

Cours de génétique quantitative

Chapitre I :

Génétique quantitative et étude des caractères multifactoriels

Introduction :

Nous avons examiné dans la génétique mendélienne des exemples de variation phénotypiques qui peuvent être classées en catégories distinctes : des plantes de pois grands ou nains, des courges sphériques, en forme de disque ou allongées, ou encore des drosophiles aux yeux rouges ou blancs . Ces types de caractères, identifiés par un petit nombre de phénotypes discrets, montrent des **variations discontinues**. Pour ce type de caractère, un génotype produit en général un seul phénotype identifiable sans ambiguïté. Toutefois, des phénomènes comme le degré de pénétrance du caractère, son niveau d'expression, la **pléiotropie** et l'**épistasie** peuvent rendre floue l'expression du génotype en un phénotype donné, même s'il s'agit de caractères simples et discontinus. Cependant, il existe aussi d'autres types de caractères plus complexes, particulièrement dans le domaine médical ou agricole. Ces caractères présentent une variation continue de leurs phénotypes associés que l'on ne peut pas facilement classer en catégories distinctes. Pour illustrer cet aspect, on peut prendre comme exemple de **variation continue**, la taille ou le poids de l'homme, la production de lait ou de viande du bétail, le rendement d'une cultivée ou encore la quantité de protéines des graines. La variation continue d'une série de phénotypes est mesurée ou décrite en termes quantitatifs et est appelée **hérédité quantitative**. Comme les différents phénotypes proviennent de l'expression de gènes présents à de nombreux loci, la génétique quantitative utilise souvent le terme de caractères **polygéniques** (littéralement "plusieurs gènes"). Pour des caractères présentant une variation continue, le génotype à la fécondation déterminera les limites au sein desquelles se situera, pour ce caractère, le phénotype d'un descendant donné. Cependant, le phénotype final est souvent influencé aussi par des facteurs environnementaux auxquels l'individu est exposé. Par exemple, la taille d'un humain est en partie génétiquement déterminée, mais elle sera aussi affectée par des facteurs environnementaux comme l'alimentation. Les termes **complexes** ou **multifactoriels** seront utilisés pour décrire ces caractères dont l'éventail de phénotypes découle de l'expression de gènes et de l'influence de l'environnement. Dans ce chapitre, nous analyserons des exemples d'hérédité quantitative et quelques techniques statistiques qui permettent de les étudier. Nous

verrons également comment les généticiens évaluent l'importance relative des effets génétiques par rapport aux influences de l'environnement et leurs interactions finales aboutissant à des phénotypes quantitatifs. Enfin, nous étudierons les moyens de cartographie et de caractériser ces caractères quantitatifs.

- **Pléiotropie :**

Jusqu'ici nous avons parlé de la génétique mendélienne comme si chaque contrôlait un seul caractère phénotypique. LA plus part des gènes cependant produisent des effets phénotypiques multiples. Cette faculté est nommée Pléiotropie. Par exemple, le même allèle produit une pigmentation anormale et un strabisme chez le tigre. Chez l'humain, les allèles à l'origine de certaines maladies héréditaires (l'anémie à hématies falciformes) provoquent le plus souvent des symptômes multiples.

- **Epistasie :**

Dans certains cas, un gène situé sur un locus donné agit sur l'expression phénotypique d'un gène, cette interaction se nomme **épistasie** (du grec action de se placer au-dessus).

Exemple : Chez la souris, le pelage noir est dominant par rapport au pelage brun. On désigne ces allèles par **N** et **n**. Pour qu'une souris ait un pelage brun, il faut que son génotype soit **nn**. En outre, un deuxième gène situé sur un autre locus détermine si ce pigment se dépose sur les poils ou non. L'allèle dominant de ce deuxième gène **C** permet au pigment de se déposer. C'est ainsi que la couleur du pelage est soit noir soit brune suivant le génotype du premier gène (NN ou nn). Mais si la souris est homozygote récessive pour le deuxième gène (**cc**), alors le pelage est blanc (albinos) quel que soit le génotype du locus brun-noir.

I. Caractères polygéniques et variations discontinues :

Au sein d'une population, lorsqu'un caractère à variation continue est mesuré, les mesures individuelles montrent une gradation continue dans les phénotypes observables. Il est possible d'identifier une variation continue de phénotypes pour un caractère à déterminisme monogénétique s'il y a différents niveaux de pénétrance pour ce gène entre les individus. Les caractères à déterminisme monogénétique peuvent aussi se révéler multifactoriels si l'interaction entre les allèles et l'environnement produit une multitude de phénotypes différents. Cependant, un

caractère quantitatif à variation continue est le plus souvent le fruit d'une hérédité polygénique et les caractères polygéniques sont fréquemment multifactoriels avec des facteurs environnementaux qui contribuent à la multitude des phénotypes observés.

En plus des caractères quantitatifs à variation continue pour lesquels les variations phénotypiques sont mesurées sur une échelle de valeurs, il existe deux autres classes de caractères polygéniques :

a-Les caractères numériques discontinus : sont ceux pour lesquels les phénotypes correspondent à des nombres entiers. Un caractère numérique sera, par exemple, le nombre de graines dans un fruit ou le nombre d'œufs pondus par une poule au cours de l'année. Il s'agit bien de caractères quantitatifs, mais ils n'ont pas une gamme infinie de variation dans les valeurs. Par exemple, on peut trouver 2, 4 ou 6 graines par fruit mais jamais 5,75.

b-Les caractères à effet de seuil :

sont polygéniques (les facteurs environnementaux affectent fréquemment les phénotypes et ils sont donc aussi multifactoriels), mais ils se distinguent des caractères continus et discontinus par un nombre limité de classes phénotypiques. Ces caractères intéressent particulièrement les généticiens car les maladies humaines sont de plus en plus souvent reliées à ce type d'hérédité polygénique. Par exemple le diabète de type II est appelé diabète de la maturité car il n'affecte les individus qu'à partir de leur quarantième anniversaire environ. Une population peut être divisée en deux catégories pour ce caractère : les individus atteints du diabète de type II et ceux qui ne le sont pas. Ainsi, à première vue, ce caractère pourrait s'assimiler à un caractère monogénique. Cependant, aucun gène responsable de ce caractère n'a été identifié. En revanche la combinaison d'allèles à différents loci dispersés dans le génome contribue à la probabilité variable, pour un individu donné, de développer la maladie. On a ainsi à une extrême, très peu de risque de développer un diabète de type II, alors qu'à l'autre extrême on a pratiquement toutes les chances d'être diabète.

comme pour la plupart des caractères à effet de seuil, les facteurs environnementaux jouent également un rôle important dans le phénotype final : le type d'alimentation et le type de vie ont des impacts significatifs sur le fait qu'un individu, *a priori* prédisposé, développera ou non une forme modérée ou sévère du diabète de type II.

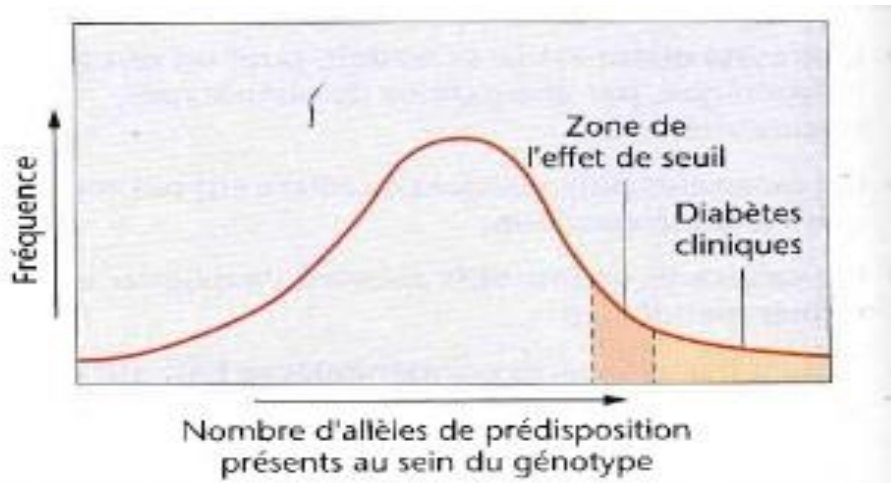


Figure 1.1. Caractères à effet de seuil

2-Explication des caractères quantitatifs en termes mendéliens :

Au début du XX^e siècle, les tentatives pour expliquer les variations phénotypiques continues par l'hérédité mendélienne ont fait l'objet de très nombreuses controverses. Certains scientifiques acceptaient volontiers l'idée que la ségrégation de caractères discontinus, avec des classes de phénotypes clairement identifiées, faisait appel aux notions d'unités génétiques ou de gènes selon la vision mendélienne ; ils n'arrivaient pas pour autant à généraliser cette notion aux caractères plus complexes et notamment aux caractères quantitatifs.

Cependant, les généticiens William et Gudny Yule, tout en restant fidèles aux théories mendéliennes, proposèrent une **hypothèse multifactorielle** ou **multi génique** dans laquelle de nombreux gènes, chacun ayant une hérédité mendélienne, pourraient au développement d'un phénotype sous une forme dite *cumulative* ou *quantitative*.

3- L'hypothèse multigénique de l'hérédité quantitative :

L'hypothèse multigénique a été fondée sur un ensemble d'expériences clés, publiées par Hermann Nilsson-Ehle en 1909, visant à tester l'hypothèse selon laquelle des effets cumulatifs de plusieurs allèles à plusieurs loci pouvaient produire une gamme de phénotype à caractères quantitatifs. L'expérience consistait à croiser du blé à grains rouges avec du blé à grains blanc (figure 1.2). Les plantes F_1 avaient des grains à phénotype intermédiaire rose ce qui pouvait s'interpréter comme une codominance de deux allèles à un même locus.

Cependant, dans la F_2 , Nilsson-Ehle n'observa pas la ségrégation 3 : l'attendue dans le cas du monohybridisme. Au lieu de cela, il observa qu'environ 15/16 des plantes avaient des grains avec divers degrés de rouge et environ 1/16 avaient des grains blanc. Un exemple plus attentif des F_2 révéla que les couleurs de grain pouvaient être classées en 4 catégories de nuances de rouge. Les ratios des phénotypes de la F_2 se comptaient en seizièmes et pouvaient donc s'expliquer par l'intervention de deux gènes, chacun portant deux allèles, ségrégant de manière indépendante et mendélienne.

Si chacun de ces gènes possède un **allèle additif** qui contribue à l'expression de la couleur rouge et un autre, **non additif**, ne produisant pas de pigment, nous pouvons en déduire une hypothèse multifactorielle pour expliquer les différentes couleurs observées. Dans la P_1 , les deux parents sont homozygotes ; le parent à grains rouges contient uniquement des allèles additifs ($AABB$ à la figure 1.2), tandis que le parent à grains blancs ne possède que des allèles non additifs ($aabb$). Les plantes F_1 sont hétérozygotes $AaBb$ et contiennent deux allèles additifs (A et B), deux allèles non additifs (a et b) et expriment un phénotype intermédiaire rose.

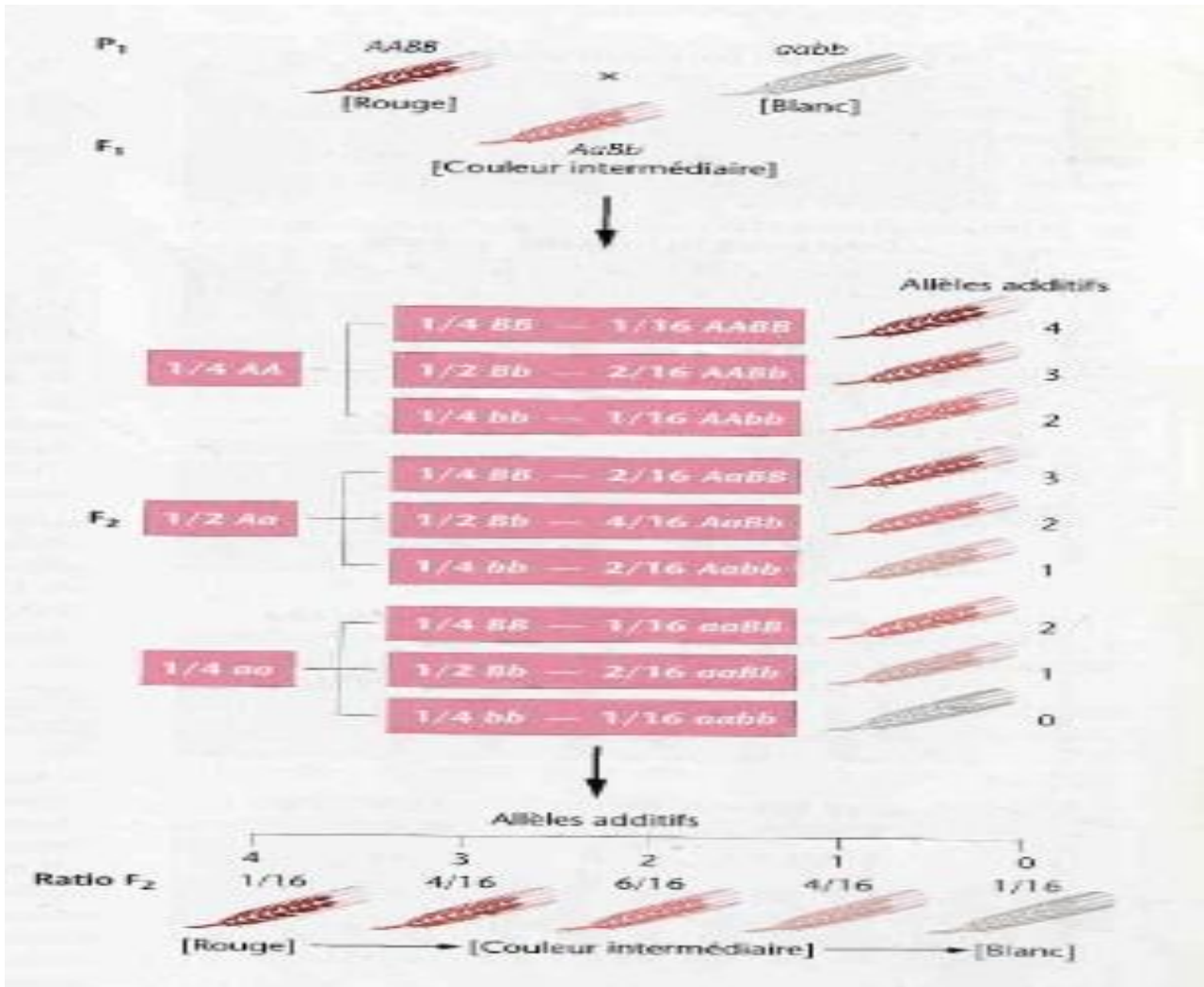


Figure 1.2. Exemple de l'hypothèse multigénique

Dans la F₂, chaque descendant peut avoir 0, 1, 2, 3, ou 4 allèles additifs. Les plantes sans allèle additif sont à grains blanc (*aabb*) comme l'un des parents P₁ et les plantes ayant les 4 allèles additifs sont à grains rouges (*AABB*) comme le deuxième parent P₁. Les plantes ayant 3, 2 ou 1 allèles additif constituent les trois autres catégories de grains rouges observées en F₂. Plus le nombre d'allèles additifs est élevé, plus la couleur rouge est intense car chaque allèle additif contribue de manière identique à la cumulée de pigment présent dans le grain.

Les résultats de Nilsson-Ehle ont donc montré que des variations continues pouvaient s'expliquer des ségrégations mendéliennes d'allèles additifs sur plusieurs loci, chacun influençant le résultat final de manière quantitative.

Le croisement de Nilsson-Ehle montre que si deux loci, comprenant chacun deux allèles, sont impliqués dans un croisement, les 5 catégories phénotypique de la F₂ seront présentes dans les proportions 1 :4 :6 :1. Par extension, on peut concevoir que lorsque 3 ou 4 loci et même plus, sont impliqués dans un croisement, ils contrôleront, de la même manière l'hérédité de caractères quantitatifs. Plus le nombre de loci impliqués est élevé, plus le nombre de classes en F₂ est théoriquement grand et les ratios attendus complexes. Un exemple de cette distribution de phénotypes F₂ pour des croisements avec 5 paires de gènes est illustré à la (figure 1.3).

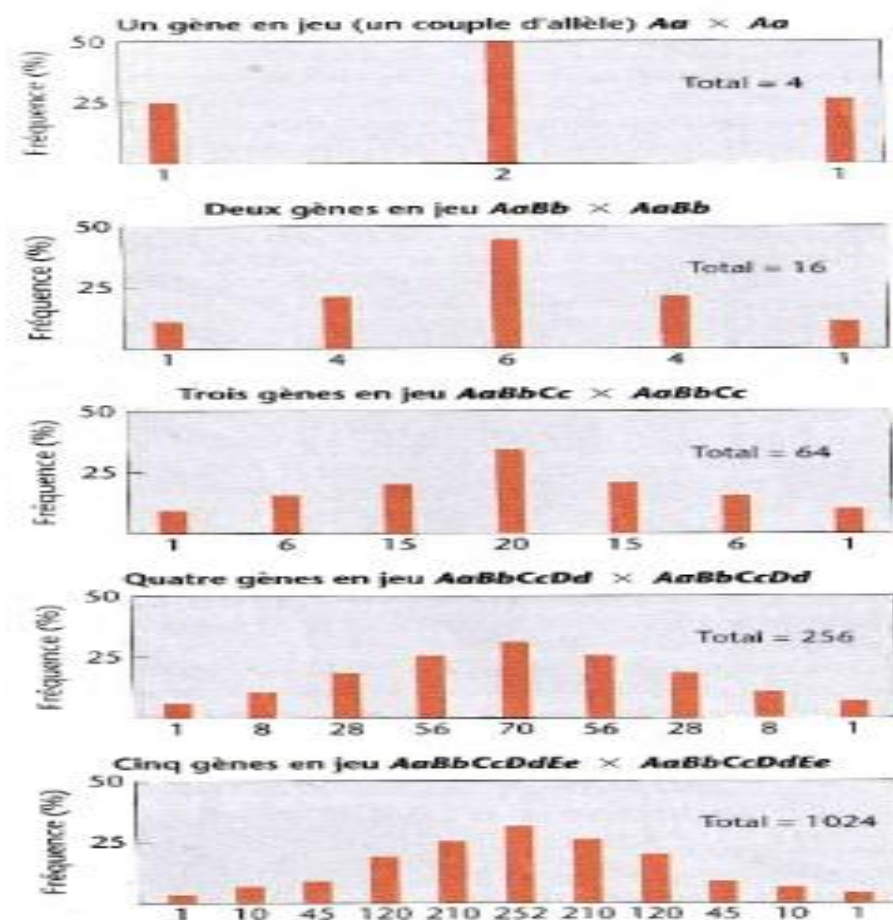


Figure 1.3. Exemple de cette distribution de phénotypes F₂ pour des croisements avec 5 paires de gènes

❖ **Les allèles additifs : la base de la variation continue :**

L'hypothèse multigénique peut se décomposer en plusieurs points importants :

1. Les phénotypes montrant des variations continues peuvent être quantifiés par des mesures, des poids, des comptages, etc.
2. Deux loci ou plus, souvent disséminés dans le génome, interviennent pour contribuer à l'expression d'un caractère le plus souvent sous une forme additive. Comme plusieurs gènes peuvent être potentiellement impliqués, ce type d'hérédité est appelé polygénique.
3. Chaque gène peut contenir un allèle additif qui contribue pour une quantité constante au phénotype, ou un allèle non additif qui ne contribue pas quantitativement à l'expression phénotypique.
4. La contribution au phénotype de chaque allèle additif même faible, est approximativement égale.
5. Un groupe d'allèles additifs impliqués dans la transmission d'un caractère quantitatif simple produit des variations phénotypiques importantes.

❖ Le calcul du nombre de polygènes :

Différentes formules ont été développées pour dénombrer le nombre de polygènes qui contribuent à un caractère quantitatif. Si on peut par exemple déterminer les proportions, dans la F₂, des phénotypes extrêmes (correspondant aux phénotypes parentaux), le nombre de gènes impliqués (*n*) peut être estimé de la manière suivante :

$1/4^n =$ proportion des individus exprimant l'un des deux caractères extrêmes

Si on reprend l'exemple des grains de blé rouges et blancs de la figure 24.2, 1/16 de la descendance présente des grains rouges ou blancs comme l'étaient les parents P₁. Ce rapport peut se substituer dans la partie droite de l'équation pour trouver la valeur de *n* :

$$\frac{1}{4^n} = \frac{1}{16}$$

$$\frac{1}{4^2} = \frac{1}{16}$$

$$n = 2$$

Le tableau (1.1) regroupe les proportions et les nombres de classes phénotypiques observés en F_2 dans des croisements impliquant jusqu'à 5 gènes. Lorsque le nombre de polygènes est faible, il peut être plus facile d'utiliser la formule $(2n + 1)$.

Si n est le nombre de loci additifs impliqués dans un caractère, alors $2n + 1$ sera le nombre total de génotypes possibles.

Par exemple, si $n = 2$, $2n + 1 = 5$ et chaque phénotype sera le résultat de l'action de 4, 3, 2, 1 ou 0 allèles additifs.

Il faut cependant rappeler que pour déterminer le nombre de polygènes impliqués dans un caractère quantitatif par ces deux méthodes simple, il faut respecter deux conditions : la contribution des allèles additifs doit être égale et l'expression phénotypique en F_2 ne doit pas être trop perturbée par les facteurs environnementaux. Nous verrons plus tard que, pour de nombreux caractères quantitatifs, ce postulat n'est pas respecté.

Tableau 1.1 : Détermination du nombre de polygènes (n) impliqués dans un caractère quantitatif

n	individus exprimant l'un des deux phénotypes extrêmes	nombre de classes phénotypiques distinctes
1	$\frac{1}{4}$	3
2	$\frac{1}{16}$	5
3	$\frac{1}{64}$	7
4	$\frac{1}{256}$	9
5	$\frac{1}{1024}$	11

Chapitre 02 :

Etude des caractères polygénique et analyses statistiques

L'une des grandes difficultés pour les généticiens travaillant sur les caractères quantitatifs est de déterminer quelle proportion de la variation phénotypique observée dans une population provient de différent entre les génotypes des individus et quelle proportion provient des effets de l'environnement. Si les méthodes pour décomposer la variance seront discutées plus en détail dans la section suivante de ce chapitre, nous allons tout de suite présenter les outils statistiques de base à utiliser avec des données quantitatives d'observations phénotypiques. Il n'est généralement pas possible de mesurer l'expression d'un caractère polygénique chez tous les individus d'une population, aussi il est nécessaire de définir un sous ensemble d'individus pris au hasard que l'on dénommera **l'échantillon**. Il est important de se souvenir que la précision des informations obtenues dépendra de la pertinence du prélèvement au hasard et du degré de représentation de la population par l'échantillon étudié. Supposons, par exemple, qu'un étudiant veuille estimer la taille moyenne des 100 étudiants de sa classe et que, pour faire cette estimation, il mesure la taille de ses deux voisins immédiats qui jouent dans l'équipe de basket de l'université. Il est peu probable que cet échantillon donne une estimation correcte de la taille moyenne des étudiants de la classe pour au moins deux raisons : premièrement, l'échantillon est trop petit et, deuxièmement, il ne correspond pas à un sous-ensemble représentatif de la classe (sauf si les 100 étudiants sont tous des joueurs de l'équipe de basket !!!).

Si l'échantillon mesuré pour déterminer un caractère quantitatif est suffisamment grand et représentatif de la population de laquelle il est issu, on trouvera alors que les données forment une **distribution normale** qui est caractérisée par une forme en cloche quand elles sont représentées à l'aide d'histogrammes (figure 2.1). Plusieurs paramètres statistiques sont utiles pour décrire une distribution normale, ce sont : la moyenne, la variance, l'écart type, l'erreur standard de la moyenne et la covariance.

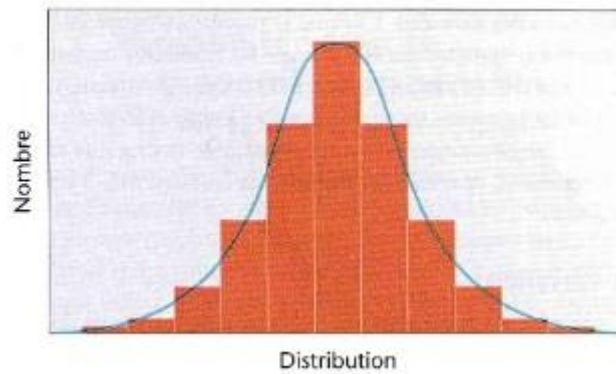


Figure 2.1. Forme d'une distribution normale

- **La moyenne**

Pour une série d'individus, la moyenne donne une information sur la localisation du point central dans la gamme de variation du caractère mesuré.

À la figure (2.4) on voit que les deux distributions de mesures phénotypiques se regroupent autour d'une valeur centrale. Ce regroupement est appelé **tendance centrale** est la **moyenne** \bar{x} .

La moyenne est la moyenne arithmétique d'une série de mesures et la formule pour la calculer est :

$$\bar{X} = \frac{\sum X_i}{n}$$

Où \bar{X} est la moyenne, $\sum X_i$ représente la somme des valeurs individuelles dans l'échantillon et n est le nombre de mesures individuelles.

La moyenne correspond à un descriptif simple de l'échantillon, mais ne renseigne pas sur l'amplitude de variation ou la dispersion des données. Comme le montre la figure (2.2), une distribution symétrique des valeurs de l'échantillon peut être centrée autour de la moyenne. Un deuxième groupe de valeurs peut avoir la même moyenne mais avec une distribution plus étalée. Une deuxième constante statistique donne une idée de la dispersion des valeurs autour de la moyenne : c'est la variance.

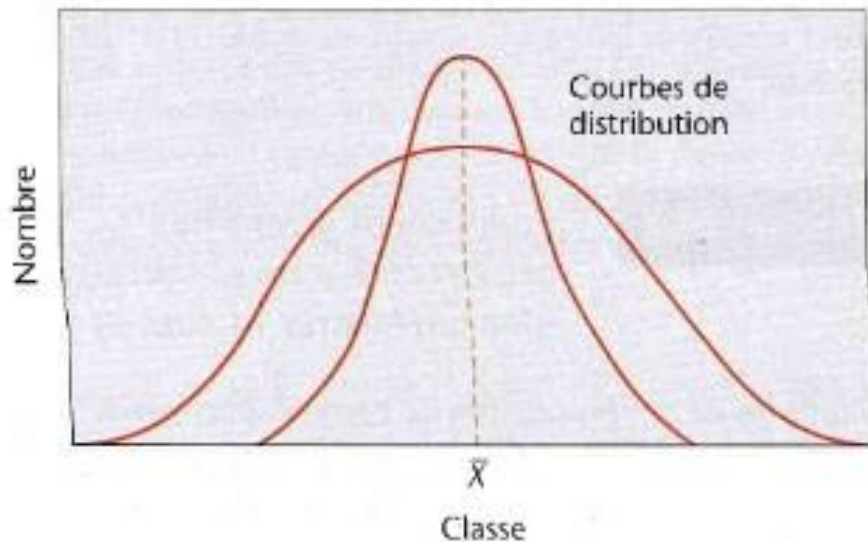


Figure 2.2. Moyenne et distribution des valeurs de l'échantillon

- **La variance**

La variance d'un échantillon est égale à la moyenne des carrés des écarts par rapport à la moyenne. Elle est calculée de la manière suivante :

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

Où la somme (\sum) des carrés à la moyenne entre chaque valeur (X_i) et la moyenne (\bar{X}) est divisée par le nombre total de mesures diminué d'une unité ($n-1$).

Comme le montre la figure (2.3), il est possible pour deux échantillons de mesures d'un caractère quantitatif d'avoir la même moyenne mais des variances différentes. Les estimations des variances peuvent être utiles pour déterminer le degré de contrôle génétique d'un caractère lorsque l'environnement immédiat influence le phénotype.

- **L'écart-type :**

Comme la variance est une valeur basée sur des carrés, ses unités de mesure sont aussi élevées au carré (m^2 , g^2 , etc). Afin d'exprimer cette variation autour de la

moyenne dans les mêmes unités que la valeur mesurée, on peut utiliser la racine carrée de la variance qui est appelée **écart-type** (s).

$$s = \sqrt{s^2}$$

Le tableau 2.2 décrit comment varient les pourcentages d'individus au sein d'une distribution normale en fonction de différentes valeurs de l'écart-type. Par exemple, 68% des valeurs de l'échantillon des individus sont centrées autour de la moyenne plus ou moins une fois l'écart-type. Une variation de deux écart-types de chaque côté de la moyenne représente 95% des valeurs. Ceci sous-entend que l'écart-type peut aussi être interprété comme une probabilité- par exemple, un échantillon pris au hasard aura 68% de chances d'avoir des valeurs dans l'amplitude située écart-type de part et d'autre de la moyenne.

Tableau 2.2 : fonction de répartition des effectifs d'un échantillon pour différentes valeurs de s

Multiples de s	Proportion de l'échantillon incluse (%)
$\bar{X} \pm 1s$	68.3
$\bar{X} \pm 1.96s$	95.0
$\bar{X} \pm 2s$	95.5
$\bar{X} \pm 3s$	99.7

- **L'erreur standard de la moyenne**

Si plusieurs échantillons différents sont prélevés au sein d'une population et que l'on mesure un même caractère quantitatif pour ces échantillons, on peut penser que leurs moyennes varieront. Théoriquement, des échantillons plus importants et plus représentatifs représenteront mieux les vraies valeurs et leurs moyennes seront plus proches les unes des autres. Afin de mesurer la précision de la moyenne estimée, on utilisera l'**erreur standard de la moyenne** ($s_{\bar{X}}$) calculée comme suit :

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Où s est l'écart-type et \sqrt{n} la racine carrée de l'effectif n . du fait l'erreur standard de la moyenne est calculée en divisant s par \sqrt{n} , sa valeur est toujours plus petite que l'écart-type.

- **L'analyse d'un caractère quantitatif**

Une lignée homozygote de tomate produit des fruits d'environ 500 grammes en moyenne alors qu'une autre lignée produit des fruits de 170 grammes en moyenne. La F_1 , obtenue en croisant les deux variétés, produit des fruits dont le poids varie entre 280 et 400 grammes. Les individus de la F_2 produisent des fruits dont le poids varie entre 170 et 500 grammes. Les résultats des deux générations sont reportés dans le tableau 2.3

La valeur moyenne du poids des fruits en F_1 peut être calculée de la manière suivante :

$$\bar{X} = \frac{\sum X_i}{n} = \frac{17754}{52} = 341,42$$

De la même, la valeur moyenne du poids des fruits en F_2 se calcule comme suit :

$$\bar{X} = \frac{\sum X_i}{n} = \frac{24731}{72} = 343,49$$

Bien que ces valeurs moyennes soient similaires, la distribution des fréquences du tableau 2.3 montre une plus grande dispersion pour les valeurs de la F_2 . Cette dispersion peut être mesurée par la variance d'échantillon s^2 calculée comme la somme des carrés des écarts à la moyenne divisée par le nombre d'observations moins 1 :

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

La variance d'échantillon obtenue est de 1039.23 pour la génération F₁ et de 3428.23 pour la F₂. Transformée en écart-type ($s = \sqrt{s^2}$), ces valeurs des poids est en F₁ de 341.42 ± 32.24 et en F₂ de 343.49 ± 58.56 .

Si on considère que les deux lignées sont homozygotes aux loci considérés et que les allèles contrôlant le poids des fruits sont additifs, on peut estimer le nombre de polygènes impliqués dans le contrôle de ce caractère. Comme 1/72 de la F₂ a un phénotype parental (72 fruits en F₂, un pèse 170 grammes et un pèse 510 grammes, voir tableau 2.3), l'utilisation de la formule $1/4^n = 1/72$ indique que n est entre 3 et 4, ce qui donne une information sur le nombre de gènes impliqués dans le contrôle du poids du fruit dans ces lignées de tomates.

Tableau 2.3 : distribution des poids dans les descendance F₁ et F₂

		Poids (en grammes)												
		170	198	227	255	284	312	340	369	397	425	454	482	510
nombre	F ₁					4	14	16	12	6				
d'individus	F ₂	1	1	2	0	9	13	17	14	7	4	3	0	1

❖ Héritabilité : mesure de la contribution génétique à la variation phénotypique :

Une question récurrente posée par les généticiens travaillant avec des caractères multifactoriels concerne la part relative des facteurs génétiques par rapport aux effets de l'environnement sur la variation phénotypique observée entre les individus.

- **Notion d'héritabilité :** Le terme **héritabilité** est utilisé pour désigner la proportion de la variation phénotypique totale qui est due aux facteurs génétiques. Pour un caractère multifactoriel d'une population donnée, une héritabilité élevée indique que la variation phénotypique est majoritairement liée à des facteurs génétiques et que l'environnement a moins d'effet sur

l'expression de ce caractère. Au contraire, une héritabilité faible traduit un impact important de l'environnement sur l'expression du caractère au sein de la population.

Le concept d'héritabilité est fréquemment mal compris et mal utilisé. Il est important de préciser qu'il n'indique pas quelle proportion d'un caractère est déterminée génétiquement ou quelle part du phénotype d'un individu est déterminée par son génotype. Ces dernières années, une mauvaise interprétation de l'héritabilité, pour des caractères quantitatifs humains, a conduit à des controverses, en particulier dans l'interprétation des mesures du quotient intellectuel ou QI. Les variations dans les estimations d'héritabilité des QI entre différents groupes raciaux ont conduit à des conclusions incorrectes qui tendaient à dire que des bases génétiques étaient à l'origine de différences de niveau d'intelligence mesurées entre les groupes raciaux. Ces insinuations interprètent mal le concept d'héritabilité et ignorent notamment l'impact de la relation génotype-environnement sur la variation phénotypique dans une population. De plus, l'héritabilité n'est pas constante pour un caractère donné. Par exemple, la valeur d'héritabilité de la production d'œufs dans un groupe de poules maintenues en cages individuelles peut s'avérer élevée, indiquant que les différences de productions d'œufs entre les individus sont largement dues à des différences génétiques dans la mesure où les poules sont maintenues dans des environnements similaires. Pour un autre groupe de poules, élevée en plein, l'héritabilité de la production en œuf peut se révéler plus faible et ces différences peuvent refléter la variation d'environnement. Ces différences peuvent inclure la quantité de nourriture que les volatiles arrivent à trouver et la manière dont ils entrent en compétition pour trouver un bon perchoir afin d'y passer la nuit. L'héritabilité nous indique donc la proportion de la variation phénotypique qui peut être attribuée à des variations génétiques pour une population donnée dans un environnement donné. Si nous mesurons l'héritabilité pour le même caractère dans plusieurs populations vivant dans des environnements différents, nous trouverons fréquemment que les valeurs d'héritabilité calculées ont des écart-type élevée. Ceci est un point important

dont il faut se souvenir, en particulier pour le calcul d'héritabilité au sein de populations humaines. Une valeur moyenne d'héritabilité de 0.65 pour la taille humaine ne signifie pas que votre taille est à 65% contrôlée par vos gènes, mais plutôt que dans l'échantillon de population étudié, en moyenne 65% de la variation de la stature peut être attribuée à des différences génotypiques entre individus.

1) Composantes de la variance phénotypique:

Nous pouvons considérer maintenant comment les généticiens séparent la variation phénotypique dans une population, en facteurs génétiques et environnementaux. Comme nous l'avons vu dans les parties précédentes, cette variation peut être quantifiée sous forme de variance d'échantillon : c'est-à-dire en prenant des mesures du caractère en question à partir d'un ensemble représentatif de la population et en déterminant la dispersion de ces mesures autour de la moyenne. Ceci nous donne une estimation de la variance phénotypique totale (V_P). Les estimations de l'héritabilité sont obtenues en utilisant différents dispositifs expérimentaux et des tests statistiques pour diviser v_p en **variance génotypique** (V_G) et **variance environnementale** (V_E).

Un troisième facteur qui contribue à des variations phénotypiques provient de la manière dont un génotype se traduit par un phénotype différent suivant l'environnement dans lequel il se réalise. Par exemple, la variété de blé A peut produire en moyenne 1260 kg/ha sur un sol pauvre alors que la variété B en produira en moyenne 1071 kg/ha. Sur un bon sol, la variété A produira 1386 kg/ha tandis que la variété B produira 1575 kg/ha. Il existe des différences de rendement entre les deux variétés génétiquement différentes donc la variation de production a une composante génétique. Les deux variétés produisent plus sur un sol de bonne qualité donc la production est aussi influencée par l'environnement. Cependant, nous voyons que les deux variétés ne répondent pas de la même manière à l'amélioration du sol. Les génotypes de la variété B ont une progression de rendement beaucoup plus forte et produisent plus que la variété A lorsque le sol est bon. Nous avons ainsi des différences d'interactions entre génotype et environnement qui contribuent aux

différences de rendements. Cette troisième composante de la variation phénotypique est la **variance d'interaction génotype-environnement** ($V_{G \times E}$) (figure 2.4).

Nous pouvons maintenant résumer toutes les composantes de la variance phénotypique totale V_P par l'équation suivante :

$$V_P = V_G + V_E + V_{G \times E}$$

En d'autres termes, la variation phénotypique totale peut être subdivisée en variance génotypique, variance environnementales et variance d'interaction génotype-environnement. Pour mesurer l'héritabilité d'un caractère multifactoriel, les chercheurs estiment souvent que la variance d'interaction génotype-environnement est faible ou assez faible pour être soit ignorée soit incluse dans la variance environnementale.

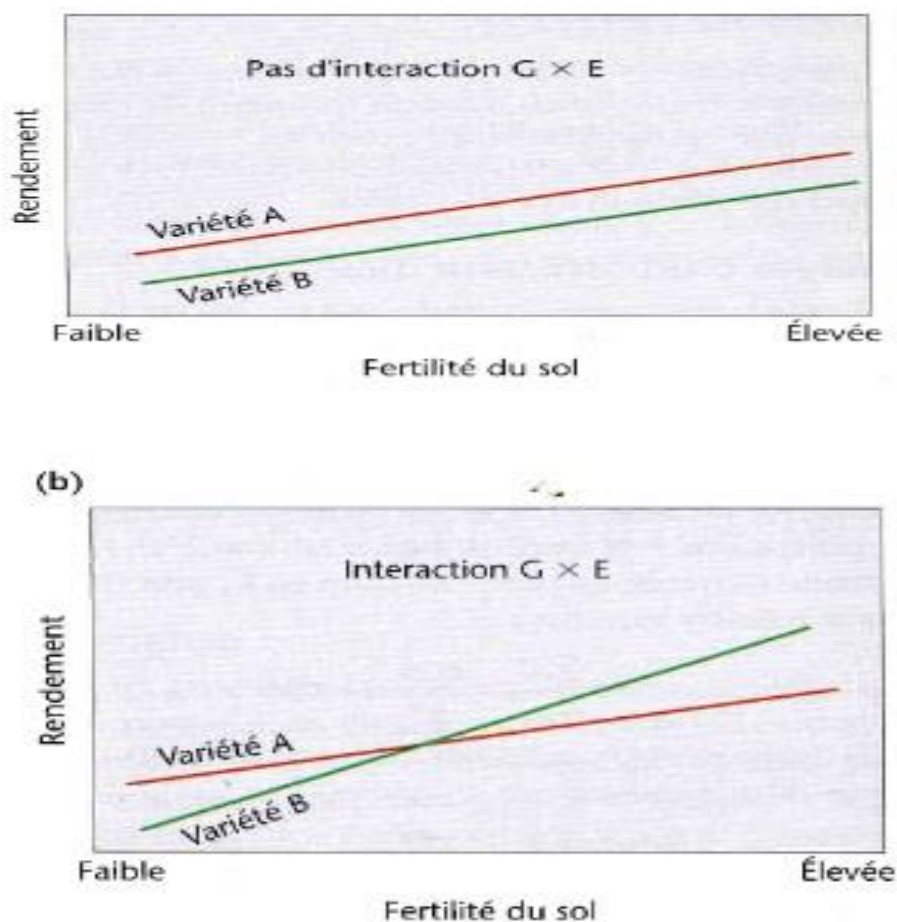


Figure 2.4. Exemple de la variance d'interaction génotype-environnement

Cependant, il faut se souvenir que ce type d'approximation justifie que l'héritabilité est estimée pour une population donnée et ne constitue pas un attribut fixe pour un caractère.

Les sélectionneurs d'animaux ou de plantes utilisent une palette de techniques expérimentales pour évaluer les héritabilités en décomposant les mesures des variances phénotypiques en leurs composantes génotypiques et environnementales. Une des approches consiste à utiliser des souches pures ne contenant que des individus génétiquement homogènes avec des génotypes très largement homozygotes pour tester l'impact de différentes conditions environnementales sur la variabilité phénotypique. Les variations entre individus provenant de souches différentes élevées dans un environnement équivalent proviennent probablement de facteurs génétiques. Au contraire, les variations entre individus d'une même lignée élevés dans des environnements différents sont plus probablement dues à des facteurs environnementaux. D'autres approches impliquent des analyses de variance pour un caractère quantitatif entre descendants de différents croisements, ou la comparaison de l'expression de différents caractères entre parents et descendants élevés dans le même environnement.

L'héritabilité au sens large :

L'héritabilité au sens large (représentée par le terme H^2) mesure la contribution de la variance génotypique à la variation phénotypique totale :

$$H^2 = \frac{V_G}{V_P}$$

Ces valeurs de l'héritabilité pour un caractère dans une population s'échelonnent entre 0 et 1. Les valeurs proches de 1 indiquent que les conditions environnementales ont peu d'effet sur la variance phénotypique qui est alors très largement due aux différences génotypiques entre individus au sein de la population. Les valeurs faibles ou proches de 0 indiquent que des facteurs environnementaux et non génétiques sont principalement responsables des différences phénotypiques observées au sein de la

population étudiée. Peu de caractères quantitatifs ont en fait des valeurs très fortes ou très faibles suggérant que l'environnement et les génotypes contribuent tous les deux à l'expression de la majorité des phénotypes.

La composante de variance génétique V_G utilisée pour mesurer l'héritabilité au sens large inclut tous les types de variation génétique dans la population : il n'y a pas de distinction entre les caractères quantitatifs polygéniques déterminés par des allèles à effets additifs, dominant ou épistatiques. L'estimation de l'héritabilité au sens large considère également que la variance génotype-environnement est négligeable. De ce fait, l'héritabilité au sens large d'un caractère est surtout intéressante du point de vue général, mais les différentes limitations énoncées font que cette mesure de l'héritabilité est peu utilisée dans les programmes de croisement. Les sélectionneurs d'animaux ou de plantes, qui veulent améliorer des caractères spécifiques chez des animaux ou augmenter le rendement de certaines plantes cultivées, ont besoin d'une estimation plus précise de l'héritabilité des caractères qu'ils souhaitent manipuler au sein d'une population. Un autre estimateur, l'héritabilité au sens étroit, a donc été élaboré.

L'héritabilité au sens étroit :

L'héritabilité au sens étroit (h^2) est la proportion de la variance phénotypique due aux seuls additifs de la variance génotypique. La variance génotypique peut être décomposée en composantes reflétant les différents modes d'action des allèles de plusieurs loci impliqués dans un caractère quantitatif. Comme tous les gènes impliqués dans un caractère quantitatif ne modifient pas le phénotype de la même manière, cette répartition permet de distinguer trois modes d'action génétique différents. La **variance additive**, V_A , est la variance génotypique liée à l'action additive des allèles aux loci quantitatifs. La **variance de dominance**, V_D , est l'écart à l'additivité lorsque l'expression du phénotype d'un hétérozygote s'écarte de la stricte somme des effets individuels des allèles des homozygotes de départ. La **variance d'interaction**, V_I , est l'écart à l'additivité lorsque deux loci ou plus

interagissent en épistasie. Le niveau de la variance d'interaction est souvent négligeable et cette composante peut être exclue des calculs de variance génotypique.

La partition de la variance génotypique V_G peut être résumée par l'équation suivante :

$$V_G = V_A + V_D + V_I$$

Et l'héritabilité au sens étroit, estimée uniquement sur la portion de variance génotypique due aux effets additifs, devient :

$$h^2 = \frac{V_A}{V_P}$$

en omettant V_I et en séparant V_P entre ses composantes de variance génotypique et environnementale, nous obtenons pour h^2 la formule suivante :

$$h^2 = \frac{V_A}{V_E + V_A + V_D}$$

Les estimations de l'héritabilité sont utilisées pour la sélection des animaux et des plantes afin d'identifier comment une population réagit à la sélection artificielle pour un caractère quantitatif. L'héritabilité au sens étroit, h^2 , donne une prédiction plus faible de la réponse à la sélection que l'héritabilité au sens large, H^2 et de ce fait h^2 est plus souvent utilisée par les sélectionneurs.

❖ La sélection artificielle :

La **sélection artificielle** est le processus qui permet de choisir quelques individus ayant des phénotypes souhaités à partir d'une population initialement hétérogène en vue d'un programme de croisement. Théoriquement, si la sélection artificielle d'un même caractère se fait sur plusieurs générations, on pourra obtenir une population ayant une fréquence élevée d'individus avec le caractère souhaité. Lorsque la sélection est effectuée sur un caractère simple, contrôlé uniquement par un ou deux gènes peu influencés par les conditions environnementales, on pourra générer facilement et rapidement la population animale ou végétale avec les caractéristiques

souhaitées. Cependant, nombre de caractères d'intérêt économique pour le bétail ou les plantes, comme la production de grain chez les plantes, le poids des grains, la production de lait pour une vache ou encore la vitesse ou la vigueur chez un cheval, sont polygéniques et fréquemment multifactoriels. La sélection artificielle pour ces caractères est évidemment plus lente et plus complexe. Les estimations de l'héritabilité au sens étroit sont utiles pour le sélectionneur de plantes ou d'animaux, car elles permettent de mesurer la proportion de la variance phénotypique totale du caractère liée à la variance additive. Les allèles additifs impliqués dans des caractères quantitatifs sont plus facilement manipulables par le sélectionneur. Les allèles qui génèrent des effets de dominance ou d'épistasie (et qui de ce fait changent les valeurs de V_D et V_I) répondent moins bien à la sélection artificielle. C'est pourquoi l'héritabilité au sens étroit, h^2 , peut être utilisée pour prédire les effets de la sélection. Plus la valeur d'héritabilité de type h^2 est élevée pour une population, plus la variation phénotypique du caractère observé par le sélectionneur ne sera grande.

La subdivision de la variance génétique en trois éléments pour calculer h^2 et prédire la réponse à la sélection est une tâche difficile qui les expériences et les analyses soient réalisées avec soin. L'approche la plus simple est de sélectionner, à partir d'une population hétérogène, des individus avec un phénotype supérieur pour le caractère considéré et de croiser leur descendance. L'intensité moyenne du caractère obtenu pourra être (1) à la population de départ (M), (2) aux individus sélectionnés et utilisés comme parent (M_1) et (3) à la descendance qui résulte du croisement entre les parents sélectionnés (M_2). La relation entre ces 3 moyennes et h^2 est donnée par la formule suivante :

$$h^2 = \frac{M_2 - M}{M_1 - M}$$

Cette équation peut être simplifiée en définissant $M_2 - M$ comme la **réponse** (R) et $M_1 - M$ comme la **sélection différentielle** (S) et h^2 reflète donc le rapport entre la réponse observée et la totalité des réponses possible. Ainsi,

$$h^2 = \frac{R}{S}$$

L'héritabilité au sens étroit calculée par cette formule, à partir de croisements sélectifs et de la mesure des réponses dans la descendance, s'appelle **l'héritabilité réalisée**.

Comme exemple de cette hérabilité réalisée, supposons que nous mesurons le diamètre des grains de maïs dans une population pour laquelle le diamètre moyen M est de 20 mm. Au sein de cette population, nous sélectionnons un groupe ayant des diamètres inférieurs, pour lequel le diamètre moyen M_1 mesuré est de 10 mm. Les plantes sélectionnées seront croisées entre elles et la moyenne M_2 des diamètres dans la descendance sera estimée à 13 mm. Nous pouvons calculer l'héritabilité réalisée h^2 pour estimer le potentiel de sélection artificielle réalisé pour la taille des grains :

$$h^2 = \frac{M_2 - M}{M_1 - M}$$

$$h^2 = \frac{13 - 20}{10 - 20}$$

$$= \frac{-7}{-10}$$

$$= 0.70$$

Cette valeur de l'héritabilité au sens étroit indique que le potentiel de sélection pour la taille des grains est assez élevé.

La sélection artificielle la plus longue et la plus connue a été menée laboratoire d'agronomie de l'état de l'Illinois. Depuis 1896, le maïs y est sélectionné pour deux caractères : fortes et basses teneurs en huile dans les grains. Après 76 générations, la sélection permet toujours une augmentation de la teneur en huile (figure 2.5).

Au cours de ces cycles de sélection, un nombre croissant de plants de maïs accumulent un pourcentage croissant d'allèles impliqués dans la production d'huile.

Par conséquent, l'héritabilité au sens étroit h^2 pour la teneur en huile dans les générations successives a baissé (voir les valeurs parentales aux générations 9, 25, 52, et 76 à la figure 2.5) en même temps que la sélection artificielle se rapproche de l'optimum du potentiel génétique pour la production d'huile. Théoriquement, le processus continuera jusqu'à ce que tous les individus de la population possèdent un génotype uniforme, comprenant tous les allèles additifs responsables de l'augmentation de la teneur en huile. A ce point, h^2 sera réduit à 0 et la réponse à la sélection artificielle cessera. La décroissance de la réponse à la sélection pour les faibles teneurs en huile montre que l'héritabilité pour ce caractère s'approche déjà de ce point.

Le tableau 2.4 donne les valeurs de l'héritabilité au sens étroit, exprimées en pourcentage pour un ensemble de caractères dans différents organismes. Comme vous pouvez le voir, les valeurs de h^2 varient énormément, mais elles ont cependant tendance à être plus faibles pour les caractères quantitatifs qui sont essentiels à la survie des organismes. Rappelons que ceci ne signifie nullement une absence de contribution génétique pour ces phénotypes observés. Au contraire, les valeurs faibles de h^2 signifient que la sélection naturelle a déjà beaucoup optimisé les composantes génétiques de ces caractères au cours de l'évolution. La production d'œufs, la taille d'une portée et le rythme de fécondation sont des exemples où des limitations physiologiques à l'évolution sont déjà en place.

Les caractères moins critiques pour la survie, comme le poids du corps, la taille de la queue, les démentions des ailes, ont des héritabilités supérieures car, pour ces caractères, les populations ont des variations génotypiques plus importantes. Rappelons enfin que l'héritabilité ne donne des informations que sur une population dans un environnement spécifique. Par conséquent, l'héritabilité au sens étroit sera une mesure plus prédictive de la sélection lorsque les données seront collectées sur plusieurs populations, plusieurs environnements et avec des phénotypes bien marqués.



Figure 2.6. Réponse du maïs à la sélection artificielle

Tableau 2.4 : estimations de l'héritabilité pour des caractères chez différents organismes :

Espèce	caractère	Héritabilité (h^2)
Souris	Longueur de la queue	60%
	poids du corps	37%
	nombre de souriceaux dans la portée	15%
Poulets	Poids du corps	50%
	production d'œufs	20%
	capacité d'éclosion	15%
bétail	Poids a la naissance	45%
	Production de lait	44%
	Taux de reproduction	3%

Chapitre:03

Introduction à l'analyse des QTLs

1. Définition

Un **locus de caractères quantitatifs** (LCQ ou **QTL** pour *quantitative trait loci*) est une région plus ou moins grande d'ADN qui est étroitement associée à un caractère quantitatif, c'est-à-dire une région chromosomique où sont localisés un ou plusieurs gènes à l'origine du caractère en question.

L'hérédité de caractères quantitatifs se rapporte à une caractéristique phénotypique qui varie par degrés, et qui peut être attribuée à l'interaction entre deux ou plusieurs gènes et leur environnement (appelé aussi hérédité polygénique).

Les QTL peuvent être identifiés « *moléculairement* » (par PCR, par exemple) pour aider à cartographier des régions du génome qui contiennent des gènes impliqués dans la spécification d'un caractère quantitatif. Ceci contribue à l'identification, l'annotation et le séquençage de ces gènes ou de gènes dits « gènes d'intérêt ».

Caractère quantitatif

Un caractère qui présente une variation continue au sein d'une population (caractère quantitatif) est supposé être contrôlé par plusieurs gènes à effet faible : modèle infinitésimal. Cependant, il a été montré que quelques gènes majeurs (QTL) pouvaient participer de façon significative à cette variabilité (modèle oligogénique).

L'analyse QTL

L'analyse QTL cherche à caractériser l'architecture génétique d'un type de caractère : c'est-à-dire à déterminer le nombre de régions du génome impliquées ainsi que leurs positions et leurs effets. Cette approche se base sur l'analyse combinée de

l'information moléculaire et d'un caractère quantitatif dans une descendance en ségrégation. Elle permet de tester statistiquement le lien entre variation génétique (comme celle de marqueurs moléculaires) et variation phénotypique. Si ce test est significatif, on met en évidence un QTL.

Partant d'une analyse simple marqueur (analyse de variance à un facteur), l'analyse QTL a évolué vers des méthodes de plus en plus complexes :

1. "l'interval mapping" (IM) : utilisant l'information de la carte génétique pour positionner les QTL le long du génome
2. le "composite interval mapping" (CIM)
3. le "multiple interval mapping" (MIM) qui prennent compte des cofacteurs (marqueurs significativement associés au caractère)

Parallèlement à ces méthodes classiques, des approches bayésiennes ont également été développées mais restent pour l'instant peu utilisées chez les plantes.

Comme pour toutes les analyses statistiques, la taille de l'échantillon est un facteur critique. Pour des petites tailles d'échantillon, le risque de ne pas détecter de QTL à effet faible est important. Cela surestime donc l'effet de ceux qui sont détectés. C'est ce que l'on appelle communément l'effet "Beavis" .

2. cartographie des caractères quantitatifs :

L'analyse de pedigree décrite précédemment n'est pas très efficace pour identifier les gènes multiples impliqués dans les caractères quantitatifs. Les effets de l'environnement, les interactions entre allèles en ségrégation et le nombre de gènes contribuant au phénotype polygénique empêchent d'isoler la contribution phénotypique d'un gène donné. Cependant, beaucoup de gènes ont une certaine importance économique et il est alors utile de les identifier et de déterminer leur localisation sur le génome. Les gènes multiples impliqués dans la formation d'un caractère quantitatif donné sont connus sous le terme de *Quantitative Trait Loci* (QTL) l'identification de ces loci donne des informations sur le nombre de gène

impliqués dans un caractère quantitatif et sur leurs contributions respectives plus ou moins fortes dans l'apparition de ce caractère. La cartographie révèle comment les QTL associés à un caractère sont regroupés sur un seul chromosome ou dispersés sur la totalité du génome.

Pour trouver et cartographier les QTL, les chercheurs essaient d'associer des séquences particulières d'ADN dans le génome avec des phénotypes qui correspondent à une classe particulière de ce phénotype quantitatif. Pour ce faire, on peut croiser des lignées homozygotes avec des phénotypes extrêmement différents pour le caractère d'intérêt et produire une génération F_1 dont les individus vont être hétérozygotes pour la quasi-totalité des loci contribuant au caractère. Des croisements supplémentaires entre individus F_1 ou entre individus F_1 et lignées parentales donneront des générations F_2 avec des degrés élevés de ségrégation pour les différents QTL et leurs phénotypes associés. Cette génération F_2 est connue sous le terme de **population de cartographie**. Les chercheurs mesurent l'expression du caractère chez tous les individus de la population de cartographie et identifient par ailleurs tous les génotypes individuels *via* l'utilisation de marqueurs moléculaires comme des RFLP et microsatellites. Les corrélations entre les marqueurs moléculaires génotypiques et les variations phénotypiques pour le caractère sont calculées à l'aide d'outils statistiques et de logiciels spécifiques. Si un marqueur ADN n'est pas lié à un QTL, alors la moyenne phénotypique pour ce caractère ne variera pas entre les individus ayant des génotypes différents au locus considéré. A l'inverse, si le marqueur est lié au QTL, l'expression du caractère considéré sera variable suivant les différents génotypes associés à ce locus. Lorsque ce cas se présente, on dit que le locus marqueur et le QTL sont en *co-ségrégation*. Des co-ségrégations véridiques démontrent la présence d'un QTL contiguë ou proche du marqueur ADN le long du chromosome : le marqueur et le QTL sont alors liés. Lorsque plusieurs QTL pour un caractère donné ont été localisés, une carte génétique est créée : elle donne les positions chromosomiques relatives des gènes impliqués.

On dispose à l'heure actuelle de très nombreux marqueurs ADN d'organisme d'importance agronomique, ce qui rend possible la cartographie systématique des QTL. Par exemple, des centaines de marqueurs RFLP ont été localisées chez la tomate. Des analyses de QTL ont été réalisées en croisant des plantes avec des phénotypes extrêmes et opposés ainsi qu'en étudiant leurs croisements sur plusieurs générations. De nombreux loci responsables de caractères quantitatifs ont été identifiés : ils contrôlent la taille du fruit, sa forme, la qualité de matières solubles et l'acidité. Ces loci sont distribués sur les 12 chromosomes représentant le génome haploïde de la plante. Plusieurs de ces chromosomes contiennent des loci contrôlant des caractères multiples.

La cartographie et la caractérisation des QTL chez la tomate ont été un des grands succès des recherches menées par Steven Tanksley et ses collègues à l'université de Cornell aux Etats Unis. Leurs observations ont montré qu'environ 30 QTL contribuent aux dimensions et à la forme de la tomate. Cependant, ces QTL n'entraînent pas forcément des effets identiques : 10 gènes, dispersés sur 7 chromosomes, suffisent pour engendrer la quasi-totalité de la variabilité phénotypique pour ces caractères. Différents allèles du même locus, *fw 2.2* sur le chromosome 2, peuvent changer la taille du fruit de 30%. *Fw2.2* a été isolé, cloné et transféré dans d'autres plantes avec des résultats tous à fait intéressants. Bien qu'une tomate puisse peser jusqu'à près de 1000 grammes, les fruits de l'espèce sauvage apparentée, probablement ancêtre de la forme cultivée moderne, ne pèsent que quelques grammes. Deux allèles distincts de *fw2.2*, identifiés grâce à des RFLP cartographiés, jouent certainement un rôle prédominant dans la tomate. Un des allèles est présent dans tous les plants de tomate sauvage à petits fruits, l'autre dans tous les plants de la tomate domestiquée à gros fruits. Lorsque l'allèle *fw2.2* donnant des petits fruits est transféré chez une plante qui normalement produit de gros fruits, la plante transformée produit des tomates dont le poids est considérablement réduit (voir figure 3.1). Chez les variétés étudiées par le groupe de Tanksley, la diminution était en moyenne de 17 grammes, une différence phénotypique significative causée par un seul QTL. D'autres analyses de *fw2.2* ont

montré que l'allèle responsable des gros fruits à ce locus code pour une protéine inhibitrice de la répression de la division cellulaire. Le résultat est une augmentation des mitoses dans le tissu cortical qui forme le fruit. Les variétés à petits fruits ont un allèle *fw2.2* qui code pour la même protéine mais qui a des mutations dans la région promotrice, ce qui entraîne la réduction du niveau de transcription de *fw2.2*, la diminution de la vitesse de la division cellulaire et au final des fruits plus petits.

La cartographie par marqueurs moléculaires permet d'identifier des régions chromosomiques contenant des QTL, mais l'identification des gènes individuels impliqués dans des caractères quantitatifs demande des recherches supplémentaires. Une des approches possibles consiste à construire des marqueurs chromosomiques étroitement liés à un gène pour ensuite réaliser du clonage positionnel et identifier le gène lui-même. Il est clair qu'isoler et caractériser l'ensemble des gènes impliqués dans un caractère quantitatif est toujours un travail long et complexe. *fw2.2* est actuellement l'un des rares gènes isolé, cloné et étudié dans le détail. Cependant, la cartographie de QTL et l'identification des gènes présents dans les régions chromosomiques correspondant à ces QTL, chez les plantes à fort intérêt agronomique, apporteront un réel progrès dans les programmes d'amélioration des plantes cultivées. Ces études ouvriront aussi la voie vers des recherches similaires sur les gènes animaux, y compris ceux de l'espèce humaine.