

Apunts de Genètica Quantitativa

J. Ignacio Lucas Lledó

27/9/2022

Introducció

El desenvolupament de la genètica quantitativa en la primera meitat del segle XX va contribuir a l'èxit de les campanyes de millora genètica, les quals, junt amb la mecanització i l'ús d'agroquímics, van possibilitar el major increment de la productivitat agropecuària de la història de la humanitat: la *revolució verda*. Actualment, l'aparell estadístic de la genètica quantitativa es combina amb l'abundància de dades genòmiques per continuar augmentant la productivitat agropecuària. El desenvolupament de mètodes com la selecció genòmica i els estudis d'associació genòmica fan que els conceptes bàsics de genètica quantitativa siguin fonamentals per entendre i participar en el progrés de la biomedicina, la biologia evolutiva i la biotecnologia.

L'esquema a seguir per introduir aquests conceptes és el següent:

1. Un gen. Amb el cas més simple possible es pretén explicar en què consisteix la descomposició del **valor genotípic**, a pesar de què la utilitat d'aquesta descomposició romandrà obscura, de moment.
2. La població. Només un incís per observar que la descomposició del valor genotípic només té sentit en referència a una població concreta. El valor reproductiu d'un individu, així com l'efecte de la dominància no són propietats inherents del gen, sinó del gen i de la població.
3. Anàlisi de la variància genètica. Una primera motivació per haver descomposat el valor genotípic és la partició de la variància genètica en variància additiva i variància de dominància.
4. Més d'un gen. Sense epístasi podem explicar casos d'efectes additius entre gens, com en l'experiment de Herman Nilsson-Ehle amb el color dels grans de blat. Amb epístasi, la cosa es complica.
5. Variància ambiental
6. Heretabilitat

Un gen

Normalment la genètica quantitativa és la genètica dels caràcters **poligènics**, tant si el fenotip es pot mesurar en una escala contínua com si és binari (e.g., malalt o sa). Però imaginem ara que el caràcter determinat per aquest únic locus bial·lèlic es mesura en una escala contínua. Si cap altre gen ni l'ambient influeixen en el caràcter, hi hauria dos (amb dominància completa) o tres fenotips possibles.

El valor del fenotip determinat pel genotip es representa amb el símbol G , i a se l'anomena *valor genotípic*, però es mesura en unitats de **fenotip**. L'escala en què es mesura el fenotip és arbitrària. Per exemple, dues escales útils per als càlculs teòrics són les següents (reproducció de la figura 4.4 de Lynch and Walsh (1998), p. 62):

| | | | |
|-----------------|----------|----------|----------|
| Genotip | B_1B_1 | B_1B_2 | B_2B_2 |
| Valor genotípic | $-a$ | d | $+a$ |

| | | | |
|-----------------|----------|----------|----------|
| Genotip | B_1B_1 | B_1B_2 | B_2B_2 |
| Valor genotípic | 0 | $(1+k)a$ | $2a$ |

En presència de dominància, la relació entre el valor genotípic i el nombre d'al·lels d'un cert tipus no és lineal. Però a Fisher (1918) se li va ocórrer la bona idea de considerar una aproximació lineal a la relació entre el contingut gènic (nombre d'al·lels B_2 , per exemple) i el valor genotípic, G . Aquesta *regressió* permet descomposar conceptualment el valor genotípic en una part atribuïble a l'efecte *additiu* o heretable dels al·lels (el valor genotípic *estimat*, \hat{G}) i una desviació causada per la dominància, δ_{ij} :

$$G_{ij} = \hat{G}_{ij} + \delta_{ij} \quad (1)$$

Una forma d'expressar la regressió és esta:

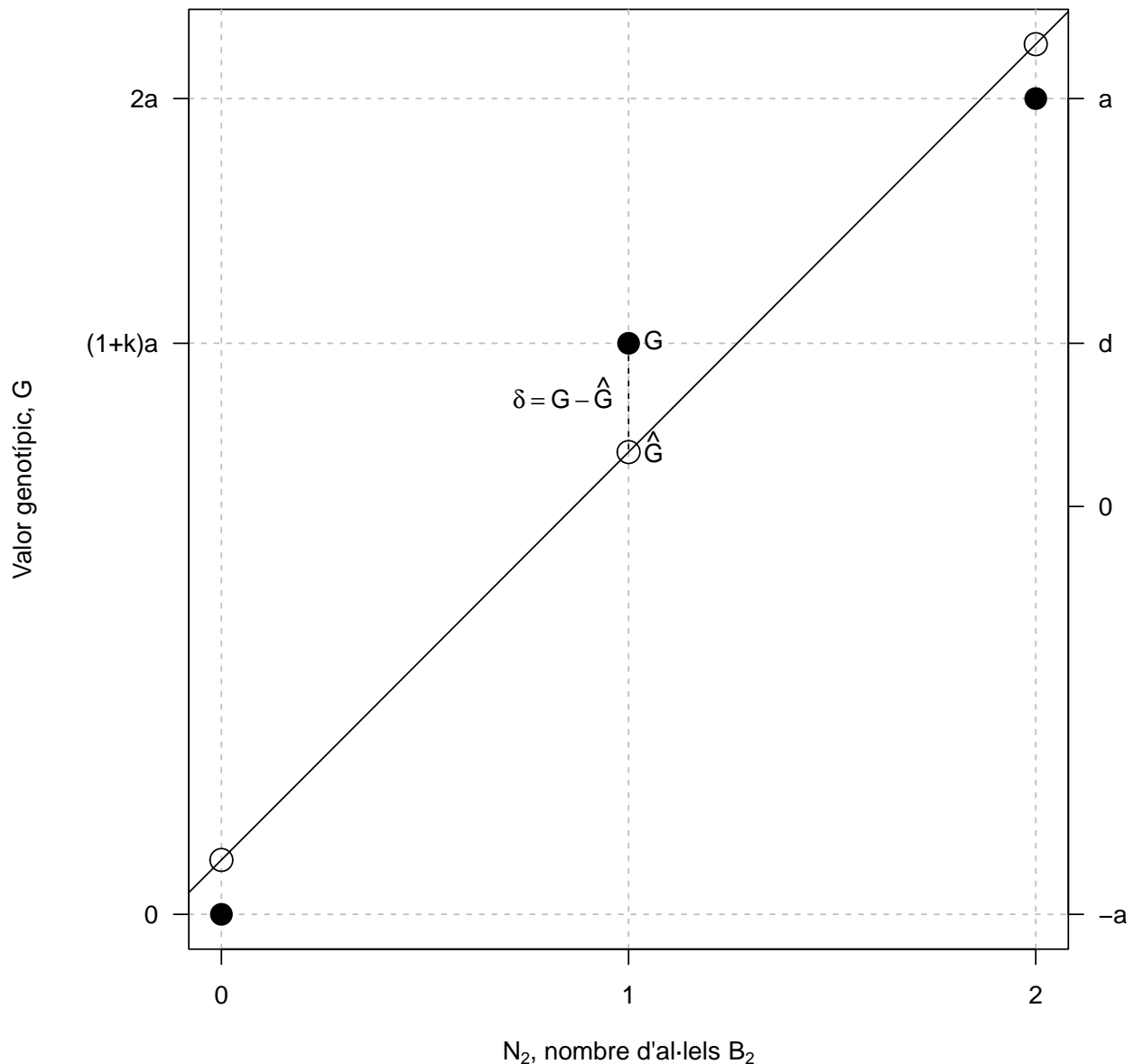
$$G_{ij} = \mu_G + \alpha_1 N_1 + \alpha_2 N_2 + \delta_{ij} \quad (2)$$

On μ_G seria el valor genotípic mig en la població, α_1 i α_2 són els coeficients de regressió parcials, corresponents als efectes dels al·lels d'un tipus (B_1) i de l'altre (B_2). I N_1 i N_2 són el nombre d'al·lels B_1 o B_2 presents al genotip; aquestes són les variables predictores. Els coeficients α_1 i α_2 es coneixen com els **efectes additius**. Si el genotip és diploide, $N_1 = 2 - N_2$, perquè només poden haver dos al·lels. Aprofitant aquesta relació, podem transformar la regressió múltiple en una regressió simple:

$$\begin{aligned} G_{ij} &= \mu_G + \alpha_1(2 - N_2) + \alpha_2 N_2 + \delta_{ij} \\ G_{ij} &= \mu_G + 2\alpha_1 + (\alpha_2 - \alpha_1)N_2 + \delta_{ij} \end{aligned}$$

$$G_{ij} = (\mu_G + 2\alpha_1) + \alpha N_2 + \delta_{ij} \quad (3)$$

Així, el pendent de la regressió és ara $\alpha = \alpha_2 - \alpha_1$, i la intersecció amb l'eix d'ordenades, $\mu_G + 2\alpha_1$, que equival al valor genotípic predit per al genotip B_1B_1 .



(Modificat la figura 4.6 de Lynch i Walsh (1998), pàgina 66.)

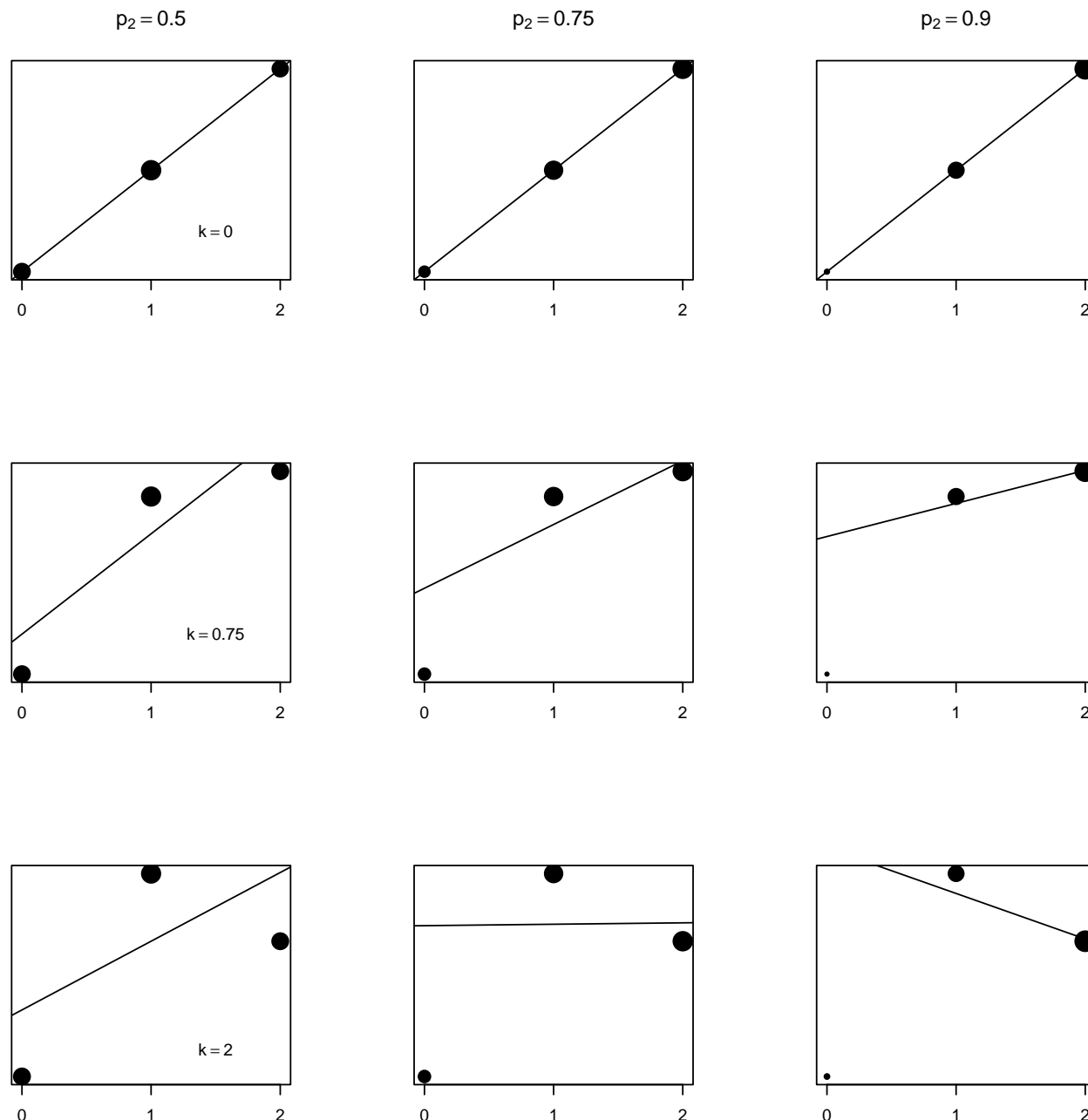
Cabria preguntar-se el per què de tot açò. Òbviament, si sabem el fenotip que correspon a cada genotip, no necessitem fer una regressió! El motiu és que en una població amb aparellament aleatori, cada progenitor transmet un únic al·lel per locus, el qual es combina a l'atzar amb qualsevol altre. En estes condicions, en la descendència es manifesten els efectes additius heretats dels progenitors. En definitiva, la regressió ens permet descomposar l'efecte genotípic en termes que ens ajuden a comprendre l'herència en el context d'una població concreta. Açò és molt important, per exemple, per a la selecció artificial on cal triar els individus que es reproduiran en cada generació i predir quina millora de la productivitat podem esperar.

La població

A la gràfica de l'apartat anterior, la regressió s'ha fet amb només tres punts, com si tinguérem només un individu de cada genotip. Però el valor genotípic mig i el pendent de la regressió, α , depenen del nombre d'individus de cada tipus. A efectes de mostrar com funciona la regressió lineal o la descomposició del valor genotípic, era suficient imaginar tres individus. Però aquesta anàlisi només pot tenir alguna utilitat pràctica si s'aplica a una població concreta. El fet que la composició genètica de la població afecte el resultat de

l'anàlisi ens indica que l'objectiu **no** és descriure una relació funcional general entre el genotip i el fenotip, aplicable a qualsevol individu o població, ni en qualsevol ambient. Una relació funcional d'aquest tipus seria descrita en tot cas per la **norma de reacció**, però no per la descomposició del valor genotípic. L'objectiu és caracteritzar l'efecte de la diversitat genètica d'una població concreta en algun fenotip.

La figura següent reproduïx la figura 4.7 de Lynch and Walsh (1998), i demostra que la descomposició del valor genotípic és una **anàlisi local**, en el sentit de què només és vàlida en una població concreta. La mida dels punts en aquesta gràfica és proporcional al logaritme del nombre d'individus de cada tipus.



El pendent de la regressió, α , és *l'efecte promig de la substitució al·lèlica*. És a dir, representa el canvi promig en el valor genotípic que resultaria quan un al·lel B_1 triat a l'atzar fóra substituït per un al·lel B_2 . Sense dominància, $\alpha = a$ (a és la meitat de la diferència de valor genotípic entre els dos homozigots). Però amb dominància α depèn de la composició genètica de la població. Amb aparellament aleatori, es pot demostrar que podem expressar α en termes dels paràmetres a i k , amb què havíem definit l'escala del valor

genotípic, així:

$$\alpha = a(1 + k(p_1 - p_2)) \quad (4)$$

On p_1 és la freqüència dels al·lels B_1 , i p_2 , la dels B_2 .

Abans havíem definit α com a $\alpha_2 - \alpha_1$, i hem dit que α_1 i α_2 són els *efectes additius*. Amb reproducció aleatòria, es pot demostrar que

$$\alpha_2 = p_1 a(1 + k(p_1 - p_2)) = p_1 \alpha \quad (5)$$

$$\alpha_1 = -p_2 a(1 + k(p_1 - p_2)) = -p_2 \alpha \quad (6)$$

Per últim cal definir el **valor reproductiu** (*breeding value*) d'un individu com a la suma dels valors additius dels seus gens. Representem el valor reproductiu amb el símbol A . En relació amb un únic locus, el valor reproductiu d'un homozigot $B_1 B_1$ serà $2\alpha_1$; el d'un heterozigot, $(\alpha_1 + \alpha_2)$; i el d'un homozigot $B_2 B_2$, $2\alpha_2$. Amb reproducció aleatòria, una conseqüència important d'aquestes definicions és que *el valor reproductiu d'un individu en una població amb aparellament aleatori és el doble de la desviació esperada entre el fenotip promig de la seua descendència i el valor fenotípic mitjà de la població*.

Anàlisi de la variància genètica

L'abstracció matemàtica de descomposar el valor genotípic en una part additiva i un efecte de la dominància, $G = \hat{G} + \delta$, ens permet descomposar la variància genètica total de la manera següent:

$$\sigma_G^2 = \sigma^2(\hat{G} + \delta) \quad (7)$$

Que d'acord amb les propietats de la variància equival a:

$$\sigma_G^2 = \sigma^2(\hat{G}) + 2\sigma(\hat{G}, \delta) + \sigma^2(\delta) \quad (8)$$

La regressió per mínims quadrats té la propietat de que la predicció **no** està correlacionada amb l'error residual. És a dir, en este cas, $\sigma(\hat{G}, \delta) = 0$. Per tant, podem dir:

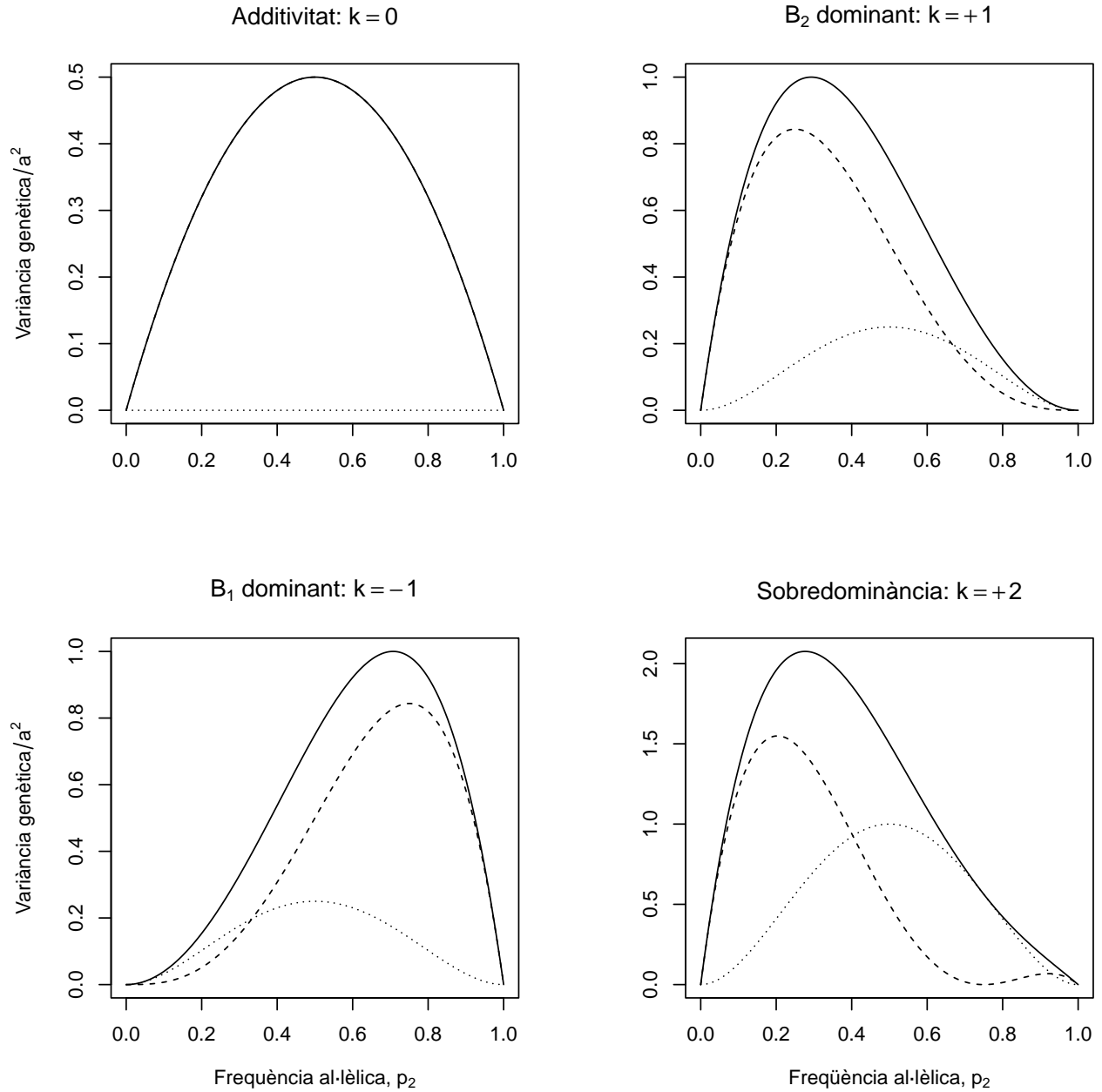
$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 \quad (9)$$

On σ_A^2 és la variància genètica explicada per la regressió sobre N_2 (seria el mateix fer la regressió sobre N_1). Se l'anomena **variància genètica additiva**. Mentre que σ_D^2 és la variància residual de la regressió o la variància genètica atribuïble a la dominància. Si la població està en equilibri de Hardy-Weinberg:

$$\sigma_A^2 = 2p_1 p_2 a^2 \quad (10)$$

$$\sigma_D^2 = (2p_1 p_2 a k)^2 \quad (11)$$

A continuació reproduïsc la figura 4.8 de Lynch and Walsh (1998), que representa les variàncies genètiques total (línia contínua), additiva (línia discontinua) i de dominància (puntejada) en diferents situacions de dominància, i amb diferents freqüències al·lel·liques.



La partició de la variància va permetre a Fisher (1918) explicar per què en una regressió, l'alçada dels progenitors no podia explicar més del 50% de la variància d'alçada entre els descendents, a pesar d'haver evidències de què la variància no explicada per la regressió no podia ser tota d'origen ambiental.

Herència poligènica

Quan hi ha dos gens influint en un caràcter, podem utilitzar la regressió lineal per separar els efectes *additius* dels dos gens de l'efecte atribuïble a la interacció entre ells. L'efecte additiu de cada gen es podia descomposar alhora en els efectes additius dels al·lels i el de la dominància.

$$G_{ijkl} = \mu_G + (\alpha_i + \alpha_j + \delta_{ij}) + (\alpha_k + \alpha_l + \delta_{kl}) + \epsilon_{ijkl} \quad (12)$$

Ací, G_{ijkl} és el valor genotípic del genotip compost pels al·lels i i j en un locus i els al·lels k i l en un altre locus. D'entrada sabem que modelar l'efecte conjunt de dos gens sobre un fenotip amb una regressió

lineal (com si els seus efectes foren additius) és una aproximació imperfecta, i en la major part dels casos esperem que el valor predit per la regressió serà diferent de l'observat. En l'equació anterior, el terme ϵ_{ijkl} (la diferència entre el valor observat i la predicció lineal) inclou els efectes possibles de totes les *interaccions* entre les variables explicatives. Per comprendre què i quines són eixes interaccions cal definir la *mitjana fenotípica condicional*. Per exemple, $G_{i...}$ seria la mitjana fenotípica condicional dels individus que porten almenys un al · lel i al primer locus, independentment de quin altre al · lel tinguen en eixe locus o de quin genotip tenen en l'altre locus. Així, podem definir tots els *efectes* inclosos en una regressió lineal. L'efecte additiu de l'al · lel i del primer locus és la diferència entre la mitjana condicional dels individus amb un al · lel i i la mitjana poblacional:

$$\alpha_i = G_{i...} - \mu_G \quad (13)$$

I de la mateixa manera, els efectes de dominància són el que queda per explicar després d'haver tingut en compte els efectes additius dels al · lels:

$$\delta_{ij} = G_{ij..} - \mu_G - \alpha_i - \alpha_j \quad (14)$$

$$\delta_{kl} = G_{..kl} - \mu_G - \alpha_k - \alpha_l \quad (15)$$

I a continuació vindrien totes les interaccions possibles, que són de tres tipus: interaccions entre els efectes additius ($\alpha\alpha$), entre efectes additius i de dominància ($\alpha\delta$), i entre efectes de dominància ($\delta\delta$):

$$\begin{aligned} (\alpha\alpha)_{ik} &= G_{i..k} - \mu_G - \alpha_i - \alpha_k \\ (\alpha\alpha)_{jk} &= G_{.jk.} - \mu_G - \alpha_j - \alpha_k \\ (\alpha\alpha)_{il} &= G_{i..l} - \mu_G - \alpha_i - \alpha_l \\ (\alpha\alpha)_{jl} &= G_{.jl.} - \mu_G - \alpha_j - \alpha_l \\ (\alpha\delta)_{ikl} &= G_{i..kl} - \mu_G - \alpha_i - \alpha_k - \alpha_l - \delta_{kl} - (\alpha\alpha)_{ik} - (\alpha\alpha)_{il} \\ (\alpha\delta)_{jkl} &= G_{.jkl.} - \mu_G - \alpha_j - \alpha_k - \alpha_l - \delta_{kl} - (\alpha\alpha)_{jk} - (\alpha\alpha)_{jl} \\ (\alpha\delta)_{ijk} &= G_{ijk.} - \mu_G - \alpha_i - \alpha_j - \alpha_k - \delta_{ij} - (\alpha\alpha)_{ik} - (\alpha\alpha)_{jk} \\ (\alpha\delta)_{ijl} &= G_{ijl.} - \mu_G - \alpha_i - \alpha_j - \alpha_l - \delta_{ij} - (\alpha\alpha)_{il} - (\alpha\alpha)_{jl} \\ (\delta\delta)_{ijkl} &= G_{ijkl} - \mu_G - \alpha_i - \alpha_j - \alpha_k - \alpha_l - \delta_{ij} - \delta_{kl} \\ &\quad - (\alpha\alpha)_{ik} - (\alpha\alpha)_{il} - (\alpha\alpha)_{jk} - (\alpha\alpha)_{jl} \\ &\quad - (\alpha\delta)_{ikl} - (\alpha\delta)_{jkl} - (\alpha\delta)_{ijk} - (\alpha\delta)_{ijl} \end{aligned}$$

Així doncs, el valor genotípic, determinat per dos (o més) gens es pot descomposar en una sèrie d'efectes additius, de dominància, additius \times additius, additius \times dominància, i de dominància \times dominància. La regressió atribueix primer la major part de la variància possible als efectes additius, i va repartint la variància que queda per explicar entre els efectes d'ordre major fins arribar a $(\delta\delta)_{ijkl}$. Si hi haguera més de dos gens influent el caràcter, es podria continuar descomposant el valor genotípic en cada vegada més efectes d'interaccions, la qual cosa es representa a continuació amb els punts suspensius:

$$\begin{aligned} G_{ijkl...} &= \mu_G + (\alpha_i + \alpha_j + \alpha_k + \alpha_l) + (\delta_{ij} + \delta_{kl}) \\ &\quad + ((\alpha\alpha)_{ik} + (\alpha\alpha)_{il} + (\alpha\alpha)_{jk} + (\alpha\alpha)_{jl}) \\ &\quad + ((\alpha\delta)_{ikl} + (\alpha\delta)_{jkl} + (\alpha\delta)_{ijk} + (\alpha\delta)_{ijl}) + (\delta\delta)_{ijkl} + \dots \end{aligned}$$

Com a conseqüència del repartiment que es fa de la variància entre els factors considerats en una regressió, tenim la garantia de què els factors o efectes, definits així a partir d'una regressió, contribueixen a la variància

fenotípica total de forma independent uns d'altres. Per tant, amb més d'un gen influent en un caràcter, també podem distingir diferents fonts de variació genètica:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 + \dots \quad (16)$$

En genètica quantitativa s'anomena epístasi a totes aquestes interaccions entre els gens, ja siguin interaccions entre els components additius o entre els additius i els de dominància o entre els de dominància. Cal tenir en compte que amb tres gens pot haver interaccions entre els efectes additius de qualsevol parell de gens, així com la interacció triple entre els components additius dels tres gens. En general, com major és el nombre de gens implicats en un caràcter quantitatiu, major és el nombre de termes no additius de la regressió. En definitiva, és molt probable que les interaccions (l'epístasi) constituïsquen una part important de la variància genètica de qualsevol caràcter quantitatiu, i en general és incorrecte assumir que els efectes epistàtics (no additius) són menypreables.

Exemple

Mitjançant una sèrie de creuaments, Doebley, Stec, and Gustus (1995) van generar un llinatge de teosinte (una espècie del genere *Zea* semblant a l'ancestre de la dacsca cultivada, *Zea mays*) amb dos petits fragments cromosòmics (en cromosomes diferents) procedents de la dacsca. Les plantes d'aquest llinatge híbrid, per tant, tenien els al·lèls propis de la dacsca U_M i B_M en els loci U i B , respectivament. Mentre que el teosinte parental tenia en eixos loci només els al·lèls propis del teosinte: U_T i B_T . En una població experimental produïda pel creuament repetit dels dos llinatges, hi havia individus amb qualsevol combinació de genotips entre els dos loci, des del genotip propi del teosinte $U_{TT} B_{TT}$, fins el genotip propi de la dacsca, $U_{MM} B_{MM}$. En un nombre elevat d'individus d'aquesta població es van genotipar els dos loci i es van mesurar les longituds entre dos nusos de la canya. La taula 1 dona les longituds mitjanes, en mil·límetres, per a cada genotip (és a dir, els valors genotípics):

Taula 1:

| Locus U | Locus B | Longitud (mm) | Freqüència |
|----------|----------|---------------|------------|
| U_{MM} | B_{MM} | 18.0 | 0.06 |
| U_{MM} | B_{MT} | 54.6 | 0.12 |
| U_{MM} | B_{TT} | 47.8 | 0.06 |
| U_{MT} | B_{MM} | 40.9 | 0.12 |
| U_{MT} | B_{MT} | 47.6 | 0.25 |
| U_{MT} | B_{TT} | 83.6 | 0.12 |
| U_{TT} | B_{MM} | 61.1 | 0.06 |
| U_{TT} | B_{MT} | 66.5 | 0.12 |
| U_{TT} | B_{TT} | 101.7 | 0.06 |

A partir d'aquestes dades, per a cada genotip fix d'un locus podem transformar les dades de fenotip a una escala entre 0 i $2a$, on l'heterozigot té un valor genotípic de $(1 + k)a$:

Els valors de a i k estimats per a un locus, quan el genotip en l'altre locus és constant poden variar perquè els valors genotípics no estiguen mesurats amb prou precisió; o bé varien perquè hi ha epístasi: els efectes additius dels al·lèls i fins i tot la dominància en un locus semblen variar en funció de quin és el genotip en l'altre locus.

Si suposem que les freqüències dels genotips són les indicades a la taula 1, es poden estimar els efectes additius i de dominància, així com les interaccions i els diferents components de la variància genètica.

Taula 2:

| Locus focal | Genotip a l'altre locus | k | a |
|-------------|-------------------------|-------|-------|
| B | U_{MM} | 14.90 | 1.46 |
| B | U_{MT} | 21.35 | -0.69 |
| B | U_{TT} | 20.30 | -0.73 |
| U | B_{MM} | 21.55 | 0.06 |
| U | B_{MT} | 5.95 | -2.18 |
| U | B_{TT} | 26.95 | 0.33 |

Experiment de Herman Nilsson-Ehle

Herman Nilsson-Ehle va creuar en 1909 una varietat de blat amb grans de color roig fosc amb una varietat amb grans blancs. La F_1 era tota homogènia, de color rosa, tal com si tots els individus de la F_1 foren heterozigots. En creuar individus de la F_1 i generar la F_2 , van aparèixer diferents tonalitats entre el blanc i el roig. La proporció de plantes amb llavors blanques en la F_2 era aproximadament $1/16$, la qual cosa suggereix l'acció de 2 gens, amb 2 al·lels cadascun en la determinació del color de l'endosperm. En classificar els grans en cinc classes de color (blanc, rosa clar, rosa, rosa fosc i roig), les proporcions eren, aproximadament, 1:4:6:4:1.

En aquest cas, no hi ha dominància (l'heterozigot en un locus té el fenotip intermig entre els homozigots del mateix locus, si l'altre locus es manté fixat en un genotip). A més, els gens actuen independentment: l'efecte d'un no depèn del genotip en l'altre. En aquest cas, el valor reproductiu és la suma dels efectes additius dels al·lels als dos loci. I la variància additiva serà la suma de les variàncies additives dels dos loci.

References

- Doebley, J, A Stec, and C Gustus. 1995. "teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance." *Genetics* 141 (1): 333–46. <https://doi.org/10.1093/genetics/141.1.333>.
- Fisher, R. A. 1918. "The Correlation Between Relatives on the Supposition of Mendelian Inheritance." *Trans. Royal Soc. Edinburgh* 52: 399–433.
- Lynch, M., and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA, U.S.A.: Sinauer Associates, Inc.