

Une notion fondamentale en génétique quantitative :

## L'HERITABILITE

La génétique quantitative a en principe un très vaste domaine d'application, puisque la très grande majorité des caractères variables dans les populations présente un déterminisme génétique complexe. Mais si effectivement cette discipline peut avoir pour objet d'étude des populations naturelles, et aborder des questions d'évolution et de sélection naturelle, c'est historiquement dans le contexte de l'amélioration des espèces animales et végétales que les principaux concepts en ont été développés.

Pour un caractère quantitatif donné, la valeur du phénotype que l'on mesure,  $P$ , est fonction de deux facteurs, la valeur du génotype de l'individu,  $G$ , d'une part, et l'ensemble des causes non génétiques et non contrôlées de variation,  $e$ , ou « micro-milieu », d'autre part. Cette fonction est sans doute toujours très complexe, mais pour des raisons pratiques c'est généralement la relation

$$P = \mu + G + e$$

qui est utilisée, où  $\mu$  est la moyenne de la population. Quoique très simplificatrice, cette relation permet des prédictions efficaces en amélioration génétique des espèces.

### L'héritabilité au sens large

Dans une population constituée d'individus de plusieurs génotypes,  $P$ ,  $G$  et  $e$  peuvent être considérées comme des variables aléatoires. Si  $G$  et  $e$  sont indépendantes, ce qui est réalisé lorsque l'espérance de  $e$  est nulle pour tout individu (c'est-à-dire que l'effet du micro-milieu ne dépend pas du génotype), la variance phénotypique de la population,  $\text{Var}P$ , est simplement (puisque  $\text{Var}\mu = 0$  et  $\text{Cov}Ge = 0$ ) :

$$\text{Var}P = \text{Var}G + \text{Var}e$$

Le rapport

$$H^2 = \frac{\text{Var}G}{\text{Var}P}$$

que l'on appelle *l'héritabilité au sens large* (également notée  $h_{SL}^2$ ), représente **la part de variance phénotypique d'origine génétique**. La valeur  $H^2 = 1$  pourrait être atteinte si tous les individus de la population se trouvaient strictement dans les mêmes conditions de milieu (ce qui est pratiquement impossible) ; la valeur 0 correspond au cas où la totalité de la variation est d'origine environnementale, ce qui peut arriver si tous les individus de la population ont le même génotype, ou bien si les locus polymorphes dans la population sont neutres vis-à-vis du caractère étudié.

Comment estimer l'héritabilité au sens large, c'est-à-dire estimer les deux composantes de la variance phénotypique ? Il faut et il suffit de disposer de *répétitions* des génotypes (clones, lignées pures) : les différences entre répétitions d'un même génotype vont nous permettre d'accéder à la part de variance qui n'est pas d'origine génétique. La méthode statistique utilisée pour cela est l'analyse de variance.

Un cas particulier important est celui où l'on dispose de deux répétitions pour toute une série de génotypes. Supposons par exemple que nous ayons mesuré un caractère sur une série de paires de jumeaux monozygotes, tirés au hasard dans une population. Si  $P$  et  $P'$  sont les valeurs phénotypiques des membres d'une paire, le coefficient de corrélation entre les valeurs des jumeaux s'écrit :

$$r = \frac{\text{Cov}PP'}{\sqrt{\text{Var}P\text{Var}P'}}$$

où  $\text{Cov}PP'$  est la covariance entre les variables  $P$  et  $P'$ . Comme  $\text{Var}P = \text{Var}P'$ , on a :

$$r = \frac{\text{Cov}PP'}{\text{Var}P}$$

Les membres d'une paire ayant le même génotype, on peut écrire :

$$\text{Cov}PP' = \text{Cov}(G + e, G + e')$$

Si l'on se rappelle que  $\text{Cov}XX = \text{Var}X$ , et *si et seulement si*  $\text{Cov}ee' = 0$ , on trouve :

$$r = \frac{\text{Var}G}{\text{Var}P} = H^2$$

Le coefficient de corrélation entre vrais jumeaux est donc une estimation de l'héritabilité au sens large, *sous réserve qu'ils n'aient pas eu un environnement commun* (c'est la condition requise pour pouvoir écrire  $\text{Cov}ee' = 0$ ). Sinon l'héritabilité est *surestimée*, puisque le numérateur contient, en plus de la composante génétique, une composante « milieu commun », qui accentue la ressemblance entre les jumeaux.

L'héritabilité au sens large est également le **coefficient de régression du génotype sur le phénotype**. Si l'on cherche en effet à prédire la valeur génotypique  $G$  connaissant la valeur phénotypique  $P$  selon un modèle linéaire, on va écrire

$$\hat{G} = b_{GP}(P - \mu)$$

où  $b_{GP}$  est le coefficient de régression. Le coefficient de régression linéaire d'une variable  $Y$  sur une variable  $X$  étant  $b = \text{Cov}XY/\text{Var}X$ , on a

$$b_{GP} = \frac{Cov(G, P - \mu)}{VarP}$$

Or  $P - \mu = G + e$ , d'où

$$b_{GP} = \frac{VarG}{VarP} = H^2$$

On voit donc que l'héritabilité au sens large est le coefficient par lequel il faut multiplier la valeur phénotypique centrée pour avoir une estimation de la valeur génotypique :

$$\hat{G} = H^2(P - \mu)$$

A moins que l'héritabilité ne soit égale à 1, les valeurs génotypiques sont donc moins dispersées que les valeurs phénotypiques : elles sont d'autant plus ramenées vers la moyenne génotypique (égale à 0 puisque l'on travaille en variables centrées) que l'héritabilité est faible. Cette plus grande dispersion des valeurs phénotypiques vient simplement des effets de milieu : à une valeur génotypique donnée, éloignée de la moyenne, peut correspondre par hasard une valeur phénotypique encore plus éloignée.

L'héritabilité au sens large a donc une triple interprétation : part de variance phénotypique d'origine génétique, coefficient de corrélation entre deux répétitions d'un même génotype, enfin coefficient de régression linéaire du génotype sur le phénotype.

## L'héritabilité au sens étroit

Le concept d'héritabilité au sens large n'est pas suffisant pour modéliser la réponse à la sélection des caractères quantitatifs. Un individu ne transmet à sa descendance qu'un des deux allèles en chacun de ses locus, si bien qu'aucun effet de dominance ne se transmet, et une petite part seulement des effets d'épistasie. Il est donc utile de décomposer la valeur génétique  $G$  elle-même en trois contributions, que l'on va encore considérer comme additives :

- l'ensemble des effets moyens des gènes pris individuellement, noté  $A$  (valeur génétique additive)
- l'ensemble des interactions entre gènes allèles, noté  $D$  (valeur de dominance)
- l'ensemble des interactions entre gènes non allèles, noté  $I$  (valeur d'épistasie)

D'où

$$G = A + D + I$$

Ce qui est transmis correspond en moyenne à la moitié de la valeur additive, soit  $A/2$ , en négligeant les effets d'épistasie (les effets en *trans* ne se transmettent pas, et les effets en *cis* sont en partie détruits par la méiose).

Etant donné qu'au sein d'une population les valeurs génétiques additives et de dominance sont non corrélées ( $CovAD = 0$ , comme on peut le démontrer), la variance génotypique peut se décomposer simplement (en négligeant la variance d'épistasie, ce que nous ferons par la suite pour des raisons de simplicité) :

$$\text{Var}G = \text{Var}A + \text{Var}D$$

où  $\text{Var}A$  est la variance génétique additive, et  $\text{Var}D$  la variance de dominance.

La variance phénotypique s'écrit donc

$$\text{Var}P = \text{Var}A + \text{Var}D + \text{Var}e$$

Le rapport de la variance génétique additive à la variance phénotypique

$$h^2 = \frac{\text{Var}A}{\text{Var}P}$$

est l'*héritabilité au sens étroit*. Ce coefficient, compris entre 0 et 1, est fondamental en génétique quantitative, car il représente la fraction de la variance phénotypique dont on peut prédire la transmission. (Par défaut, lorsque l'on parle d'héritabilité, il s'agit d'ailleurs du sens étroit.)

L'estimation de cette héritabilité s'appuie sur une caractéristique fondamentale que présentent les caractères quantitatifs, la *ressemblance entre apparentés*. Il est en effet d'expérience courante que pour de nombreux caractères, deux frères se ressemblent plus que deux cousins, qui eux-mêmes se ressemblent plus que deux individus quelconques de la population. C'est la *covariance entre apparentés* qui va donc servir à estimer les composantes de la variance génotypique. Prenons l'exemple d'une série de couples pères-fils tirés au hasard dans une population, et mesurés au même âge pour un caractère donné. Appelons  $\text{Cov}P_pP_f$  la covariance entre la valeur phénotypique du père  $P_p$  et celle de son fils  $P_f$ . On a

$$\text{Cov}P_pP_f = \text{Cov}(A_p + D_p + e_p, A_f + D_f + e_f)$$

Les variables  $A$  et  $D$  n'étant pas corrélées, et sous l'hypothèse d'absence de milieu commun, on peut écrire

$$\text{Cov}P_pP_f = \text{Cov}A_pA_f + \text{Cov}D_pD_f$$

La composante de dominance de la covariance est ici nulle, puisqu'un père et son fils ne peuvent posséder les mêmes couples d'allèles (les mères, choisies au hasard dans une population de grande taille, ne sont pas apparentées aux pères). En revanche un fils possède la moitié des gènes de son père. On peut écrire  $A_f = \frac{1}{2} A_p + \frac{1}{2} A_m$  (indice  $m$  pour mère), si bien que la composante additive non nulle de la covariance s'écrit

$$\text{Cov}(A_p, \frac{1}{2} A_p) = \frac{1}{2} \text{Var}A_p = \frac{1}{2} \text{Var}A$$

(les pères ont été tirés au hasard dans la population, d'où  $\text{Var}A_p = \text{Var}A$ ).

Le coefficient de régression enfant-parent est donc

$$b_{fp} = \frac{\text{Cov}P_pP_f}{\text{Var}P_p} = \frac{\frac{1}{2} \text{Var}A}{\text{Var}P}$$

d'où

$$2b_{fp} = h^2$$

De la même manière, on peut montrer que le coefficient de régression enfant-parent moyen (moyenne des valeurs phénotypiques des deux parents) est

$$b_{pf} = h^2$$

La régression enfant-parent, ou enfant-parent-moyen, constitue donc un moyen simple d'estimer l'héritabilité. Mais il faut insister sur les hypothèses faites : que l'environnement commun des pères et des fils n'influence pas le caractère étudié ( $Cov e_p e_f = 0$ ), que les effets des gènes soient indépendants de l'environnement ( $Cov G e = 0$ ), enfin que les unions des mères et des pères se fassent au hasard. Il est intéressant de savoir qu'historiquement, le terme « régression » vient précisément de l'observation selon laquelle les valeurs phénotypiques des enfants sont moins dispersées que celles de leurs parents, en d'autres termes qu'elles présentent une « régression » vers la moyenne (Galton, 1886) : seule une fraction  $h^2$  des différences phénotypiques parentales se transmet à la descendance.

Le coefficient de régression linéaire,  $b_{fp}$ , est aussi égal au coefficient de corrélation parent-enfant. En effet :

$$r_{pf} = \frac{Cov P_p P_f}{\sqrt{Var P_p Var P_f}} = \frac{Cov P_p P_f}{Var P} = b_{pf}$$

Plus généralement la *corrélation entre apparentés* fournit un moyen d'estimer l'héritabilité, à partir d'une formule générale de décomposition de la covariance qui tient compte du degré d'apparentement entre membres des paires d'individus choisis :

$$Cov P_1 P_2 = 2\phi Var A + u Var D$$

où  $\phi$  est la probabilité qu'un gène tiré à un locus chez l'individu 1 soit identique par descendance à un gène tiré au même locus chez l'individu 2 (c'est le coefficient de parenté), et  $u$  est la probabilité que les deux gènes d'un locus chez l'individu 1 soient identiques aux deux gènes du même locus chez l'individu 2 (c'est à cette seule condition que la composante de dominance ne sera pas nulle). Par exemple pour des demi-frères-sœurs, on a  $Cov P_1 P_2 = 1/4 Var A$ , et donc

$$r = \frac{\frac{1}{4} Var A}{Var P} = \frac{1}{4} h^2$$

d'où

$$h^2 = 4r$$

Pour des pleins-frères-sœurs, la composante de dominance n'est pas nulle, puisque deux frères peuvent avoir hérité de leurs parents les mêmes gènes ( $u = 1/4$ ). On a

$$Cov(P_1 P_2) = \frac{1}{2} Var A + \frac{1}{4} Var D$$

et donc

$$r = \frac{1}{2} h^2 + \frac{\frac{1}{4} Var D}{Var P}$$

La valeur  $2r$  surestime donc l'héritabilité.

L'autre manière d'estimer les composantes de la covariance consiste à appliquer l'analyse de variance à un plan de croisement. Par exemple en sélection animale il est fréquent de croiser un certain nombre de mâles à plusieurs femelles, ce qui produit des familles de pleins-frères et des familles de demi-frères de pères. Ce dispositif hiérarchique permet d'estimer les covariances entre pleins-frères et entre demi-frères, et de là d'estimer les variances additives et de dominance à partir des formules ci-dessus, et donc bien sûr l'héritabilité.

L'héritabilité au sens étroit est également le coefficient de régression de la valeur additive,  $A$ , sur le phénotype. Dans une démarche semblable à celle décrite pour l'héritabilité au sens large, si l'on cherche à prédire la valeur génotypique  $A$  connaissant la valeur phénotypique  $P$  selon un modèle linéaire, on va écrire

$$\hat{A} = b_{AP}(P - \mu)$$

d'où

$$b_{AP} = \frac{\text{Cov}(A, P - \mu)}{\text{Var}P}$$

Or  $P - \mu = A + D + e$ , d'où

$$b_{AP} = \frac{\text{Cov}(A, A + D + e)}{\text{Var}P} = \frac{\text{Var}A}{\text{Var}P} = h^2$$

et

$$\hat{A} = h^2(P - \mu)$$

Cette présentation de l'héritabilité est importante dans le contexte de la prédiction de la réponse à la sélection. Supposons que dans une population de grande taille distribuée normalement, nous retenions pour constituer la génération suivante un groupe d'individus de moyenne  $\mu_S$ , différente de la moyenne  $\mu$  de la population. La formule précédente permet d'écrire l'espérance de la valeur additive  $A_S$  du groupe d'individus sélectionnés :

$$E(\hat{A}_S) = h^2(\mu_S - \mu)$$

Comme la valeur additive est la seule qui se transmette à la descendance, cette espérance exprime également l'écart entre la valeur phénotypique moyenne des descendants des individus sélectionnés,  $\mu_R$ , et celle d'individus tirés au hasard dans la population ( $\mu$ , car la moyenne de la population sans sélection est constante par hypothèse). L'écart  $\Delta G = \mu_R - \mu$ , appelé *progrès génétique*, est donc

$$\Delta G = h^2(\mu_S - \mu)$$

que l'on écrit aussi classiquement

$$R = h^2S$$

Expérimentalement, le rapport  $R/S$  est donc un moyen (peu puissant), pour estimer l'héritabilité, que l'on appelle *l'héritabilité réalisée*.

En résumé, l'héritabilité au sens étroit est un concept fondamental pour décrire la transmission héréditaire des variations quantitatives, et dont la signification peut être formulée de trois façons. C'est la part héritable de la variabilité phénotypique, c'est le coefficient de régression enfant-

parent-moyen, et enfin, en tant que coefficient de régression de la valeur additive sur la valeur phénotypique, c'est le paramètre qui permet de prédire la réponse à la sélection.

### **L'héritabilité, un concept à manier avec précaution**

Il faut insister pour finir sur diverses difficultés liées au concept d'héritabilité.

Tout d'abord l'héritabilité d'un caractère donné n'est en aucun cas un paramètre biologique immuable attaché à ce caractère. Sa valeur dépend d'une part de l'environnement qu'ont connu les individus (toutes choses égales par ailleurs, plus la variance environnementale est élevée, plus l'héritabilité est faible), d'autre part de la population considérée, car la variance génétique additive dépend de la fréquence des gènes. Schématiquement, l'héritabilité est plus élevée pour les fréquences intermédiaires que pour les fréquences extrêmes (si les gènes susceptibles d'influencer la variation du caractère sont fixés, l'héritabilité sera nulle).

En second lieu le terme héritabilité est ambigu : sa proximité avec le terme *hérédité* entraîne des confusions fréquentes (le plus souvent par ignorance, parfois volontairement chez les eugénistes). Un caractère *héréditaire* est un caractère qui est génétiquement déterminé, tandis qu'un caractère *héritable* est un caractère qui, dans une population donnée, a une variance génétique additive non nulle, la première propriété n'entraînant pas nécessairement la seconde. Par exemple le caractère « nombre d'oreilles » chez l'homme est un caractère *héréditaire* car, pleinement contrôlé par nos gènes de développement, il se transmet à travers les générations, mais il est *non héritable*, car à notre connaissance il ne présente pas de variabilité génétique dans les populations humaines.

Enfin une différence entre les moyennes de deux populations peut être entièrement d'origine génétique, alors que l'héritabilité calculée dans chacune d'entre elles est nulle ; à l'inverse ce n'est pas parce que deux populations présentent chacune une héritabilité élevée pour un caractère qu'une différence entre leurs moyennes sera d'origine génétique : cette différence peut entièrement provenir de conditions environnementales différentes. Ces précisions sont importantes dans le contexte de certains débats sur le quotient intellectuel et autres caractères de comportement chez l'homme.