza, n = familias

Observamos que tenemos tres ecuaciones para estimar tres parámetros VA, VD y VM, por lo tanto sí podemos estimar la VD. No obstante las generaciones de autofecundaciones no son útiles para estimar VD porque cada generación de autofecundación reduce a la mitad los heterocigotos que son donde se manifiesta la dominancia. Es común que la VD en estos casos no sea significativa.

De una forma general podemos calcular la relación entre las varianzas de las familias y los componentes genético y ambiental para sucesivas generaciones de autofecundación (Tabla 2.30)

Tabla 2.30. Componentes de varianza en las distintas generaciones

Nivel	F ₂	F ₃	F ₄	F ₅
1	VA* + VD* + VM	VA* + ¼ VD*	VA* + 1/16 VD*	VA* + 1/64 VD*
2		½ VA* + ½ VD* + VM	½ VA* + 1/8 VD*	½ VA* + 1/32 VD*
3			¼VA* + ¼ VD* + VM	¼ VA* + 1/16VD*
4				1/8VA* + 1/8VD* + VM

VA = varianza aditiva, VD = varianza dominante, VM = varianza ambiental

Se observa que dentro de niveles el coeficiente de VA permanece constante mientras que el coeficiente de la varianza dominante se reduce en ¼. Los primeros componentes genéticos de cada nivel son exactamente la mitad de los primeros componentes del nivel anterior. Por lo tanto en cada autofecundación la variación genética de las familias se reduce a la mitad. Al igual que ocurre con el diseño de medias generacionales con más generaciones se puede estimar la varianza epistática. Si seguimos autofecundando conseguiremos un conjunto de líneas completamente homocigóticas que suelen denominarse líneas recombinantes. También se pueden obtener un conjunto de líneas homocigóticas utilizando la técnica de dobles haploides pero hay que tener en cuenta que mientras que para obtener las primeras ha habido

recombinación en cada una de las generaciones de autofecundación, en las líneas dihaploides sólo se ha producido una recombinación.

2.1.3.2.7. Métodos de estimación de los componentes de varianza

La estimación de componentes de varianza se puede realizar por los métodos de momentos (ANOVA), estimación cuadrática insesgada de mínima varianza (MIVQUE), máxima verosimilitud (ML) y máxima verosimilitud restringida (REML). Tradicionalmente, los mejoradores de plantas han estimado los componentes de varianza basados en el análisis de varianza de mínimos cuadrados (ANOVA), igualando los cuadrados medios observados a sus esperanzas matemáticas y resolviendo el sistema de ecuaciones resultantes para los diferentes componentes de varianza. De hecho los diseños explicados anteriormente están expuestos para que se pueda aplicar este método ya que los estimadores basados en el método ANOVA aplicado a conjuntos de datos balanceados poseen una serie de características que los hace muy interesantes: no están sesgados y tienen la mínima varianza entre todos los estimadores no sesgados. El problema surge cuando los datos no son balanceados por ejemplo cuando tenemos datos perdidos.

La estimación por el método de MIVQUE son funciones de los datos y de los valores previos, y son de varianza mínima solamente cuando los valores previos son iguales a los valores verdaderos de los componentes de varianza. Los estimadores de los componentes de varianza obtenidos por los métodos de ML Y REML tienen propiedades asintóticas que los hacen preferibles sobre los estimadores obtenidos con otros métodos. Cuando el tamaño de muestra es considerablemente grande los estimadores de ML Y REML son consistentes y no requieren de conjuntos de datos balanceados para mantener estas propiedades; además permiten el establecimiento de intervalos de confianza y pruebas de hipótesis acerca de los componentes de la varianza. Para tamaños de muestra pequeños los estimadores de máxima verosimilitud

son generalmente sesgados, puesto que al estimar los efectos aleatorios no se toma en cuenta la pérdida en grados de libertad que resulta de la estimación de los efectos fijos. Para solucionar este problema, el método REML, utiliza combinaciones lineales de los elementos del vector de datos de tal manera que esas combinaciones no contienen efectos fijos. Esas combinaciones generan una función de verosimilitud que no depende de los efectos fijos, posteriormente la función de verosimilitud restringida se maximiza con respecto a cada uno de los componentes de varianza. Los valores que maximizan la función son los respectivos estimadores de REML de los componentes de varianza y covarianza.

Para conjuntos de datos completamente balanceados y cuando no hay estimaciones negativas de los componentes de varianza los estimadores de componentes por el método ANOVA son idénticos a los de REML. Sin embargo, en conjuntos de datos no balanceados el método REML es más adecuado siempre y cuando el tamaño de muestra sea elevado (Galán et al., 2003).

2.1.3.2.8. Varianzas genéticas. Estimaciones experimentales

Existe un gran número de artículos en los que se estiman la varianza genética y la partición en sus componentes (utilizando como criterio de búsqueda en la ISI Web Knowledge 'componentes de la varianza genética' se encontraron, en abril de 2012, más de 6.000 citas) en todos los cultivos y para todo tipo de caracteres. A modo de ejemplo comentaremos algunos artículos recientes.

El guandú (*Cajanus cajan* L.) es un vegetal muy apreciado en el Caribe. En el pasado, los programas de mejora se centraron fundamentalmente en la producción despreocupándose de los caracteres relacionados con la calidad. No obstante antes de iniciar un programa de mejora para estos caracteres Beekham y Humaharan (2010)

estudiaron su herencia utilizando un dialelo y un diseño II de Carolina del Norte. La varianza aditiva es el componente más importante para las características físicas excepto para el peso de semilla donde también los efectos no aditivos tienen importancia. Por contra en los caracteres bioquímicos (azúcares, almidón, proteínas) los efectos no aditivos son los más importantes.

El diseño I se utilizó para ver como evolucionaban las varianzas genéticas tras tres ciclos de selección para mejorar la resistencia del maíz a la plaga del taladro (Sandoya et al., 2009) encontrando que todavía existe varianza aditiva para continuar con el proceso de selección. La herencia de la resistencia a plagas también fue el objetivo del trabajo de Ojwang et al. (2011), en concreto se estudió la resistencia de la judía a la mosca de los sembrados utilizando un dialelo 8 × 8 encontrando que, para la mayor parte de los caracteres de daño, la aptitud combinatoria general era más importantes que la específica, aunque no se puede rechazar el papel de los efectos no aditivos en la herencia de la resistencia.

El rendimiento ha sido uno de los principales objetivos en la mayor parte de los programas de mejora en todos los cultivos del mundo pero cada vez más se buscan características que le den un nuevo valor añadido al cultivo, tal es el caso de las frambuesas donde se intentan mejorar los caracteres saludables del fruto (fenoles, actividad antioxidante, antocianos). No obstante antes es necesario conocer la herencia de estos caracteres. Así, utilizando un diseño de hermanos completos Stephens et al. (2009) estudiaron la herencia de estos caracteres bioquímicos y el rendimiento, sugiriendo que si mejoramos por rendimiento podemos reducir la cantidad de sustancias beneficiosas.

También en maíz se estudió la herencia del rendimiento y de procesos fisiológicos relacionados con él. En este caso se utilizó un diseño II que no fue efectivo para diseccionar cuantitativamente el rendimiento en distintos mecanismos fisiológicos (Lee et al., 2005). Esto puede ser debido a que los efectos genéticos significativos son pequeños o que las diferencias entre híbridos se dividieron en cuatro aptitudes combinatorias generales y dos aptitudes combinatorias específicas por lo que cada uno de ellos es un efecto muy pequeño.

Por último un ejemplo de utilización del diseño III en la estimación de varianzas genéticas lo encontramos en Wolf et al. (2000) que estudia la herencia del rendimiento en una población F_2 en maíz

2.2. Selección

2.2.1. Introducción

Estimar la respuesta genética a la selección, antes de realizarla, permite valorar si la mejora que se espera conseguir compensa el esfuerzo que es necesario realizar. También permitiría elegir entre diferentes poblaciones de partida aquella en la que se obtendría mayor respuesta o elegir el método más eficaz para realizar la mejora. Dentro de la respuesta a la selección hay que distinguir entre la respuesta a corto-medio plazo, hasta 10-15 generaciones, y la respuesta a largo plazo, a partir de 20 generaciones. La selección genética depende de la correlación entre el fenotipo y el genotipo. Es decir, si no hubiera correlación no sería posible hacer la selección y, por el contrario cuando la correlación es completa la efectividad de la selección es máxima. Este es el caso de los caracteres monogénicos, es decir, que dependen de un solo gen y en los que el efecto ambiental no enmascara los efectos genéticos. En estos caracteres el valor del carácter de todos los individuos portadores de un alelo determinado no se solapa con el valor del carácter de todos los individuos portadores de otro alelo

diferente. La selección es sencilla: se eligen los individuos con el valor más adecuado del carácter y tenemos la garantía de que los descendientes también tendrán dicho valor. Sin embargo, en los caracteres poligénicos la distribución es continua y, si consideramos un gen cualquiera, el valor de los individuos portadores de un alelo determinado se solapa con los valores de los individuos portadores de otro alelo.

2.2.2. Respuesta a la selección a corto plazo

2.2.2.1. Especies autóгамas.

El método más eficaz de selección en una especie autógamma en la que se parte del cruzamiento de líneas es el método genealógico porque este método utiliza tanto la variación entre familias como la variación dentro de familias. Para comprender la eficacia de este método es muy útil la división de la varianza aditiva y dominante a largo de los ciclos de selección (Tabla 2.31).

Tabla 2.31. Varianza genética en diferentes generaciones de autofecundación.

Generación	Entre familias		Dentro de familia	
	V_A^*	V_D	V_A	V_D
1	1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{3}{2}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{4}$
3	$\frac{7}{4}$	$\frac{7}{64}$	$\frac{1}{8}$	$\frac{1}{8}$
4	$\frac{15}{8}$	$\frac{15}{256}$	$\frac{1}{16}$	$\frac{1}{16}$
Infinito	2	0	0	0

* V_A = varianza aditiva, V_D = varianza dominante.

En primer lugar observamos que la varianza aditiva es siempre mayor entre familias que dentro de familias por lo que siempre la selección entre familias es de esperar que sea más efectiva que la selección dentro de familias. Además, la diferencia entre la varianza aditiva de la selección entre familias y la varianza aditiva de la selección

dentro de familias se va incrementando con las generaciones de autofecundación por lo que cuanto más avanzan las generaciones de autofecundación más efectiva es la selección entre familias con respecto a la selección dentro de familias. Es obvio que cuando se alcanza la fijación u homocigosis completa ya no es posible realizar la selección dentro de familias porque ya no hay variación genética. A su vez también vemos que, dado que en la selección entre familias la varianza aditiva aumenta con las generaciones de autofecundación, también aumenta la efectividad de la selección que prácticamente se dobla en 4 generaciones de autofecundación (de 1 a 1,9). La ventaja de la selección entre familias con respecto a la selección dentro de familias es mayor cuando el carácter tiene una heredabilidad baja (gráfica 8.1 en Bernardo, 2002).

A la hora de seleccionar un carácter, aunque sea poligénico, y no seamos capaces de identificar genes individuales, el objetivo es combinar todos los alelos favorables de los padres en una única línea. Si un carácter está controlado, por ejemplo, por 100 genes que están segregando entre dos líneas parentales, la probabilidad de obtener un número determinado de alelos favorables solamente por azar se puede calcular mediante la distribución binomial. En la Tabla 2.32 se muestra la probabilidad de obtener 50 o más alelos favorables, 55 o más alelos favorables, etc.

Tabla 2.32. Probabilidad de obtener un número determinado de alelos favorables, de un total de 100 genes, de acuerdo con la distribución binomial.

Número de alelos favorables	Probabilidad
50	$4,6 \cdot 10^{-1}$
55	$1,4 \cdot 10^{-1}$
60	$1,8 \cdot 10^{-2}$
65	$9,0 \cdot 10^{-4}$
70	$1,6 \cdot 10^{-5}$
75	$9,1 \cdot 10^{-8}$
100	$7,9 \cdot 10^{-31}$

Para que pueda actuar la selección primero tenemos que tener los genotipos con las combinaciones de genes apropiadas sobre los que actuar. La probabilidad de la tabla nos permite conocer el número de individuos que necesitaríamos obtener para conseguir un genotipo con un número determinado de alelos favorables. Por ejemplo, si queremos conseguir dos genotipos con al menos 60 alelos favorables ($P = 0,02$, aproximadamente) necesitaríamos alrededor de 100 individuos. Lo que nos indica la tabla es que para conseguir genotipos con 70 o más alelos favorables los tamaños de población que tenemos que manejar son muy elevados. En poblaciones de tamaño manejable, en ausencia de selección, los individuos que escojamos tendrán alrededor de 50 alelos favorables, aunque en la población habrá individuos de hasta 65 alelos favorables, aunque a una frecuencia baja. La cuestión es si seremos capaces de seleccionar esos individuos o no, o en otras palabras, cuan efectiva será la selección. En primer lugar hay que considerar que en los caracteres poligénicos el efecto de la selección en cada gen individual es pequeño debido a que el efecto del gen también es pequeño. En segundo lugar hay que considerar que debido a la autofecundación el efecto de la deriva es muy elevado lo que interfiere en el efecto de la selección, es decir, la fijación ocurre demasiado rápido, de forma que la selección no tiene tiempo de actuar. Bailey y Comstock (1976) estimaron la probabilidad de fijación de un alelo favorable como $P = 0,5(1 + 0,5 s)$, siendo s el coeficiente de selección. Basado en datos empíricos, Comstock (1974) especula, que, para caracteres cuantitativos, s varía entre 0,088 (cuando se seleccionan 10% de los individuos) y 0,334 (cuando se seleccionan 1 % de los individuos). Con estos valores la probabilidad de fijación aumenta de 0,5 (sin selección) a 0,52 (10% de individuos) o a 0,58 (1% de individuos). Esto nos muestra que la mejora genética de caracteres cuantitativos puede ser lenta o difícil de conseguir. De todos modos hay que puntualizar que datos recientes muestran que la distribución de los efectos génicos en los caracteres cuantitativos no es uniforme, sino que hay unos pocos genes con mayor efecto y progresivamente mayor número de genes con efecto cada vez más pequeño (Dekkers, 2012). La mayor dificultad radicaría en conseguir la fijación de los alelos favorables en estos genes de efecto pequeño.

2.2.2.2. Especies alógamas

En las especies alógamas, si el número de individuos de la población es elevado, no hay deriva y la selección puede actuar de forma más efectiva que en las especies autógamias. Como ya vimos la probabilidad de tener todos los alelos favorables en un individuo de una población es muy pequeña, pero la solución es ir aumentando de forma paulatina la frecuencia de alelos favorables a través de repetidos ciclos de selección y recombinación en lo que se conoce como selección recurrente.

Si consideramos una especie alógama en la que los individuos se aparean al azar y no hay deriva genética, la efectividad de la selección se puede estimar mediante la regresión del valor de los padres y del valor de los hijos. Algunos autores han denominado a esta fórmula o a modificaciones de la misma la ecuación del mejorador. A partir del coeficiente de regresión y el valor de los padres se estima el valor esperado de los hijos, es decir, la respuesta a la selección. Siendo y la variable dependiente y x la variable independiente el coeficiente de regresión b se puede expresar como la covarianza de las dos variables (x e y) dividida por la varianza de x ($b = \text{Cov}_{xy}/\text{Var}_x$). De esta forma la respuesta esperada se puede expresar de forma equivalente como $R = S \text{Cov}_{PH}/\text{Var}_P$, donde R es la media de los hijos y S la media de los padres seleccionados estando los valores ajustados por la media de la población sin seleccionar. A su vez Cov_{PH} es la covarianza de padres e hijos y Var_P es la varianza de los padres. Este método de estimar la respuesta a la selección puede ser generalizado considerando diferentes unidades de selección (individuos, familias de medios hermanos, etc.), siendo entonces $R = S \text{Cov}_{SM}/\text{Var}_S$. Cov_{SD} es la covarianza entre los individuos seleccionados y los individuos de la población mejorada y Var_S la varianza de los individuos seleccionados. La covarianza entre parientes puede ser estimada, en determinados casos, en la población de partida como combinación lineal de componentes de varianza, fundamentalmente varianza aditiva y dominante, mediante una serie de diseños que han sido descritos en apartados anteriores. De esta forma podemos estimar la respuesta esperada con distintos métodos de selección y elegir el

más adecuado para nuestra población. A la hora de elegir el método más adecuado también hay que tener en cuenta si el objetivo de la selección es la mejora de las poblaciones *per se* (mejora intrapoblacional) o la mejora del cruzamiento de poblaciones (mejora interpoblacional). En la Tabla 2.33 se muestran las ganancias teóricas esperadas de diferentes métodos de selección en función de las varianzas en las poblaciones base.

Tabla 2.33. Predicción teórica de la respuesta a la selección con diferentes esquemas de selección (Sprague y Eberhart, 1977).

	Ganancia esperada*
Selección intrapoblacional	
Masal	$K(1/2)V_A/\sqrt{V_w + V_{AE} + V_{DE} + V_A + V_D}$
Medios hermanos	$K(1/4)V_A/\sqrt{(V_w/rm + 1/4V_{AE}/m + 1/4V_A)}$
Hermanos completos	$K(1/2)V_A/\sqrt{(V_w/rm + (1/2V_{AE} + 1/4V_{DE})/m + (1/2V_A + 1/4V_D))}$
Familias S ₁	$KV_A/\sqrt{(V_w/rm + (V_{AE} + 1/4V_{DE})/m + (V_A + 1/4V_D))}$
Familias S ₂	$K(3/2)V_A/\sqrt{(V_w/rm + (3/2V_{AE} + 3/16V_{DE})/m + (3/2V_A + 3/16V_D))}$
Selección interpoblacional	
Medios hermanos	$(K(1/4)V_{A(1)}/\sqrt{V_{w(1)}/rm + 1/4V_{AE(1)}/m + 1/4V_{A(1)}}) + (K(1/4)V_{A(2)}/\sqrt{V_{w(1)}/rm + 1/4V_{AE(2)}/m + 1/4V_{A(2)}})$
Hermanos completos	$K(1/2)V_A/\sqrt{(V_w/rm + (1/2V_{AE} + 1/4V_{DE})/m + (1/2V_A + 1/4V_D))}$

*V_A = varianza aditiva, V_D = varianza dominante, V_W = varianza del error dentro de ambiente, V_{AE} y V_{DE} = interacción aditiva y dominante por ambiente, k = selección diferencial estandarizada, r = número de repeticiones dentro de ambiente y m = número de ambientes. En la selección interpoblacional los componentes de varianza son diferentes para las dos poblaciones y se definen en función del cruzamiento de poblaciones.

En la población Iowa Stiff Stalk Synthetic se estimaron las varianzas genéticas y no genéticas del carácter rendimiento que sirvieron para predecir la respuesta a la selección (Tabla 2.34).

Tabla 2.34. Predicción de la ganancia a la selección con distintos métodos de selección en la población Iowa Stiff Stalk Synthetic (datos de Hallauer y Miranda, 1988)

Método de selección	Incremento del rendimiento (t ha ⁻¹)	
	Por año	Por ciclo
Masal	0,054	0,054
Medios hermanos	0,175	0,350
Hermanos completos	0,225	0,451
Familias S ₁	0,361	0,722
Familias S ₂	0,328	0,985

De acuerdo con estos datos, el método de selección más adecuado para aumentar el rendimiento de la población sería el de familias S₁. Sin embargo, en otra población en la que las varianzas son diferentes el método idóneo podría ser otro, por lo que para predecir correctamente el progreso de la selección habría que estimar las varianzas independientemente para cada población. Por ejemplo, en la respuesta a la selección de la población BS11 el método más eficaz fue el de familias S₂, no el de familias S₁ (Weyhrich et al., 1998). La estimación de la varianzas de la población base supone un gasto de tiempo y un trabajo considerable de campo adicional al propio proceso de selección. Por ello, algunos autores proponen estimar directamente las varianzas mediante los datos que son generados directamente en los programas de mejora (Bernardo, 2002; Hill, 2011). Con la filosofía de optimizar los recursos uno de los mayores avances en el campo de la predicción de la respuesta a la selección y de la predicción de los genotipos con mejor valor genético (o aditivo) es el modelo animal en los que el fenotipo de cada individuo se define en términos de efectos y la estructura genética es incorporada en las varianzas y covarianzas de estos efectos (Hill, 2010). Los

efectos fijos (ambientales) y aleatorios (genéticos) se incorporan en un modelo mixto en el que los parámetros (p. ej. los valores aditivos) son estimados mediante el método “Best Linear Unbiased Prediction” (Hill, 2012). Estos modelos han sido utilizados con éxito en mejora animal y desde hace relativamente poco tiempo han empezado a ser utilizados en mejora vegetal, siendo un tema de investigación de gran interés como se refleja en los numerosos artículos publicados (p. ej., Piepho et al., 2008, Piepho y Mohring, 2011, etc.).

Mediante el análisis de los factores que intervienen en la fórmula del mejorador es posible valorar que factores puede controlar el mejorador, además del método de selección, para incrementar la respuesta a la selección, aunque para ello es mejor expresar la ecuación de manera ligeramente diferente. Si tenemos en cuenta que la heredabilidad es equivalente al coeficiente de regresión de padres e hijos, entonces $R = ih^2V_P$, donde i es la intensidad de selección que depende de la proporción de individuos seleccionados, h^2 es la heredabilidad y V_P es la varianza fenotípica. Cuanto menor es la proporción de individuos seleccionados mayor es la intensidad de selección y por tanto, la respuesta es mayor. Para disminuir la proporción de individuos seleccionados podemos aumentar el tamaño de la población lo cual presenta el problema de que aumenta el número de individuos que tenemos que evaluar. En este caso, el principal limitante será la posibilidad de evaluar un número elevado de individuos o familias. También podemos mantener el tamaño de población y disminuir el número de individuos seleccionados. Sin embargo, si seleccionamos un número muy bajo de individuos aumenta la deriva genética y la consanguinidad que disminuye la respuesta a la selección a medio o largo plazo. La heredabilidad se puede expresar como el cociente entre la varianza aditiva y la varianza fenotípica. La heredabilidad se puede aumentar disminuyendo la varianza fenotípica. Esto se puede conseguir, en primer lugar, evaluando familias en vez de individuos porque las medias de las familias tienen menor varianza fenotípica que los valores de los individuos. Siendo V_p la varianza fenotípica de las medias de las familias, $V_P = V_{EF} + V_{FAM \times AMB}/a +$

$V_E/ar + V_{DF}/arn$, donde V_{EF} es la varianza entre familias, $V_{FAM \times AMB}$ es la interacción familias por ambiente, V_E es el error ambiental, V_{DF} es la varianza dentro de familias, a es el número de ambientes, r el número de repeticiones dentro de ambiente y n el número de plantas dentro de cada familia. De esta fórmula se concluye que para disminuir la varianza fenotípica en primer lugar se puede disminuir la varianza del error mediante un diseño experimental adecuado (ver Capítulo 4). Además, aumentando el número de ambientes y el número de repeticiones en las que las familias son evaluadas también se consigue una disminución de la varianza fenotípica. De nuevo el limitante es la capacidad de trabajo del mejorador. El último término de la ecuación del mejorador, la varianza fenotípica, se

puede aumentar incrementando el número de individuos de la población de partida o mediante la elección de poblaciones con mayor variabilidad. Algunos autores proponen recombinar distintas poblaciones antes de comenzar la selección para maximizar la varianza y, por tanto, la respuesta a la selección (Kutka y Smith, 2007). Al considerar la recombinación de poblaciones hay que valorar, no solo el incremento de varianza, sino también la posible reducción de la media del compuesto con respecto a la media de alguna de las poblaciones individuales.

2.2.3. Respuesta a la selección a largo plazo

Hay numerosos experimentos que confirman que la ecuación del mejorador predice adecuadamente la respuesta de la selección a corto plazo (Walsh and Lynch, 2010). Sin embargo, la ecuación no es válida cuando el número de generaciones de selección es elevado (más de 10 o 20) porque el valor de h^2 cambia y por tanto la respuesta a la selección ya no es proporcional a la h^2 de la población base. Predecir la respuesta a largo plazo es más complicado porque depende de múltiples factores como el número de genes, frecuencias alélicas, distribución de efectos de los genes, etc. Sin embargo, asumiendo el modelo infinitesimal, es decir, que el carácter está controlado por un

número infinito de genes no ligados de efecto pequeño, Robertson (1960) demostró que la respuesta a largo plazo es $2N_e$ veces la respuesta en la generación inicial, siendo N_e el censo efectivo. El valor de N_e de una generación cualquiera es el número de individuos seleccionados en la generación anterior. Entonces si mantenemos el número de individuos totales evaluados y queremos aumentar N_e , para aumentar la respuesta a largo plazo, tenemos que reducir la intensidad de selección lo que reduce la selección a corto plazo. Es decir, si aumentamos la selección a largo plazo disminuimos la selección a corto plazo. La proporción de individuos seleccionados que da la máxima respuesta a largo plazo es 50%, lo cual supone una respuesta muy pequeña a corto plazo. Desde el punto de vista del mejorador la conclusión práctica es que una intensidad muy fuerte de selección puede limitar la respuesta a la selección en futuras generaciones. Por ejemplo, en los años 20-30 hubo una selección muy fuerte entre variedades de maíz de Estados Unidos y solo unas pocas contribuyeron con líneas puras a las siguientes generaciones. Esto puede suponer un problema a largo plazo y, conscientes de este problema, instituciones públicas y privadas de este país han realizado un esfuerzo muy importante para identificar nuevo germoplasma que permita ampliar la base genética del material utilizado en los programas de mejora (Salhuana y Pollak, 2006). Sin embargo, hay que decir que en casi 100 años de selección intensa en Estados Unidos no hay indicios de una disminución en la respuesta a la selección (Duvick, 2005). Por el contrario, algunos experimentos de selección a largo plazo alcanzaron un límite en el que la respuesta a la selección se detuvo (Hill, 2011) probablemente debido a que la varianza de la población base se había agotado. La diferencia puede ser debida a la mutación que, cuando la varianza inicial se ha agotado, puede generar nueva variabilidad sobre la que la selección puede actuar (Keightley, 2004). Weber (2004) discute inteligentemente la relación entre el tamaño de la población y la respuesta a la selección combinando argumentos teóricos y experimentales. En principio el coeficiente de selección que actúa sobre cada gen individual depende del efecto del gen, pero también del tamaño de la población. Si el tamaño de la población es pequeño el efecto de la deriva es tan grande que contrarresta el efecto de los genes que se comportan como si fueran neutros. Esto lleva a que haya

fijación de genes desfavorables y a que disminuya la respuesta final a la selección. Otra ventaja de las poblaciones grandes es que aumenta el número de mutaciones, simplemente porque hay más individuos, lo que incrementa la varianza debida a la mutación y la respuesta a la selección. Weber (2004) da referencias interesantes en las que se confirma, mediante datos empíricos, que la respuesta a la selección es mayor en las poblaciones grandes y que, a idéntica intensidad de selección, la respuesta es más rápida en dichas poblaciones. El efecto beneficioso de las poblaciones grandes es mayor cuanto más poligénico es el carácter. En uno de los trabajos más completos hasta el momento sobre la selección a largo plazo Burke et al. (2010) estudian el cambio de frecuencias de 24.000 SNPs en poblaciones de 1.000 individuos de *Drosophila* que han experimentado 600 generaciones de selección por desarrollo acelerado. Los autores encontraron que en la población sometida a selección los individuos se desarrollaban un 20% más rápido que en la población control. Además, más de 20.000 SNPs cambiaron significativamente de frecuencia debido a la selección, pero no encontraron evidencia de fijación de ninguno de los alelos.

Como conclusión general hay que decir que, si bien el conocimiento de los caracteres complejos o cuantitativos a nivel de genes concretos está resultando más complicado de lo esperado, en cambio, la mejora de dichos caracteres se puede considerar que ha sido muy efectiva. La genética cuantitativa aporta el conocimiento necesario para llevar a cabo la mejora de los caracteres de manera eficaz y continúa, siendo un área muy activa de investigación, en particular, los diseños y métodos analíticos relacionados con la mejora facilitada por marcadores, como se refleja en el Capítulo 7.

2.3. Referencias

Audilakshmi S, Mall AK, Swarnalatha M, Seetharama N. 2010. Inheritance of sugar concentration in stalk (brix), sucrose content, stalk and juice yield in sorghum. *Biomass. Bioenerg.* 34: 813-820.

Bailey TB, Comstock RE. 1976. Linkage and the synthesis of better genotypes in self-fertilizing species. *Crop Sci.* 16: 363-370.

Beekham A, Umaharan P. 2010. Inheritance and combining ability studies of pod physical and biochemical quality traits in vegetable pigeonpea (*Cajanus cajan* L. Mill sp). *Euphytica* 176: 37-47.

Bernardo R. 2002. *Breeding for Quantitative Traits in Plants*. Stemma Press, Minnesota, EEUU.

Bnejdi F, Saadoun M, El Gazzah M . 2011. Genetic adaptability of the inheritance of the resistance to different levels of aggressiveness of *Septoria tritici* isolates in durum wheat. *Crop Protection* 30:1280-1284.

Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR. 2010. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467: 587-590.

Comstock RE. 1974. Consequences of genetic linkage. pp 353-364. En: Proc. 1st World Cong. Genet. Appl. Anim. Prod., Madrid.

Comstock RE, Robinson HF. 1948. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4: 254-266.

Dashti H, Naghavi MR, TajabadipourA. 2010. Genetic analysis of salinity tolerance in a bread wheat cross. *J. Agr. Sci. Tech.* 12: 347-356.

Dekkers J. 2012. Application of genomics tools to animal breeding. En: B Ordás (ed.), *Genetic Dissection of Complex Traits in the Genomic Era*, Current Genomics special issue (en prensa).

Duvick DN. 2005. Genetic progress in yield of United States maize (*Zea mays* L.). *Maydica* 50: 193-202.

- Eshghi R, Akhundova E. 2009. Genetic analysis of grain yield and some agronomic traits in hulless barley. *Afric. J. Agric.* 4: 1464-1474.
- Fakhfakh MM, Yahyaoui A, Rezgui S, Elias EM, Daaloul A. 2011. Inheritances of Fusarium head blight resistance in a cross involving local and exotic durum wheat cultivars. *Crop Sci.* 51: 2517-2524.
- Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*, 4^a Ed. Longmans Green, Harlow, Essex, UK.
- Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh* 52: 399-433.
- Galán J, Efrán P, Cervantes C. 2003. Estimación por métodos de verosimilitud restringida de componentes de varianza y covarianza de múltiples características bajo los diseños I y II de Carolina del Norte. *Rev. Fitotec. Mex.* 26: 53-66.
- Gardner CO, Eberhart SA. 1966. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* 22: 439-452.
- Ghaderi M, Pahlevani M, Razavi SE. 2011. Inheritance of resistance to resistance to *Pythiummultimum* in safflower determined by generation means analysis. *Aust. J. Crop Sci.* 5: 439-446.
- Griffing B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 9: 463-493.
- Gutierrez OA, Robinson AF, Jenkins JN, McCarty JC, Wubben MJ, Callahan FE, Nichols RL. 2011. Identification of QTL regions and SSR markers associated with resistance to reniform nematode in *Gossypium barbadense* L. accession GB713. *Theor. Appl. Genet.* 122: 271-280.
- Hallauer RA, Miranda JB. 1988. *Quantitative Genetics in Maize Breeding*. Iowa State University, Ames, EEUU.

Hill WG. 2010. Understanding and using quantitative genetic variation. *Phil. Trans. R. Soc. B.* 365: 73-85.

Hill WG. 2011. Can more be learned from selection experiments of value in animal breeding programs? Or is it time for an obituary? *J. Anim. Breed. Genet.* 128: 87-94.

Hill WG. 2012. Quantitative genetics in the genomic era. En: B Ordás (ed.), *Genetic Dissection of Complex Traits in the Genomic Era, Current Genomics special issue* (en prensa).

Lee EA, Ahmadzadeh A, Tollenaar M. 2005. Quantitative genetic analysis of the physiological processes underlying maize grain yield. *Crop Sci.* 45: 981-987.

Keightley PD. 2004. Mutational variation and long-term selection response. *Plant Breed. Rev.* 24: 227-248.

Kersey MJ, Jinks JL. 1968. A general method of detecting additive, dominance and epistatic variation for metrical traits: I theory. *Heredity* 23: 403-409.

Kutka FJ, Smith ME. 2007. How many parents give the highest yield in predicted synthetic and composite population of maize? *Crop Sci.* 47: 1905-1913.

Lyimo HJF, Pratt RC, Mnyuku RSOW. 2011. Heritability and gene effect estimates for components of partial resistance to grey leaf spot of maize by generation mean analysis. *Plant Breed.* 130: 633-639.

Mather K. 1949. *Biometrical Genetics*. London, Methuen & Co.

Mather K, Jinks JL. 1982. *Biometrical Genetics*, 3rd Edition, London: Chapman and Hall.

Ojwang PPO, Melis R, Githiri MS, Songa, JM. 2011. Genetic analysis for resistance to bean fly (*Ophiomyia phaseoli*) and seed yield among common bean genotypes in a semi-arid environment. *Field Crop Res.* 120: 223-229.

- Ordás B, Serrano L, Ordas A, Butron A, Revilla P. 2009. Transition between vegetative phases in maize: genetic effects and variances and associated markers. *J. Agric. Sci.* 147: 547-554.
- Piepho HP, Mohring J. 2011. On estimation of genotypic correlations and their standard errors by multivariate REML using the MIXED procedure of the SAS system. *Crop Sci.* 51: 2449-2454.
- Piepho HP, Mohring J, Melchinger AE, Buchse A. 2008. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161: 209-228.
- Robertson A. 1960. A theory of limits in artificial selection. *Proc. Royal Soc. B* 153: 234-249.
- Salhuana W, Pollak L. 2006. Latin American Maize Project (LAMP) and Germplasm Enhancement of Maize (GEM) project: Generating useful breeding germplasm. *Maydica* 51: 339-355.
- Sandoya G, Malvar RA, Revilla P, Butrón A. 2009. Effects of selection for maize resistance to *Sesamia nonagrioides* on the additive and dominant components of genetic variance. *Plant Breed.* 128: 244-248.
- Sprague GF, Eberhart SA. 1977. Corn breeding. pp 305-362. En: Sprague (ed.), *Corn and Corn Improvement*. 2nd ed., ASA, Madison, EE.UU.
- Stephens MJ, Scalzo J, Alspach PA, Beatson RA, Connor AM. 2009. Genetic variation and covariation of yield and phytochemical traits in a red raspberry factorial study. *J. Am. Soc. Hortic. Sci.* 134: 442-452.
- Walsh B, Lynch M. 2010. Genetics and analysis of quantitative traits. Volume 2: Advances topics in breeding and evolution. http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html.
- Weber K. 2004. Population size and long-term selection. *Plant Breed. Rev.* 24: 249-268.

Weyhrich RA, Lamkey KR, Hallauer AR. 1998. Responses to seven method of recurrent selection in the BS11 maize population. *Crop Sci.* 38: 308-321.

Wolf DP, Peternelli LA, Hallauer, AR. 2000. Estimates of genetic variance in an F-2 maize population. *J. Hered.* 91: 384-391.

Wright S. 1935. The analyses of variance and the correlation between relatives with respect to desviations from an optimun. *J. Genet.* 30: 243-256.