

# La genèse et le développement des concepts de l'évaluation génétique classique

D. LALOË

INRA, UMR1313 Génétique Animale et Biologie Intégrative, F-78352 Jouy-en-Josas, France  
AgroParisTech, Génétique Animale et Biologie Intégrative, 16 rue Claude Bernard, F-75231 Paris, France  
Courriel : denis.laloe@jouy.inra.fr

L'évaluation génétique «classique» est souvent décrite comme une simple application de la génétique quantitative théorique, se bornant à l'utilisation d'un modèle de type «boîte noire». C'est pourtant un questionnement tourné vers la hiérarchisation des individus selon leur valeur transmissible qui a permis de définir les concepts-clés de phénotype et de génotype ainsi que le modèle polygénique qui traite des caractères continus. A travers quelques grandes étapes, on essaye d'en dégager les aspects fondamentaux.

L'une des caractéristiques de la domestication est la modification génétique des espèces domestiquées, via le contrôle de la reproduction, ce qui a profondément changé leur anatomie, leur physiologie et leur comportement (Denis 2004). Dès les origines, les éleveurs ont, au moins inconsciemment, évalué les animaux et les ont choisis comme reproducteurs en fonction des caractéristiques qu'ils voulaient garder ou renforcer. Cette sélection s'est progressivement rationalisée avec la notion de race, apparue dans l'espèce bovine à la fin du 18<sup>ème</sup> siècle en France (Denis et Avon 2010), puis avec l'essor de la génétique et l'approche quantitative de l'hérédité.

De Pythagore à Darwin, les plus grands esprits ont d'abord tenté d'expliquer les mécanismes physiques de l'hérédité au moyen de théories plus ou moins hasardeuses. Dans la deuxième partie du 19<sup>ème</sup> siècle, Mendel inventa une approche phénoméniste (au sens du grec *phainomenon*, manifestation) radicalement originale, en s'intéressant à la manifestation de l'hérédité et en la quantifiant statistiquement, sans plus se préoccuper de la nature concrète des facteurs héréditaires (Pichot 1999). Cette conception phénoméniste apparaît clairement dans la citation suivante : «Il faut traiter le gène comme une unité de comptage ou de calcul. Nous n'avons aucunement le droit de définir le gène comme une structure morpho-

logique, au sens des «gemules» de Darwin, des «biophores», des «déterminants» ou de toute autre sorte de concept morphologique»<sup>1</sup> (d'après Johannsen 1911).

La génétique quantitative, et encore plus les méthodes liées à l'évaluation génétique, s'inscrivent pleinement dans cette conception du gène comme unité de calcul.

Cet article est un essai sur les concepts de l'évaluation génétique, en ce sens qu'il expose ces concepts en liaison avec une pratique de plus de deux décennies dans cette activité. On veut montrer ici que les concepts fondateurs de la génétique quantitative sont liés à l'évaluation génétique des individus. Pour ce faire, nous avons isolé trois grandes étapes :

- les expériences de Johannsen au tout début du 20<sup>ème</sup> siècle sur les lignées pures d'haricots Princesse, qui fixent les concepts-clés de gène, génotype et phénotype ;

- l'article de 1918 de Fisher à propos des corrélations entre apparentés, qui montre que l'hérédité des caractères à variation continue n'est pas incompatible avec les mécanismes mendéliens de transmission ;

- les apports d'Henderson, dans les années 70, qui ont révolutionné la pratique de l'évaluation génétique.

## 1 / L'origine des concepts

On situe en règle générale la naissance de la génétique moderne au début du 20<sup>ème</sup> siècle, avec la «redécouverte» des lois de Mendel. Si celles-ci concernent au premier chef les caractères qualitatifs, les caractères quantitatifs à variation continue (croissance, taille, poids...) sont de première importance pour l'économie des productions agricoles, animales comme végétales. Dans ce domaine, la discorde règne (Verrier *et al* 2001, Morrison 2002), avec en arrière-plan des divergences sur les mécanismes de l'évolution et des rivalités entre biologistes et mathématiciens. En effet, alors que la théorie de Darwin postule une évolution graduelle sous l'effet de la sélection naturelle, on considère alors que les lois de Mendel, fondées sur la ségrégation de facteurs discontinus, ne peuvent pas expliquer les variations continues des caractères quantitatifs. Deux écoles s'affrontent :

- les «mutationnistes», représentés notamment par Wilse Bateson (Bateson 1901), défendent les lois de Mendel et mettent ainsi en cause la théorie de Darwin ;

- les «biométriciens», avec comme chef de file Karl Pearson, inventeur entre autres de l'analyse en composantes principales (Pearson 1901), et de la corrélation du même nom. Ils soutiennent la théorie gradualiste de Darwin, et

<sup>1</sup> Les citations en italique sont traduites par l'auteur.

privéliegient une approche populationnelle (Pearson 1904), par l'étude de liaisons entre les valeurs mesurées chez des individus apparentés (Verrier *et al* 2001).

Une controverse d'une autre nature oppose ces biométriciens et un botaniste danois, Wilhelm Johannsen (Roll-Hansen 2009). Celui-ci considère que les mécanismes de l'hérédité doivent être étudiés en eux-mêmes, indépendamment de la théorie de l'évolution. Pour celui-ci, l'approche populationnelle de Pearson est insuffisante, et il préconise une approche fondée sur l'individu : «*la théorie statistique de l'hérédité, telle qu'elle est développée par Galton et Pearson, s'intéresse aux groupes d'individus ou aux populations... L'étude de l'hérédité ne peut se satisfaire d'une telle recherche essentiellement statistique... Une population peut contenir de nombreux types indépen-*

*dants, différant sensiblement les uns des autres, que l'observation de tables ou de courbes de fréquences empiriques ne peut absolument pas déceler*» (cité par Yule 1903).

Johannsen est guidé par des motivations pratiques de hiérarchisation des individus selon leur «force héréditaire», en d'autres termes, une évaluation génétique : «*On doit décider si la "force héréditaire" (en français dans le texte), comme Vilmorin la nomme, est grande ou petite*» (cité par Yule 1903).

## 2 / Un génotype, plusieurs phénotypes : les lignées pures de Johannsen

Johannsen (Johannsen 1903) met en place au début du 20<sup>ème</sup> siècle une série d'expériences sur des haricots Princesse

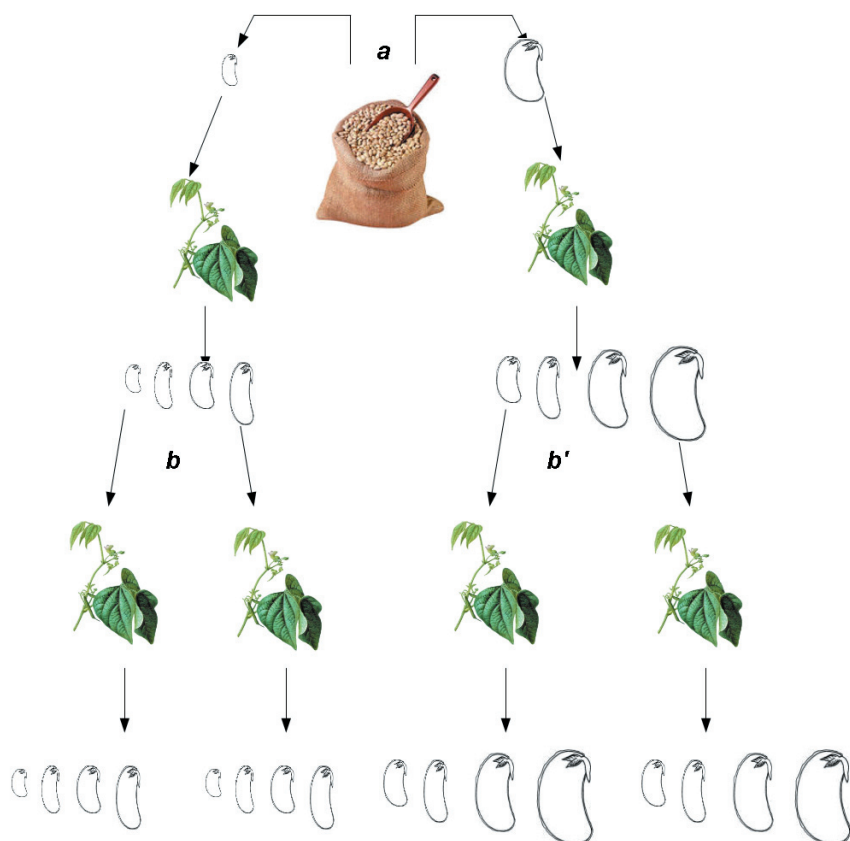
(une variété de *Phaseolis Vulgaris*). Ces plantes, autogames, se reproduisent par autofécondation. A partir d'une population commerciale de plantes dont la récolte comporte des graines de poids différents, Johannsen sème ces graines après les avoir pesées. Il observe que les plantes issues de petites graines donnent des graines plus petites. Cette première sélection est efficace. Mais cette efficacité ne dure pas : si, parmi les graines de la plante issue de la plus grosse graine du lot initial, l'on choisit la plus petite et la plus grosse, les plantes qui en sont issues donnent des graines de même poids moyen, et dont les poids individuels se répartissent de la même façon (figure 1). De même pour les autres plantes de la première génération de sélection : leurs descendances sont identiques entre elles et, quelle que soit la génération, la valeur moyenne du poids des graines d'une plante ne dépend que de la plante prise pour ancêtre. La population initiale, provenant d'une variété commerciale non sélectionnée pour le poids des graines, constituait un mélange de formes génétiquement différentes les unes des autres. Le haricot étant autogame, la sélection sur la taille des graines donne naissance à des lignées pures, homozygotes, sur lesquelles une sélection ultérieure est inefficace (Valdeyron 1961).

Ces expériences permettent à Johannsen (Johannsen 1911) de poser la pierre angulaire de la génétique quantitative, c'est-à-dire la distinction claire entre «phénotype» et «génotype» (c'est lui qui introduit ces termes). Plusieurs génotypes coexistent dans le lot initial de graines, ces génotypes se transmettent, d'où les différences transmises en première génération. Les lignées pures n'ont qu'un seul génotype. Les phénotypes correspondant à ce génotype particulier peuvent être différents (Johannsen observe que leur distribution empirique est gaussienne), mais les variations phénotypiques observées pour un même génotype ne se transmettent pas.

## 3 / Le modèle polygénique de Fisher

En 1918, Ronald Fisher résout la contradiction entre mutationnistes et biométriciens dans un célèbre article (Fisher 1918) où, au passage, le terme «variance» est employé pour la première fois. Il reprend une idée de Mendel, selon laquelle un caractère continu (la couleur d'une fleur de haricot issu d'un croisement entre variété à fleurs blanches et variété à fleurs rouges, allant du pourpre au violet pâle, s'interprète comme mélange de caractères élémentaires) (Verrier *et al* 2001). Fisher

Figure 1. Expérience de Johannsen (1903).



En **a**, on choisit la plus grosse graine et la plus petite. La sélection est efficace : la plante issue de la plus grosse graine donne en moyenne des grosses graines, et la plante issue de la plus petite graine donne en moyenne des petites graines. En **b** et en **b'**, on recommence, mais cette fois sans résultat : les plantes issues des graines de **b**, grosses ou petites, donnent des graines de même taille moyenne ; de même pour **b'**. La sélection est inefficace parce qu'on s'est adressé à des lignées pures, homozygotes. En **a**, nous avons affaire à un mélange de lignées pures, que la sélection individuelle a eu pour effet de séparer. Ceci n'est valable que dans le cas d'une plante autogame.

Le phénotype est ce qui est observé : c'est la somme de l'effet du génotype et de l'effet de l'environnement. Les variations entre les moyennes de poids des graines des lignées **b** et **b'** est due aux différences de génotype entre les deux lignées. Ces variations sont transmissibles. Les graines d'une même lignée (**b** ou **b'**), ont le même génotype. La variation du poids intralignée est donc due au seul effet de l'environnement, et n'est pas transmissible (d'après Valdeyron 1961).

considère que la «valeur génétique» attachée à un caractère quantitatif continu (comme le poids à un âge donné) est la somme des effets de nombreux «facteurs mendéliens». Il divise cette valeur génétique en valeur génétique additive, «qui reflète la nature génétique sans distorsion», et un résidu (effets de dominance et d'épistasie).

On peut alors considérer que cette valeur génétique additive est distribuée selon une loi de Gauss, en appliquant le théorème de limite centrale : toute somme de variables aléatoires indépendantes et identiquement distribuées tend vers une variable distribuée selon une loi de Gauss. De ce fait, la loi de Gauss, et les lois associées (Student, Fisher,  $\chi^2$ ) jouent un rôle majeur en statistique. Comme son nom l'indique, il s'agit d'un théorème traitant d'une limite, et non d'une situation réelle. En pratique, même les mathématiciens les plus stricts considèrent que l'approximation gaussienne est licite pour des situations où aucune variable n'est prépondérante (Dacunha-Castelle et Duflo 1982), le nombre de ces variables dépassant la trentaine. Enfin, Fisher calcule dans ce cadre les covariances théoriques entre apparentés.

L'intérêt de ce modèle est multiple :

- il rejoint Johannsen en raisonnant au niveau de l'individu : la valeur génétique d'un individu est intrinsèquement aléatoire car elle est la somme d'effets des allèles portés par l'individu, et que ces allèles résultent eux-mêmes d'un tirage aléatoire parmi l'ensemble des allèles. Fisher l'affirme d'ailleurs explicitement : «Les formulations vagues sur le pourcentage de causalité obscurcissent la distinction essentielle entre individu et population, et devraient être soigneusement évitées» ;

- il justifie la distribution de la valeur génétique selon une loi de Gauss, dans le cadre d'un modèle polygénique, où le caractère résulte d'une somme d'un grand nombre de gènes ;

- il fournit un cadre théorique de calcul des corrélations entre apparentés, là où l'approche populationnelle de Pearson ne pouvait qu'estimer des corrélations empiriques.

#### 4 / Le modèle d'évaluation génétique

A partir de ces contributions, on peut écrire en termes «actuels» le modèle d'expression des phénotypes comme suit, en supposant que les phénotypes (P) s'expriment dans un environnement homogène et qu'ils sont la somme des effets du génotype (G) et de l'environnement (E) :

**Tableau 1.** Quelques valeurs du coefficient de parenté (Wright) entre un animal et un apparenté.

| Animal | Père | Plein Frère | Grand-père | Demi-Frère |
|--------|------|-------------|------------|------------|
| 1      | 0,5  | 0,5         | 0,25       | 0,25       |

Additivité du génotype et de l'environnement :  $P = G + E$

Indépendance entre phénotype et environnement :

$$\text{Var}(P) = \text{var}(G) + \text{var}(E) ;$$

$$\text{cov}(G,E) = 0.$$

La covariance entre individus est uniquement d'origine génétique. Elle est fonction du degré d'apparentement entre ces individus :

$$\text{Cov}(P_i,P_j) = A(i,j) * \text{Var}(G)$$

où  $A(i,j)$  est le coefficient de parenté (au sens de Wright, i.e. le double du coefficient de parenté au sens de Malécot, en absence de consanguinité, (Ollivier 2002)) entre les individus  $i$  et  $j$ , et dont on trouvera quelques valeurs dans le tableau 1.

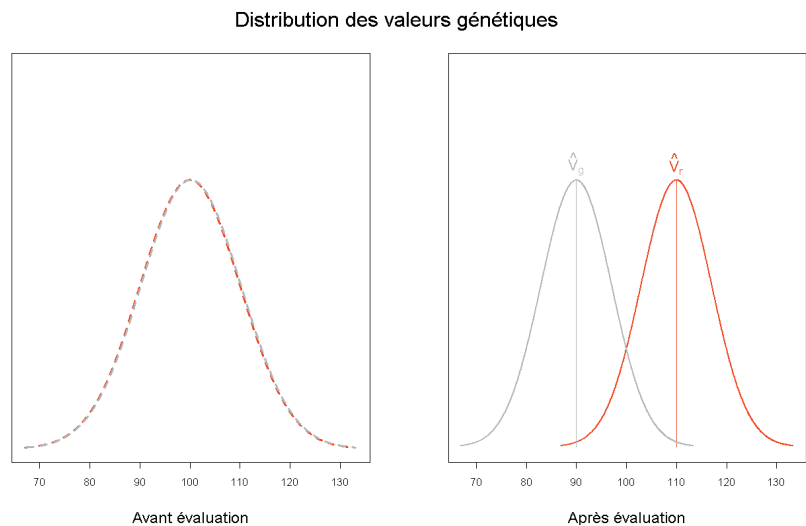
Les variances phénotypique et génétique sont les paramètres génétiques. Le rapport entre variances génétique et phénotypique est l'héritabilité, notée  $h^2$ . Les méthodes d'estimation de ces paramètres sortent du cadre de cet article. On en trouvera une présentation succincte dans (Hill et Kirkpatrick 2010). L'ensemble des

paramètres génétiques est supposé connu préalablement à l'évaluation génétique.

#### 5 / L'évaluation génétique : une espérance conditionnelle

L'évaluation génétique repose sur le modèle établi ci-dessus. Elle consiste, une fois connus les paramètres généraux d'une population (variances génétique et phénotypique, corrélation entre apparentés (via les généalogies), à utiliser les phénotypes des individus pour «prédire» leur valeur génétique. Avant l'évaluation génétique, tous les animaux se valent (l'espérance de leur valeur génétique est nulle), et on ne sait rien sur eux (la variance de la valeur génétique est égale à la variance génétique). L'évaluation génétique substitue à la distribution initiale des valeurs génétiques, d'espérances nulles et de variance égale à la variance génétique, la distribution conditionnelle aux phénotypes, avec une variance plus faible et d'espérances différentes (les valeurs génétiques «prédites») selon les animaux (figure 2). Le gain d'informa-

**Figure 2.** Le principe de l'évaluation génétique.



On ne connaît jamais avec certitude la valeur génétique d'un individu. La valeur génétique d'un individu est représentée par sa distribution théorique, qu'on résume par deux paramètres, l'espérance et la variance. La valeur génétique prédite est égale à l'espérance, tandis que la variance mesure le degré d'incertitude sur cette valeur.

On considère une population de deux individus, rouge (r) et gris (g).  
 A gauche : avant évaluation, on considère que tous les individus sont égaux (l'espérance de leur valeur génétique est égale à la moyenne de la population), et la variance de la valeur génétique est égale à la variance génétique du caractère : les distributions (courbes pointillées) sont identiques.  
 A droite : l'évaluation génétique consiste, grâce à la prise en compte des performances, à substituer à ces distributions des distributions de variances plus faibles (réduction de l'incertitude) et d'espérances différentes : les distributions des valeurs génétiques des individus sont différentes. On choisira ainsi l'individu rouge car sa valeur génétique prédite  $\hat{V}_r$  est supérieure à  $\hat{V}_g$ , la valeur génétique prédite de l'individu gris.



tion est mesuré par le CD, qui est la baisse relative des variances avant et après évaluation. Dans le cas où les différentes variables suivent une loi de Gauss, cette espérance conditionnelle *a posteriori* est une simple fonction linéaire des phénotypes, analogue à une équation de régression. Les coefficients de cette équation s'expriment en fonction des variances et covariances génétiques.

## 6 / Une révolution technique : Henderson et le modèle mixte

Le modèle ci-dessus présente deux inconvénients :

Il nécessite l'homogénéisation des conditions de milieu. Johannsen avait déjà noté dans ces expériences que la moyenne des tailles de graines dépendait de l'année de production des plants. Plus généralement, une performance est influencée par tout un ensemble de facteurs, année de naissance, élevage, sexe, etc. On peut s'en tirer en travaillant non pas sur les performances brutes, mais sur les déviations de ces performances par rapport à ces facteurs. Mais cela revient à ne pas pouvoir comparer les niveaux génétiques d'animaux nés dans des années différentes : on ne peut pas savoir dans quelle proportion l'évolution des performances au cours du temps est dû au progrès génétique et au progrès dans la conduite des élevages, puisque les moyennes des animaux nés une même année sont annulées. Pour estimer les progrès génétiques, il faut alors mettre en œuvre des expérimentations lourdes. Par exemple, Tribout *et al* (2010) compare des groupes d'animaux en inséminant des truies avec de la semence congelée de verrats nés à des périodes différentes. La comparaison des performances des deux groupes d'animaux permet d'estimer les évolutions génétiques réalisées au cours de la période.

Le calcul des espérances conditionnelles nécessite le calcul explicite des covariances entre valeur génétique de l'animal et phénotypes des apparentés. Cela réduit en pratique l'utilisation des apparentés à l'animal (sélection individuelle), à ses descendants (sélection sur descendance) ou ses proches collatéraux. Les méthodes combinant les différents apparentements sont vite compliquées à formaliser.

### 6.1 / Le BLUP d'Henderson

Le premier problème est résolu dans les années 1970 par le «BLUP» (*Best Linear Unbiased Prediction*)

(Henderson 1975) : les différences de milieu sont explicitement incluses dans le modèle, qui s'écrit alors :

$$\text{Phénotype} = \text{Effets de milieu} + \text{Génotype} + \text{Environnement}$$

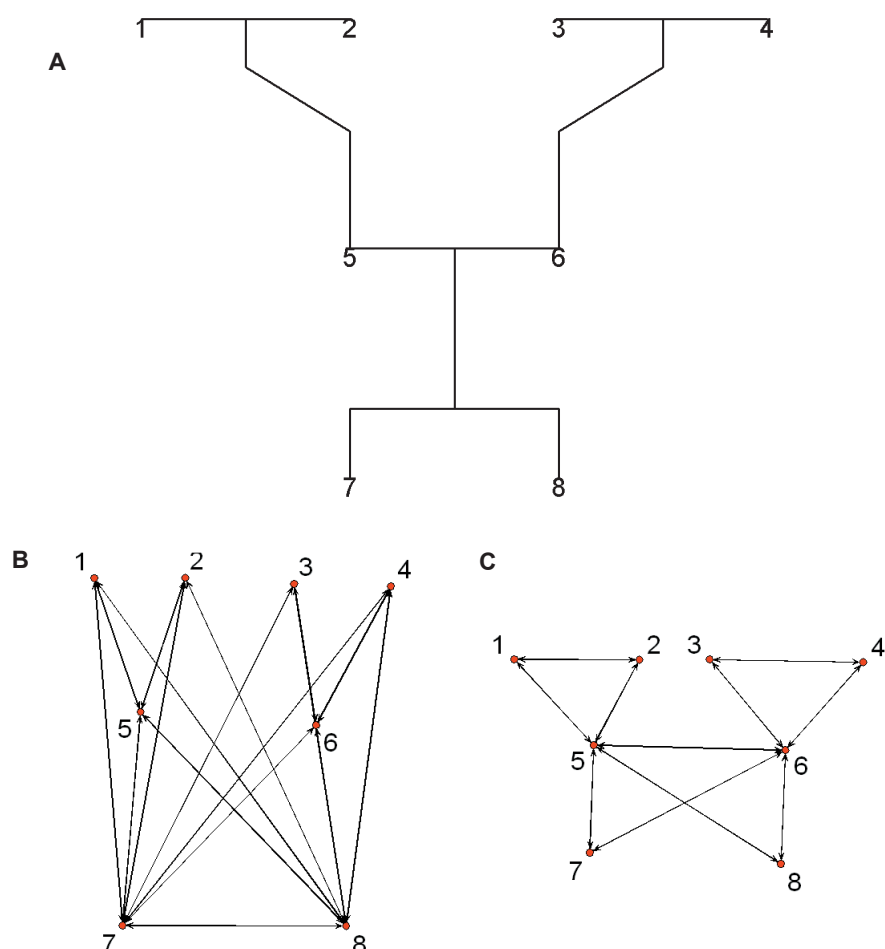
Par la résolution simultanée des effets de milieu et des effets génétiques, on peut comparer alors des animaux élevés dans différents environnements ou nés à des périodes différentes, pourvu que des liens génétiques existent entre ces environnements ou années (Foulley *et al* 1990, Laloë 1993). Ces équations ne font pas intervenir la matrice **A** des corrélations entre apparentés, mais la matrice inverse, notée **A**<sup>-1</sup>.

### 6.2 / La matrice inverse de parenté. Corrélations partielles

Le second problème est résolu encore une fois par Henderson dans les années 1970. Henderson montre que les seuls coefficients non nuls de la matrice **A**<sup>-1</sup>

relient les parents entre eux, et les parents aux descendants (Henderson 1976). Son calcul est très simple, pourvu que l'ensemble des généalogies de la population soit prise en compte. Cette particularité est à première vue tout à fait étonnante. On peut l'expliquer si l'on considère que **A**<sup>-1</sup> s'interprète en termes de corrélations partielles entre individus, qui mesurent l'information marginale qu'apporte la connaissance de la valeur génétique d'un individu *i* sur un individu *j*, quand tous les autres individus de la population sont connus. Ainsi, si on connaît la valeur génétique du père, connaître les grands-parents paternels ou les demi-frères paternels ne procure aucune information supplémentaire (figure 3). Les corrélations partielles sont souvent préférées aux corrélations car elles sont moins sujettes à la contingence (du type «le nombre de naissances en Alsace est corrélé au nombre de cigognes observées») et traduisent plus des relations de causalité. Ce sujet a été amplement débattu dès les années 20, dans un contexte génétique

Figure 3. Corrélations et corrélations partielles entre apparentés.



En haut, généalogie de 8 animaux. En bas à gauche, les corrélations (matrice de parenté) entre les individus. En bas à droite, les corrélations partielles (matrice inverse) entre les individus. Les corrélations partielles relient uniquement les parents aux descendants directs, et les conjoints entre eux.

**Tableau 2.** Evaluations génétiques réalisées par le département de Génétique Animale de l'Inra. (Source, Département de Génétique animale, Inra 2009).

| Espèce                          | Evaluation                                     | Fréquence /an | Races     | Nombre maximal d'animaux (données) utilisés dans une évaluation | Nombre de caractères |
|---------------------------------|--|---------------|-----------|---|----------------------|
| Porc (Large White type femelle) | Reproduction                                   | 12            | 2         | ~500 000  | 4                    |
|                                 | Production                                     | 12            | 4         | ~500 000  | 11                   |
| Bovin Laitier (Prim' Holstein)  | Production                                     | 3             | 8         | (18 400 000)  | 6                    |
|                                 | Conformation                                   | 3             | 8         | 6 340 000   | 28                   |
|                                 | Longévité                                      | 2             | 8         | (10 870 000)  | 1                    |
|                                 | Condition de naissance / Mortinatalité         | 3             | 8         | (12 800 000)  | 4                    |
|                                 | Mammites cliniques                             | 3             | 8         | 3 029 000   | 1                    |
|                                 | Fertilité                                      | 3             | 9         | 17 500 000  |                      |
|                                 | Index de synthèse                              | 3             | 8         | 14 100 000  | 12                   |
|                                 | Evaluation assistée par marqueurs              | 12            | 3         | 26 500  | 15                   |
| Ovin Laitier (Lacaune)          | Production & cellules                          | 3             | 5         | 1 390 000   | 6                    |
|                                 | Morphologie                                    | 3             | 1         | 582 000   | 3                    |
| Caprin laitier (toutes races)   | Production                                     | 3             | 3         | 2 760 000   | 5                    |
| Caprin laitier (Alpine)         | Morphologie                                    | 2             | 2         | 158 000   | 4                    |
| Caprin angora                   | Production                                     | 1             | 1         | 31 300  | 1                    |
| Cheval                          | Conformation                                   | 1             | 4         | ~12 000   | 33                   |
|                                 | Trot   | 1             | 1         | 417 000   | 1                    |
|                                 | Saut d'obstacles                               | 1             | Multirace | 1 970 000   | 2                    |
|                                 | Concours Complet                               | 1             | Multirace | 1 860 000   | 2                    |
|                                 | Dressage                                       |               | Multirace | 1 920 000   | 2                    |
|                                 | Condition de naissance                         | 1             | 9         | 7 024 000   | 2                    |
| Bovin allaitant (Charolais)     | Croissance                                     | 1             | 9         | 3 970 000   | 2                    |
|                                 | Morphologie                                    | 1             | 9         | 4 044 000   | 19                   |
|                                 | Carcasse                                       | 1             | 5         | 3 360 000   | 5                    |
|                                 | Proliférite, Aptitudes maternelles, Croissance | 10            | 99        | 2 450 000   | 3                    |
| Lapins                          | Production                                     | 8             | 2         | 123 000   | 3                    |
|                                 | Aptitudes maternelles                          | 8             | 2         | 54 500  | 3                    |

par Fisher (1925) et un autre généticien, Sewal Wright, via les «path coefficients» (Wright 1921). Il est de nouveau d'actualité avec les techniques les plus récentes d'inférence sur les réseaux de gènes (de la Fuente *et al* 2004, Blum *et al* 2010).

Ces apports ne remettent pas en cause les concepts de base évoqués plus haut. En revanche, ils ont révolutionné la pratique de l'évaluation génétique. Des algorithmes efficaces (Poivey 1986, Schaeffer et Kennedy 1986, Ducrocq *et al* 1990), couplés avec la montée en puissance de l'in-

formatique ont permis les évaluations actuelles, pouvant inclure plusieurs dizaines de millions d'animaux (tableau 2).

## 7 / Des caractères particuliers

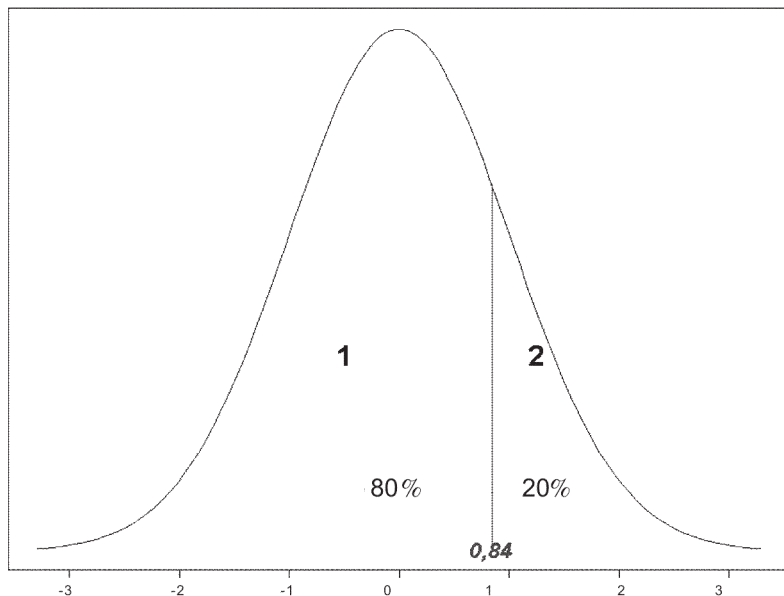
Tous les caractères d'intérêt ne suivent pas une loi de Gauss, ou ne sont pas à variation continue. Ainsi, la gemellité, caractère discret, ou les rangs d'arrivée dans les compétitions. On peut néanmoins se ramener à des lois de

Gauss pour traiter ces caractères. C'est le cas des variables de rang dans les compétitions hippiques (Tavernier 1991) ou des caractères discrets analysés par des modèles à seuil (Foulley et Manfredi 1991) (figure 4). D'autres extensions concernent les variables «censurées» comme la longévité (Ducrocq 1994), ou l'hétérogénéité de variances (San Cristobal *et al* 1993, Foulley et Gianola 1996, Robert-Granié *et al* 1997).

Ce cadre conceptuel est donc très souple et permet de traiter l'ensemble des caractères d'intérêt zootechnique.

**Figure 4.** Modélisation d'une variable dichotomique ordonnée  $X$  par une variable normale  $Y$  sous-jacente.

La probabilité que  $X$  égale 1 est de 0,8. Le seuil correspondant pour la variable  $Y$  est égal à 0,84.



## Conclusion

Cette présentation ne prétend pas retracer fidèlement l'histoire, et elle néglige l'apport d'acteurs importants, tels que Lush ou Kempthorne (Ollivier 2002). Elle tente de montrer qu'une modélisation souvent présentée comme simple, voire simpliste (Le terme «boîte noire» est souvent utilisée à son sujet), est plus subtile qu'il n'y paraît. C'est guidé par des objectifs d'évaluation génétique (c'est-à-dire de discrimination entre individus), que les concepts-clés de génotype et phénotype, définis au niveau de l'individu, ont été élaborés. Le contraste est saisissant entre la stabilité de ces concepts et la révolution permanente qui caractérise la génétique moléculaire dans la seconde moitié du 20<sup>ème</sup> siècle (Pearson 2006). Mais stabilité ne signifie pas immobilisme, et l'évaluation génétique a profondément évolué, du fait des progrès en statistique, en calcul numérique et de l'explosion de la puissance informatique.

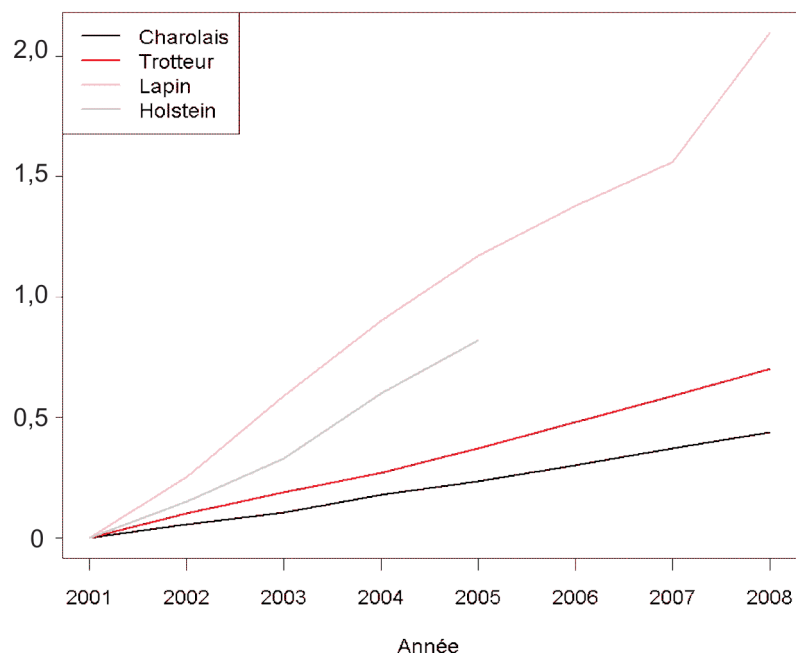
On peut considérer qu'il s'agit d'un premier exemple de biologie prédictive, domaine appelé à prendre une grande importance (Inra 2010).

Il est de bon ton d'affirmer «Prédire n'est pas expliquer» (Thom 1999). Il est en effet aventureux d'accorder trop de crédit à la réalité physique d'un grand nombre de gènes exerçant chacun des effets indépendants de même faible amplitude sur un caractère donné, même si, *a contrario*, les études récen-

tes d'association réalisées au niveau du génome (GWAS, «Genome-Wide Association Studies»), semblent le justifier (Manolio *et al* 2009).

Il n'en reste pas moins que ce modèle est un modèle instrumentaliste, et sa portée est uniquement pragmatique.

**Figure 5.** Moyennes des valeurs génétiques prédites par année, de 2001 à 2008, exprimées en écart type génétique.



En rose, poids à 63 jours chez le lapin (lignée AGP39) ; en rouge, Courses au trot chez le Trotteur Français ; en noir, le poids à 210 jours en Charolais (moyenne des taureaux utilisés une année donnée) ; en gris la longévité combinée en race Holstein (moyenne des taureaux nés une année donnée).

En ce sens, il a pleinement réussi : l'évaluation génétique permet aujourd'hui de prédire les valeurs génétiques de dizaines de millions d'individus dans les espèces d'intérêt zootechnique (tableau 2). La capacité prédictive de ces méthodes est amplement prouvée par l'observation du progrès génétique réalisé dans ces espèces (Hill et Kirkpatrick 2010) (figure 5 pour quelques exemples français).

## Remerciements

Je remercie Anne Barbat, Christine Blouin, Vincent Ducrocq, Marie-Noëlle Fouilloux, Hervé Garreau, Armelle Gion, Hélène Larroque, Isabelle Palhière, Anne Ricard, Christèle Robert-Granié, Thierry Tribout et Eric Venot pour m'avoir fourni les données nécessaires à l'établissement du tableau 2 et de la figure 5.

## Références

- Bateson W., 1901. Heredity, differentiation and other conceptions of biology: a consideration of professor Karl Pearson's paper «On the principle of homotyposis». *Proc. Roy. Soc., Serie B, Biol. Sci.*, 69, 193-205.
- Blum Y., Le Mignon G., Lagarrigue S., Causeur D., 2010. A factor model to analyze heterogeneity in gene expression. *BMC Bioinformatics*, 11-368.
- Dacunha-Castelle D., Duflo M., 1982. *Probabilités et statistiques. Problèmes à temps fixe*. Masson, Paris, France, 214p.
- de la Fuente A., Bing N., Hoeschele I., Mendes P., 2004. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, 20, 3565-3574.
- Denis B., 2004. La domestication : un concept devenu pluriel. *INRA Prod. Anim.*, 17, 161-166.
- Denis B., Avon L., 2010. *Races bovines. Castor et Pollux*, Paris, France, 321p.
- Ducrocq V., 1994. Statistical analysis of length of productive life for dairy-cows of the normande breed. *J. Dairy. Sci.*, 77, 855-866.
- Ducrocq V., Boichard D., Bonaiti B., Barbat A., Briand M., 1990. A pseudo-absorption strategy for solving animal-model equations for large data files. *J. Dairy Sci.*, 73, 1945-1955.
- Fisher R.A., 1918. The correlation between relatives on the supposition of mendelian inheritance. *Trans. Roy. Soc.*, 52, 399-433.
- Fisher R.A., 1925. *Statistical methods for research workers*. Oliver and Boyd, London, Royaume-Uni, 239p.
- Foulley J.L., Gianola D., 1996. Statistical analysis of ordered categorical data *via* a structural heteroskedastic threshold model. *Genet. Sel. Evol.*, 28, 249-273.
- Foulley J.L., Manfredi E., 1991. Statistical approaches to genetic evaluation for threshold binary traits. *Genet. Sel. Evol.*, 23, 309-338.
- Foulley J.L., Bouix J., Goffinet B., Elsen J.M., 1990. Connectedness in genetic evaluation. In: *Advances in statistical methods for genetic improvement of livestock*. Gianola D, Hammond K. (Eds). Springer, Heidelberg, Allemagne, 302-337.
- Henderson C.R., 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31, 423-447.
- Henderson C.R., 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, 32, 69-83.
- Hill W.G., Kirkpatrick M., 2010. What animal breeding has taught us about evolution. *Ann. Rev. Ecol. Evol.*, 41, 1-19.
- INRA, 2010. Document d'orientation Inra 2010-2020. Une science pour l'impact. Ed. Inra. Paris, France, 60p.
- Johannsen W., 1903. *Über Erbllichkeit in Populationen und reinen Linen. Eine Beitrag zur Beleuchtung schwebender Selektionfragen*. Gustav Fischer, Jéna, Allemagne. 68p.
- Johannsen W., 1911. The genotype conception of heredity. *Am. Nat.*, 45, 129-159.
- Laloë D., 1993. Precision and information in linear models of genetic evaluation. *Genet. Sel. Evol.*, 25, 557-576.
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorf L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon L.R., Chakravarti A., Cho J.H., Guttmacher A.E., Kong A., Kruglyak L., Mardis E., Rotimi C.N., Slatkin M., Valle D., Whittemore A.S., Boehnke M., Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F.C., McCarroll S.A., Visscher P.M., 2009. Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.
- Morrison M., 2002. Modelling populations: Pearson and Fisher on mendelism and biometry. *Brit. J. Philos. Sci.*, 53, 39-68.
- Ollivier L., 2002. *Eléments de génétique quantitative*. Inra Editions, Paris, France, 184p.
- Pearson H., 2006. What is a gene? *Nature*, 441, 398-401.
- Pearson K., 1901. On lines and planes of closest fit to systems of points in space. *Philos. Mag.*, 2, 559-572.
- Pearson K., 1904. Mathematical contributions to the theory of evolution, XII: On a generalized theory of alternative inheritance with special reference to Mendel's laws. *Philos. Trans. Royal. Soc. A*, 203, 53-86.
- Pichot A., 1999. *Histoire de la notion de gène*. Flammarion, Paris, France. 174p.
- Poivey J.P., 1986. Simplified computing method for estimation of the breeding values of females with a relationship matrix. *Genet. Sel. Evol.*, 18, 321-331.
- Robert-Granié C., Ducrocq V., Foulley J.L., 1997. Heterogeneity of variance for type traits in the Montbeliarde cattle breed. *Genet. Sel. Evol.*, 29, 545-570.
- Roll-Hansen N., 1990. Sources of Wilhelm Johannsen's genotype theory. *J. Hist. Biol.*, 42, 457-493.
- San-Cristobal M., Foulley J.L., Manfredi E., 1993. Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genet., Sel. Evol.*, 25, 3-30.
- Schaeffer L.R., Kennedy B.W., 1986. Computing strategies for solving mixed model-equations. *J. Dairy Sci.*, 69, 575-579.
- Tavernier A., 1991. Genetic evaluation of horses based on ranks in competitions. *Genet. Sel. Evol.*, 23, 159-173.
- Thom R., 1999. *Prédire n'est pas expliquer*. Flammarion, Paris, France. 174p.
- Tribout T., Caritez J.C., Gruand J., Bouffaud M., Guillouet P., Billon Y., Pery C., Laville E., Bidanel J.P., 2010. Estimation of genetic trends in French Large White pigs from 1977 to 1998 for growth and carcass traits using frozen semen. *J. Anim. Sci.*, 88, 2856-2867.
- Valdeyron G., 1961. *Génétique et amélioration des plantes*. J.B. Baillièere et Fils, Paris, France, 374p.
- Verrier E., Brabant P., Gallais A., 2001. *Faits et concepts de base en génétique quantitative*. Polycopié INA Paris-Grignon, 134p.
- Wright S., 1921. Correlation and causation. *J. Agric. Res.*, 20, 557-585.
- Yule G.U., 1903. Professor Johannsen's experiments in heredity: A review. *New. Phytol.*, 2, 235-242.

## Résumé

**On décrit les aspects fondamentaux de l'évaluation génétique à travers quelques grandes étapes :**

- l'explicitation des concepts-clés de génotype et phénotype,
- le modèle polygénique de Fisher, et la formulation de la covariance entre apparentés,
- les apports d'Henderson qui, couplés avec l'explosion des capacités informatiques, ont permis l'évaluation génétique simultanée de millions d'animaux.

**Ces concepts, modèles et méthodes s'inscrivent dans une approche instrumentaliste qui ne prétend pas refléter la réalité. Leur efficacité a été prouvée à travers la réalisation d'un progrès génétique dans toutes les espèces évaluées.**

## Abstract

---

### *The origin and the development of the concepts of classical genetic evaluation*

Fundamental aspects of genetic evaluation are addressed through some milestones:

- Genesis of paradigmatic concepts of genotype and phenotype,
- The polygenic model of Fisher, and the covariance between relatives,
- Henderson's works, which, hand in hand with the dramatic increase of computing facilities, have enabled us to perform genetic evaluation at a population level.

These concepts, models and methods have been developed within a pragmatic approach. They have been proved efficient thanks to the response to selection in most breeding programs.

LALOË D., 2011. La genèse et le développement des concepts de l'évaluation génétique classique. In : Numéro spécial, Amélioration génétique. Mulsant P., Bodin L., Coudurier B., Deretz S., Le Roy P., Quillet E., Perez J.M. (Eds). INRA Prod. Anim., 24, 323-330.