**Introduction**

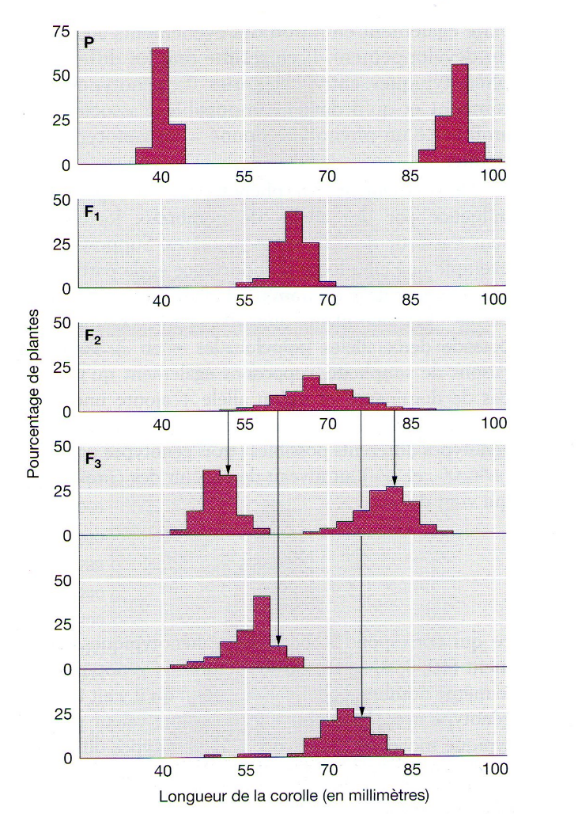
**Place de la génétique quantitative dans les sciences de l’hérédité**

La **génétique quantitative** est la génétique des caractères qui peuvent donner lieu à des [mesures](https://fr.wikipedia.org/wiki/Mesure), que ce soient des caractères à variation continue (tels que le poids ou la taille d'un organisme) ou discontinue. La génétique quantitative s’appuie sur la [**génétique des populations**](https://fr.wikipedia.org/wiki/G%C3%A9n%C3%A9tique_des_populations) et les [**statistiques**](https://fr.wikipedia.org/wiki/Statistique)**.**

**Gregor Mendel**avait observé que, chez le haricot, le croisement de variétés à fleurs blanches et à fleurs rouges donnait « *toute une série de couleurs allant du pourpre au violet pâle et blanc* ». Pour expliquer ce phénomène, Mendel proposa de considérer que la couleur de la fleur du haricot est un mélange de caractères élémentaires, la transmission de chacun suivant les mêmes règles que celles mises en évidence chez le pois, suggérant ainsi ce qu’aujourd’hui nous décrivons comme l’intervention de plusieurs**loci**pour rendre compte de la variation continue de certains caractères. Plus tard, des sélectionneurs de plantes comme l’Américain East (1903) ou le Suédois Nielson Ehle (1908) firent des expériences étayant cette hypothèse multifactorielle. Il fallut cependant attendre 1918 pour que le statisticien anglais [Ronald Fisher](https://fr.wikipedia.org/wiki/Ronald_Fisher) propose un modèle de synthèse, rendant à la fois compte des lois de Mendel et des relations biométriques entre apparentés. L'analyse de [loci de caractères quantitatifs](https://fr.wikipedia.org/wiki/Locus_de_caract%C3%A8res_quantitatifs) (QTL) est un ajout plus récent à l'étude de la génétique quantitative.

Jusqu’à présent, nous avons examiné des exemples de variations 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 typede 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 plante cultivée ou encore la quantité de protéines dans les 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. 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 gènes et les caractères quantitatifs**



**La variation des caractères quantitatifs augmente dans la descendance**

**Figure 1.**Lesrésultats de croisements entre des souches de *Nocotiana longiflora* (une variété de tabac réputée pour le parfum de ses fleurs) qui diffèrent par la longueur de leur corolle. Les graphes représentent (de haut en bas) la distribution des fréquences de longueur de la corolle dans les deux souches parentales (P), dans la F1, dans la F2 et dans quatre croisements produisant une F3, réalisés à partir de parents issus des quatre parties indiquées dans la distribution de la F2. (D’après K.Mather, Biometrical Genetics. Methuen, 1959).

L’expérience décrite dans la figure 1est un exemple classique des résultats de croisements entre des souches différant au niveau d’un caractère quantitatif. La longueur de la corolle a été mesurée chez de nombreuses plantes différentes issues de deux lignées pures de *Nicotiana longiflora,* une espèce apparentée au tabac. La distribution des longueurs des corolles de deux lignées parentales est indiquée dans le schéma du haut de la figure. La différence entre les deux lignées est génétique mais la variation entre les plantes d’une même lignée est due à une variation environnementale incontrôlée et au bruit de fond développemental. Les plantes de la F1 dont la longueur moyenne de la corolle est très proche de la moyenne entre les deux lignées parentales. Varient également de l’une à l’autre en raison d’une variation environnementale et développementale. Dans la F2, la longueur moyenne de la corolle reste quasiment inchangée par rapport à celle de la F1 mais il y a une forte augmentation de la variation car il y a à présent une ségrégation des différences génétiques héritées des deux lignées parentales d’origine. On voit dans la F3 qu’une partie au moins de cette variation résulte de différences génétiques entre les plantes de la F2. Des pires distinctes de plantes parentales ont été choisies à partie de quatre zones de distribution de la F2 et ont été croisées pour produire la génération suivante, F3. Dans chaque cas, la moyenne de la F3 était proche de la valeur de la zone de distribution des F2 à partir de laquelle les plants parentaux avaient été prélevés.

Le résultat du croisement est très différent des résultats obtenus lorsqu’on effectue un croisement entre individus portant des allèles distincts d’un gène dont les effets phénotypiques se distinguent nettement. Les descendants ne se répartissent pas en rapports mendéliens nets : 1 :2 :1 et il y a de plus en plus de variation individuelle à chaque génération de descendants. Mendel a obtenu des résultats relativement simples parce qu’il a utilisé des variétés horticoles de pois se distinguant les unes des autres par des différences alléliques simples à effets tranchés sur le phénotype. Les résultats des croisements présentés dans la figure n’ont rien d’exceptionnel. Ils sont de ce type pour la plupart des caractères dans la majorité des espèces. Si Mendel avait mené ses expériences sur la variation naturelle des mauvaises herbes de son jardin au lieu de variétés sélectionnées de pois, il n’aurait sans doute jamais découvert les lois sur l’hérédité qui portent son nom. D’une façon générale, des caractères comme la taille, la forme, la couleur, l’activité physiologique et le comportement ne ségrègent pas de façon simple lors des croisements.

Le fait que la plupart des caractères varient de façon continue ne signifie pas que leur variation résulte de mécanismes génétiques différents de ceux qui s’appliquent aux gènes mendéliens. La continuité du phénotype est en fait le résultat de deux phénomènes. Tout d’abord chaque génotype ne détermine pas une expression phénotypique unique mais plutôt une expression qui dépend de l’environnement dans lequel se développe l’organisme. Le phénotype varie également en raison des évènements aléatoires du développement. Il en résulte que les différences phénotypiques entre catégories de génotypes s’estompent et qu’il devient impossible d’associer sans ambiguïté un phénotype donné à un génotype particulier.

Deuxièmement, un phénotype donné peut être influencé par des allèles de nombreux loci pouvant s’assortir de différentes manières. Supposons par exemple que cinq loci d’importance égale participent à la détermination du nombre de fleurs qui se développeront chez une plante annuelle et que chaque locus possède deux allèles (+ et -). Pour des raisons de simplicité, supposons également qu’il n’y a pas de dominance et qu’un allèle + joute une fleur, tandis que – n’ajoute rien. Il y a donc 35 = 243 génotypes différents possibles au niveau de chacun des cinq loci qui vont de

+++++/+++++ Jusqu’à -----/----- ce qui correspond à onze classes phénotypiques seulement (10, 9, 8,….0), car de nombreux génotypes ont le même nombre d’allèles + et -. Bien qu’il y ait un seul génotype avec 10 allèles + et donc une valeur phénotypique moyenne de 10, et il y a 51 génotypes différents ayant tous 5 allèles + et 5 allèles -, par exemple :

++++-/+---- et ++-+-/++---

Ainsi de nombreux génotypes différents peuvent avoir le même phénotype moyen. Par ailleurs en raison de la variation environnementale, deux individus de génotype identique peuvent ne pas avoir le même phénotype. Cette absence de correspondance directe entre génotype et phénotype masque le mécanisme mendélien sous-jacent.

Si l’on ne peut étudier le comportement des facteurs mendéliens qui contrôlent directement ce type de caractères, que pouvons-nous apprendre sur leur constitution génétique ? à l’évidence, les méthodes utilisées pour étudier des caractères qualitatifs – telles que l’examen des rapport des descendants issus d’un croisement génétique – ne fonctionnent pas pour des caractères quantitatifs. Il faut utiliser à la place des méthodes statistiques pour établir des prévisions sur la transmission des phénotypes si l’on ignore tout sur les génotypes sous-jacents. Cette approche s’appelle la génétique quantitative.

La génétique quantitative tente de répondre aux questions suivantes :

1. La variation observée pour un caractère est d’une façon ou une autre influencée par la variation génétique ? l’ensemble de la variation est-il dû simplement à la variation environnementale et au bruit de fond développemental ? ou bien certains des allèles ségregeant dans la population ont-ils une autre influence sur le caractère,
2. S’il existe une variation génétique, quels sont les phénotypes des différents génotypes dans des environnements distincts ?
3. Quelle est l’importance de la variation génétique dans la variation phénotypique totale ? la quasi-totalité de la variation résulte-t-elle de différences de milieu et des perturbations au cours du développement, ou bien la variation génétique est-elle prédominante ?
4. De nombreux loci (ou seulement quelques-uns) contribuent-ils à la variation du caractère ? comment sont-ils distribués dans le génome ? quelle est l’origine moléculaire de leur influence sur le phénotype ?

Finalement on se pose ces questions pour essayer de prédire les types de descendants qui résulteront de croisements entre différents phénotypes.

Le degré de précision avec lequel ces questions peuvent être formulées et résolues est très variable. Chez les organismes expérimentaux, il est relativement simple de déterminer si une influence génétique s’exerce, mais la localisation des gènes (même approximative) est extrêmement laborieuse. Chez l’homme, il est même très difficile de répondre ne serait-ce qu’à la question de savoir si une influence génétique s’exerce sur un caractère, car il est presque impossible de distinguer les effets du milieu, des effets génétiques sur un organisme qui ne peut pas être manipulé expérimentalement. On dispose par conséquent de nombreuses données sur la génétique du nombre de soies chez la drosophile, alors qu’on ignore pratiquement tout de la génétique de caractères humains complexes, si ce n’est qu’un petit nombre d’entre eux (comme la couleur de la peau) sont clairement soumis à l’influence de gènes tandis que d’autres (comme les langues parlées) ne le sont pas. Ce module a pour but de développer les concepts statistiques et génétiques fondamentaux nécessaires pour répondre à ces questions.

1. **La variabilité génétique dans les populations naturelles**

Une particularité du monde vivant est la variabilité des phénotypes individuels. A l'intérieur d'une espèce, il n'existe pas deux individus ayant exactement les mêmes caractéristiques phénotypiques : l'individu est unique. Si pour une espèce donnée on peut noter l'absence de variations pour certains caractères essentiels, il existe toujours de nombreux autres caractères pour lesquels des variations entre individus sont observées. Certaines de ces variations s'expriment au niveau phénotypique (morphologie, physiologie, comportement, etc) mais les autres restent "cachées" et leur mise en évidence nécessite l'utilisation de techniques adaptées (variabilité des protéines ou des séquences d'ADN). Cette variation est nécessaire pour permettre aux plantes et aux animaux de s’adapter à un environnement physique et biologique continuellement en changement.

Au niveau du phénotype, le support de la variabilité est le **caractère**qui peut prendre plusieurs **états.** Par exemple, si l’on parle du caractère « couleur de l’œil », ce caractère peut prendre les états « vert », « bleu », « gris », « marron », « rouge » (pour les drosophiles ou les lapins albinos).

La **variation** est donc toute différence acquise au cours de son développement ou que présente au moment de l’observation une unité biologique par rapport aux autres unités du même groupe. Il peut s’agir soit des cellules d’un même organisme ; soit des individus d’une même population ; soit des populations d’une même espèce. Chaque espèce vivante possède une réserve de variations et l’objet de la biologie a été toujours d’en déceler le sens.

Les variations du phénotype sont dues pour partie à des facteurs environnementaux (alimentation, climat, interactions avec les autres espèces, etc) et pour partie à des différences entre les génotypes individuels, transmissibles à la descendance. **Le phénotype** dépend donc à la fois **du génotype et du milieu.**Les variations de l’un et de l’autre peuvent provoquer des changements phénotypiques. Dans la plupart des cas, ces deux causes de variation interagissent fortement (= interactions génotype-environnement), et il est difficile de mesurer leur part relative dans la variation phénotypique globale.

**Les caractères qualitatifs**

Ainsi appelés parce qu’ils comportent un petit nombre de formes facilement discernables, mais difficilement mesurables : couleur des fleurs, sensibilité ou résistance nette à une maladie. Ces caractères sont fréquemment appelés mendéliens simples car ils correspondent aux premiers traits étudiés par le généticien Mendel (caractère ridé ou lisse du pois). Ils permettent une classification en phénotypes bien séparés. Chacun de ces caractères discontinus est contrôlé par un faible nombre de gènes. La génétique qualitative concerne l’étude de la variabilité du caractère qualitatif, dont le déterminisme génétique est connu et implique peu de loci.

Exemple de caractères qualitatifs et discontinus : les groupes sanguins chez l’homme. Caractères qualitatifs et continus : la couleur d’une fleur qui peut prendre toutes les nuances possibles.

**Les caractères quantitatifs**

Ils sont ainsi nommés car leur variation s’apprécie par des mesures. Ces caractères présentent des variations **continues** telles qu’il est possible de définir des phénotypes bien séparés, il existe une gamme continue de phénotypes avec toutes intermédiaires d’un type à l’autre. Les caractères quantitatifs sont qualifiés de **polygéniques,** car ils sont généralement gouvernés par plusieurs **gènes** à effets individuels faibles. [C'est-à-dire, que chacun des gènes joue dans le déterminisme du phénotype un rôle si discret qu’il ne peut être mis en évidence par la méthode mendélienne]. De plus la variabilité phénotypique de ces caractères est **multifactorielle**c'est-à-dire qu’elle résulte de différents facteurs génétiques et environnementaux.

**Caractères ou traits continus quantitatifs chez les humains**

- La taille des humains n’est généralement pas contrôlée par un seul gène (même s’il existe un gène dominant du nanisme). Le déterminisme génétique de la taille humaine est clair : les enfants ont tendance à avoir des tailles proches de celle de leurs parents cependant les enfants mal nourris ne grandissent pas comme les enfants bien nourris donc une malnutrition ou une famine pendant l’enfance limite la croissance à une taille inférieure à la capacité génétique. La plupart des gens mesurent entre 125cm et 200cm, mais peuvent avoir n’importe quelle taille à l’intérieur de cette fourchette, certaines races sont grandes comme les Masaï en Afrique et certaines petites comme les pygmées en Afrique.

- La couleur de la peau,

- La prédisposition aux maladies cardiovasculaires et au diabète,

- L’intelligence et la personnalité.

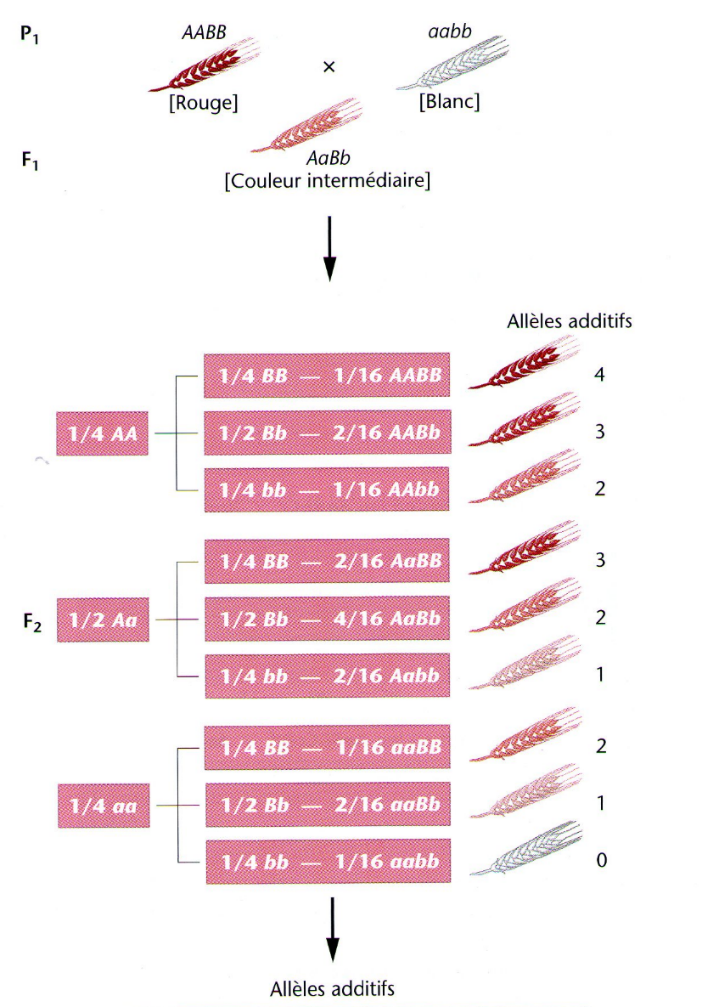
Les rendements et les qualités de tous les produits agricoles (lait, viande, fruits, céréales…) font également partie des traits continus variables. Il est important de connaître les contributions respectives des gènes et de l’environnement dans le but de maximiser la production par sélection sur élevage pour améliorer le génome et en modifiant les pratiques culturales pour améliorer l’environnement.

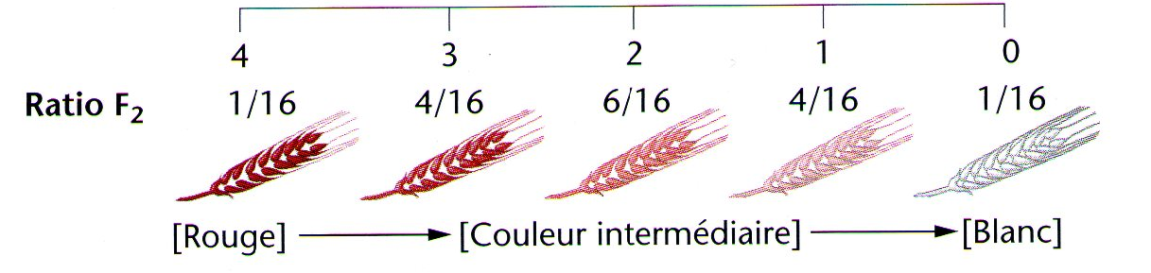
**Le nombre de gènes et les caractères quantitatifs**

On suppose parfois que la variation continue d’un caractère est nécessairement due à la ségrégation d’un grand nombre de gènes. La variation continue et alors présentée comme un élément de preuve en faveur d’un contrôle du caractère par plusieurs gènes. Cette **hypothèse des facteurs multiples (**le fait que de nombreux gènes dont chacun n’a qu’un faible effet ségrégent pour créer une variation quantitative) a longtemps servi de modèle élémentaire à la génétique quantitative, même si cette hypothèse n’est pas forcément vraie. Si la différence entre les moyennes génotypiques est faible comparée à la variance due au milieu, alors des cas aussi simples que ceux d’un gène à deux allèles peuvent conduire à une variation phénotypique continue.

1. **Explication des caractères quantitatifs en termes mendéliens**

**II.1.L’hypothèse multigénique de l’hérédité quantitative**

****

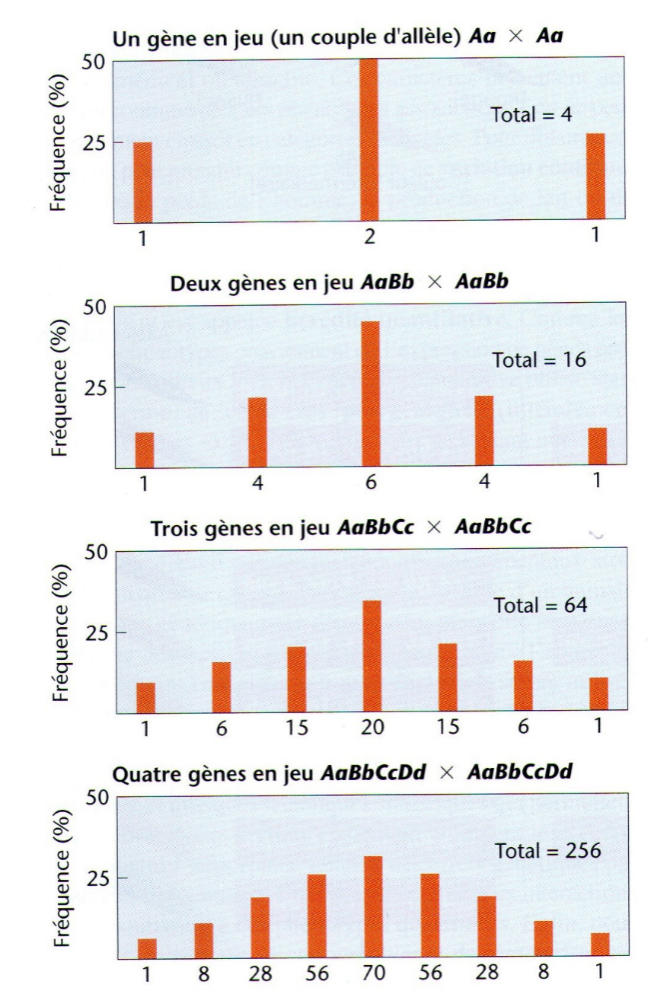
****

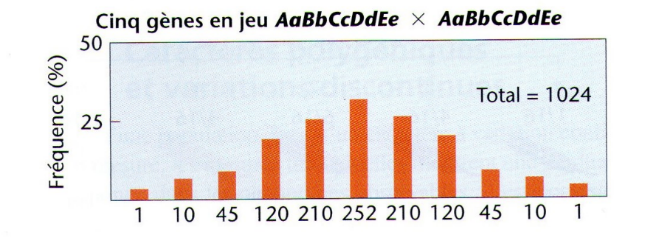
**Figure 2 :** la figure montre comment l’hypothèse multifactorielle permet d’expliquer les ratios 1 :4 :6 :4 :1 attendus dans une descendance F2 entre deux génotypes de blé, l’un à grains rouges et l’autre à grains blancs. Les lettres en majuscules correspondent à des allèles additifs qui contribuent de manière égale à l’élaboration de la pigmentation

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 phénotypes à caractères quantitatifs. L’expérience consistait à croiser du blé à grains rouges avec du blé à grains blancs (figure). Les plantes F1 avaient des grains à phénotype intermédiaire rose ce qui pouvait ‘interpréter comme une codominance de deux allèles à un même locus. Cependant, dans la F2, Nilsson-Ehle n’observa pas la ségrégation 3 :1 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 blancs. Un examen plus attentif des F2 révéla que les couleurs des grains pouvaient être classées en quatre catégories de nuances de rouge. Les ratios des phénotypes de la F2 se comptaient en seizièmes et pouvaient donc s’expliquer par l’intervention de deux gènes, chacun portant deux allèles, ségregeant 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 P1 les deux parents sont homozygotes ; le parent à grains rouges contient uniquement des allèles additifs AABB, tandis que le parent à grains blancs ne possède que les allèles non additifs aabb. Les plantes F1 sont hétérozygotes AaBb et contiennent deux allèles additifs (A et B), deux allèles non additifs (a et b) et expriment un phénotype intermédiaire rose. Dans la F2, chaque descendant peut avoir 0, 1, 2, 3 ou 4 allèles additifs. Les plantes sans allèle additif sont à grains blancs (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 allèle 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 quantité cumulée de pigment présent dans le grain.

Les résultats de Nilsson-Ehle ont donc montré que des variations continues pouvaient s’expliquer par 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, comprennent chacun deux allèles, sont impliqués dans un croisement, les 5 catégories phénotypiques de la F2 seront présentes dans les proportions 1 :4 :6 :4 :1. Par extension, on peut concevoir que lorsque 3 ou 4 loci et même plus, sont impliqués dans un croisement, ils contrôlent 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 3.

****

****

**Figure 3** : résultats de croisements entre deux hétérozygotes dans le cas d’une hérédité polygénique impliquant de 1 à 5 gènes. Chaque barre des histogrammes indique un phénotype F2 de classe distincte entre un extrême (à gauche) et l’autre (à droite). Chaque phénotype résulte d’un nombre différent d’allèles additifs.

**II.2. 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 quantifiées 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.

**II.3.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 :

1/4n = proportion des individus exprimant l’un des deux caractères extrêmes.

Si on reprend l’exemple des grains de blé rouges et blancs, 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 :

1/4n=1/16 ¼2 =1/16 n=2

Le tableau I regroupe les proportions et les nombres de classes phénotypiques observées en F2 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 simples, il faut respecter deux conditions : la contribution des allèles additifs doit être égale et l’expression phénotypique en F2 ne doit pas être trop perturbée par les facteurs environnementaux.

**Tableau I**. 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 | 1/4 | 3 |
| 2 | 1/16 | 5 |
| 3 | 1/64 | 7 |
| 4 | 1/256 | 9 |
| 5 | 1/1024 | 11 |

**II.4.Caractères polygéniques et variations discontinues**

Au sein d’une population, lorsqu’un caractère à variation continue est mesuré, les mesures individuelles montrent une graduation continue dans les phénotypes observables. Il est possible d’identifier une variation continue de phénotypes pour un caractère à déterminisme monogénique s’il y a différents niveaux de pénétrance pour ce gène entre les individus. Les caractères à déterminisme monogénique 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. 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. 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 et eux 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.

1. **Etude des caractères polygéniques et analyses statistiques**

**Quelques notions de statistique**

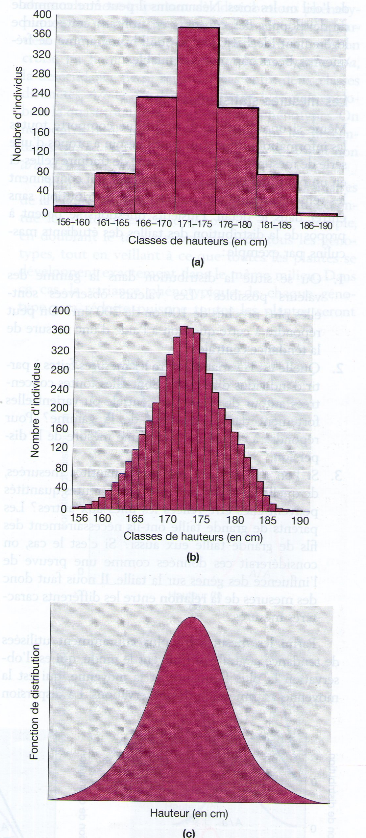
Pour répondre aux questions portant sur les types les plus courants de variation phénotypique nous devons d’abord faire connaissance avec certains outils statistiques essentiels à l’étude de la génétique quantitative.

**Les distributions statistiques**

Dans le cas d’une variation simple qui dépend exclusivement de différences alléliques au niveau d’un locus unique, les descendants d’un croisement se répartissent dans plusieurs classes phénotypiques distinctes. Par exemple, on peut s’attendre à ce qu’un croisement entre une plante à fleurs rouges et une plante à fleurs blanches aboutisse exclusivement à des plantes à fleurs rouges ou s’il s’agit d’un croisement en retour d’une plante de la F1 avec le parent à fleurs blanches , à ½ de plantes à fleurs rouges e ½ de plantes à fleurs blanches. Il nous faut en revanche un autre mode de description pour les caractères quantitatifs. Si la taille d’un grand nombre d’étudiants masculins est mesurée à 5 centimètres près, elle variera par exemple entre 145 cm et 195 cm, mais la majorité des étudiants se répartiront entre les catégories intermédiaires (disons entre 170, 175 et 180 cm) plutôt qu’entre les deux extrêmes. Une telle description d’un ensemble de mesures quantitatives s’appelle **la distribution statistique.**

Nous pouvons représenter chaque classe de mesures sous la forme d’une barre, dont la hauteur est proportionnelle au nombre d’individus de cette classe, comme dans la figure **4.** Un graphique de ce type représentant le nombre d’individus observés par classe de mesure porte le nom d’**histogramme de fréquences**.Supposonsà présent que cinq fois plus d’individus soient mesurés, au centimètre près cette fois, de sorte qu’ils soient répartis en classe de mesures encore plus limitées. Ceci produirait un histogramme comme celui de la figure b. si l’on continue le processus en affinant la mesure tout en augmentant proportionnellement le nombre d’individus mesurés, alors l’histogramme tend vers une forme continue de ce type s’appelle la **fonction de distribution** des tailles de la population.

La fonction de distribution est une courbe idéalisée de la véritable distribution des fréquences des tailles dans une population réelle, puisque aucune mesure ne peut être prise avec une précision infinie, ni inclure un nombre illimité d’individus. En outre, le caractère mesuré peut lui-même être intrinsèquement discontinu, comme dans le cas où il représente la somme d’un certain nombre d’objets distincts, tels que les facettes de l’œil ou les soies. Néanmoins, il peut être commode de développer des concepts en utilisant cette courbe légèrement idéalisée, plutôt que l’histogramme de fréquences beaucoup moins maniable.



**Figure 4**. Les distributions des fréquences de la taille des étudiants masculins. a) un histogramme de fréquences établi à partir d’intervalles de 5 cm entre les classes. b) un histogramme établi à partir d’intervalles de 1 cm entre les classes. c) graphique d’une distribution continue.

**Les mesures statistiques**

Même si une distribution statistique comporte toutes les informations dont nous avons besoin à propos d’une série de mesures, il est souvent utile de traduire celles-ci en un petit nombre de caractéristiques qui contiennent les renseignements nécessaires sur la distribution, sans entrer dans le détail. Plusieurs questions se posent à propos de la distribution des tailles des étudiants masculins, par exemple :

1. Où se situe la distribution dans la gamme des valeurs possibles ? les valeurs observées sont-elles plus proches de 100 ou de 200 cm ? on peut répondre à cette question à l’aide d’une mesure de la **tendance centrale**.
2. Quelle est la variation entre les mesures faites à partir d’individus différents ? sont-elles toutes concentrées autour de la mesure centrale ou varient elles fortement dans l’ensemble de la gamme ? pour répondre à cela, il nous faut une mesure de la **dispersion**.
3. Si l’on s’intéresse à plusieurs quantités mesurées, de quelle façon les valeurs des différentes quantités peuvent-elles être reliées les unes aux autres ? les parents de grande taille ont-ils nécessairement des fils de grande taille eux aussi ? si c’est le cas, on considérerait ces données comme une preuve de l’influence des gènes sur la taille il nous faut donc des mesures de la relation entre les différents caractères mesurés

Parmi les mesures les plus couramment utilisées de tendance centrale, se trouvent le **mode** qui est l’observation la plus fréquente, et la **moyenne**, qui est la moyenne arithmétique des observations la dispersion d’une distribution est presque toujours mesurée par la **variance,** qui est la moyenne des carrés des écarts séparant les observations de leur moyenne ou par l’écart-type qui est la racine carrée de la variance la relation entre les différentes variables est mesurée par leur **covariance** et leur **corrélation**. La **covariance** entre deux variables est l’écart entre une variable et sa moyenne multiplié par l’écart entre l’autre variable et sa propre moyenne. La **corrélation** est la covariance entre deux variables divisée par le produit des écarts types des deux variables. En tant que mesure de la relation entre deux variables, la corrélation présente un avantage : elle varie de +1 pour les variables parfaitement corrélées positivement à 0 dans le cas de variables qui n’ont en moyenne aucune relation, jusqu’à -1 pour des variables parfaitement corrélées négativement.

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 page 625).** 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.

**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. A la figure **24.5** on voit que les deux distributions de mesures phénotypiques se regroupent autour d’une valeur centrale. Ce regroupement est appelé **tendance centrale** et le point central est la moyenne x̄. la moyenne est la moyenne arithmétique d’une série de mesures et la formule pour la calculer est :

 x̄ =

La moyenne correspond à un descriptif simple de l’échantillon, mais ne renseigne pas sur l’amplitude de variation ou la dispersion des données. Une deuxième constante statistique donne une idée de la dispersion des valeurs autour de la moyenne : c’est la variance.

**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 :

s²**=**

Où la somme (Ʃ) des carrés des écarts à la moyenne entre chaque valeur et la moyenne est divisée par le nombre total de mesures diminué d’une unité.

**L’écart type**

Comme la variance est une valeur basée sur des carrés, ses unités de mesures sont aussi élevées au carré (m², g², 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 =

**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 x̄) calculée comme suit :

s x̄=

ù s est l’écart type et la racine carrée de l’effectif n.

**La covariance**

Souvent, les généticiens travaillant avec des caractères quantitatifs analysentdeux caractères phénotypiques simultanément. Par exemple, un sélectionneur de volailles pourra rechercher la covariance entre le poids du corps et la production d’œufs des poules, car les oiseaux plus gros peuvent pondre plus d’œufs.

La covariance mesure la part de variation commune à deux caractères quantitatifs. Elle est calculée en prenant les écarts à la moyenne pour chaque caractère (comme pour la variance) et pour chaque individu de l’échantillon. Ceci donne des couples de valeurs qui sont multipliées entre elles ; la somme de ces produits individuels est alors divisée par l’effectif diminué d’une unité. Ainsi, la covariance, covxy de deux ensembles de mesures, x et y est donnée par la formule :

covxy =

La covariance peut ensuite être standardisée sous forme d’une autre valeur statistique : le coefficient de corrélation (r). son calcul se fait suivant la formule :

r = covxy /Sx-Sy

où Sx est l’écart type du premier jeu de données quantitatives x et Sy celui du deuxième jeu de données y. les valeurs du coefficient de corrélation r varient de -1 à 1. Une corrélation positive signifie que lorsqu’une valeur augmente pour un caractère elle augmente aussi pour l’autre caractère. Une corrélation négative signifie que lorsque les valeurs d’un caractère augmentent celles de l’autre diminuent.

1. **Héritabilité : mesure de la contribution génétique à la variation phénotypique**

La 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. 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 su. 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. 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 production 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, élevées en plein air, l’héritabilité de la production en œufs peut se révéler plus faible et ces différences peuvent refléter la variation de l’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-types élevés. Ceci est un point important dont il faut se souvenir, en particulier pour le calcul de l’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.

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. Nous voyons que les deux variétés ne répondent pas de la même manière à l’amélioration du sol. Les génotypes de l 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 (VGxE).**

Nous pouvons maintenant résumer toutes les composantes de la variance phénotypique totale Vp = VG + VE + VGxE

**L’héritabilité au sens large**

L’héritabilité au sens large mesure la contribution de la variance génotypique à la variance phénotypique totale : H² = . Ces valeurs d’héritabilité ‘échelonnent entre 0 et 1 pour un caractère donné. 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. Les valeurs faibles ou proches de 0 indiquent que des facteurs environnementaux et non génétiques sont principalement responsables des différences phénotypiques observées.

La composante de variance génétique VGutilisée 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, dominants ou épistatiques. L’estimation de l’héritabilité au sens large considère également que la variance génotype-environnement est négligeable. Les sélectionneurs d’animaux ou de plantes qui veulent améliorer de caractères spécifiques 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 (h²) est la proportion de la variance phénotypique due aux seuls aspects 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,** VA, est la variance génotypique liée à l’action additive des allèles aux loci quantitatifs. La **variance de dominance**, VD 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**, VI, est l’écart à l’additivité lorsque deux loci ou plus interagissent en épistasie. Le niveau e la variance d’interaction est souvent négligeable et cette composante peut être exclue des calculs de variance génotypique.

La partition de la variance génotypique VG peut être résumée par l’équation suivante :

VG = VA + VD + VI

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

En omettant VI et en séparant VP entre ses composantes de variance génotypique et environnementale, nous obtenons pour h² la formule suivante :

H² =

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.

1. **Cartographie des caractères quantitatifs**

L’analyse de pedigree 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 **Quantitative Trait Loci** (QTL). L’identification de ces loci donne des informations sur le nombre de gènes 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 caractèresont 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 F1 dont les individus vont être hétérozygotes pour la quasi-totalité des loci contribuant au caractère. Des croisements supplémentaires entre individus F1 ou entre individus F1 et lignées parentales donneront des générations F2 avec des degrés élevés de ségrégation pour les différents QTL et leurs phénotypes associés. Cette génération F2 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 des microsatellites.

Les corrélations entre les marqueurs moléculaires génotypiques et les variations phénotypiques pour le caractère sont calculés 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 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’organismes 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 leur croisement sur plusieurs générations. De nombreux loci responsables de caractères quantitatifs ont été identifiés : ils contrôlent la taille du fruit, sa forme, la quantité de matières solubles et l’acidité. Ces loci sont distribués sur les 12 chromosomes représentant le génome haploïde de la plante.

La cartographie par marqueurs moléculaires permet d’identifier des régions chromosomiques contenant des QTL, mais l’identification des gènes individuels impliqués dans les 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.