

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Ibn Khaldoun –Tiaret–
Faculté Sciences de la Nature et de la Vie
Département de Nutrition et Technologies Agro-Alimentaires

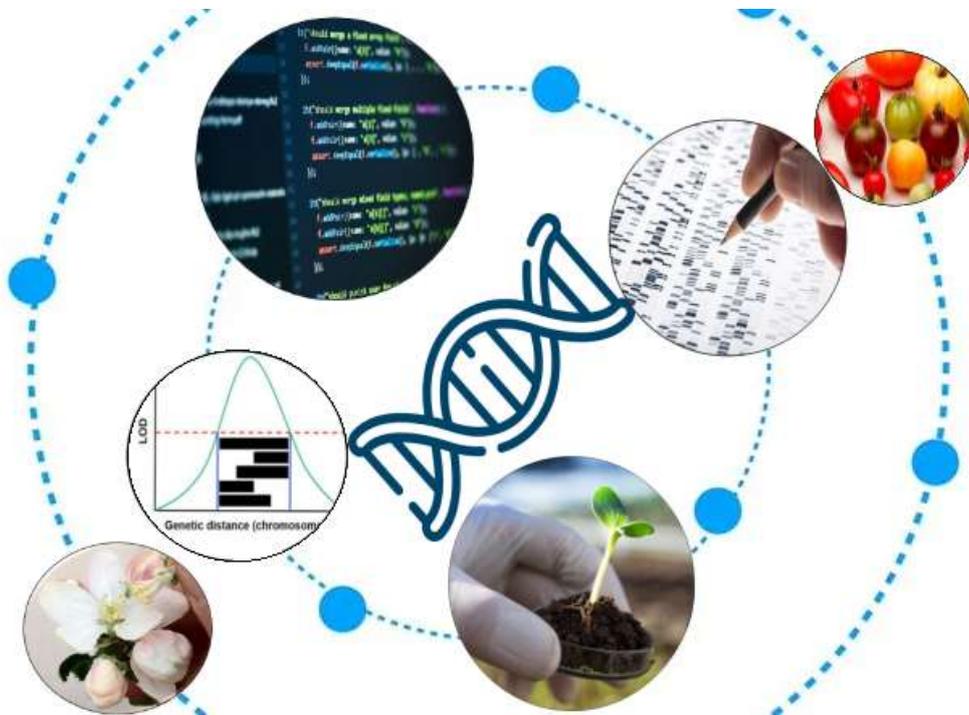


Génétique quantitative

Polycopié de cours dédié aux étudiants de Première Année Master en Génétique
Moléculaire et Amélioration des Plantes

Préparé par :

Dr. DAHLIA Fatima



Année universitaire 2023-2024

Liste des figures

Figure 1 : Distribution normal des variables quantitatives continus.	11
Figure 2 : Distributions asymétrique (A) et distribution bimodale (B).	12
Figure 3 : différents types de corrélation entre caractères quantitatives	13
Figure 4 : Corrélation parabolique entre deux traits.	14
Figure 5 : Corrélation linéaire positive parfaite entre les variables x et y.	15
Figure 6 : Distribution des génotypes et du phénotype.	18
Figure 7 : Extrêmes de la taille chez les êtres humains.	20
Figure 8 : Valeurs génotypiques en fonction du dosage en structure A.	23
Figure 9 : Représentation d'un cas d'additivité parfaite.	24
Figure 10 : Régression et estimation de la dominance et de l'additivité	25
Figure 11 : Dominance et définition de l'hétérozygote	25
Figure 12 : Les divers types d'effets (direct et en interaction) intervenant dans l'expression génotypiques : cas de deux paires de gènes A_i , A_j , B_k et B_l	27
Figure 13 : Régression des valeurs mesurées des descendants rapportés à celle des moyennes des deux parents pour un caractère à héritabilité étroite de 0,5.	35
Figure 14 : Représentation graphique de la répartition des trois génotypes autour du centre.	48
Figure 15 : Représentation graphique de la répartition des génotypes A_1A_1 , A_1A_2 et A_2A_2 autour leur centre.	49
Figure 16 : Différentes qualités de balance génétique	55
Figure 17 : Tests de descendance maternelles	58
Figure 18 : Plan systématique et poly-cross pour 6 individus.	59
Figure 19 : Croisement hiérarchique	61
Figure 20 : Croisement diallèle complet.	62
Figure 21 : Croisement demi-diallèle : (A) croisement demi-diallèle non incluant les autofécondations ; (B) croisement diallèle incluant les autofécondations.	63
Figure 22 : La localisation de QTL impliqués dans la variation de la dureté du grain de blé	68
Figure 23 : Exemple d'une carte génétique : Carte génétique de <i>Podospora anserina</i>	71
Figure 24 : Principe de la méthode de séquençage de nouvelle génération (NGS)	73
Figure 25 : Analyse d'acide nucléique par puce à ADN.	75
Figure 26 : Trois applications des puces à ADN dans l'analyse microbiologique.	76
Figure 27 : Sélection assistée par marqueurs moléculaire, Exemple de l'introgression du gène B_t chez le maïs	84

Liste des tableaux

Tableau 1 : Principales différences entre les caractères qualitatifs et quantitatifs.....	4
Tableau 2 : Rendement moyen des culture blé de dix fermes.	5
Tableau 3 : Tailles moyennes et fréquences des génotypes.....	6
Tableau 4 : Notations statistiques pour une population et un échantillon.....	17
Tableau 5 : Résultats des moyennes et fréquences des génotypes (aa ; Aa et AA) de la hauteur de la tige.	31
Tableau 6 : Résultats des moyennes et fréquences des génotypes (AA ; AA ₁ et A ₁ A ₁) de la hauteur de la tige.	32
Tableau 7 : Proportions relatives de la variance additive et de la variance due à la dominance continue dans la covariance génétique entre individus apparentés.	34
Tableau 8 : Résultats du croisement panmictiques des gamètes A1 et A2.....	43
Tableau 9 : Niveau d'hétérozygotie après plusieurs générations de consanguinité dans le cas de deux systèmes d'appariement.....	44
Tableau 10 : Effets des environnements sur les génotypes.....	47
Tableau 11 : Effets des caractères sur les génotypes.	47
Tableau 12 : Hauteurs moyennes des tiges des trois génotypes d'orge.	48
Tableau 13 : Détermination de la moyenne de la population.....	49
Tableau 14 : Génération F ₁ obtenue par croisement aléatoires des individus des deux populations.	51
Tableau 15 : fréquences alléliques des valeurs des individus de la F ₁	51
Tableau 16 : Génération F ₂ obtenue par croisement aléatoires des individus de la F ₂	53
Tableau 17 : stratégies de sélection en fonction des variances d'aptitude à la combinaison..	57

Table des matières

Liste des figures	i
Liste des tableaux	ii
Table des matières	iii
Introduction	1
Chapitre 1 : Définitions et concepts	3
1. La différence entre caractères quantitatifs et caractères qualitatifs	3
2. Particularités des traits quantitatifs	4
3. Rappel de quelque notion de statistiques	5
3.1. Statistiques de position	5
3.1.1. Moyenne	5
3.1.2. Médiane	6
3.1.3. Mode	7
3.2. Statistiques de dispersion	8
3.2.1. Variance et écart type	8
3.2.2. Erreur standard	9
3.2.3. Coefficient de variation	10
3.3. Modèle de distribution	11
3.4. Les corrélations	12
3.4.1. Généralités	12
3.4.2. Corrélation et identité	14
3.5. La régression	15
3.6. Echantillons et populations	16
3.7. Génotypes et distribution des phénotypes	18
Chapitre 2 : Héritéité polygénique	20
1. Généralités	20
2. Les relations entre les gènes	22
2.1. Définition des effets moyens (additivité)	22
2.2. Effet de la dominance	24
2.3. Les formes d'épistasie	26
2.3.1. Épistasie dominante	26
2.3.2. Épistasie récessive	27
Chapitre 3 : Héritabilité d'un caractère	28
1. Définition	28
2. Parenté et héritabilité	28
3. Comment quantifier l'héritabilité ?	29
3.1. Par les variances	29
3.1.1. Analyse approfondie de la variance (variance additive et variance due à l'environnement)	31
3.1.2. Estimation des composantes de la variance génétique	34
3.2. Autre méthode d'estimation de la valeur de l'héritabilité	36
Chapitre 4 : Hétérozygotie	39
1. Généralités	39

2.	Hétérozygotie et fréquence allélique	40
2.1.	Calcul de l'hétérozygotie à partir des fréquences alléliques	40
2.2.	Relation entre l'hétérozygotie et l'homéostasie génétique	40
3.	Les lois de Hardy-Weinberg	41
3.1.	Présentation des lois de Hardy-Weinberg	41
3.2.	Conditions d'équilibre de Hardy-Weinberg	41
4.	Evolution de l'hétérozygotie au cours des générations.....	43
5.	Facteurs influençant l'hétérozygotie	44
5.1.	Mutation génétique	44
5.2.	Flux génétique (migration de gènes)	44
5.3.	Sélection naturelle	44
5.4.	Dérive génétique.....	45
5.5.	Consanguinité	45
5.6.	Ségrégation génétique.....	45
	Chapitre 5 : Le phénomène de l'hétérosis.....	46
1.	Définition	46
2.	Hypothèses ou mécanismes	46
2.1.	Dominance.....	46
2.2.	Superdominance	47
3.	Evolution de l'hétérosis au cours des générations	48
3.1.	Moyenne d'une population.....	48
3.2.	Evolution de l'hétérosis proprement dit	50
	Chapitre 6 : La valeur d'un individu au croisement	54
1.	Généralités	54
2.	Les balances génétiques	54
2.1.	La balance interne et la balance de relation.....	54
2.2.	Réalité biologique de la balance et représentation par l'ensemble des génotypes	56
3.	Les aptitudes à la combinaison	56
3.1.	Définitions	56
3.2.	Intérêts des variances liées aux aptitudes de la combinaison	57
4.	Méthodes permettant d'apprécier les valeurs d'un individu en croisement	57
4.1.	Les fécondations libres (descendances maternelles)	57
4.2.	Le test top-cross.....	58
4.3.	Le test poly-cross.....	59
4.4.	Croisement hiérarchique.....	60
4.5.	Croisement diallèle	60
	Chapitre 7 : QTL (Quantitative trait loci)	64
1.	Introduction aux QTL	64
1.1.	Définition et concepts clés.....	64
1.2.	Applications des QTL en recherche agricole, médicale et écologique.....	64
1.2.1.	Recherche agricole	64
1.2.2.	Recherche médicale.....	65
1.2.3.	Recherche écologique	66
1.3.	Limites et défis de l'utilisation des QTL.....	66

2.	Techniques pour la détection des QTL	67
2.1.	Analyse de liaison et cartographie génétique	67
2.1.1.	Analyse de liaison	67
2.1.2.	Cartographie génétique	67
2.2.	Analyse d'association et étude d'association pangénomique	69
2.2.1.	Analyse d'association	69
2.2.2.	Étude d'association pangénomique.....	69
2.3.	Utilisation de marqueurs moléculaires pour la cartographie de QTL	70
2.4.	Techniques de génotypage haut débit pour la cartographie de QTL.....	72
2.4.1.	Séquençage de nouvelle génération (NGS).....	72
2.4.2.	Les puces à ADN.....	74
2.4.3.	Avantages et limites des techniques de génotypage haut débit pour la cartographie de QTL	77
3.	Analyse statistique des QTL	78
3.1.	Modèles de régression et d'analyse de variance	78
3.2.	Modèles mixtes linéaires pour la détection de QTL.....	79
3.3.	Correction de l'effet de population et de l'effet environnemental.....	80
3.4.	Évaluation de la signification statistique des QTL.....	81
4.	Applications des QTL dans la recherche agricole	82
4.1.	Sélection assistée par marqueurs pour la sélection de caractères complexes.....	82
4.2.	Cartographie de QTL pour la résistance aux maladies et aux ravageurs.....	83
4.3.	Cartographie de QTL pour la qualité des cultures et des animaux.....	85
4.4.	Amélioration de la qualité nutritionnelle des aliments.....	86
5.	Applications des QTL en médecine et en écologie.....	87
5.1.	Cartographie de QTL pour les maladies humaines	87
5.2.	Identification des QTL impliqués dans la résistance aux maladies chez les animaux sauvages et domestiques.....	88
5.3.	Identification des QTL impliqués dans l'adaptation aux changements environnementaux	89
6.	Techniques avancées en génétique quantitative	89
6.1.	Analyse d'interaction génétique.....	89
6.2.	Modèles de régression bayésienne	90
6.3.	Méthodes de sélection assistée par marqueurs avancées.....	91
7.	Développements futurs dans le domaine des QTL	92
7.1.	Évolution des technologies de séquençage haut débit.....	92
7.2.	Nouvelles approches pour l'analyse de données à haut débit	93
7.3.	Perspectives pour l'application des QTL dans la sélection assistée par marqueurs et la biologie de synthèse.....	94
	Références bibliographiques	96

Introduction

Introduction

La génétique quantitative est une branche de la génétique qui étudie la variation des caractéristiques complexes chez les individus et les populations, comme la taille, le poids, la croissance, la production laitière, la teneur en huile chez les plantes, etc. Ces caractéristiques sont également appelées traits quantitatifs, car elles varient continuellement plutôt que de se répartir en catégories distinctes.

La génétique quantitative étudie la manière dont les traits quantitatifs sont hérités et influencés par l'environnement. Les caractéristiques quantitatives sont souvent polygéniques, ce qui signifie qu'elles sont influencées par plusieurs gènes. De plus, les caractéristiques quantitatives peuvent également être affectées par l'environnement, comme l'alimentation, la température, la lumière, le stress, etc.

Les méthodes de la génétique quantitative permettent de mesurer l'héritabilité des traits quantitatifs, c'est-à-dire la proportion de la variance phénotypique observée qui est due à des différences génétiques entre individus. Les méthodes de la génétique quantitative peuvent également être utilisées pour identifier les régions du génome qui sont associées à des variations de traits quantitatifs, appelées QTL (Quantitative Trait Loci). La caractérisation des QTL permet de mieux comprendre les mécanismes génétiques sous-jacents à la variation des traits quantitatifs.

Son histoire remonte au début du 20^e siècle, lorsque des scientifiques tels que Ronald Fisher, J.B.S. Haldane et Sewall Wright ont commencé à développer des méthodes statistiques pour analyser les données relatives à des caractéristiques complexes. Ils ont proposé que les traits quantitatifs étaient régis par de nombreux gènes ayant des effets minuscules, et que ces gènes étaient hérités de manière polygénique.

Au fil du temps, des méthodes statistiques plus avancées ont été développées pour étudier les traits quantitatifs, notamment la régression linéaire et l'analyse de variance (ANOVA). Ces méthodes ont permis de mieux comprendre la manière dont les traits quantitatifs étaient hérités et comment ils étaient influencés par l'environnement.

Dans les années 1980 et 1990, l'utilisation de marqueurs génétiques a permis d'identifier les régions du génome qui étaient associées à des variations de traits quantitatifs, appelées QTL (Quantitative Trait Loci). Les QTL ont permis de mieux comprendre les mécanismes génétiques sous-jacents à la variation des traits quantitatifs et ont ouvert la voie à la sélection assistée par

marqueurs (MAS), une méthode de sélection d'individus pour des caractéristiques spécifiques en utilisant des marqueurs génétiques.

Avec l'avènement des technologies de séquençage de l'ADN à haut débit au début du 21^e siècle, la génétique quantitative a connu une nouvelle avancée. Ces technologies ont permis d'identifier des millions de variations génétiques dans le génome, et de découvrir de nouveaux gènes et de nouvelles régions du génome qui influencent les traits quantitatifs.

Aujourd'hui, la génétique quantitative est utilisée dans de nombreux domaines de la biologie, notamment en agriculture, en élevage, en médecine et en écologie. Les avancées récentes dans le domaine de la génétique quantitative ont ouvert de nouvelles perspectives pour la compréhension de la complexité des traits quantitatifs et pour leur utilisation dans la sélection assistée par marqueurs et la biologie de synthèse.

Chapitre 1 :
Définitions et concepts

Chapitre 1 : Définitions et concepts

La génétique quantitative étudie des différences entre individus. Ces différences sont quantitatives et non qualitatives.

Tout organe ou n'importe quelle espèce montre des différences éventuelles. Ex : la différence de taille, de résistance aux maladies ou du rendement. Ces différences individuelles peuvent soit former des séries continues et croissantes d'un extrême à un autre, soit former des catégories distinctes. Dans le premier cas, c'est des caractères quantitatifs et dans le second cas, ce sont des caractères qualitatifs.

1. La différence entre caractères quantitatifs et caractères qualitatifs

Les traits qualitatifs et les traits quantitatifs sont deux types de caractéristiques qui peuvent être héritées et transmises de génération en génération. Voici les principales différences entre ces deux types de traits :

- *Nature des caractéristiques* : Les traits qualitatifs sont des caractéristiques qui sont facilement observables et qui peuvent être classées en catégories distinctes, comme la couleur des yeux, le groupe sanguin, la présence ou l'absence d'une maladie, etc. Les traits quantitatifs, en revanche, sont des caractéristiques qui varient en continu et qui peuvent être mesurées avec des unités numériques, comme la taille, le poids, la productivité, etc.
- *Nombre de gènes impliqués* : Les traits qualitatifs sont souvent régis par un petit nombre de gènes, souvent avec un effet dominant ou récessif, alors que les traits quantitatifs sont généralement influencés par de nombreux gènes, chacun ayant un effet relativement petit.
- *Environnement* : Les traits qualitatifs sont souvent peu influencés par l'environnement, tandis que les traits quantitatifs sont souvent fortement influencés par l'environnement. Par exemple, la couleur des yeux est généralement déterminée par les gènes, tandis que la taille peut être influencée par de nombreux facteurs environnementaux tels que la nutrition, l'exercice physique, etc.
- *Héritabilité* : Les traits qualitatifs ont souvent une héritabilité élevée, ce qui signifie qu'une grande partie de la variation dans ces traits est due aux différences génétiques entre les individus. En revanche, les traits quantitatifs ont souvent une héritabilité plus faible, car ils sont fortement influencés par l'environnement.

- *Analyses statistiques* : pour les traits qualitatifs, on peut uniquement faire une analyse des proportions, en revanche, pour les traits quantitatifs (mesurables), on peut faire toutes séries d'analyses statistiques (moyennes, variances, covariances, ...etc.).

Le tableau suivant résume les principales différences entre les caractères qualitatifs et quantitatifs.

Tableau 1 : Principales différences entre les caractères qualitatifs et quantitatifs.

Caractères quantitatifs	Caractères qualitatifs
Variabilité continue de la majorité des caractères. Ex : rendement, croissance, taille ...etc.	Variabilité discrète (phénotype) Ex : couleur des feuilles, des fleurs ...etc.
Déterminisme polygénique	Déterminisme monogénique ou oligogénique
Analyse d'une série de descripteurs statistiques : moyennes, variance, écart type	Analyse des proportions

2. Particularités des traits quantitatifs

Grâce à des techniques expérimentales établis, les généticiens sont amenés à répondre sur des questions sur le déterminisme génétique d'un caractère à variation continue dans une population. Ces questions qui sont à la base de l'étude de la génétique quantitative sont :

- La variation observée du caractère est-elle d'une façon ou d'une autre influencée par la variation génétique ? La ségrégation d'allèles dans la population est-elle responsable des différences dans le caractère ou bien toute cette variation n'est-elle que le résultat d'une variation du milieu et de la fluctuation fortuite au cours de développement ?
- Si une variation génétique existe, quelles sont les normes de réaction des différents génotypes ?
- Quelle est l'importance de la variation génétique dans la variation phénotypique totale ? Les normes de réaction et les effets du milieu sont-ils tel que presque toute la variation résulte de différence du milieu et de perturbations dans le développement ou bien la variation génétique est-elle prédominante ?
- De nombreux loci ou seulement quelques-uns affectent ils le caractère ? comment sont-ils répartis dans le génome ?

- Comment les différents loci interagissent-ils pour influencer le caractère ? a-t-on affaire à des cas de dominance ou d'épistasie ?
- Sommes-nous en présence d'hérédité non nucléaire (effet maternelle) ?

Chez des organismes expérimentaux, il est relativement simple d'établir si une influence génétique s'exerce, mais la localisation de ces gènes même approximative est extrêmement laborieuse.

3. Rappel de quelque notion de statistiques

3.1. Statistiques de position

3.1.1. Moyenne

La moyenne est une mesure de tendance centrale. Elle représente la "valeur typique" ou "centrale" des données et est souvent utilisée pour décrire la moyenne des performances des individus d'une population ou d'un échantillon. La moyenne est souvent représentée par la lettre grecque mu (μ) et peut être calculée pour des variables continues ou discrètes. Cependant, il est important de noter que la moyenne peut être influencée par les valeurs extrêmes ou aberrantes, ce qui peut avoir un impact sur son interprétation.

Le calcul de la moyenne diffère selon que les proportions sont égales ou non égales.

- *Proportions égales* : Si les proportions sont égales, la moyenne peut être calculée en additionnant toutes les valeurs d'une variable et en divisant le résultat par le nombre total d'observations ($\mu = \Sigma x_i/n$). Avec Σx_i est la somme valeur des variables et n est le nombre total d'observations.

Exemple : Prenons l'exemple d'une étude menée dans une région donnée pour mesurer le rendement moyen des cultures de blé. Les données ont été collectées sur les rendements de blé pour 10 fermes différentes, avec les résultats suivants (en tonnes par hectare) :

Tableau 2 : Rendement moyen des culture blé de dix fermes.

Fermes	1	2	3	4	5	6	7	8	9	10
RDT (Tonne/ha)	6	5	5,5	7	5,5	6,5	6	6,5	7	7,5

Moyenne = $\Sigma x_i/n$

$$= (6 + 5 + 5,5 + 7 + 5,5 + 6,5 + 6 + 6,5 + 7 + 7,5) / 10$$

$$= 6,3$$

- *Proportions inégales* : Si les proportions sont non égales, la moyenne doit être calculée en pondérant chaque observation par sa proportion respective. Pour ce faire, on multiplie chaque valeur de l'observation par sa proportion, on additionne les produits et on divise le résultat par la somme des proportions ($\mu = [\sum(x_i * p_i) / \sum p_i]$). Avec x_i sont les variables, p_i les fréquences des variables. La somme des fréquences est égale à 1. Donc $\mu = \sum(x_i * p_i)$.

Exemple : Prenons le cas de deux allèles (A) et (a) à un locus, contrôlant la taille, ségrégent dans les milieux où la population se trouve, les phénotypes moyens (hauteur de la tige) et la fréquence des trois génotypes sont les suivants :

Tableau 3 : Tailles moyennes et fréquences des génotypes.

	aa	Aa	AA
Valeurs (cm)	10	18	20
Fréquences	0,36	0,48	0,16

$$\begin{aligned} \text{La moyenne} &= \sum(x_i * p_i) / \sum p_i \\ &= [10(0,36) + 18(0,48) + 20(0,16)] / 1 \\ &= 15,44 \text{ cm} \end{aligned}$$

3.1.2. Médiane

La médiane est une mesure de tendance centrale qui représente la valeur centrale d'un ensemble de données triées par ordre croissant ou décroissant. La médiane divise l'ensemble de données en deux parties égales, de sorte que la moitié des observations sont inférieures ou égales à la médiane et l'autre moitié sont supérieures ou égales à la médiane.

La médiane est moins sensible aux valeurs aberrantes que la moyenne. Elle est souvent utilisée pour décrire les performances centrales d'un échantillon ou d'une population et peut être utilisée pour comparer des groupes ayant des distributions différentes.

La médiane est souvent notée par la lettre M ou Md et peut être calculée pour des variables continues ou discrètes. Cependant, il est important de noter que la médiane peut ne pas être représentative de la distribution des données dans certains cas, comme lorsque la distribution est biaisée ou fortement asymétrique.

Pour calculer la médiane, les données doivent être triées par ordre croissant ou décroissant :

- Si l'ensemble de données comporte un nombre impair d'observations, la médiane est la valeur de l'observation centrale. Par exemple, si nous avons les données suivantes : 2, 5, 8, 10, 15, la médiane est 8.
- Si l'ensemble de données comporte un nombre pair d'observations, la médiane est la moyenne des deux observations centrales. Si l'ensemble de données comporte un nombre pair d'observations, la médiane est la moyenne des deux observations centrales. Par exemple, si nous avons ces données suivantes : 3, 6, 7, 11, la médiane est $(6 + 7) / 2 = 6,5$.

3.1.3. Mode

Le mode est une mesure de tendance centrale qui représente la valeur la plus fréquente d'un ensemble de données. C'est la valeur qui apparaît le plus souvent dans un ensemble de données.

Le mode est la valeur qui apparaît le plus souvent dans l'ensemble de données. Il peut y avoir un mode unique (une valeur qui apparaît le plus souvent) ou plusieurs modes (plusieurs valeurs qui apparaissent également souvent).

Le mode est une mesure de tendance centrale qui peut être utilisée pour décrire les performances centrales d'un échantillon ou d'une population. Cependant, le mode n'est pas toujours une mesure utile de la tendance centrale, car il peut ne pas être représentatif de la distribution des données dans certains cas, comme lorsque les données sont uniformément réparties ou lorsque les données sont fortement asymétriques.

Le mode est souvent noté par la lettre M_o ou Mod et peut être calculé pour des variables continues ou discrètes.

Exemple : Supposons que nous avons collecté les mesures de la hauteur de plantes de blé dans une parcelle de terrain. Les données sont les suivantes (en cm) :

50, 45, 52, **48**, 54, 60, 42, 44, 47, 53, **48**.

Toutes les valeurs apparaissent une seule fois sauf la valeur **48** qui apparaît deux fois, ce qui en fait la valeur la plus fréquente dans l'ensemble de données. Ainsi, le mode de la hauteur des plantes de blé dans cette parcelle est de 48 cm. Cela signifie que la hauteur de 48 cm est la valeur la plus fréquente dans l'ensemble de données.

Remarque : Il est important de noter que la moyenne, la médiane et le mode sont des mesures différentes de la tendance centrale et qu'elles peuvent être utilisées en conjonction pour obtenir une meilleure compréhension de la distribution des données.

3.2. Statistiques de dispersion

3.2.1. Variance et écart type

La variance et l'écart type sont des mesures de dispersion ou de variation des données. Ils permettent de mesurer à quel point les données sont éloignées de la moyenne.

La variance est une mesure de la dispersion des données autour de la moyenne. Elle est calculée en faisant la somme des carrés des écarts entre chaque valeur et la moyenne, puis en divisant cette somme par le nombre total de valeurs dans l'échantillon moins un. La variance est souvent notée par le symbole σ^2 .

$$\sigma^2 = \frac{1}{n-1} \sum (x_i - \mu)^2$$

σ^2 : variance ;

$\sum (x_i - \mu)^2$: la somme des écarts au carré ;

x_i : la valeur de chaque observation ;

μ : la moyenne ;

n : le nombre total d'observations.

Il existe une autre formule pour calculer la variance pour les ensembles de données avec un grand effectif ($n > 30$) :

$$\sigma^2 = \frac{1}{n-1} \left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right]$$

σ^2 : variance ;

$\sum x_i$: la somme de toutes les observations

$\sum x_i^2$: la somme des carrés de toutes les observations

n : le nombre total d'observations

L'écart type est la racine carrée de la variance. Il mesure la dispersion des données autour de la moyenne en termes de l'unité de mesure originale. L'écart type est souvent noté par le symbole σ .

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

L'écart type est une mesure de dispersion plus couramment utilisée que la variance car il est exprimé dans les mêmes unités que les données, ce qui facilite l'interprétation des résultats. Il est également plus facile à manipuler mathématiquement que la variance.

Les valeurs élevées de l'écart type indiquent une grande dispersion des données, tandis que les valeurs faibles indiquent une faible dispersion. L'écart type est souvent utilisé pour décrire la dispersion des données dans une distribution normale.

3.2.2. Erreur standard

L'erreur standard est une mesure de la variabilité des données d'un échantillon par rapport à la moyenne de l'échantillon. Elle est également connue sous le nom *d'écart-type de l'échantillon* et est souvent utilisée pour estimer la précision d'une moyenne échantillonnée par rapport à la moyenne réelle de la population.

Plus précisément, l'erreur standard est calculée en divisant l'écart-type de l'échantillon par la racine carrée de la taille de l'échantillon. Elle représente la variation attendue entre les moyennes de tous les échantillons possibles d'une population donnée de la même taille.

$$\sigma_{\mu} = \sigma / \sqrt{n}$$

L'erreur standard est couramment utilisée en statistique pour calculer les intervalles de confiance et pour déterminer si les différences entre les moyennes de deux groupes sont significatives. Elle est également utilisée dans l'analyse de régression pour évaluer la précision des coefficients de régression.

Exemple : Supposons que nous ayons mesuré la hauteur de 20 plants de blé et que nous voulions estimer l'erreur standard de la moyenne de la hauteur pour cette population de plantes. Voici les données :

50, 52, 53, 51, 49, 52, 48, 50, 52, 51, 50, 53, 54, 50, 52, 51, 49, 50, 49, 53

Pour calculer l'erreur standard de la moyenne, nous avons besoin de la formule suivante:

$$\sigma_{\mu} = \sigma / \sqrt{n}$$

Pour calculer l'écart-type de l'échantillon, nous pouvons utiliser la formule suivante :

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

$$\sigma^2 = 1/(n-1) * [\sum(X_i - \mu)^2]$$

Où X_i est la $i^{\text{ème}}$ donnée, μ est la moyenne des données et n est la taille de l'échantillon.

Calculons d'abord la moyenne de ces données :

$$\mu = (50+52+53+51+49+52+48+50+52+51+50+53+54+50+52+51+49+50+49+53)/20 = 51,1$$

Maintenant, nous pouvons calculer l'écart-type de l'échantillon :

$$\sigma^2 = 1/(20-1) * [(50 - 51,1)^2 + (52 - 51,1)^2 + \dots + (53 - 51,1)^2] = 2,56$$

$$\sigma = \sqrt{2,56} = 1,6$$

Enfin, nous pouvons calculer l'erreur standard de la moyenne :

$$\sigma_{\mu} = \sigma/\sqrt{n} = 1,60/\sqrt{20} = 0,36$$

La moyenne peut donc s'écrire : 51,1±0,36

3.2.3. Coefficient de variation

Le coefficient de variation est une mesure de la dispersion relative d'un ensemble de données. Il est défini comme le rapport de l'écart-type à la moyenne de l'ensemble de données, multiplié par 100 pour obtenir un pourcentage :

$$\text{Coefficient de variation} = (\text{écart-type} / \text{moyenne}) \times 100\%$$

Le coefficient de variation est utilisé pour comparer la dispersion relative de deux ensembles de données ayant des moyennes différentes :

→ Plus le coefficient de variation est élevé, plus la dispersion relative de l'ensemble de données est grande. Par conséquent, un coefficient de variation élevé peut indiquer que l'ensemble de données est très hétérogène ou très dispersé.

→ En revanche, un coefficient de variation faible indique une faible dispersion relative de l'ensemble de données, ce qui peut indiquer que les données sont plus homogènes ou plus similaires les unes aux autres.

Exemple : Supposons avoir mesuré la teneur en azote de différents échantillons de sol provenant de différentes parcelles et avoir obtenu les résultats suivants :

Parcelle 1 : 0,28%

Parcelle 2 : 0,32%

Parcelle 3 : 0,35%

Parcelle 4 : 0,31%

Parcelle 5 : 0,33%

Pour calculer le coefficient de variation (CV) de ces données, nous avons besoin de la formule suivante : $CV = (\sigma/\mu) \times 100\%$

$$\mu = (0,28 + 0,32 + 0,35 + 0,31 + 0,33)/5 = 0,318$$

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

$$\sigma^2 = [1/(n-1) * (\sum (X_i - \mu)^2)]$$

$$\sigma^2 = 1/4 [(0,28 - 0,318)^2 + (0,32 - 0,318)^2 + (0,35 - 0,318)^2 + (0,31 - 0,318)^2 + (0,33 - 0,318)^2]$$

$$= 0,000729$$

$$\sigma = \sqrt{(\sigma^2)} = \sqrt{0,000729} = 0,027$$

$$CV = (\sigma/\mu) \times 100\%$$

$$= (0,027/0,318) \times 100\%$$

$$= 8,49\%$$

Le coefficient de variation de la teneur en azote pour ces échantillons de sol est donc de 8,49%. Cette valeur nous donne une idée de la variation relative des teneurs en azote entre les différentes parcelles de sol, ce qui peut être utile pour évaluer la stabilité ou l'uniformité des conditions de sol d'une zone donnée.

3.3. Modèle de distribution

La distribution normale est utilisée pour représenter la variation des caractères quantitatifs dans une population. Elle est caractérisée par une courbe symétrique en forme de cloche (Fig. 1), avec la moyenne (valeur centrale) représentant la tendance centrale de la population et l'écart type représentant la dispersion ou la variabilité des valeurs autour de la moyenne. Deux distributions de même moyenne peuvent diffuser notablement par la façon dont les mesures sont groupées autour de la moyenne.

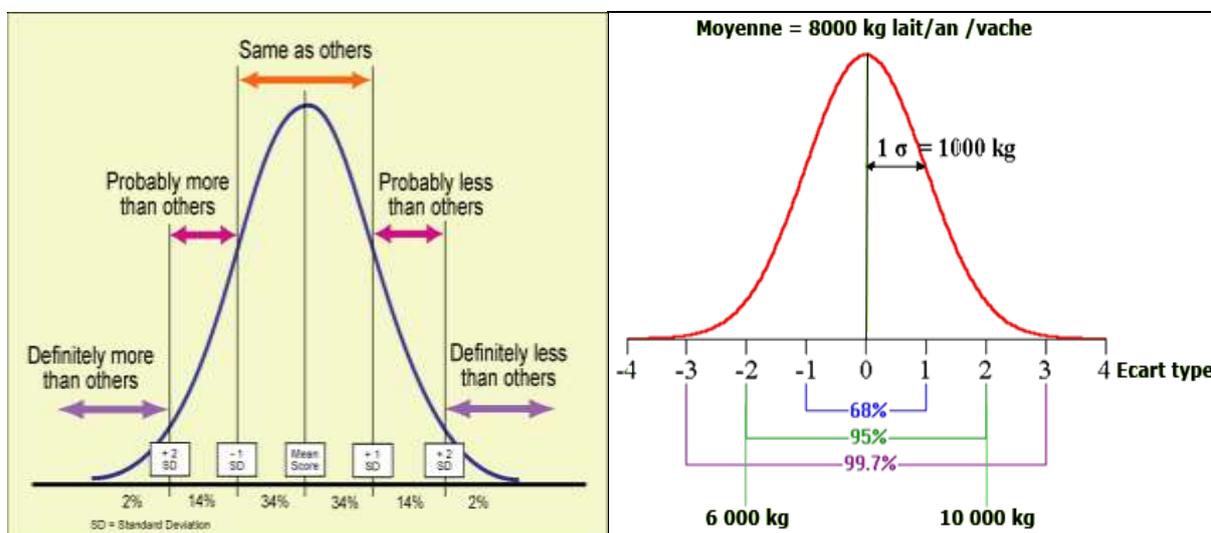


Figure 1 : Distribution normal des variables quantitatives continues.

La distribution normale en génétique quantitative est basée sur l'hypothèse que les effets des nombreux gènes contribuant à un trait donné sont indépendants et se combinent de manière additive. Selon cette hypothèse, la variation observée dans un trait est le résultat de la somme des effets génétiques et environnementaux, et suit donc une distribution normale.

Des exceptions par rapport à cette configuration peuvent se rencontrer. Des distributions asymétriques ou bimodale peuvent être rencontrés (Fig. 2).

Une distribution asymétrique peut indiquer la présence de valeurs aberrantes (beaucoup plus élevées ou beaucoup plus faibles que les autres variables). Une distribution bimodale peut indiquer que la population étudiée pourrait en fait être considérée comme un mélange de deux populations possédant chacune son propre mode (par exemple, la mesure de la taille d'une population d'étudiants et d'étudiantes).

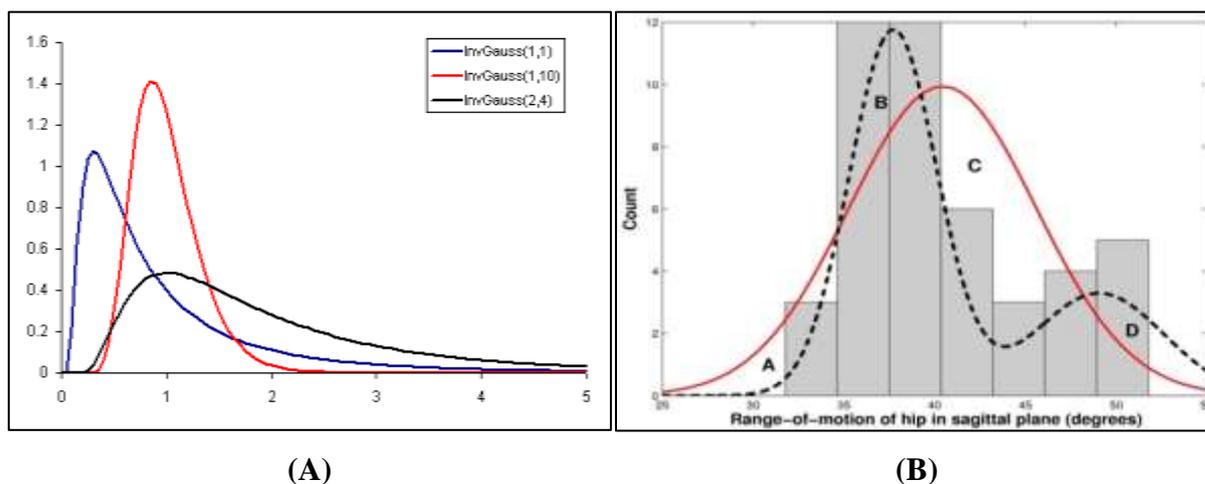


Figure 2 : Distributions asymétrique (A) et distribution bimodale (B).

3.4. Les corrélations

3.4.1. Généralités

Une autre notion statistique utilisée en génétique quantitative est la corrélation entre variables. On peut avoir des relations très étroites imprécises ou inexistantes.

La corrélation entre caractères quantitatifs, également appelée corrélation phénotypique, mesure la relation statistique entre deux traits ou caractères quantitatifs dans une population (Fig. 3). Elle permet de déterminer dans quelle mesure les variations dans un trait sont associées aux variations dans un autre trait, et si ces variations sont positives (les traits augmentent ou diminuent ensemble) ou négatives (les traits varient en sens inverse l'un de l'autre).

La corrélation entre caractères quantitatifs est généralement exprimée par le coefficient de corrélation (r), qui varie de -1 à $+1$. Une valeur de r proche de 1 indique une corrélation positive forte, ce qui signifie que les deux traits ont tendance à augmenter ou diminuer ensemble. Une valeur de r proche de -1 indique une corrélation négative forte, ce qui signifie

que les deux traits ont tendance à varier en sens inverse l'un de l'autre. Enfin, une valeur de r proche de 0 indique une absence de corrélation linéaire entre les deux traits.

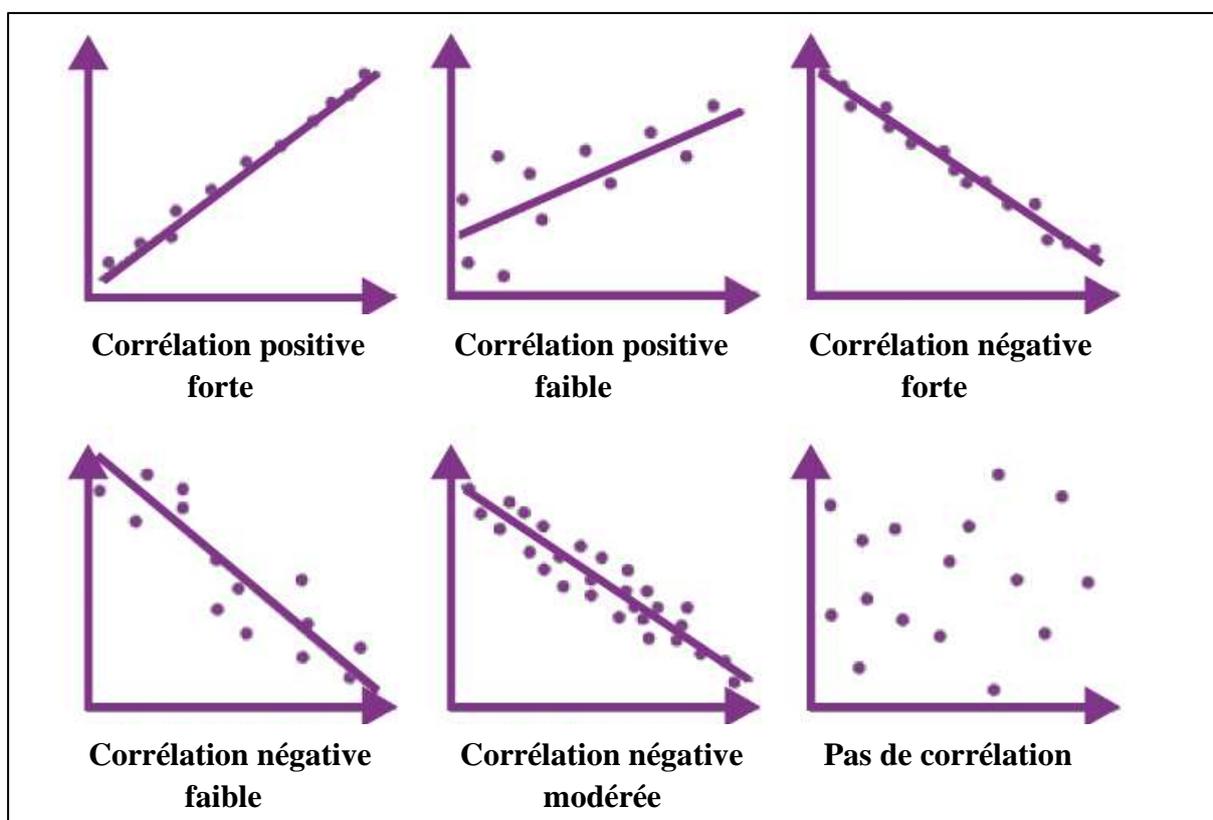


Figure 3 : différents types de corrélation entre caractères quantitatives

Le coefficient de corrélation « r_{xy} » est la grandeur communément utilisée pour estimer le degré de corrélation entre deux variables x et y . Il se calcule à partir des écarts de chaque valeur x à la moyenne des valeurs y .

$$\text{Cov}_{xy} = \frac{1}{n-1} [\sum (X_i - \bar{X})(Y_i - \bar{Y})] \dots \dots \dots (1)$$

$$\text{Cov}_{xy} = \frac{1}{n-1} \left[\sum (X_i Y_i) - \frac{\sum X_i \sum Y_i}{n} \right] \dots \dots \dots (2)$$

$$r_{xy} = \frac{\text{Cov}_{xy}}{\sigma_{xy}} \dots \dots \dots (3)$$

Dans la formule (3), les produits des écarts sont divisés par le produit des écarts type de x et de y . Cette normalisation par les écarts type a pour effet de rendre r_{xy} un nombre sans dimension indépendant des unités utilisées pour mesurer x et y ainsi définit r_{xy} .

Remarque

Parfois, il n'y a pas de relations linéaires entre deux variables mais qu'une relation régulière non linéaire existe entre elles, la valeur d'une variable peut parfaitement se prédire à partir de l'autre.

Exemple : Relation parabolique.

Au niveau d'une parabole, chaque valeur de **Y** peut se prédire à partir de la valeur de **X** sans qu'il y ait corrélation linéaire entre elles (Fig. 4).

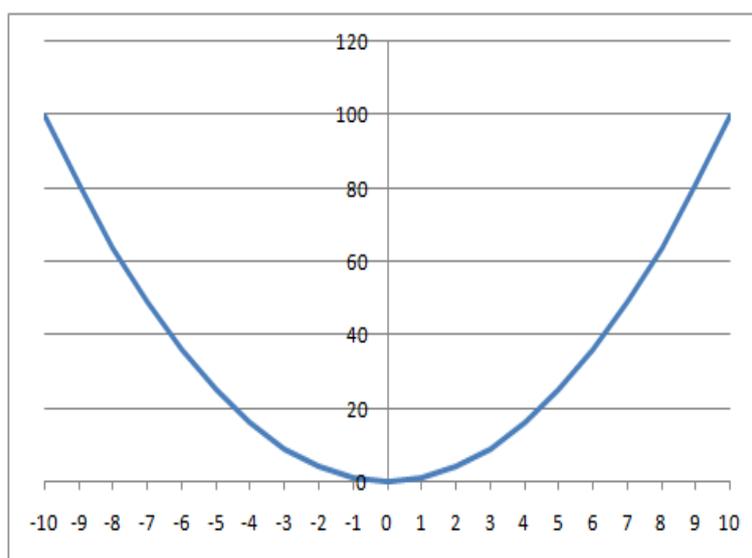


Figure 4 : Corrélation parabolique entre deux traits.

La corrélation peut se réaliser entre deux caractères d'un même individu mais aussi entre par exemple, la taille d'un parent (**x**) et celle d'un descendant (**y**) ou entre les tailles d'un membre âgé (**x**) un membre jeune (**y**) d'une famille. Cette utilisation de la corrélation est celle qui est pratiquée dans les problèmes de la génétique quantitative.

3.4.2. Corrélation et identité

Il est important de souligner que la corrélation n'équivaut pas l'identité. Des valeurs peuvent être très bien corrélées sans être identiques.

Exemple 1 : Les variable **x** (1 ; 2 ; 3 ; 4 ; 5 ; 6 ; 7 ; 8 ; 9 ; 10) et **y** (20 ; 22 ; 24 ; 26 ; 28 ; 30 ; 32 ; 34 ; 36 ; 38) au sein des couple sont parfaitement corrélés ($r = 1$) quoique chaque valeur de **y** est environ plus grande de 20 unités que la valeur correspondante de **x** (Fig. 5).

Deux valeurs sont parfaitement corrélées si pour une augmentation de l'une se produit une augmentation constante de l'autre ou diminution constante si r est négatif. Ici lorsque x augmente par une unité, y augmente par 2 unités.

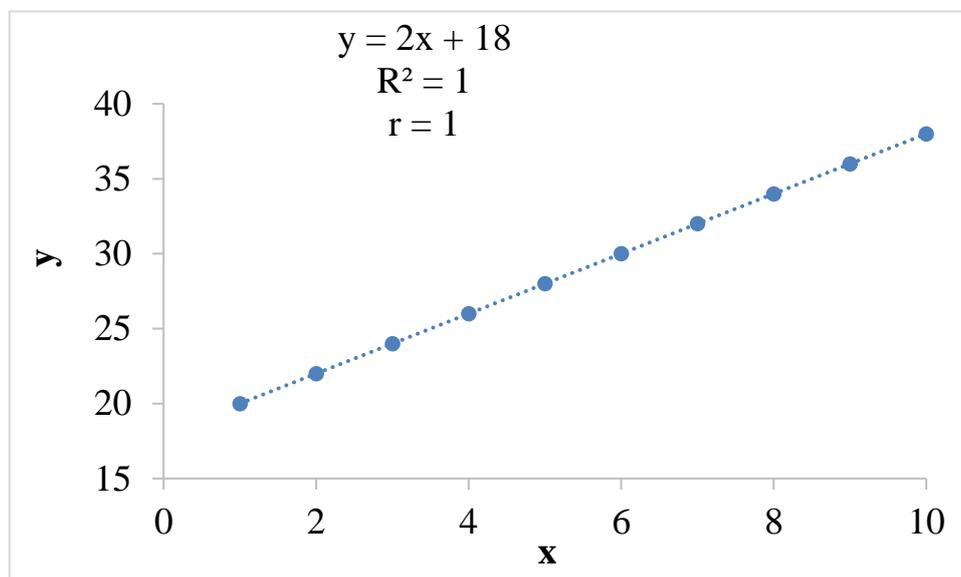


Figure 5 : Corrélation linéaire positive parfaite entre les variables x et y .

Exemple 2 : Des parents et leurs descendances peuvent être très bien corrélés pour une caractéristique comme la taille ; mais à cause d'une modification du milieu d'une génération à l'autre, chacun des enfants peut être plus grand que le parent.

Ce phénomène apparaît dans des études sur l'adoption où les enfants quoique corrélés à leurs parents biologiques peuvent en moyenne en être très différents en raison d'un changement de leur situation sociale.

3.5. La régression

La mesure de la corrélation ne fournit qu'une estimation de la précision de la relation existante entre deux variables. Le problème consiste à prédire la valeur d'une variable d'après la valeur prise par l'autre. Si x augmente de deux unités, de combien y augmentera-t-elle ?

La régression est une technique statistique utilisée pour modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes. Elle permet d'estimer et de prédire la valeur d'une variable dépendante en fonction des valeurs des variables indépendantes.

La régression peut être utilisée pour analyser des caractères quantitatifs et déterminer la relation fonctionnelle entre eux. Dans le cadre de la régression linéaire, la relation entre la

variable dépendante (y) et une variable indépendante (x) est modélisée par une équation linéaire de la forme :

$$y = a + bx$$

où (a) représente l'ordonnée à l'origine (intercept) et (b) représente la pente (coefficient de régression). L'équation de régression permet de prédire la valeur de (y) en fonction des valeurs de (x).

$$\text{la pente } b = \frac{\text{Cov } xy}{\sigma^2 x} \quad \text{et} \quad a = y - bx$$

La régression peut également être utilisée pour des modèles plus complexes, tels que la régression multiple, où plusieurs variables indépendantes sont incluses dans le modèle. Dans ce cas, l'équation de régression prend la forme :

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

où x_1, x_2, \dots, x_n représentent les différentes variables indépendantes et b_1, b_2, \dots, b_n sont les coefficients de régression correspondants.

L'équation ne peut pas prévoir exactement y pour une valeur donnée de x à cause de la dispersion existante autour de la droite.

3.6. Echantillons et populations

Lorsqu'on étudie un échantillon d'une population, on veut estimer ou connaître les caractéristiques de la population globale à travers l'échantillon.

On définit une population ou un échantillon par des caractéristiques statistiques (moyenne, variance, écart type, etc.). Ce sont des données certaines pour une population (paramètres) et des valeurs aléatoires pour un échantillon (caractéristiques). Les caractéristiques de chaque échantillon ne sont pas identiques à celle de l'ensemble globale mais varie d'un échantillon à un échantillon. Cette différence est traduite par des notations avec des symboles différents : des lettres grecques pour la population et des lettres latines pour l'échantillon (tableau 4).

Exemple : deux échantillons de luzerne, extrait de l'ensemble de luzerne, n'auront pas exactement la même moyenne et variance.

\bar{x} est en fait une estimation non biaisée de μ . Si un grand nombre d'échantillons sont utilisés et \bar{x} calculé pour chacun d'eux dans ce cas, la moyenne des valeurs \bar{x} sera μ .

Tableau 4 : Notations statistiques pour une population et un échantillon

Paramètre ou caractéristiques	Population	Echantillon
Moyenne	μ	\bar{x}
Variance	σ^2	s^2
Ecart type	σ	s
Proportion	π ou p	$f^{(1)}$
Effectif	N	n
Coefficient de régression	α, β	a, b
Coefficient de corrélation	ρ	r

⁽¹⁾ pour un échantillon on dira la proportion observée ou une fréquence.

Mais s^2 n'est pas une estimation non biaisée de σ^2 , elle a tendance à être un peu plus petite et la moyenne des valeurs de s^2 est inférieure à σ^2 . Comme estimation de l'ensemble représenté par les échantillons, la valeur appropriée à utiliser est :

$$\begin{aligned} \sigma^2 &= \frac{N}{N-1} [s^2] \\ \Rightarrow \sigma^2 &= \frac{N}{N-1} \left[\frac{1}{N} \left[\sum (xi - \bar{x})^2 \right] \right] \text{ (lorsque } N < 30) \\ \Rightarrow \sigma^2 &= \frac{N}{N-1} \left[\frac{1}{N} \left[\sum (xi - \bar{x})^2 \right] \right] \\ \Rightarrow \sigma^2 &= \frac{1}{N-1} \left[\sum (xi - \bar{x})^2 \right] \end{aligned} \quad \left| \quad \begin{aligned} \sigma^2 &= \frac{N}{N-1} [s^2] \\ \Rightarrow \sigma^2 &= \frac{N}{N-1} \left[\frac{1}{N} \left[\sum x^2 i - \frac{\sum (xi)^2}{N} \right] \right] \text{ (lorsque } N > 30) \\ \Rightarrow \sigma^2 &= \frac{N}{N-1} \left[\frac{1}{N} \left[\sum x^2 i - \frac{\sum (xi)^2}{N} \right] \right] \\ \Rightarrow \sigma^2 &= \frac{1}{N-1} \left[\sum x^2 i - \frac{\sum (xi)^2}{N} \right] \end{aligned}$$

avec N : nombre d'individus de l'échantillon.

Pour estimer l'écart type, qui est simplement la racine carrée de la variance, à partir d'un échantillon, la formule suivante est utilisée :

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N-1} \left[\sum xi - \bar{x} \right]^2} \text{ (si } N < 30) \text{ OU } \sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N-1} \left[\sum x^2 i - \frac{\sum (xi)^2}{N} \right]} \text{ (si } N > 30)$$

Ceci est une estimation non biaisée de σ^2 . Le choix entre les deux formules dépend de la préoccupation du chercheur (l'échantillon ou l'ensemble).

Toutes ces considérations sur les déviations s'appliquent aussi à la covariance de l'échantillon mais pas aux coefficients de corrélation car le facteur $n/n-1$ apparaîtra tant au numérateur qu'au dénominateur et donc ne jouera aucun rôle dans les calculs.

Il est important de noter que pour les données de la population entière, la variance est calculée en utilisant n à la place de $n-1$ dans la formule. La raison de l'utilisation de $n-1$ pour les échantillons est due à la correction de Bessel, qui prend en compte le fait que l'estimation

de la variance à partir d'un échantillon est plus incertaine que l'estimation de la variance de la population entière.

3.7. Génotypes et distribution des phénotypes

Supposant qu'une population de plante contient trois (03) génotypes, chacun d'eux ayant un effet différent sur le taux de croissance. En outre, supposons qu'une variation due au milieu existe de plante à plante en raison de propriétés hétérogènes du sol sur lequel la population se développe et qu'ils se produisent des fluctuations au cours du développement. Pour chaque génotype, nous aurons une distribution de phénotype distincte avec une moyenne et un écart type qui dépend du génotype et des facteurs de l'environnement (Fig. 6).

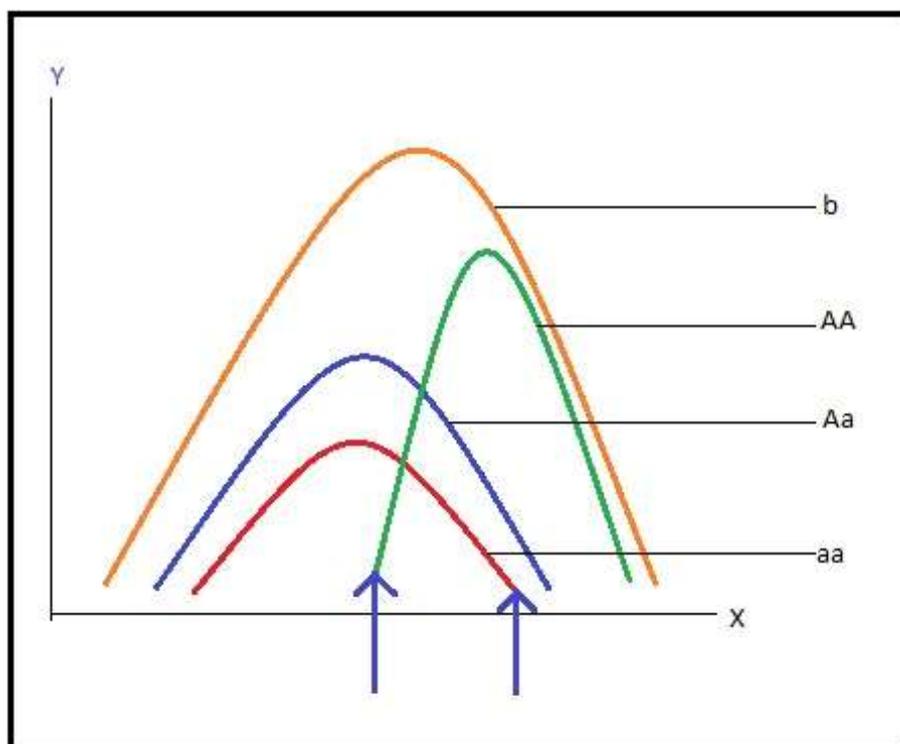


Figure 6 : Distribution des génotypes et du phénotype.

(aa : génotypes aa ; Aa : génotype Aa, AA : génotype AA ; b : phénotype)

Si les proportions sont de $1/aa$, $2/Aa$, $3/AA$, la distribution phénotypique des individus de la population globale sera de type « b » résultant de la somme des trois distributions génotypiques.

La moyenne de cette distribution globale est la moyenne des trois moyennes génotypiques. La variance de la distribution globale est produite en partie par la variation due à l'environnement et en partie par la moyenne légèrement différente des trois génotypes.

Propriétés de la distribution globale

- Malgré l'existence de trois distributions génotypiques distinctes, il n'y a qu'un seul mode ;
- N'importe quel individu dans la hauteur se situe entre les deux flèches (du graphe) pourrait provenir de n'importe lequel des trois génotypes en raison de leur chevauchement donc nous ne pouvons pas mener une analyse mendélienne simple pour déterminer le génotype d'un individu.

Un caractère quantitatif se définit par le fait que les différences phénotypiques moyennes entre génotypes sont faibles par rapport à la variation entre individus de même génotype.

Chapitre 2 :
Hérédité poly-génétique

Chapitre 2 : Hérité polygénique

1. Généralités

Les caractères continus sont les caractères pour lesquels les phénotypes varient graduellement d'un extrême à l'autre (ex. taille des humains). Ces caractères varient plutôt le long d'un continuum. Mais on ne peut pas classer les humains seulement selon qu'ils sont "grands" ou "petits" (Fig. 7).



Figure 7 : Extrêmes de la taille chez les êtres humains.

En général, les caractères continus dépendent de plus d'un gène. Les caractères qui sont régis par plusieurs gènes sont appelés des **caractères polygéniques**.

L'hérité polygénique, également connue sous le nom d'hérité quantitative ou hérabilité polygénique, fait référence à l'influence de plusieurs gènes sur un trait ou une caractéristique donnée. Contrairement à l'hérité monogénique, où un seul gène détermine en grande partie le trait, l'hérité polygénique implique que de nombreux gènes interagissent pour influencer la variation d'un trait donné.

De nombreux traits sont déterminés par la contribution de plusieurs gènes. Par exemple, la hauteur des plantes, la croissance des racines, le rendement des cultures, la résistance aux maladies, la teneur en éléments nutritifs, la tolérance au stress environnemental, la couleur des fleurs, etc., sont souvent des caractéristiques polygéniques.

L'hérité polygénique des traits peut également être influencée par des facteurs environnementaux. Les conditions de croissance, y compris le sol, l'humidité, la température,

l'exposition à la lumière, etc., peuvent interagir avec les gènes pour influencer l'expression des traits polygéniques chez les plantes.

La compréhension de l'hérédité polygénique est d'une grande importance pour la sélection et l'amélioration des cultures. En identifiant les gènes et les allèles qui contribuent aux traits souhaités, les sélectionneurs peuvent utiliser ces informations pour développer des variétés de plantes présentant des caractéristiques améliorées, telles que le rendement élevé, la résistance aux maladies, etc.

Exemple : La couleur de la peau chez l'humain.

La couleur de la peau résulte des interactions de plusieurs facteurs déterminés par des paires différentes de gènes. Certains gènes pourraient agir sur le métabolisme des mélanocytes de la peau et en modifier le taux de production de mélanine. Si les mélanocytes produisent plus de mélanine, la peau sera plus foncée. D'autres gènes peuvent déterminer la distribution de la mélanine dans l'épaisseur de la peau (chez les noirs, la mélanine est dispersée dans toute l'épaisseur de l'épiderme, ce qui n'est pas le cas chez les blancs). Certains gènes pourraient déterminer les quantités relatives de chacun des deux types possibles de mélanine (on connaît deux types de mélanine, l'eumélanine, noire, et la phéomélanine, jaune-rouge). D'autres pourraient affecter la production de certaines hormones intervenant dans l'activité des mélanocytes.

Supposons, par exemple, que la couleur de la peau soit régie par 4 paires de gènes différents. Supposons également que chacun des gènes impliqués n'existe que sous deux formes alléliques : A et a, B et b, C et c, D et d (la réalité est sans doute beaucoup plus complexe).

Supposons encore que les allèles A, B, C et D déterminent des facteurs responsables d'une couleur plus foncée de la peau et les allèles a, b, c et d déterminent des facteurs responsables d'une couleur claire et qu'il n'y a pas de dominance entre les allèles. Un individu AA BB CC DD aurait une couleur de peau très foncée alors qu'un individu aa bb cc dd aurait une couleur très pâle.

On comprend aussi qu'on peut obtenir toutes les nuances entre ces deux extrêmes selon les combinaisons possibles (Aa BB CC dd donnerait une couleur particulière alors que aa Bb cc DD en donnerait une autre).

Un couple dont les deux parents seraient AaBbCcDd pourrait avoir des enfants possédant toutes les nuances possibles de couleur (de « AABBCcDD » à « aabbccdd »).

2. Les relations entre les gènes

L'ensemble de caractères mesurables sur les êtres vivants (plantes ou animaux) sont régies par un système polygénique. Ces gènes peuvent avoir des effets différents sur les caractères étudiés. En effet, les modes d'action de ces gènes sont l'additivité, la dominance et l'épistasie.

2.1. Définition des effets moyens (additivité)

Les effets moyens, également connus sous le nom d'additivité, font référence à la contribution individuelle de chaque allèle (forme alternative d'un gène) à un trait donné dans le cadre de l'hérédité polygénique.

Dans le contexte de l'hérédité polygénique, chaque gène impliqué dans la détermination d'un trait particulier peut avoir plusieurs allèles différents. Chaque allèle peut avoir un effet spécifique sur le trait, et ces effets s'additionnent pour influencer l'expression phénotypique du trait.

Dans une population que nous supposons panmictique (croisements au hasard) un caractère génotypique donné est généralement conditionné par plusieurs structures homologues ($A_1, A_2, A_3, \dots, A_i, \dots, A_n$). Chacune de ces (n) structures ayant sa fréquence relative ($P_1, P_2, P_3, \dots, P_i, \dots, P_n$). Un individu (ig) de fréquence relative ($P_i P_g$) dans cette population a donc le génotype $A_i A_g$. Sa valeur génotypique correspondante pourrait s'écrire $[A_i A_g]$.

L'appréciation de la valeur moyenne $[A_i]$ apportée par une structure (A_i) peut être faite à partir de tous les génotypes contenant (A_i). Donc :

$$[A_i] = \frac{P_1 P_i [A_1 A_i] + P_2 P_i [A_2 A_i] + P_3 P_i [A_3 A_i] + \dots + P_i P_i [A_i A_i] + \dots + P_n P_i [A_n A_i]}{P_1 P_i + P_2 P_i + P_3 P_i + \dots + P_i P_i + \dots + P_n P_i}$$

Cette équation se simplifie par P_i . En notant que $P_1 + P_2 + P_3 + \dots + P_i + \dots + P_n = 1$

$$\text{Donc : } [A_i] = \frac{P_i [P_1 [A_1 A_i] + P_2 [A_2 A_i] + P_3 [A_3 A_i] + \dots + P_i [A_i A_i] + \dots + P_n [A_n A_i]]}{P_i (P_1 + P_2 + P_3 + \dots + P_i + \dots + P_n) \Rightarrow = 1}$$

$$[A_i] = \sum P_n [A_i P_n]$$

Lorsque les effets moyens sont présents, cela signifie que chaque allèle contribue indépendamment et de manière additive à la variation du trait. Par conséquent, la somme des effets de tous les allèles présents chez un individu détermine en grande partie la valeur phénotypique du trait observé.

Cette expression peut être appliquée en cas simple où la population ne comprend que deux structures A et B de fréquences respectives p et q. Cette situation se rencontre dans la multiplication panmictique d'un hybride entre deux lignées homozygotes.

♀ ↙	A (p)	B (q)
♂ ↘	A (p)	B (q)
A (p)	AA (p ²)	AB (pq)
B (q)	AB (pq)	BB (q ²)

D'où :

AA	AB	BB
p ²	2pq	q ²

$$\text{Avec : } p^2[AA] + 2pq[AB] + q^2[BB] = 1$$

D'où nous tirons les effets moyens des structures :

$$[A] = p^2[AA] + pq[AB] \quad \text{et} \quad [B] = q^2[BB] + pq[AB]$$

On peut considérer que les trois génotypes [BB], [AB] et [AA] marquent une graduation de type 0, 1 et 2 (Fig. 8).

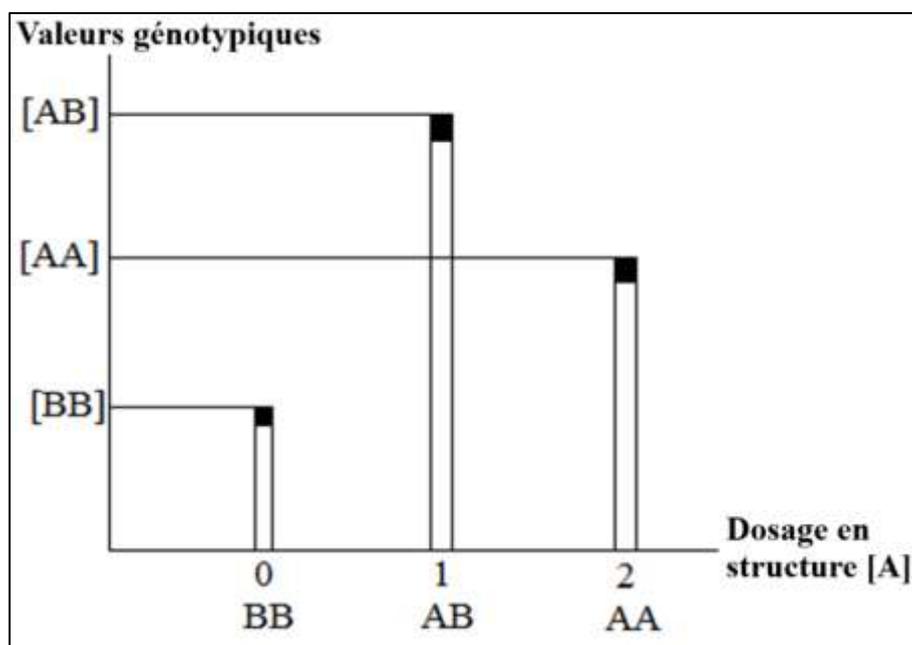


Figure 8 : Valeurs génotypiques en fonction du dosage en structure A

Si les apports des structures A et B dans la réalisation du phénotype comprennent des effets uniquement additifs [A'] et [B'], le génotype [AB] devrait avoir une valeur phénotypique intermédiaire entre [AA] et [BB] (F). Dans ce cas, les trois génotypes seraient alors alignés sur une droite (Δ') en des ordonnées $2[B']$, $2[A']$ et $[A'] + [B']$.

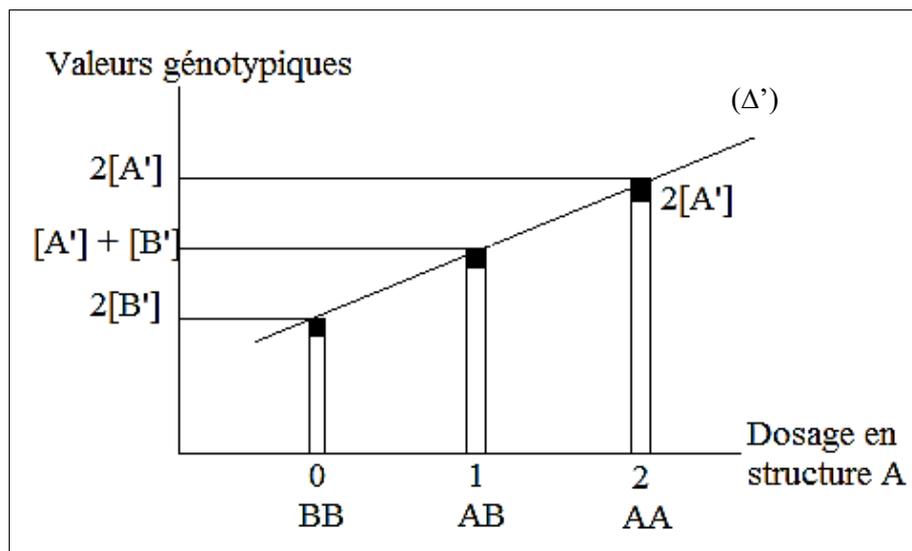


Figure 9 : Représentation d'un cas d'additivité parfaite.

Les effets moyens ne sont pas toujours strictement additifs. Dans la pratique, ce cas est très rarement réalisé, il peut y avoir des interactions entre les allèles, ce qui signifie que l'effet d'un allèle peut être modifié par la présence d'autres allèles. Ces interactions peuvent être synergiques (effets combinés plus importants que l'addition) ou antagonistes (effets combinés moins importants que l'addition).

Parmi les effets multiples qui concourent à la réalisation des valeurs génotypiques réelles [BB], [AB] et [AA], la part appelée additivité sera celle qui est définie par la droite de régression (Δ) la mieux ajustée aux valeurs réelles, donc par la droite de régression entre ces valeurs génotypiques (Fig. 10).

Sur cette droite correspond respectivement aux abscisses, des dosages 0, 1 et 2 de structure A et se trouve sur les points d'ordonnées $2[B']$, $[A'] + [B']$ et $2[A']$ qui représente la part d'additivité pour chaque valeur phénotypique.

Les effets moyens sont souvent utilisés pour quantifier la contribution de chaque allèle à la variation d'un trait et pour prédire les performances phénotypiques des individus dans le cadre de la sélection et de l'amélioration génétique.

2.2. Effet de la dominance

L'effet de la dominance est un concept important dans la génétique qui décrit l'interaction entre les allèles d'un gène donné et leur influence sur l'expression phénotypique d'un trait. Il concerne spécifiquement les cas où un allèle masque ou prédomine sur un autre allèle dans la détermination du trait.

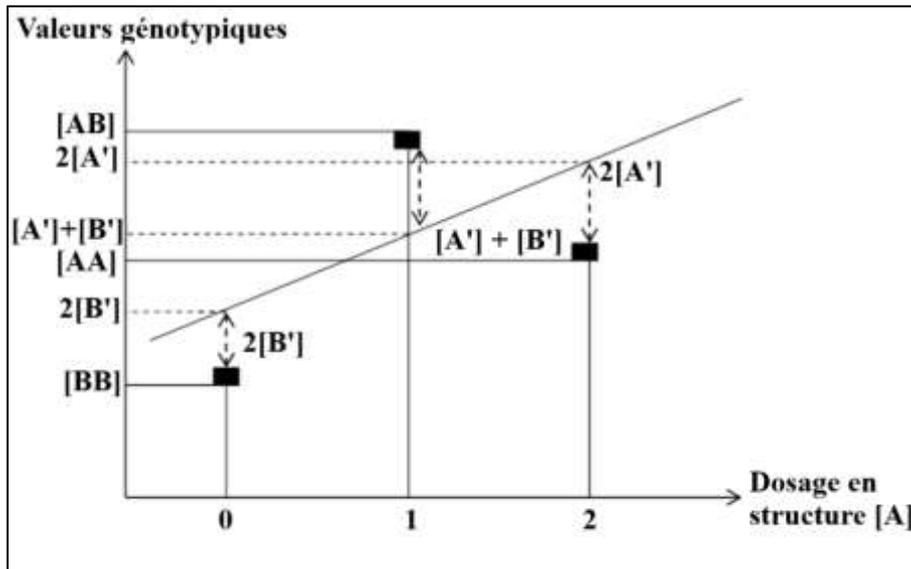


Figure 10 : Régression et estimation de la dominance et de l'additivité

On considère la figure 10. On peut remarquer pour les zones chromosomiques considérées que les effets de l'additivité ne couvraient qu'une partie de la valeur génotypique. L'écart entre ces valeurs réelles observées et la valeur additive définit les effets de la dominance. *La dominance apparaît donc comme une somme d'interaction entre les ségrégents homologues.*

Chez une plante diploïde, pour une zone chromosomique donnée, si l'hétérozygote a un génotype parfaitement intermédiaire entre celui des deux homozygotes, il y a une additivité parfaite (Fig. 11).

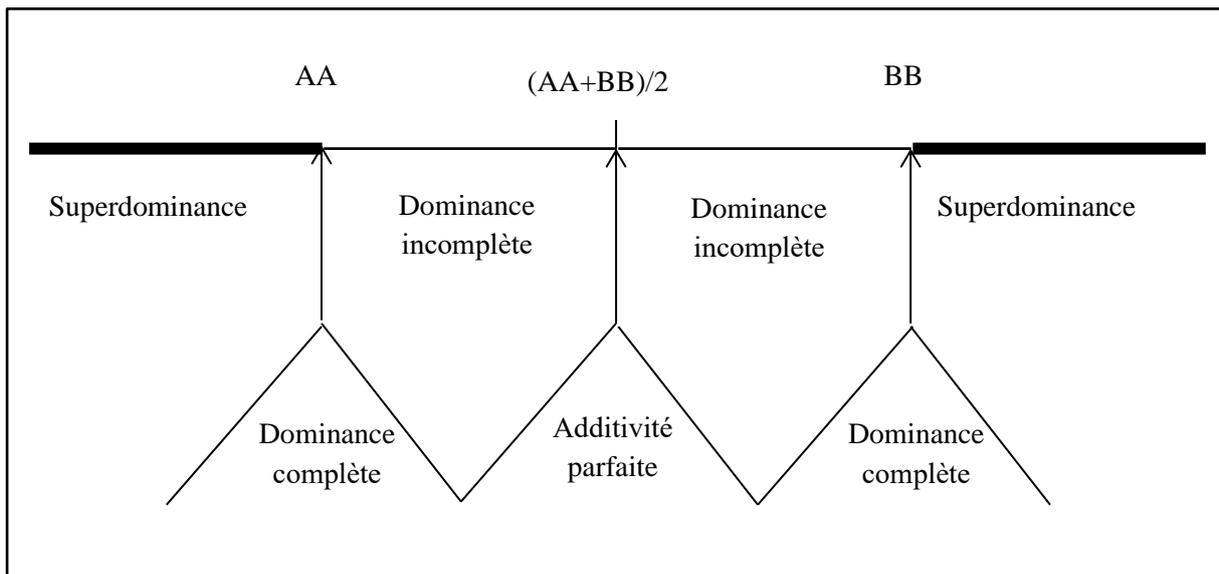


Figure 11 : Dominance et définition de l'hétérozygote

Les autres situations se définissent comme suit :

- $[AB] = [AA]$ ou $[BB] \Rightarrow$ Dominance complète ;
- $[AB] < [AA]$ ou $[BB] \Rightarrow$ Dominance incomplète ;
- $[AB] > [AA]$ ou $[BB] \Rightarrow$ Superdominance.

La dominance et l'additivité ne sont pas mutuellement exclusives et peuvent coexister dans les systèmes génétiques. Dans certains cas, les traits peuvent être influencés à la fois par des allèles dominants et récessifs, ainsi que par des effets additifs de plusieurs gènes.

2.3. Les formes d'épistasie

Pour la plupart des caractères complexes, il n'y a pas que deux segments homologues qui interviennent ; mais plusieurs zones en divers points du génome ont pu décomposer la valeur génotypique.

En effet, l'additivité est en interaction avec la dominance pour des segments homologues ; mais il existe différents types d'interactions entre segments non homologues dont la somme constitue les effets d'épistasie.

L'épistasie est un concept en génétique qui décrit l'interaction entre les gènes, où l'effet d'un gène masque ou modifie l'effet d'un autre gène dans la détermination d'un trait phénotypique. L'épistasie peut se produire lorsque deux gènes ou plus interagissent pour influencer un même trait.

Dans le contexte de l'épistasie, un gène est considéré comme épistatique s'il masque ou modifie l'effet d'un autre gène, appelé gène hypostatique. Le gène épistatique peut inhiber ou bloquer l'expression d'un trait phénotypique, indépendamment des allèles présents dans le gène hypostatique.

Il existe différents types d'épistasie, notamment l'épistasie dominante et l'épistasie récessive

2.3.1. Épistasie dominante

Dans ce cas, un seul allèle dominant du gène épistatique est suffisant pour masquer l'expression du gène hypostatique. Par conséquent, même si le gène hypostatique possède des allèles récessifs qui pourraient normalement influencer le trait, ils seront masqués par l'allèle dominant du gène épistatique.

2.3.2. Épistasie récessive

Dans ce cas, les deux allèles récessifs du gène épistatique doivent être présents pour masquer l'expression du gène hypostatique. Si au moins un allèle dominant est présent dans le gène épistatique, l'effet du gène hypostatique sera observé.

L'épistasie peut être illustrée par divers exemples. Par exemple, dans le cas de la couleur des cheveux chez les mammifères, le gène responsable de la production de pigment mélanique peut être affecté par un gène épistatique qui bloque la production de pigment, indépendamment des allèles présents dans le gène hypostatique. Ainsi, même si le gène hypostatique a des allèles qui déterminent la couleur des cheveux, ils ne seront pas exprimés en présence de l'allèle dominant du gène épistatique.

L'épistasie peut compliquer la prédiction des phénotypes en génétique, car elle modifie les ratios attendus de certains croisements génétiques. Comprendre les interactions épistatiques est essentiel pour une meilleure compréhension de la génétique des traits complexes.

Ces interactions non homologues peuvent intervenir entre des effets additifs ($A \times A \times A \dots$), entre des effets de dominance ($D \times D \times D \dots$) ou être mixte ($A \times D \times A \times D \dots$) (Fig. 12).

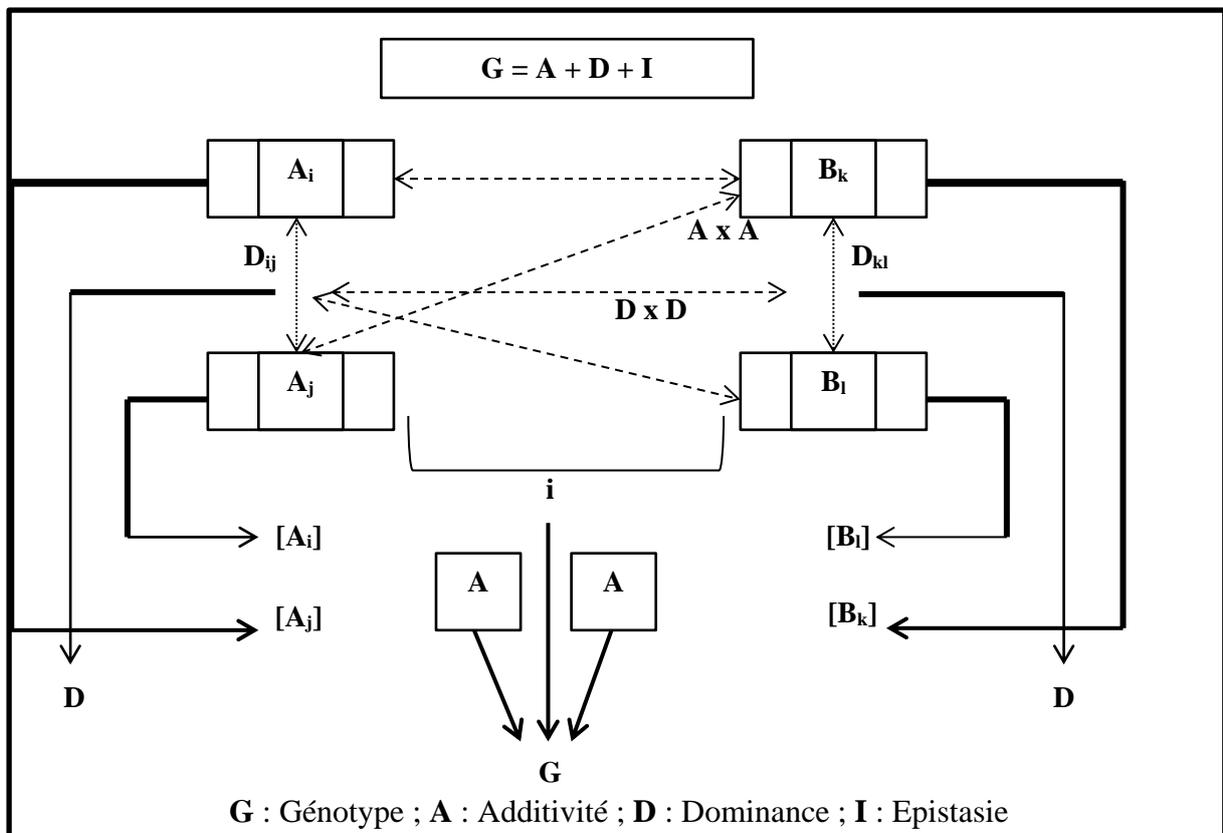


Figure 12 : Les divers types d'effets (direct et en interaction) intervenant dans l'expression génotypiques : cas de deux paires de gènes A_i , A_j , B_k et B_l

Chapitre 3 :
Héritabilité d'un caractère

Chapitre 3 : Héritabilité d'un caractère

1. Définition

La question de savoir si un caractère est oui ou non héréditaire pose celle du rôle joué par les différences génétiques dans les différences phénotypiques existant entre individus ou groupes d'individus.

L'héritabilité d'un caractère est une mesure de la proportion de la variation phénotypique d'un trait dans une population qui est attribuable à des différences génétiques. En d'autres termes, il s'agit de quantifier dans quelle mesure les différences observées dans un trait entre les individus sont dues à des variations génétiques plutôt qu'à des facteurs environnementaux.

2. Parenté et héritabilité

La parenté et l'héritabilité sont deux concepts liés à la génétique et à la transmission des traits d'une génération à l'autre, mais ils ont des significations différentes.

La parenté fait référence au degré de relation génétique entre deux individus. Les individus qui partagent un ancêtre commun plus récent ont une parenté plus élevée que ceux qui partagent un ancêtre commun plus éloigné.

Des individus biologiquement apparentés doivent se ressembler d'avantage que des individus non apparentés. Cette ressemblance se reflètera comme une corrélation positive entre parents et descendants, ou entre individus issus des mêmes parents.

Exemples :

- ✓ Des parents plus grands que la moyenne devraient avoir une descendance plus grande que la moyenne.
- ✓ Plus une plante produit de graines, plus ses descendants devraient en produire.

La parenté fait référence au degré de relation génétique entre les individus, tandis que l'héritabilité mesure la proportion de la variation phénotypique d'un trait dans une population qui est attribuable à des différences génétiques.

La parenté peut influencer l'héritabilité d'un trait, car des traits héréditaires ont plus de chances d'être transmis entre des individus étroitement apparentés. Cependant, la parenté seule ne détermine pas l'héritabilité d'un trait. L'héritabilité est une mesure statistique qui est

déterminée par des études génétiques et dépend de la variation génétique observée dans une population spécifique.

Des caractères sont dits familiaux, si les membres d'une même famille les partagent quel que soit la raison. Ces caractères ne sont héréditaires que si leur similitude est due à des génotypes communs. Chez les organismes se prêtant à l'expérimentation, il n'y a pas de problèmes pour distinguer les similitudes dues au milieu de celle dues à l'hérédité.

Exemples :

- ✓ Les descendances d'une plante sensible à une maladie et celle d'une plante résistante peuvent être cultivées ensemble dans le même milieu pour déterminer si malgré la similitude du milieu, chacune d'entre elles ressemble à son propre parent (différence génétique).
- ✓ Le fait que des parents parlent une langue et que leurs enfants parlent la même langue ne permet pas de dire que c'est un caractère héréditaire mais c'est un caractère familial. S'il y a immigration, les enfants parleront la langue du pays d'accueil.

3. Comment quantifier l'héritabilité ?

Si un caractère montre une certaine héritabilité dans une population, il est alors possible de quantifier son degré d'héritabilité.

3.1. Par les variances

La variation au sein des phénotypes d'une population découle de deux sources. D'une part, les différences moyennes entre les génotypes et d'autre part, chaque génotype exprime une variance phénotypique suite à la variance du milieu.

La variance phénotypique totale (notée σ_p^2) peut être divisée en deux parties :

- Les variances au sein des valeurs génotypiques moyennes (σ_g^2) ;
- La variance du milieu (σ_e^2).

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

L'héritabilité au sens large (H^2) est une mesure de la proportion de la variation phénotypique d'un trait dans une population qui est attribuable à des différences génétiques. Elle inclut à la fois les effets directs des gènes (héritabilité additive) et les effets des interactions entre les gènes (héritabilité non additive), tels que les effets de dominance et d'épistasie. En

d'autres termes, H^2 capture la contribution totale des facteurs génétiques à la variation du trait et le degré d'héritabilité peut être défini comme la contribution de la variance génétique à la variance totale :

$$H^2 = \frac{\sigma^2_g}{\sigma^2_p} = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e} \quad \text{héritabilité au sens large}$$

En revanche, l'héritabilité au sens étroit (h^2) se réfère spécifiquement à la proportion de la variation phénotypique d'un trait qui est attribuable aux seuls effets des gènes additifs. Elle ne prend pas en compte les effets non additifs, tels que la dominance ou l'épistasie. Ainsi, h^2 représente une estimation plus restreinte de la contribution génétique à la variation du trait.

$$h^2 = \frac{\sigma^2_{ad}}{\sigma^2_p} \quad \text{héritabilité au sens étroit}$$

La distinction entre H^2 et h^2 est importante car elle permet de mieux comprendre les mécanismes génétiques sous-jacents à la variation des traits. Si h^2 est significativement élevée, cela suggère que les différences génétiques entre les individus contribuent de manière importante à la variation du trait. En revanche, si H^2 est élevée par rapport à h^2 , cela indique que les effets non additifs, tels que la dominance, jouent un rôle important dans la variation du trait.

Remarques

L'héritabilité d'un caractère est variable selon la population et selon l'ensemble des milieux dans lesquelles celle-ci se développe.

L'héritabilité n'est pas une caractéristique fixée d'un caractère ; mais dépend de la population au sein de laquelle la mesure est effectuée de même que l'ensemble des milieux dans lesquels, la population évolue.

- Dans une population, un caractère peut être influencé par les allèles ségrégant à de nombreux loci. Dans une autre, ces loci peuvent très bien être **homozygotes (lignée pure)** et dans ce cas, l'héritabilité des caractères sera nulle puisque la variance génotypique (σ^2_g) = 0 avec une $H^2 = 0$; une telle valeur ne signifie pas que les gènes ne jouent aucun rôle dans le développement du caractère ; mais bien que la variation existante entre les individus de cette population ne peut pas être attribué à la variation génétique.
- Une population qui se développe dans un milieu très homogène se caractérisera par un caractère donné, par une variance du milieu plus petite qu'une population évoluant dans un

milieu plus variable. La variance de milieu sera nulle et donc $\sigma_g^2 = \sigma_p^2$ avec une $H^2 = 1$. L'absence d'hétérogénéité du milieu exprimera une valeur plus élevée de l'héritabilité ; mais cette valeur ne signifie nullement que le caractère est insensible à l'action du milieu.

3.1.1. Analyse approfondie de la variance (variance additive et variance due à l'environnement)

La subdivision de la variance phénotypique peut fournir une information intéressante pour les sélectionneurs et les éleveurs.

Exemple 1 : Si deux allèles A et a à un locus, contrôlant la taille, ségrégent dans les milieux où la population se trouve, les phénotypes moyens (hauteur de la tige) et la fréquence des trois génotypes sont les suivants :

Tableau 5 : Résultats des moyennes et fréquences des génotypes (aa ; Aa et AA) de la hauteur de la tige.

	aa	Aa	AA
Phénotypes	10	18	20
Fréquences	0,36	0,48	0,16

La variance additive se calcule à partir du carré de l'écart à la moyenne de l'effet moyen de A par rapport à la moyenne pondérée à sa moyenne.

$$s^2_{ad} = 2[f_A(A - \bar{m})^2 + f_a(a - \bar{m})^2]$$

La fréquence de l'allèle a: $f_a = \frac{2(aa)+1(Aa)}{2} = \frac{2(0,36)+1(0,48)}{2} = 0,60$

avec : 2 car il y a deux allèles a dans le génotype homozygote aa.

1 car il y a un allèle a dans le génotype hétérozygote Aa.

La fréquence de l'allèle A : $f_A = \frac{2(AA)+1(Aa)}{2} = \frac{2(0,16)+1(0,48)}{2} = 0,40$

L'effet moyen de l'allèle a est noté \bar{a} .

$$\bar{a} = \frac{2(0,36)(10)+1(0,48)(18)}{2(0,36)+1(0,48)} = 13,20 \text{ cm}$$

L'effet moyen de l'allèle A est noté \bar{A} .

$$\bar{A} = \frac{2(0,16)(20) + 1(0,48)(18)}{2(0,16) + 1(0,48)} = 18,80 \text{ cm}$$

Cette différence moyenne de 5,6 cm entre les effets dus aux allèles A et a ($18,80 - 13,20 = 5,60$) rend compte d'une partie de la variance phénotypique mais pas dans son ensemble car l'hétérozygote n'est pas exactement intermédiaire par rapport aux homozygotes, il y a un effet de dominance.

$$\text{La moyenne phénotypique} = 10(0,36) + 18(0,48) + 20(0,16) = 15,44 \text{ cm}$$

Donc :

$$s_{ad}^2 = 2[0,4(18,80 - 15,44)^2 + 0,6(13,20 - 15,44)^2] = 15,05 \text{ cm}^2$$

La variance génétique totale associée à ce locus peut être divisée en :

- Une variance génétique additive (variance génétique due à la substitution de A par a),
- Une variance due à la dominance (variance génétique que résulte de la dominance partielle de A sur a chez les hétérozygotes).

La variance génétique totale issue de la variation parmi les phénotypes moyens des 3 génotypes est de :

$$s_g^2 = 0,36(10 - 15,44)^2 + 0,48(18 - 15,44)^2 + 0,16(20 - 15,44)^2 = 17,13 \text{ cm}^2$$

D'où :

$$s_d^2 = s_g^2 - s_{ad}^2 = 17,13 \text{ cm}^2 - 15,05 \text{ cm}^2 = 2,08 \text{ cm}^2$$

Cette subdivision de la variance génétique est utile car elle permet de prédire l'effet de la sélection.

Exemple 2 : Considérons le même exemple précédant sauf que les allèles qui contrôlent la taille des plantes au locus donné sont A et A₁.

Tableau 6 : Résultats des moyennes et fréquences des génotypes (AA ; AA₁ et A₁A₁) de la hauteur de la tige.

	AA	AA ₁	A ₁ A ₁
Phénotypes	10	12	10
Fréquences	0,25	0,50	0,25

L'effet moyen de l'allèle A_1 .

$$\bar{A}_1 = \frac{2(0,25)(10) + 1(0,50)(12)}{2(0,25) + 1(0,50)} = 11 \text{ cm}$$

L'effet moyen de l'allèle \bar{A} .

$$\bar{A} = \frac{2(0,25)(10) + 1(0,50)(12)}{2(0,25) + 1(0,50)} = 11 \text{ cm}$$

La différence moyenne est de 0 cm entre les effets dus aux allèles A et a ($11 - 11 = 0$).

La moyenne phénotypique = $10(0,25) + 12(0,50) + 10(0,25) = 11 \text{ cm}$

La variance génétique totale issue de la variation parmi les phénotypes moyens des 3 génotypes est de : $s^2_g = 0,25(10 - 11)^2 + 0,50(12 - 11)^2 + 0,25(10 - 11)^2 = 1 \text{ cm}^2$

La variance additive

$$s^2_{ad} = 2[fA(A - \bar{m})^2 + fa(a - \bar{m})^2]$$

La fréquence de l'allèle a : $fA_1 = \frac{2(aa) + 1(Aa)}{2} = \frac{2(0,25) + 1(0,50)}{2} = 0,50$

La fréquence de l'allèle A : $fA = \frac{2(AA) + 1(Aa)}{2} = \frac{2(0,25) + 1(0,50)}{2} = 0,50$

Donc : $s^2_{ad} = 2[0,5(11 - 11)^2 + 0,5(11 - 11)^2] = 0 \text{ cm}^2$

D'où : $s^2_d = s^2_g - s^2_{ad} = 1 \text{ cm}^2 - 0 \text{ cm}^2 = 1 \text{ cm}^2$

Interprétation

Il n'y a pas de différence entre les effets des allèles A et A_1 puisque chacun d'eux détermine un effet de 11 unités, il n'y a pas de variance génétique additive quoiqu'une variance due à la dominance soit présente, les individus les plus grands sont hétérozygotes.

Si un sélectionneur s'efforce d'augmenter la taille de cette population par sélection, l'union de ces hétérozygotes ne donnera mieux qu'à la population originale. La sélection sera inopérante (n'aboutit à rien, on reviendra aux parents).

L'effet de la sélection dépend de la part due à la variance génétique additive et non de la variance génétique en générale. Il en résulte que c'est l'héritabilité au sens étroit et non l'héritabilité au sens large qui permet de prédire, la réponse à la sélection.

3.1.2. Estimation des composantes de la variance génétique

Les composantes génétiques de la variance sont estimées à partir de la covariance entre individus apparentés (tableau 7).

Tableau 7 : Proportions relatives de la variance additive et de la variance due à la dominance continue dans la covariance génétique entre individus apparentés.

Individus apparentés	Proportions estimées de :	
	S^2_{ad}	S^2_d
Cov (jumeaux identiques)	1	1
Cov (parents – enfants)	1/2	00
Cov (demi consanguin)	1/4	00
Cov (consanguin)	1/2	1/4

Ces relations combinées avec la variance phénotypique totale permettent d'estimer l'héritabilité au sens étroit.

Exemple

La covariance entre parents et enfants contient une moitié de la variance additive donc :

$$\text{Cov (parent – enfant)} = \frac{S^2_{ad}}{2} \Rightarrow 2 \text{ Cov (parent – enfant)} = S^2_{ad}$$

D'où :

$$\frac{2 \text{ Cov (p – e)}}{S^2_p} = \frac{S^2_{ad}}{S^2_p} = 2 \text{ corrélation} = h^2$$

Il en découle que le double de la corrélation parent-enfant fournit une estimation de l'héritabilité au sens étroit.

Il y a une autre façon d'estimer l'héritabilité au sens étroit. Si nous portons en graphique les phénotypes des descendants par rapport aux phénotypes moyens de leurs parents (Fig. 13). Il est possible d'obtenir les relations suivantes :

La droite de régression passera par la moyenne de tous les parents et la moyenne de tous les descendants qui sont égales puisque aucun changement ne s'est produit dans la population d'une génération à l'autre.

En outre, des parents plus grands ont des enfants plus grands et des parents plus petits ont des enfants plus petits de tel sorte que la pente de la droite de régression est positive mais la pente n'est pas égale à 1.

Des parents très petits ont des enfants légèrement plus grands et de très grands parents ont des enfants légèrement plus petits qu'eux même. Cette pente inférieure à 1 de la droite de régression provient de ce que l'héritabilité est loin d'être parfaite.

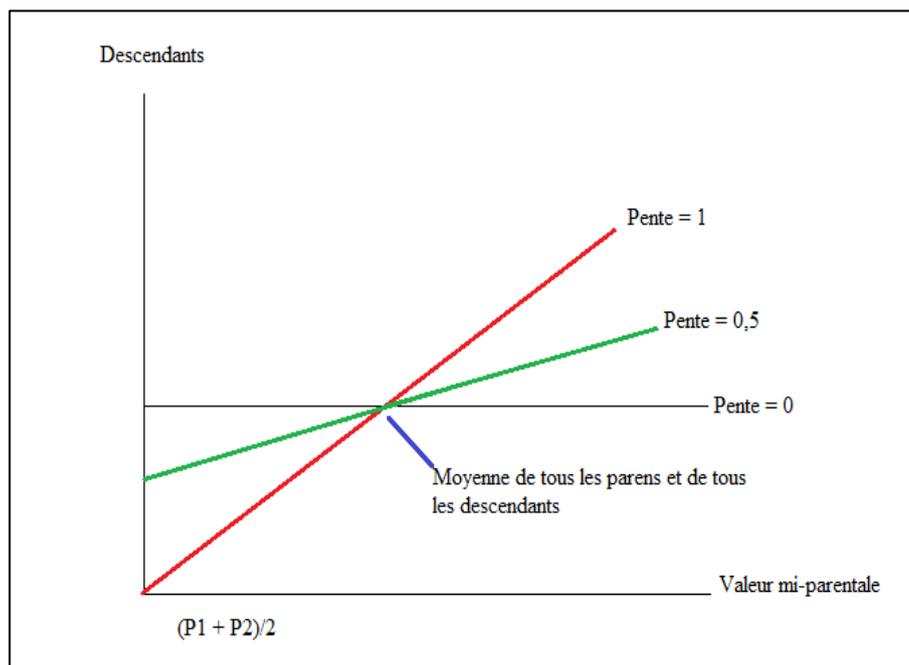


Figure 13 : Régression des valeurs mesurées des descendants rapportées à celle des moyennes des deux parents pour un caractère à héritabilité étroite de 0,5 (droite verte). La droite rouge illustre la pente de régression obtenue si le caractère était parfaitement héréditaire.

Si un phénotype était hérité additivement avec une fidélité parfaite, la taille des descendants serait identique à la valeur mi- parentale et la pente de la droite serait égale à 1.

Par contre, si les descendants ne représentaient aucune similarité avec leurs parents, tous les parents auraient des descendants de la même taille moyenne et la pente de la droite de régression serait égale à 0.

Ceci suggère que la pente de la droite de régression rapportant la valeur de la descendance à la valeur mi- parentale constitue une estimation de l'héritabilité additive (au sens étroite) et de fait, cette relation est précise.

Par définition, la régression $b = \frac{Cov\ xy}{s^2_x}$ (de la valeur des descendants sur la variance des parents). A partir du tableau des proportions des variances additives et de dominances, on aura :

$$Cov (P-E) = \frac{S^2_a}{2}$$

Il est à noter que les valeurs moyennes varient moins que leurs composantes, donc on posant s^2_p est la variance phénotypique des parents, nous obtenons ce qui suit :

$$2b = 2\left(\frac{\frac{S^2_a}{2}}{S^2_P}\right) = \frac{S^2_a}{S^2_P} = h^2$$

Le fait que la pente correspond à l'héritabilité additive nous permet de l'utiliser pour prévoir les effets de la sélection artificielle.

3.2. Autre méthode d'estimation de la valeur de l'héritabilité

Supposant que nous choisissons comme parent de la génération suivante ceux qui ont en moyenne deux unités de valeurs au-dessus de la moyenne générale de la population au sein de laquelle ils ont été choisis.

Si l'héritabilité au sens étroit égale à **0,5**, la descendance des parents sélectionnés de la population se situe à **0,5*2=1** unité au-dessus de la moyenne de la population actuelle puisque le coefficient de régression prédit qu'une augmentation de **y** sera due à l'augmentation d'une unité de **x**.

Nous pouvons définir :

→ *L'écart de la sélection comme l'écart entre les parent sélectionnés et la moyenne générale de la population sélectionnée) ;*

→ *La réponse de la sélection comme l'écart entre leurs descendants et la génération du départ.*

$$\text{Réponse à la sélection} = h^2 * \text{écart de la sélection}$$

$$h^2 = \frac{\text{Réponse à la sélection}}{\text{Ecart de la sélection}}$$

Cette deuxième expression nous donne un autre moyen d'exprimer l'héritabilité au sens étroit en sélectionnant pendant une génération et en comparant la réponse avec l'écart de la sélection. Le plus souvent ceci se fait pendant plusieurs générations et on détermine la réponse moyenne.

3.3. L'utilisation de h^2 en élevage et en amélioration

L'importance de h^2 est grande pour les sélectionneurs même si sa valeur ne s'applique qu'à une population particulière et une gamme de milieux.

Exemple : Un généticien de volailles désireux d'augmenter le taux de croissance, ne se sent pas concerné par la variance génétique de toutes les basses-cours possibles distribuées dans tous les milieux possibles.

La question est de savoir si un schéma de sélection peut être conçu afin d'augmenter la vitesse de croissance et si oui, dans quelle mesure ?

- Si un groupe présente une **variance génétique élevée et une autre** une variance **très faible**, l'éleveur choisira la **première** pour sa sélection.
- Si l'héritabilité (h^2) du groupe choisi est très élevée, la moyenne de la population répondra rapidement à la pression de sélection imposée, puisque la majeure partie de la supériorité des parents choisis se traduira dans la descendance. Plus h^2 est élevée, plus la corrélation parent-enfant le sera aussi.
- Si h^2 est faible alors une petite fraction seulement de l'augmentation du taux de croissance des parents sélectionnés se retrouvera dans la génération suivante.
- Si H^2 et h^2 sont faibles, cela implique que l'importance de la variance due au milieu est grande par rapport à la variance génétique. Peut-on diminuer la variance du milieu (S^2_e) ? Il existe plusieurs méthodes pour diminuer (S^2_e) :
 - ✓ Modifier les conditions d'élevage de telle sorte que la variance due au milieu décroisse ;
 - ✓ Utiliser la **sélection par famille**. Plutôt que de choisir les meilleurs individus, le sélectionneur produit à partir de divers couples plusieurs descendances et **l'accouplement** se décide sur la base de la performance moyenne des descendances. En établissant une moyenne à partir des descendances, les fluctuations incontrôlées dues au milieu ou au développement sont éliminées et une meilleure estimation de la différence génotypique entre couples peut être établie et permet donc de choisir les meilleurs partenaires comme parents de la génération suivante.
- **Si h^2 est faible mais H^2 est élevée**, cela signifie que le rôle de la variance due au milieu est minime. La faible héritabilité au sens étroit est le résultat d'une faible variance génétique

additive par rapport à la variance due à la dominance et aux interactions. Une telle situation conduit à faire appel à des schémas de sélection particuliers basés sur la variance non-additive. C'est **la méthode des hybrides contrôlés** largement utilisés chez le maïs, la tomate. Un grand nombre de lignées pures sont produites par autofécondation. Elles sont ensuite croisées entre elles dans toutes les combinaisons possibles (si cela se justifie économiquement) et l'on choisit le croisement fournissant le meilleur hybride. De nouvelles lignées pures sont produites à partir de cet hybride supérieur et des croisements sont à nouveau réalisés afin de déceler au cours de ce second cycle le meilleur hybride. Ce schéma sélectionne en fait les effets de dominance, parce qu'il se base sur les meilleurs hétérozygotes.

Chapitre 4 :
Hétérozygotie

Chapitre 4 : Hétérozygotie

1. Généralités

L'hétérozygotie est un concept en génétique qui fait référence à la présence de deux allèles différents pour un gène particulier chez un individu. Un allèle est une version alternative d'un gène qui peut influencer un trait spécifique ou une caractéristique chez un organisme.

L'hétérozygotie joue un rôle crucial dans la génétique quantitative pour ces raisons :

- **Diversité génétique** : L'hétérozygotie est un indicateur de la diversité génétique au sein d'une population. Une population présentant une hétérozygotie élevée est généralement considérée comme plus diversifiée génétiquement. Cette diversité génétique est importante car elle offre une plus grande variabilité dans les traits et les caractères, ce qui peut favoriser l'adaptation et la survie face à des changements environnementaux.
- **Potentiel évolutif** : L'hétérozygotie est liée au potentiel évolutif d'une population. Lorsqu'une population présente une hétérozygotie élevée, elle dispose d'un réservoir génétique plus riche. Cela signifie qu'elle a une plus grande capacité à générer de nouvelles combinaisons génétiques par recombinaison lors de la reproduction, ce qui peut conduire à l'émergence de nouvelles variations phénotypiques et favoriser l'évolution de la population.
- **Stabilité génétique** : L'hétérozygotie joue un rôle dans le maintien de la stabilité génétique au sein des populations. Lorsque les individus sont hétérozygotes pour un gène donné, cela signifie qu'ils portent deux allèles différents pour ce gène. Cette diversité allélique accrue peut contribuer à maintenir un équilibre génétique et à prévenir l'accumulation de mutations délétères.
- **Amélioration génétique** : En génétique quantitative, l'hétérozygotie est souvent recherchée dans le cadre de programmes d'amélioration génétique des plantes et des animaux. Lorsqu'il y a une hétérozygotie élevée, cela peut indiquer une plus grande variabilité génétique et un potentiel d'amélioration plus élevé pour les caractères souhaités, tels que la productivité, la résistance aux maladies ou la qualité des produits.
- **Étude des maladies génétiques** : L'hétérozygotie est également importante dans l'étude des maladies génétiques. Dans certains cas, être hétérozygote pour un gène peut conférer un avantage en termes de résistance à certaines maladies. Par exemple, dans le cas des maladies récessives, être hétérozygote peut offrir une protection contre la maladie tout en maintenant la diversité génétique de la population.

2. Hétérozygotie et fréquence allélique

L'hétérozygotie est souvent mesurée en calculant le taux d'individus hétérozygotes dans une population donnée. Par exemple, si une population compte 100 individus et que 20 d'entre eux sont hétérozygotes pour un gène spécifique, le taux d'hétérozygotie serait de 20%.

2.1. Calcul de l'hétérozygotie à partir des fréquences alléliques

Dans une population diploïde, la fréquence allélique peut être calculée à partir de l'hétérozygotie et vice versa. Supposons qu'il y ait deux allèles pour un gène, A et a, et que p représente la fréquence de l'allèle A et q la fréquence de l'allèle a dans la population. Alors, la fréquence de l'homozygotie AA serait p^2 , la fréquence de l'homozygotie aa serait q^2 , et la fréquence de l'hétérozygotie Aa serait $2pq$. De plus, $p + q = 1$, car les fréquences alléliques totales doivent se sommer à 1.

L'hétérozygotie peut être calculée à partir des fréquences alléliques en utilisant la formule suivante :

$$\text{Hétérozygotie} = 2pq$$

où **p** représente la fréquence de l'allèle A et q représente la fréquence de l'allèle a dans la population.

Exemple : Supposons que dans une population, la fréquence de l'allèle A soit de 0,6 ($p = 0,6$) et la fréquence de l'allèle a soit de 0,4 ($q = 0,4$).

$$\text{Hétérozygotie} = 2pq = 2(0,6)(0,4) = 0,48$$

Donc, dans cette population, l'hétérozygotie est de 0,48, ce qui signifie que 48% des individus sont hétérozygotes pour ce gène.

Remarque : Cette formule est basée sur certaines hypothèses, notamment l'équilibre de Hardy-Weinberg, qui suppose l'absence de forces évolutives telles que la sélection, la dérive génétique, la migration et la mutation. Dans la réalité, ces forces peuvent influencer les fréquences alléliques et l'hétérozygotie d'une population.

2.2. Relation entre l'hétérozygotie et l'homéostasie génétique

L'hétérozygotie fait référence à la présence de deux allèles différents pour un gène chez un individu. Elle est souvent associée à une plus grande variabilité génétique au sein d'une population. Lorsque les individus sont hétérozygotes pour un gène, cela peut contribuer à maintenir un équilibre génétique et à prévenir l'accumulation de mutations délétères.

L'hétérozygotie peut également favoriser l'adaptation et la survie face à des changements environnementaux, car elle offre une plus grande diversité génétique et un potentiel évolutif plus élevé.

L'homéostasie génétique, en revanche, se réfère à la tendance des populations à maintenir la stabilité des fréquences alléliques au fil du temps, en l'absence de forces évolutives majeures. L'homéostasie génétique est basée sur le principe de l'équilibre de Hardy-Weinberg, qui spécifie que les fréquences alléliques restent constantes d'une génération à l'autre si certaines conditions sont remplies. Dans cet équilibre, les fréquences alléliques et l'hétérozygotie peuvent être prédites en fonction des fréquences alléliques initiales.

Ainsi, l'hétérozygotie peut contribuer à l'homéostasie génétique en maintenant un équilibre entre les allèles et en évitant que les fréquences alléliques ne dévient significativement de leurs valeurs initiales. Si les individus hétérozygotes ont un avantage sélectif par rapport aux homozygotes pour un gène donné, cela peut conduire à un équilibre stable où ces deux types d'allèles sont maintenus dans la population.

Cependant, il est important de noter que l'homéostasie génétique n'est pas un état permanent et peut être perturbée par des forces évolutives, telles que la sélection naturelle, la dérive génétique, la migration et la mutation. Ces forces peuvent modifier les fréquences alléliques et l'hétérozygotie d'une population au fil du temps.

3. Les lois de Hardy-Weinberg

3.1. Présentation des lois de Hardy-Weinberg

Les lois de Hardy-Weinberg décrivent les principes fondamentaux de l'équilibre génétique dans les populations mendéliennes, en l'absence de forces évolutives majeures. Ces lois ont été formulées indépendamment par les généticiens britannique Godfrey Hardy et allemand Wilhelm Weinberg au début du XXe siècle.

3.2. Conditions d'équilibre de Hardy-Weinberg

Dans une population formée d'effectifs illimités, non soumise à la sélection ni aux mutations et/ou migrations et où les unions se font au hasard, les fréquences géniques restent constantes de génération en génération, de même que les fréquences génotypiques, celles-ci pouvant se déduire de la connaissance des fréquences géniques. Une population est dite en équilibre de Hardy-Weinberg lorsqu'elle satisfait à toutes les conditions de validité à cette loi.

- ✓ **Population fermée** : cette condition exclu les migrations et les croisements entre populations qui peuvent modifier les fréquences géniques,
- ✓ **Effectif illimité** : les fréquences géniques et génotypiques peuvent varier d'une génération à l'autre du seul fait du hasard mais ces variations tendent à s'accumuler dès que l'effectif de la population devient grand
- ✓ **Absence de sélection** : on suppose que le nombre de descendants d'un individu est indépendant de son génotype,
- ✓ **Absence de mutation** : les mutations sont les événements aléatoires de très faible probabilité qu'ils font apparaître des gènes nouveaux ou font disparaître des gènes existants dans la population et modifier donc les fréquences géniques.
- ✓ **Le régime de panmixie** : union au hasard s'oppose au système suivant :
 - **Homogamie** : quand les unions se font préférentiellement entre individus qui se ressemblent phénotypiquement (homogamie phénotypique) ou génotypiquement (homogamie génotypique),
 - **Hétérogamie** : quand les unions se font préférentiellement entre individus dissemblables,
 - **Endogamie** : quand les unions se font préférentiellement entre individus apparentés (consanguinité),
 - **Exogamie** : quand les unions se font préférentiellement entre individus non apparentés.

Démonstration

Considérant le cas d'un locus avec deux allèles et une population infinie où les fréquences des 3 génotypes ont des valeurs quelconques à la génération N.

$$A_1A_1 = P ; A_1A_2 = 2Q \text{ et } A_2A_2 = R$$

$$P + 2Q + R = 1$$

La génération suivante sera obtenue dans les conditions de la panmixie en supposant les 3 génotypes également fertiles (absence de sélection) et non susceptibles de muter par combinaison au hasard des gamètes porteurs de l'allèle A_1 dont la fréquence est $P + Q = p$ et des gamètes porteurs de l'allèle A_2 dont la fréquence $Q + R = q$.

Tableau 8 : Résultats du croisement panmictiques des gamètes A1 et A2

Gamètes	A ₁ (p)	A ₂ (q)
A ₁ (p)	A ₁ A ₁ (p ²)	A ₁ A ₂ (pq)
A ₂ (q)	A ₁ A ₂ (pq)	A ₂ A ₂ (q ²)

$$p^2 + 2pq + q^2 = 1$$

La fréquence de l'allèle A₁ dans cette génération est = p²+pq

La fréquence de l'allèle A₁ dans cette génération = p(p + q) et comme p + q = 1

La fréquence de l'allèle A₁ dans cette génération = p

La fréquence de l'allèle A₂ dans cette génération est = q²+pq

La fréquence de l'allèle A₂ dans cette génération = q(p + q) et comme p + q = 1

La fréquence de l'allèle A₂ dans cette génération = q

Donc les fréquences géniques dans une population ne varient pas. Comme on a les mêmes fréquences géniques, alors les fréquences génotypiques de la génération n+2 resteront également inchangées.

4. Evolution de l'hétérozygotie au cours des générations

Si l'individu de départ à la première génération est homozygote pour un locus, il en résultera que tous ces descendants obtenus par autofécondation seront homozygotes et identiques pour ce locus.

Si par contre, l'individu du départ est hétérozygote (Aa), l'autofécondation produira ¼ d'homozygote AA, ¼ d'homozygote aa et ½ d'hétérozygote Aa. Si un seul individu est choisi dans la génération suivante pour assurer la propagation de la lignée, il existe 50% de chance (½) qu'il soit à son tour hétérozygote, après 2 générations 25% (¼), après 3 générations 1/8 et à la n^{ème} génération $H_n = 1/2^n$ et H tend vers 0 (H_n la proportion des locus hétérozygote à la n^{ème} génération).

Lorsque l'autofécondation n'est pas possible, les unions frères et sœurs aboutissent aux mêmes résultats quoique plus lentement (tableau 9).

Tableau 9 : Niveau d'hétérozygotie après plusieurs générations de consanguinité dans le cas de deux systèmes d'appariement.

Génération	Niveau d'hétérozygotie	
	Autofécondation	Croisement frère-sœur
0	1,000	1,000
1	0,500	0,750
2	0,250	0,625
3	0,125	0,500
4	0,062	0,406
10	0,000977	0,114
20	$1,05 \cdot 10^{-6}$	0,014
n	$H_n = \frac{1}{2} H_{n-1}$	$H_n = \frac{1}{2} H_{n-1} + \frac{1}{4} H_{n-2}$

5. Facteurs influençant l'hétérozygotie

L'hétérozygotie, qui fait référence à la présence de deux allèles différents pour un gène chez un individu, peut être influencée par plusieurs facteurs. Voici quelques-uns des principaux facteurs qui peuvent affecter l'hétérozygotie :

5.1. Mutation génétique

Les mutations génétiques, qui sont des changements aléatoires dans la séquence d'ADN, peuvent influencer l'hétérozygotie en introduisant de nouveaux allèles dans une population. Les mutations peuvent augmenter l'hétérozygotie en créant de nouveaux allèles et en ajoutant à la diversité génétique de la population.

5.2. Flux génétique (migration de gènes)

Le flux génétique, qui correspond au mouvement des individus ou des gènes entre les populations, peut influencer l'hétérozygotie. Si des individus porteurs d'allèles différents migrent d'une population à une autre, cela peut augmenter l'hétérozygotie dans la population receveuse.

5.3. Sélection naturelle

La sélection naturelle peut jouer un rôle important dans le maintien ou la réduction de l'hétérozygotie dans une population. Si les individus hétérozygotes ont un avantage sélectif par rapport aux homozygotes (ce qu'on appelle l'hétérozygotie avantageée), cela peut favoriser

l'hétérozygotie. En revanche, si les homozygotes ont un avantage sélectif, cela peut conduire à une réduction de l'hétérozygotie.

5.4. Dérive génétique

La dérive génétique, qui est le changement aléatoire des fréquences alléliques au fil des générations, peut influencer l'hétérozygotie. Dans de petites populations, la dérive génétique peut entraîner une fluctuation aléatoire des fréquences alléliques, ce qui peut réduire l'hétérozygotie au fil du temps.

5.5. Consanguinité

La consanguinité, qui correspond à la reproduction entre individus étroitement apparentés, peut réduire l'hétérozygotie. Lorsque des individus consanguins se reproduisent, il y a une plus grande probabilité que les allèles identiques se rencontrent, ce qui réduit l'hétérozygotie et augmente l'homozygotie.

5.6. Ségrégation génétique

Lors de la formation des gamètes, les allèles sont répartis de manière aléatoire, ce qui peut influencer l'hétérozygotie. Si les allèles sont répartis de manière équitable et indépendante lors de la méiose, cela peut favoriser l'hétérozygotie en générant une variété de combinaisons alléliques dans les gamètes.

Chapitre 5 :
Phénomène de l'hétérosis

Chapitre 5 : Le phénomène de l'hétérosis

1. Définition

L'hétérosis a été défini au niveau d'un croisement de deux lignées homozygotes comme la supériorité de l'hybride F_1 par rapport au meilleur des parents en passant de la F_1 à la F_2 par autofécondation en fécondation libre.

L'hétérosis peut également être défini au niveau du croisement entre populations comme la supériorité de l'hybride résultant par rapport à la meilleure population.

En général, l'hétérosis est beaucoup plus fort chez les plantes allogames que chez les plantes autogames. Les caractères les plus affectés sont les caractères complexes.

2. Hypothèses ou mécanismes

Deux grandes hypothèses ont été formulées : l'Hypothèse de dominance, et de superdominance

2.1. Dominance

Pour un caractère quantitatif contrôlé par plusieurs gènes, l'avantage assez systématique du croisement entre plantes non apparentées par rapport au croisement entre plantes apparentées peut être dû à la réunion dans un même génotype d'un grand nombre de gènes dominants favorables.

Exemple

$$L_1 : AABBccdd \quad x \quad L_2 : aabbCCDD$$

$$\text{Hybride :} \quad AaBbCcDd$$

L'hybride réuni en une seule génération, les 4 gènes favorables dans un même génotype. Dans cette situation, il est possible de concevoir un génotype homozygote aussi bon que la F_1 , l'hétérosis est dit fixable.

L'autofécondation des plantes F_1 fait apparaître les récessifs défavorables d'où la chute de la vigueur de la F_1 à la F_2 . Cette hypothèse suppose que la dominance soit dans le sens favorable.

2.2. Superdominance

Dans cette hypothèse, la supériorité de l'hybride par rapport aux parents viendrait de la supériorité de l'état hétérozygote à un certain nombre de loci. La combinaison des deux gènes à l'état hétérozygote entraînerait une nouvelle potentialité supérieure à celle des homozygotes.

Exemple 1

Soient 2 gènes à deux allèles A_1 et A_2 . A_1 efficace dans le milieu E_1 et A_2 efficace dans le milieu E_2 . E_1 et E_2 sont des milieux successifs dans la vie de l'individu (tableau 10).

Tableau 10 : Effets des environnements sur les génotypes.

Génotypes	Milieux		Bilan
	E_1	E_2	
A_1A_1	4	1	5
A_1A_2	4	4	8
A_2A_2	1	4	5

Dans chaque milieu, il y a seulement dominance mais sur l'ensemble des milieux il y a super dominance, on parle de super dominance marginale.

Exemple 2

Soient deux gènes allèles à effet pléiotropique (gène peut coder pour plusieurs caractères) peut coder pour C_1 et C_2 . A_1 est dominant pour C_1 et A_2 est dominant pour C_2 (tableau 11).

Tableau 11 : Effets des caractères sur les génotypes.

Génotypes	Caractères		Bilan	
	C_1	C_2	Somme	Produit
A_1A_1	4	1	5	4
A_1A_2	4	4	8	16
A_2A_2	1	4	5	4

L'hétérozygotie, pour le caractère complexe, présente donc un net avantage sans qu'il y ait superdominance pour chacun des caractères. L'hétérosis dans le cas de superdominance est infixable.

3. Evolution de l'hétérosis au cours des générations

Dans ce chapitre, on s'intéresse à l'hétérosis manifesté par le croisement de deux lignées ou deux populations qui ne possèdent aucune origine commune. Dans ce cas, l'étude de la théorie de l'hétérosis se fera à partir des fréquences des gènes dans les deux lignées.

3.1. Moyenne d'une population

Considérons l'effet d'un seul locus avec deux allèles dont les valeurs des 3 génotypes sont désignés par : +a ; b et -a (Fig. 14).

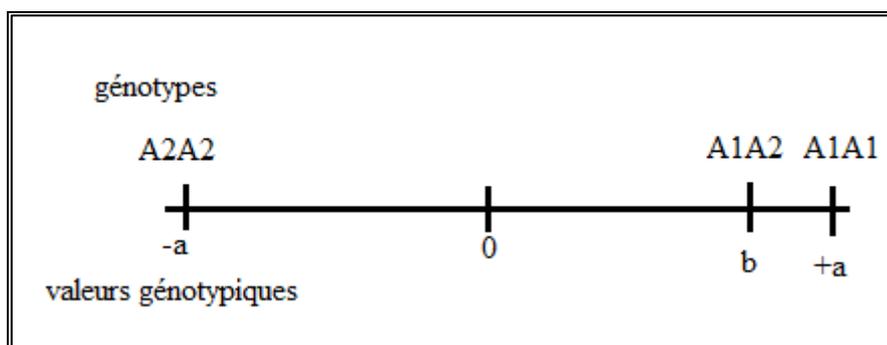


Figure 14 : Représentation graphique de la répartition des trois génotypes autour du centre.

La valeur b de l'hétérozygote dépend du degré de dominance :

- ✓ S'il n'y a pas de dominance : $b = 0$;
- ✓ Si A_1 domine A_2 : $b > 0$;
- ✓ Si A_2 domine A_1 : $b < 0$;
- ✓ Si la dominance est complète : $b = +a$ ou $b = -a$;
- ✓ S'il y a superdominance : $b > +a$ ou $b < -a$;
- ✓ Le degré de dominance = b/a

Exemple : Considérons deux allèles A_1 et A_2 contrôlant la hauteur des tiges des plantes d'orges. Les hauteurs moyennes pour les 3 génotypes sont représentées dans le tableau suivant :

Tableau 12 : Hauteurs moyennes des tiges des trois génotypes d'orge.

	Génotypes		
	A_1A_1	A_1A_2	A_2A_2
Hauteurs	100 cm	85 cm	50 cm

La moyenne de la population se détermine selon la démarche expliquée dans le tableau suivant :

Tableau 13 : Détermination de la moyenne de la population

Génotypes	Fréquences	Valeurs	Fréquences x valeur
A_1A_1	p^2	+a	ap^2
A_1A_2	$2pq$	b	$2bpq$
A_2A_2	q^2	-a	$-aq^2$
Moyenne de la population			$a(p - q) + 2bpq$

$$\begin{aligned}
 \text{Moyenne de la population} &= ap^2 + 2bpq - aq^2 \\
 &= a(p^2 - q^2) + 2bpq \\
 &= a[(p+q)(p-q)] + 2bpq \\
 &= \mathbf{a(p - q) + 2bpq}
 \end{aligned}$$

$a(p - q)$ est attribuable aux homozygotes et $2bpq$ est attribuable aux hétérozygotes

Exemple

Supposant que le gène de nanisme A_2 soit présent à la fréquence de 0,1 et que les homozygotes sont dans le cas d'accouplement au hasard (panmixie) (Fig. 15).

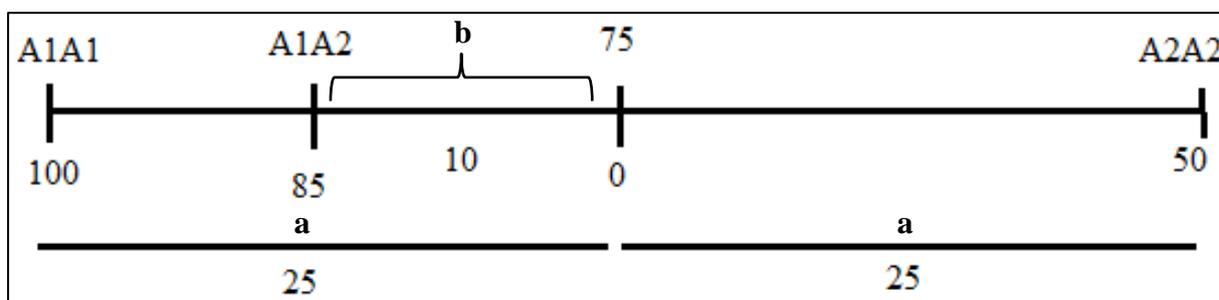


Figure 15 : Représentation graphique de la répartition des génotypes A_1A_1 , A_1A_2 et A_2A_2 autour leur centre.

Les valeurs à introduire dans l'équation sont :

- Présence du gène du nanisme $q = 0,1$
- Absence du gène de nanisme $p = 1 - 0,1 = 0,9$
- $b = 10$ cm ($((100+50)/2 = 75$ $b = 85 - 75 = 10)$)
- $a = 25$ (le nombre moyen d'unité entre 1 100 et le 50 est 25 c'est-à-dire $(100 - 50) / 2 = 25$).

La moyenne centrée = $25(0,9 - 0,1) + 2(10)(0,9)(0,1)$

$$= 21,8 \text{ (c'est une valeur centrée et non pas la vraie moyenne)}$$

Cette valeur de la moyenne est estimée à partir du point m qui est de 75 donc la valeur réelle de la population est :

Moyenne réelle = Moyenne centrée + la valeur du centre

$$= 21,8 + 75 = 96,8 \text{ cm (c'est la vraie moyenne)}$$

3.2. Evolution de l'hétérosis proprement dit

Considérant deux populations que l'on appellera populations parentales, chacune est le résultat d'accouplement au hasard mais elles ne sont pas nécessairement grandes. Ces populations parentales sont croisées pour donner une F1 au croisement de 1^{re} génération. Les individus de la F1 sont croisés entre eux au hasard pour donner la F2 au croisement de la 2^{ème} génération.

La valeur de l'hétérosis présentée par la F1 et la F2 est mesurée par l'écart avec la valeur du parent moyen c'est-à-dire par rapport à la moyenne des 2 populations.

Hétérosis F1 = $\overline{F_1} - \overline{P}$ $\overline{F_1}$: valeur moyenne de la 1^{re} génération, $\overline{F_2}$ de la 2^{ème}

Hétérosis F2 = $\overline{F_2} - \overline{P}$ \overline{P} : Parent moyen

Considérant l'effet d'un seul locus avec deux allèles dont les fréquences sont p et q dans une population et p' et q' dans l'autre.

$$y = p - p' = q' - q$$

Le calcul est simplifié en mettant les fréquences des gènes p' et q' dans la seconde population dans la forme (p - y) et (q + y). On désigne toujours par +a, b et -a les valeurs génotypiques, valeurs que l'on suppose être la même dans les deux populations (épistasie n'étant pas envisagée).

Les moyennes des parents Mp1 et Mp2 sont :

$$Mp_1 = a(p - q) + 2bpq \dots \dots \dots (1)$$

$$Mp_2 = a(p - y - q - y) + 2b(p - y)(q + y)$$

d'où :

$$Mp_2 = a(p - q - 2y) + 2b[pq + yp - yq - y^2]$$

$$Mp_2 = a(p - q - 2y) + 2b[pq + y(p - q) - y^2]$$

$$\overline{MP} = \frac{1}{2} (Mp_1 + Mp_2)$$

$$\overline{Mp} = \frac{1}{2} [a(p - q) + 2bpq + a(p - q - 2y) + 2b(pq + y(p - q) - y^2)]$$

$$\overline{Mp} = \frac{1}{2} [a(p - q) + a(p - q - 2y) + 2bpq + 2b[(pq + y(p - q) - y^2)]$$

$$\overline{Mp} = \frac{1}{2} [2a(p - q - y) + 2bpq + 2b[(pq + y(p - q) - y^2)]$$

$$\overline{Mp} = [a(p - q - y) + bpq + b[(pq + y(p - q) - y^2)]$$

$$\overline{Mp} = [a(p - q - y) + b[(pq + pq) + y(p - q) - y^2]]$$

$$\overline{Mp} = a(p - q - y) + b[(2pq + y(p - q) - y^2)] \dots \dots \dots (2)$$

Lorsque les deux populations sont croisées pour produire la F_1 , des individus pris au hasard dans une des populations sont accouplées avec des individus pris dans la seconde (tableau 14).

Tableau 14 : Génération F_1 obtenue par croisement aléatoires des individus des deux populations.

Gamètes de P_2 et fréquences	Gamètes de P_1 et fréquences			
	A1 (p)		A2 (q)	
A1 (p - y)	A1A1	p(p - y)	A1A2	q(p - y)
A2 (q + y)	A1A2	p(q + y)	A2A2	q(q + y)

Ce qui équivaut à prendre des gènes au hasard dans les deux populations, la F_1 est alors constituée de la façon suivante (tableau 15) :

Tableau 15 : fréquences alléliques des valeurs des individus de la F_1 .

Génotypes	A1A1	A1A2	A2A2
Fréquences	p(p - y)	2pq + y(p - q)	q(q + y)
Valeurs génotypiques	+a	b	-a

Les moyennes génotypiques de la F₁ sont alors comme suit :

$$\bar{m} F_1 = ap(p - y) + b[2pq + y(p - q)] + (-a)q(y + q)$$

$$= a(p^2 - py) + b[2pq + y(p - q)] - a(q^2 + qy)$$

$$= a[p^2 - py - q^2 - qy] + b[2pq + y(p - q)]$$

Et comme $p^2 - q^2 = (p + q)(p - q)$ et $p + q = 1$

$$= a[p - q - y(p + q)] + b[2pq + y(p - q)]$$

$$\mathbf{m \bar{F}_1 = a(p - q - y) + b[2pq + y(p - q)] \dots\dots\dots(3)}$$

La valeur de l'hétérosis exprimée par la différence entre la F₁ et le parent moyen s'obtient en soustrayant l'équation 2 de l'équation 3 donc HF₁ (hétérosis) :

$$\mathbf{HF_1 = \bar{m} F_1 - \bar{mp} = by^2}$$

b : Dominance

y: différence de fréquences entre les 2 populations

Ainsi l'existence de l'hétérosis de même que celle de la dépression de la consanguinité dépend de la dominance. Les loci sans dominance ($d = 0$) ne présente ni dépression de consanguinité ni hétérosis.

La valeur de l'hétérosis qui se manifeste après croisement entre 2 lignées particulières ou 2 populations dépend du carré de la différence de la fréquence des gènes (y) entre les deux populations. Si les populations que l'on croise ne diffèrent pas par leurs fréquences génétiques, il n'y aura pas d'hétérosis.

L'hétérosis sera maximum lorsqu'un allèle sera fixé dans une population et un autre allèle dans une autre population.

Si on considère les effets conjugués de tous les loci pour lesquels les deux populations de parents diffèrent et que les valeurs génotypiques dues aux loci séparés se combinent de façon additive, nous pouvons présenter l'hétérosis produit par l'effet de tous les loci réunis comme la somme de leurs effets individuels.

$$\mathbf{HF_1 = \Sigma dy^2}$$

Si certains loci sont dominant dans une direction et certains dans une autre direction, leurs effets tendront à s'accumuler et on n'observera pas d'hétérosis bien qu'il y ait dominance pour les loci pris un à un.

La présence d'hétérosis dans un croisement est donc comme la dépression de la consanguinité sous la dépendance de la dominance orientée. L'absence de l'hétérosis n'est pas suffisante pour conclure que les loci pris individuellement ne présentent pas de dominance.

Considérons maintenant la F_2 d'un croisement particulier de 2 populations de parents (tableau 16). Cette F_2 est obtenue en accouplant au hasard les individus de la F_1 . Comme conséquence de cette méthode d'accouplement au hasard, les fréquences génotypiques de la F_2 seront les fréquences de Hardy-Weinberg correspondant aux fréquences des gènes dans la F_1 . On peut alors calculer la moyenne génotypique de la F_2 par l'équation (1). La fréquence des gènes dans la F_1 étant la moyenne des fréquences des gènes dans les 2 populations des parents.

Tableau 16 : Génération F_2 obtenue par croisement aléatoires des individus de la F_2 .

Fréquences	Allèles	
	A ₁	A ₂
P ₁	p	q
P ₂	p - y	q + y
Somme	2p - y	2q + y
Moyenne	(2p - y)/2	(2q + y)/2
Moyenne	p - 1/2 y	q + 1/2 y

Si on remplace p et q de l'équation (1), on a la valeur génotypique moyenne de la F_2 :

$$MF^2 = [a(p - \frac{1}{2} y - (q + \frac{1}{2} y))] + 2b(p - \frac{1}{2} y)(q + \frac{1}{2} y)$$

Après simplification on aura :

$$MF_2 = [a(p - \frac{1}{2} y - q - \frac{1}{2} y)] + 2b(pq + \frac{1}{2} py - \frac{1}{2} qy - \frac{1}{4} y^2)$$

$$MF_2 = [a(p - q - y)] + b(2pq + py - qy - \frac{1}{2} y^2)$$

$$MF_2 = a(p - q - y) + b[2pq + y(p - q) - \frac{1}{2} y^2]$$

Alors $HF_2 = MF_2 - mp = \frac{1}{2} by^2$ donc :

$$HF_2 = \frac{1}{2} HF_1$$

La valeur de l'hétérosis présentée par la F_2 est seulement la moitié de celle présentée par la F_1

Chapitre 6 :
Valeur d'un individu au
croisement

Chapitre 6 : La valeur d'un individu au croisement

1. Généralités

La valeur d'un individu au croisement, également appelée valeur de croisement, est une mesure de la contribution génétique d'un individu à la génération suivante lorsqu'il est utilisé comme parent dans un programme d'amélioration. Cette valeur est basée sur les performances attendues de ses descendants pour des caractères d'intérêt spécifiques.

La valeur de croisement est utilisée pour évaluer et sélectionner les individus qui contribueront le plus favorablement aux objectifs de sélection. Elle repose sur l'idée que la performance d'un individu dans un environnement donné est le résultat de l'interaction entre ses gènes et son environnement.

Il est question de rechercher les facteurs qui conditionnent la valeur d'un individu en tant que géniteur. Cette analyse a une importance considérable en amélioration des plantes où tout sélectionneur sait empiriquement que tel génotype marque d'une qualité donnée tous ses descendants en croisement ; alors que tel autre, au contraire, semble ne rien apporter.

Cette appréciation est empirique et imprécise. Les notions d'héritabilité, de balance génétique et des aptitudes à la recombinaison vont permettre de préciser les différentes relations entre les géniteurs et les descendants.

2. Les balances génétiques

2.1. La balance interne et la balance de relation

L'expression balance exprime ici une adaptation de la constitution génétique à un milieu donné. Lorsque le fonctionnement des ensembles polygéniques est bon, on dit que le génotype est bien balancé. On parle au contraire de structure déséquilibrée lorsque l'ensemble est mal régulé donnant une impression de dysharmonie dans le développement. Selon LERNER, la balance d'un ensemble polygénique peut être atteinte de deux manières (Fig. 16):

- Par une balance interne jouant à l'intérieur d'une séquence génétique pouvant entraîner un arrangement allélique équilibré ;
- Par une hétérozygotie à chaque locus pouvant entraîner un bon équilibre de relation entre allèles. Il s'agit dans ce cas de balance de relation.

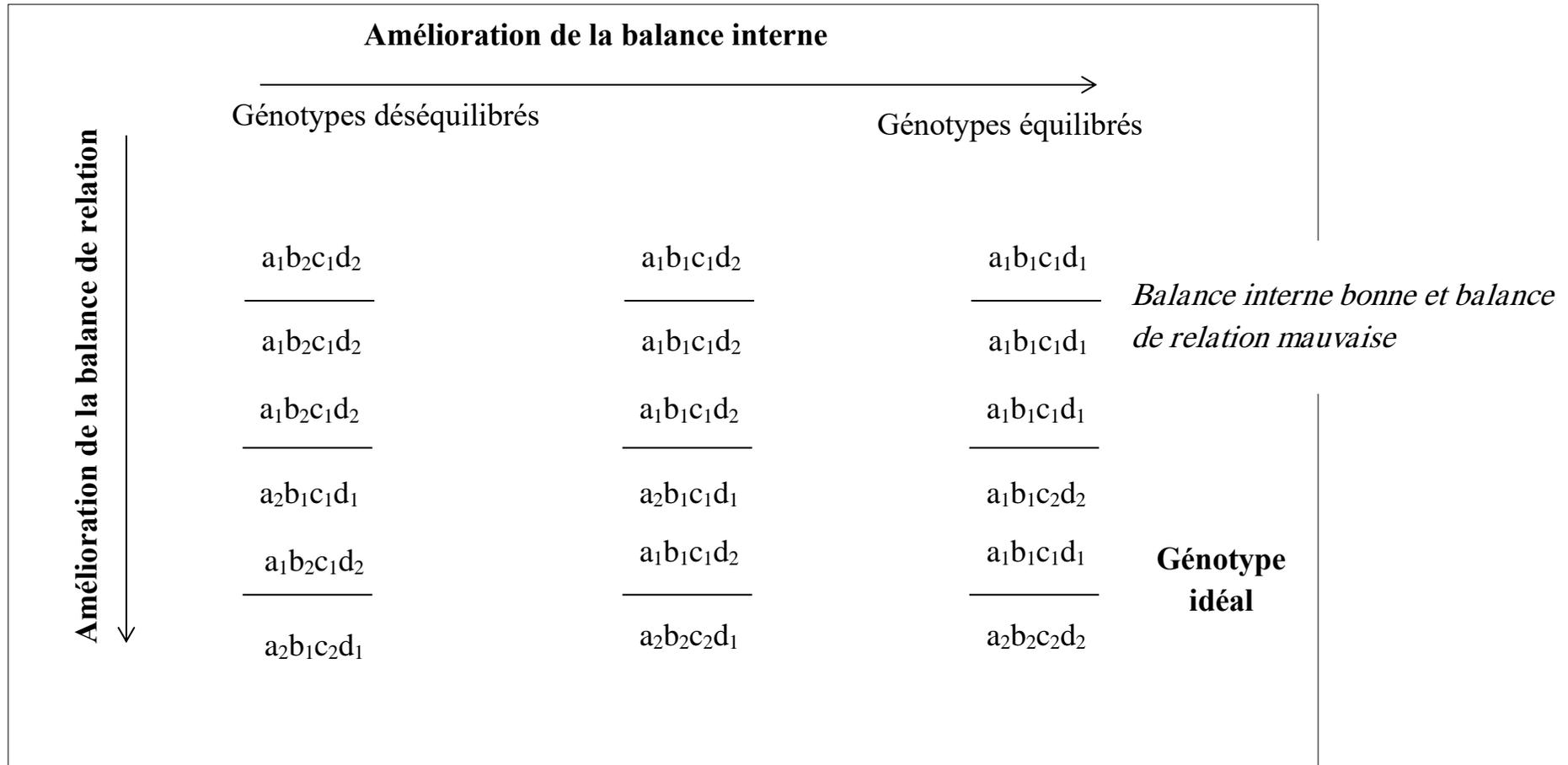


Figure 16 : Différentes qualités de balance génétique.

Les lettres a, b, c, et d représentent les gènes. Ces gènes ont une bonne balance de relation lorsque dans un locus les deux homologues ont des indices différents. La bonne balance interne est représentée par l'ensemble des lettres minuscules ayant toutes les mêmes indices.

2.2. Réalité biologique de la balance et représentation par l'ensemble des génotypes

Les états de balance interne sont construits ou détruits par le jeu des crossing-over qui modifient l'arrangement et la distribution allélique au long du chromosome.

La réalité biologique des balances internes repose fortement sur la notion de linkats, c'est-à-dire d'une part, sur la continuité chromosomique et d'une tension épistasique entre les gènes qui se suivent le long du chromosome et d'autre part, sur l'existence d'une continuité d'expression entre les ensembles dupliqués juxtaposés.

La balance interne est déterminée par la qualité de l'interaction entre linkats homologues. Elle est construite ou détruite à chaque génération.

3. Les aptitudes à la combinaison

On distingue deux types d'aptitude à la combinaison qui sont l'aptitude générale et spécifique à la combinaison.

3.1. Définitions

Par un caractère donné, on estime l'aptitude générale à la combinaison (AGC : general combining ability) d'une *structure parentale* (individu, lignée, ... etc.) à partir de la valeur moyenne des descendants lorsque ce parent est croisé avec un certain nombre de partenaires. C'est donc la moyenne des effets gamétiques d'un individu. C'est par conséquent, la mesure de la valeur du gamète moyen d'un parent.

Au niveau de l'ensemble de la descendance XY issus de croisement de deux structures parentales, on doit donc retrouver la somme des aptitudes générale des parents (XX) x (YY). En fait, la valeur des individus XY (les hybrides) présente un écart par rapport aux prévisions d'additivité des aptitudes générales.

Cet écart qui caractérise spécifiquement le croisement (XX) x (YY) est appelé *Aptitude spécifique à la combinaison* (ASC : specific recombining ability). L'aptitude spécifique à la combinaison est une caractéristique du zygote et non pas du parent haploïde.

Exemple : la hauteur de la tige (h) d'un individu issu de croisement entre 2 individus (XX)x(YY).

$$H_{XY} = AGC_{xx} + AGC_{yy} + ASC_{XY}$$

3.2. Intérêts des variances liées aux aptitudes de la combinaison

La part prise par les effets des aptitudes générales et spécifiques à la combinaison dont les différences de valeurs entre des descendance est importante à déterminer. C'est la comparaison des variances liées à l'aptitude générale et l'aptitude spécifique qui est essentielle. Elle détermine, en effet, la stratégie dans le programme d'amélioration pour une qualité (tableau 17).

Tableau 17 : stratégies de sélection en fonction des variances d'aptitude à la combinaison

Variance de l'aptitude générale à la combinaison (AGC)		Variance de l'aptitude spécifique à la combinaison (ASC)	
Faible	Peu de choix dans les formules parentales.	Forte	Faire de nombreux croisements et choisir ensuite.
Forte	Choix efficace dans les formules parentales.	Faible	Le choix des parents avant les croisements reste prioritaire.

4. Méthodes permettant d'apprécier les valeurs d'un individu en croisement

Plusieurs systèmes de croisements peuvent être utilisés pour apprécier la qualité du génotype en tant que géniteurs, selon que l'espèce considérée est autogame ou allogame, les techniques utilisées ne sont évidemment pas les mêmes.

Chez une espèce autogame, toute hybridation systématique nécessite une intervention (castration et pollinisation) souvent longue et délicate ce qui limite l'utilisation des tests décrits ci-dessous.

4.1. Les fécondations libres (descendances maternelles)

Réservée aux espèces allogames, ce système consiste à comparer les descendants de chaque génotype lorsqu'il est librement pollinisé. La comparaison des structures (homozygotes ou hétérozygote) ce fait comme suit (Fig. 17).

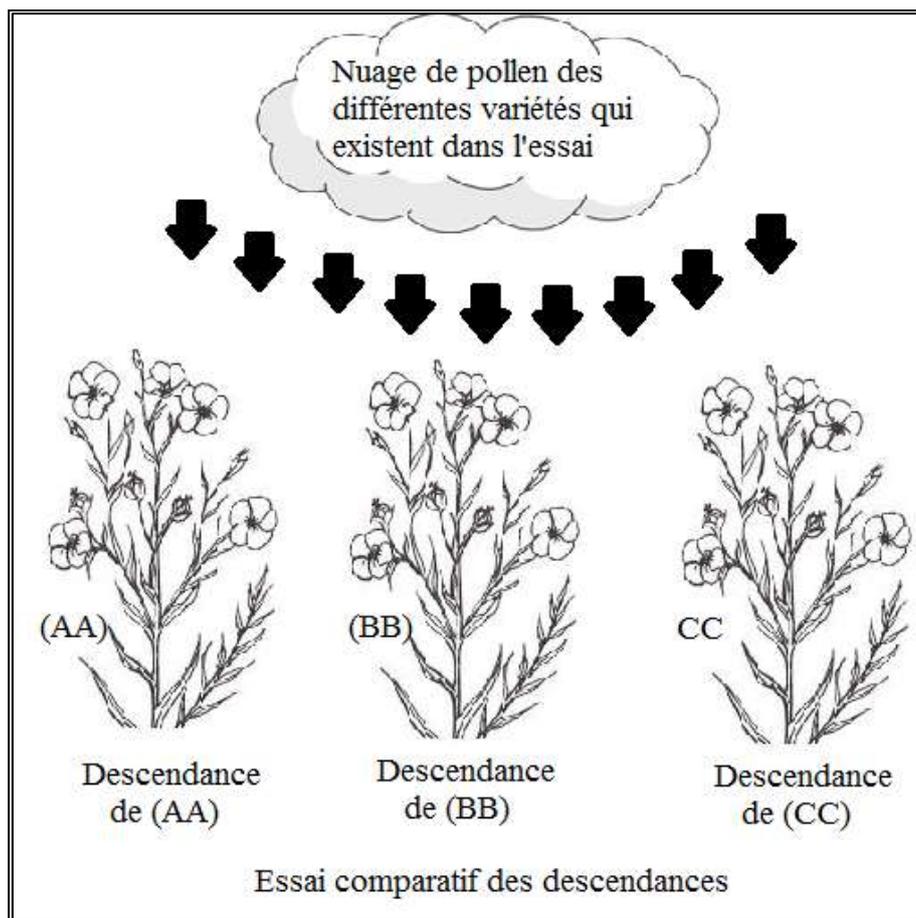


Figure 17 : Tests de descendance maternelles

Les limites de cette méthode, dont le grand avantage est la simplicité, sont :

- L'information qu'elle apporte ne concerne que les aptitudes générales à la combinaison donc ne touche essentiellement que les effets additifs ;
- La population pollinique qui féconde un génotype est rarement la même que celle qui féconde un autre génotype, car le pollen d'une plante se distribue autour de cette plante selon un gradient très marqué que la plante soit anémophile ou entomophile. Les insectes pollinisateurs choisissent certains phénotypes qu'ils pollinisent préférentiellement. Les ajustements de précocité entre maturité de pollen et récessivité des stigmates sont très étroits chez certaines espèces.

4.2. Le test top-cross

Les structures génétiques à étudier (famille, lignées ou clones) sont pollinisées par un testeur commun qui peut être une variété, une population ou un hybride. Dans ce cas, les structures à tester sont en général représentées par plusieurs individus ce qui permet une

meilleure exploitation dans l'espace et dans le temps de la population gamétique mâle. L'hybridation peut être effectuée de manière contrôlée. Le testeur peut être une structure bien définie et équilibrée qui est donc très stable.

Lorsque le testeur est une structure génétique à base large, l'estimation obtenue est une aptitude générale à la combinaison. Si, par contre, le testeur est de structure étroite, les résultats du top-cross ne reflètent qu'une aptitude spécifique vis-à-vis des testeurs.

4.3. Le test poly-cross

Dans le poly-cross, c'est la population pollinique des structures en études qui constitue le testeur (Fig. 18). Le poly-cross est la fécondation libre entre-elles de (n) structures (clones, lignées inbred, familles, ... etc.), dont on veut comparer les aptitudes générales à la combinaison.

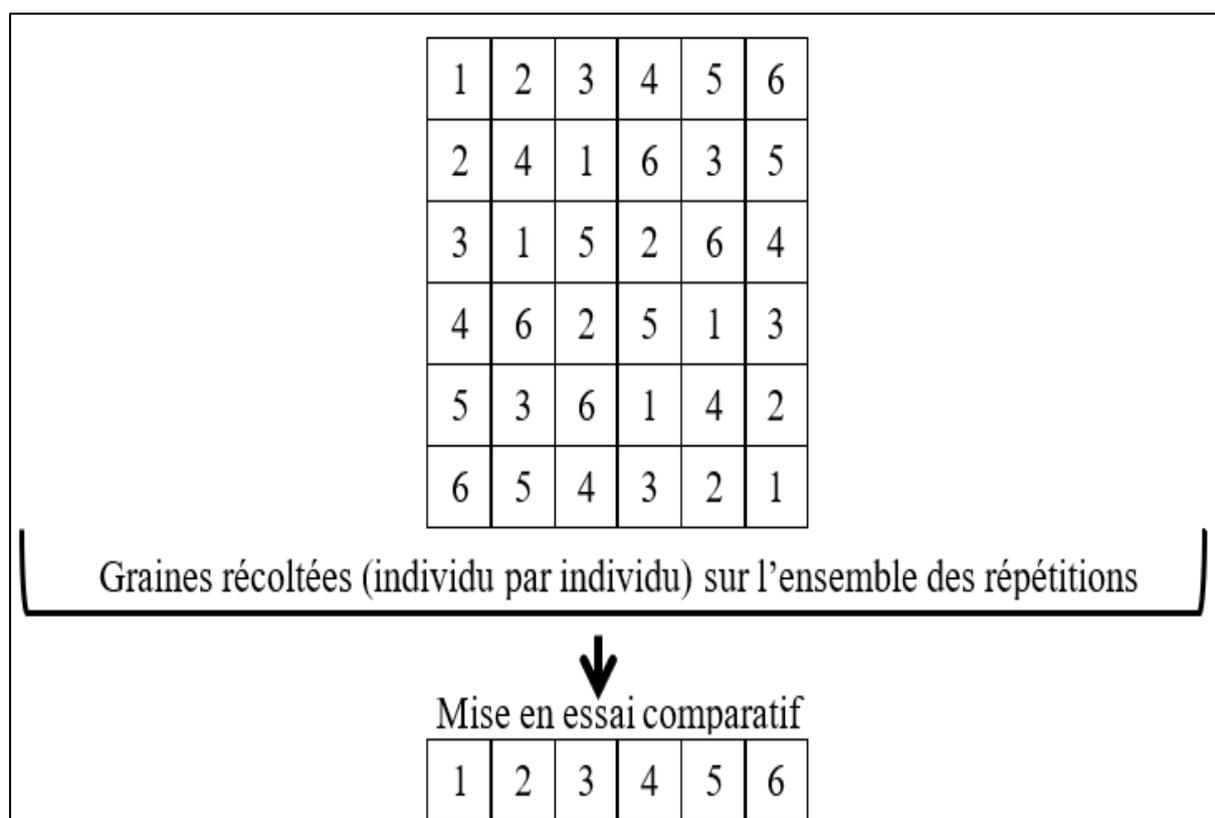


Figure 18 : Plan systématique et poly-cross pour 6 individus.

Dans ce croisement, chaque parent femelle est croisé artificiellement (pollinisation croisée¹) avec un mélange de pollen provenant d'un certain nombre suffisant de parents mâles. Pour faire une analyse de descendance, chaque arbre femelle doit être pollinisé par le même mélange de pollen et en même quantité.

4.4. Croisement hiérarchique

Un certain nombre de parents P_1, P_2, \dots, P_j sont choisis au premier niveau et on les croise chacun à des groupes différents p_1, p_2, \dots, p_j qui peuvent avoir ou non des éléments communs et qui constituent un deuxième niveau (Fig. 19).

Ce dispositif permet de tester un grand nombre de parents avec des groupes de partenaires relativement faibles. Il y a de nombreux avantages pratiques (facilité d'association par groupes de précocité, plus faible nombre d'hybridation par plante). Il donne une bonne information sur la variance de l'aptitude générale à la combinaison.

4.5. Croisement diallèle

Dans tous ces tests, seule l'aptitude générale à la combinaison (AGC) était accessible. Les croisements diallèles par contre, permettent une interprétation plus poussée en termes d'aptitude à la combinaison.

Le croisement diallèles s'applique, à la fois, aux espèces autogames et allogames. C'est un ensemble d'hybridations dirigées entre structures à étudier comprenant systématiquement toute une série de combinaisons (les graines issues de chaque parent mâle étant individualisées sur chaque parent femelle).

On appelle croisement diallèle, des systèmes de croisement systématiques combinant (n) géniteurs. L'ensemble des opérations peut être écrit sous forme d'un tableau ou matrice diallèle (Fig. 20).

¹ **La pollinisation croisée** : est le fait de planter, pas trop loin l'une de l'autre, deux variétés de fruits dont les pollens sont compatibles, car certaines variétés ont besoin d'être fécondées.

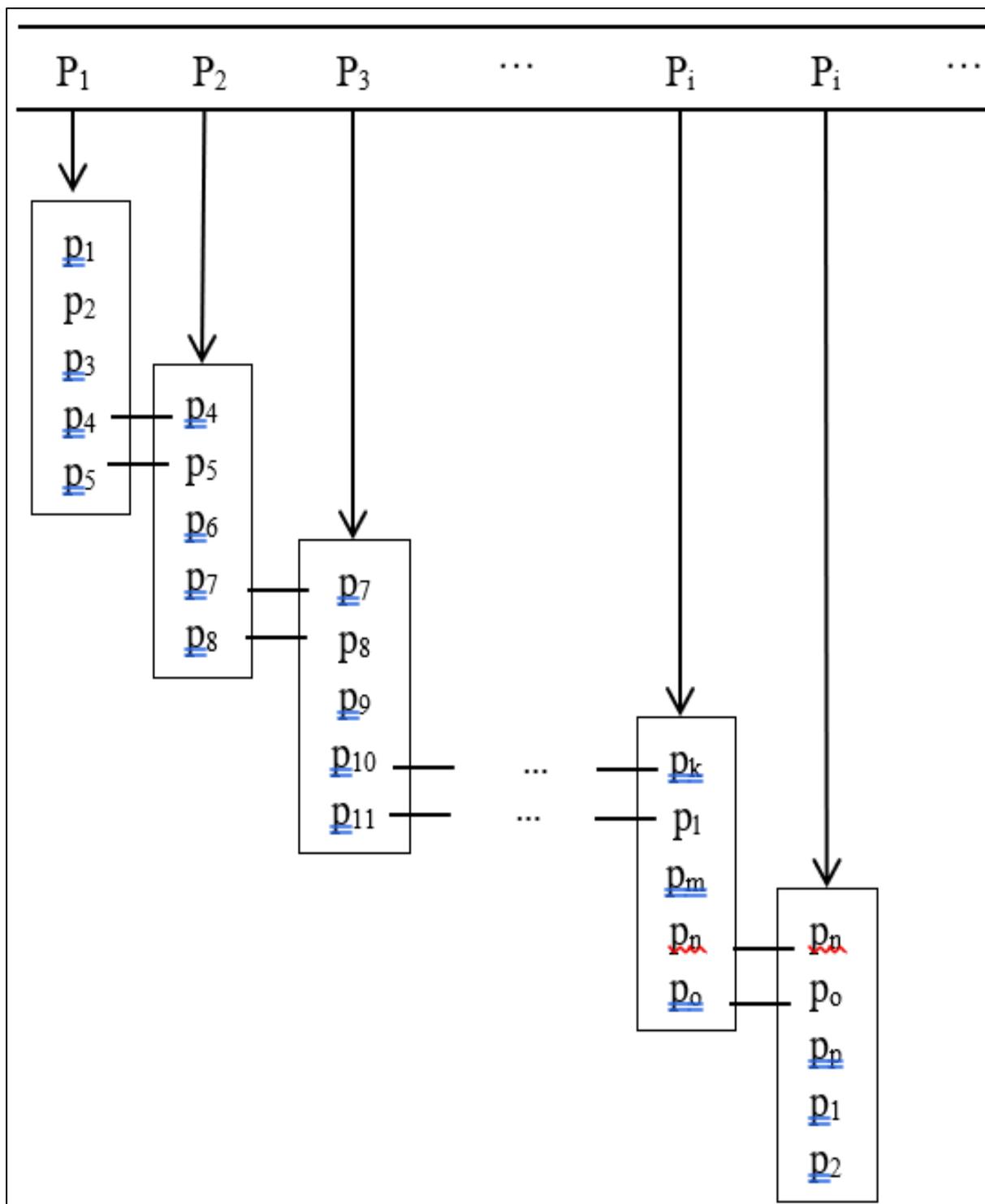


Figure 19 : Croisement hiérarchique

♂

	1	2	3	4	5	6	7	8	9	10
1	X	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X	X	X
4	X	X	X	X	X	X	X	X	X	X
5	X	X	X	X	X	X	X	X	X	X
6	X	X	X	X	X	X	X	X	X	X
7	X	X	X	X	X	X	X	X	X	X
8	X	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X	X	X
10	X	X	X	X	X	X	X	X	X	X

♀

Figure 20 : Croisement diallèle complet. Tous les géniteurs sont croisés à chaque autre géniteur dans toutes les combinaisons possibles, auto-croisement inclus. Le diallèle complet de 10 parents exige donc : $10 \times 10 = 100$ croisements

- Le diallèle peut être complet et comprendre le carré des combinaisons deux à deux. Le nombre de croisements = $(n)^2$; avec (n) = nombre des parents ou structures à étudier ;
- Le diallèle peut couvrir $n(n - 1)$ hybrides. Ce sont les croisements inter-géniteurs (les autofécondations sont exclues) ;
- Le diallèle peut comprendre $\frac{p(p+1)}{2}$ types. Un seul sens de croisement incluant les autofécondations, il s'appelle aussi le demi-diallèle (Fig. 21) ;
- Le diallèle peut comprendre $\frac{p(p-1)}{2}$ types. Un seul sens de croisement sans les autofécondations (Fig. 21).

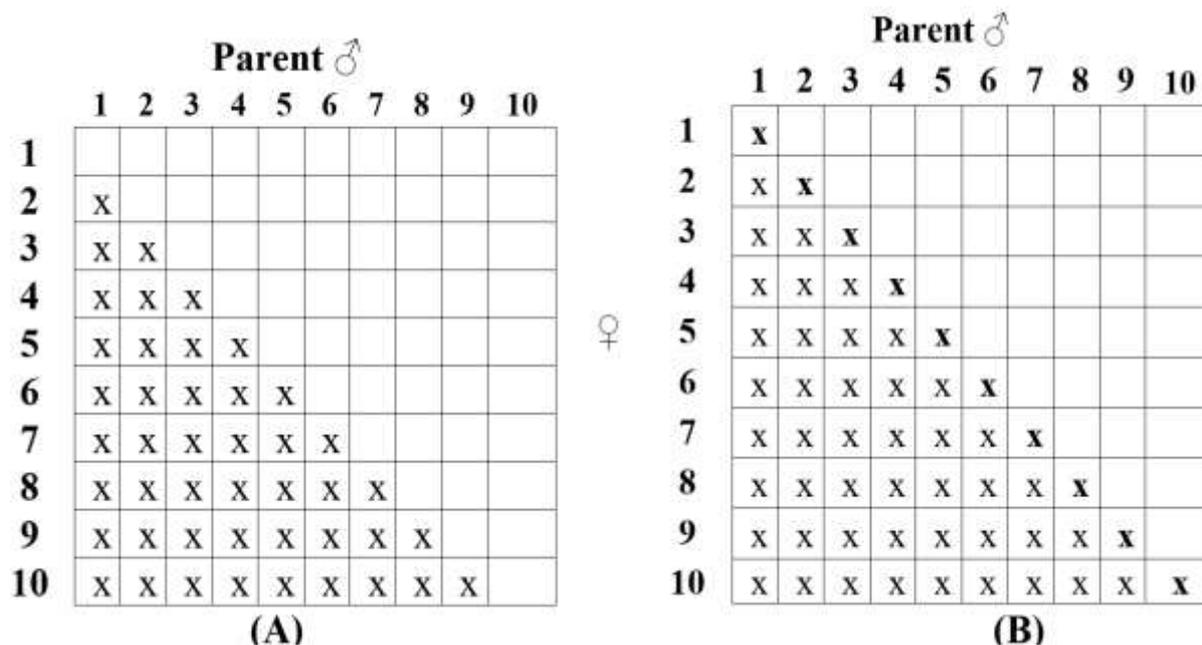


Figure 21 : Croisement demi-diallèle : (A) croisement demi-diallèle non incluant les autofécondations ; (B) croisement diallèle incluant les autofécondations.

Chapitre 7 :

QTL

Chapitre 7 : QTL (Quantitative trait loci)

1. Introduction aux QTL

1.1. Définition et concepts clés

Les QTL (Quantitative Trait Loci) sont des Loci de caractères quantitatifs. Ils sont responsables du contrôle des caractères à effet quantitatif, c'est-à-dire ayant une variation continue de leur valeur (rendement, taille...). Les QTL sont des régions du génome qui sont associées à la variation quantitative de traits complexes, tels que la taille, le poids, la production de lait, la résistance aux maladies, la teneur en nutriments, la qualité de la viande, etc. Les QTL sont souvent recherchés en utilisant des méthodes d'analyse de liaison ou d'association, qui comparent les génotypes et les phénotypes de populations de plantes ou d'animaux.

Les QTL sont importants pour comprendre la génétique de caractères complexes, car ils peuvent aider à identifier les gènes qui contribuent à la variation quantitative des traits. En effet, les QTL sont des marqueurs génétiques qui permettent de localiser des régions du génome qui sont liées à un caractère donné. Les QTL peuvent ainsi aider à identifier les gènes qui contrôlent un caractère, ainsi que les interactions entre les gènes et l'environnement.

Les QTL peuvent être détectés à l'aide de marqueurs moléculaires tels que les SNPs (Single Nucleotide Polymorphisms) et les microsatellites. Les marqueurs moléculaires sont utilisés pour cartographier les QTL en comparant les génotypes et les phénotypes de populations de plantes ou d'animaux.

Les QTL sont utilisés dans de nombreuses applications en recherche agricole, médicale et écologique. Par exemple, les QTL peuvent être utilisés pour la sélection assistée par marqueurs. Les QTL peuvent également être utilisés pour étudier la génétique de maladies humaines et animales, ainsi que pour étudier l'adaptation des espèces à des environnements changeants.

1.2. Applications des QTL en recherche agricole, médicale et écologique

Les QTL (Quantitative Trait Loci) ont de nombreuses applications en recherche agricole, médicale et écologique. Voici quelques exemples d'applications des QTL :

1.2.1. Recherche agricole

a) Sélection assistée par marqueurs

Les QTL peuvent être utilisés pour la sélection assistée par marqueurs, qui permet de sélectionner des variétés de plantes ou d'animaux ayant des caractéristiques améliorées, telles

que la résistance aux maladies, la qualité des produits et la productivité. La sélection assistée par marqueurs permet de réduire le temps et les coûts associés à la sélection conventionnelle basée sur des caractéristiques phénotypiques.

b) Cartographie de QTL pour la résistance aux maladies et aux ravageurs

Les QTL peuvent être utilisés pour cartographier les gènes impliqués dans la résistance aux maladies et aux ravageurs, ce qui peut aider à développer des variétés résistantes et réduire l'utilisation de pesticides.

c) Amélioration de la qualité nutritionnelle des aliments

Les QTL peuvent être utilisés pour cartographier les gènes qui contrôlent la teneur en nutriments dans les plantes, ce qui peut aider à développer des variétés ayant une teneur accrue en nutriments tels que les vitamines et les minéraux.

1.2.2. Recherche médicale

a) Cartographie de QTL pour les maladies humaines

Les QTL peuvent être utilisés pour cartographier les régions du génome associées à des maladies complexes telles que le diabète, la maladie d'Alzheimer et le cancer. Cela peut aider à comprendre la génétique de ces maladies et à développer de nouveaux traitements.

b) Identification des QTL impliqués dans la résistance aux maladies chez les animaux sauvages et domestiques

Les QTL peuvent être utilisés pour identifier les gènes qui contribuent à la résistance aux maladies chez les animaux, ce qui peut aider à développer des variétés plus résistantes et à réduire l'utilisation d'antibiotiques.

c) Identification des QTL impliqués dans l'adaptation aux changements environnementaux

Les QTL peuvent être utilisés pour comprendre comment les espèces s'adaptent aux changements environnementaux tels que le changement climatique, ce qui peut aider à prédire comment les espèces réagiront à l'avenir.

1.2.3. Recherche écologique

a) Étude de l'adaptation des espèces aux environnements changeants

Les QTL peuvent être utilisés pour comprendre comment les espèces s'adaptent à des environnements changeants, ce qui peut aider à prédire comment les espèces réagiront à l'avenir.

b) Étude de la biodiversité

Les QTL peuvent être utilisés pour comprendre la diversité génétique dans les populations sauvages, ce qui peut aider à protéger la biodiversité et à préserver les espèces menacées d'extinction.

c) Étude de la génétique des interactions entre les espèces

Les QTL peuvent être utilisés pour comprendre comment les interactions entre les espèces sont régies par la génétique, ce qui peut aider à comprendre les écosystèmes et les processus écologiques.

1.3. Limites et défis de l'utilisation des QTL

Bien que les QTL (Quantitative Trait Loci) soient une méthode puissante pour étudier les caractéristiques complexes, leur utilisation présente certaines limites et défis :

- **Coût élevé** : L'identification et la cartographie des QTL nécessitent des ressources importantes, notamment en termes de coûts financiers, de temps et de personnel qualifié.
- **Complexité des données** : Les données génétiques générées lors de l'étude des QTL sont souvent très complexes et difficiles à interpréter, nécessitant une expertise en statistiques et en bio-informatique.
- **Variabilité environnementale** : Les caractéristiques phénotypiques peuvent être influencées par des facteurs environnementaux tels que la température, l'humidité et les nutriments du sol, ce qui peut rendre la cartographie des QTL plus difficile.
- **Effets de liaison et interactions génétiques complexes** : Les QTL sont souvent liés à d'autres QTL et à d'autres gènes, ce qui peut rendre difficile l'identification de l'effet spécifique de chaque QTL sur la caractéristique étudiée. De plus, les interactions entre les gènes peuvent être complexes et difficiles à modéliser.
- **Éthique** : L'utilisation des QTL dans la sélection assistée par marqueurs soulève des questions éthiques sur les implications sociales et environnementales de la manipulation génétique des plantes et des animaux.

- **Propriété intellectuelle** : L'utilisation des QTL dans la sélection assistée par marqueurs soulève également des questions de propriété intellectuelle, car les entreprises peuvent revendiquer la propriété des marqueurs génétiques utilisés pour la sélection.

En somme, l'utilisation des QTL présente des avantages significatifs pour la recherche agricole, médicale et écologique, mais elle nécessite également une attention particulière à certaines limites et défis.

2. Techniques pour la détection des QTL

2.1. Analyse de liaison et cartographie génétique

La détection des QTL (Quantitative Trait Loci) peut être réalisée en utilisant deux techniques principales : l'analyse de liaison et la cartographie génétique.

2.1.1. Analyse de liaison

L'analyse de liaison est une technique qui cherche à établir une relation entre un trait quantitatif et une région spécifique du génome. Elle implique généralement la comparaison des marqueurs génétiques avec les caractéristiques phénotypiques dans des populations de lignées consanguines ou de famille d'animaux.

La technique consiste à établir une corrélation entre la présence de marqueurs génétiques et la présence d'un trait spécifique. Si un marqueur est présent chez des individus qui présentent le trait, alors le marqueur est considéré comme étant lié au trait. En d'autres termes, la fréquence d'un marqueur spécifique est comparée à la fréquence de la caractéristique phénotypique étudiée, ce qui permet de déterminer si ces deux caractéristiques sont corrélées.

2.1.2. Cartographie génétique

La cartographie génétique est une technique qui utilise des marqueurs génétiques pour cartographier la position des QTL sur le génome. Cette technique utilise des marqueurs génétiques pour cartographier les positions des gènes sur les chromosomes. Ces marqueurs peuvent être des séquences d'ADN simples ou des variants d'ADN tels que des SNPs (Single Nucleotide Polymorphisms).

Les marqueurs génétiques sont cartographiés en utilisant des cartes génétiques. Ces cartes sont créées en examinant la fréquence des recombinaisons qui se produisent pendant la méiose, lors de la formation des gamètes. En utilisant ces cartes, les QTL peuvent être cartographiés en relation à des marqueurs génétiques connus.

La combinaison de l'analyse de liaison et de la cartographie génétique permet de localiser précisément les QTL et d'identifier les marqueurs génétiques qui sont étroitement liés à ces QTL. Ces marqueurs peuvent ensuite être utilisés pour la sélection assistée par marqueurs, qui permet de sélectionner des individus présentant les caractéristiques souhaitées en utilisant des marqueurs génétiques plutôt que des caractéristiques phénotypiques.

Exemple : La localisation de QTL impliqués dans la variation de la dureté du grain de blé (Fig. 22).

Un des facteurs affectant la qualité boulangère des blés cultivés est la dureté des grains. L'établissement de cartes génétiques sur le blé a permis d'identifier plusieurs QTL impliqués dans la variation de la dureté du grain. Un des QTL majeurs est situé sur le chromosome 5. L'infographie associée représente une carte génétique d'une partie du chromosome 5. En abscisse sont données les distances entre les marqueurs RFLP le long de ce bras de chromosome. A la taille des barres est associé un test de présence du QTL : la barre la plus grande correspond à la position statistique la plus probable du QTL. Ainsi, un QTL de la dureté du grain a été trouvé à proximité du marqueur Xmta9. C'est un marqueur du QTL.

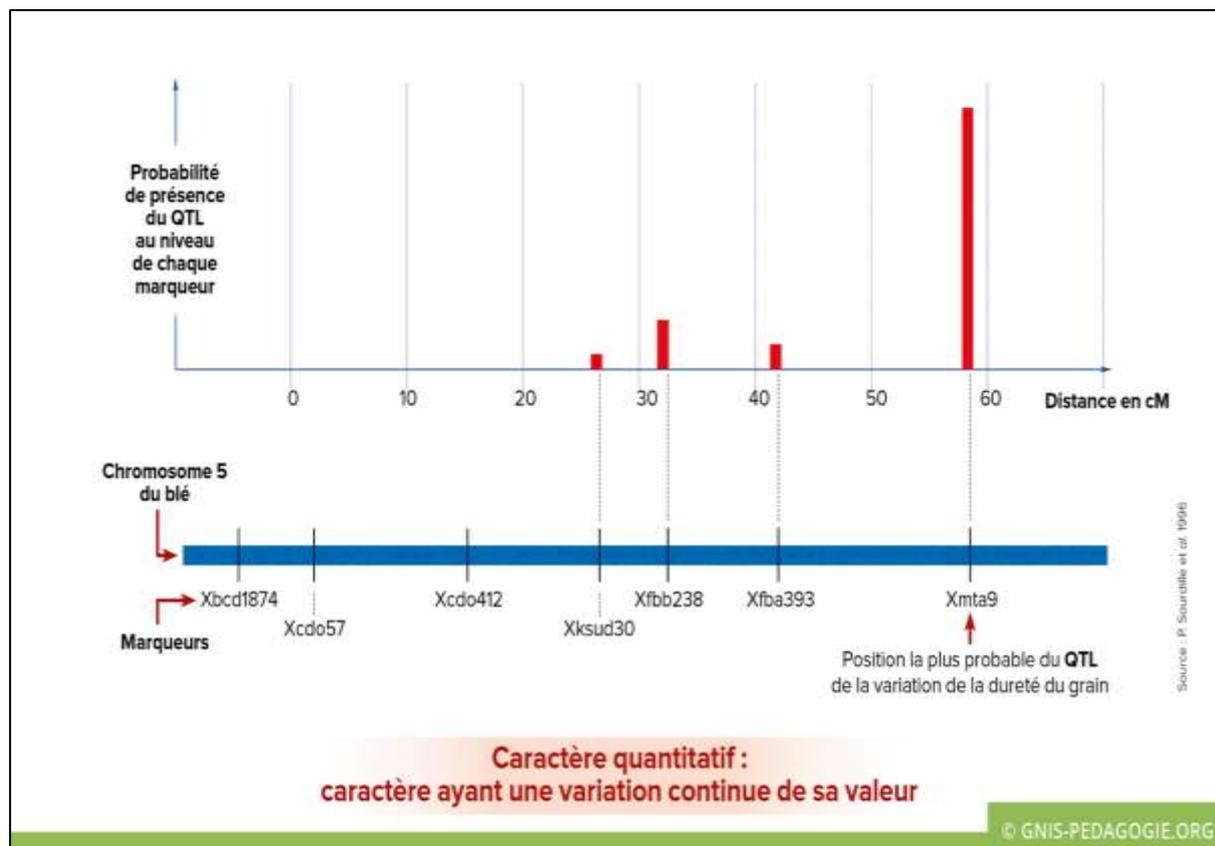


Figure 22 : La localisation de QTL impliqués dans la variation de la dureté du grain de blé

2.2. Analyse d'association et étude d'association pangénomique

Outre l'analyse de liaison et la cartographie génétique, deux autres techniques pour la détection des QTL (Quantitative Trait Loci) sont largement utilisées : l'analyse d'association et l'étude d'association pangénomique.

2.2.1. Analyse d'association

L'analyse d'association est une technique qui utilise des marqueurs génétiques pour étudier les relations entre des traits quantitatifs et des variations génétiques. Elle diffère de l'analyse de liaison en ce sens qu'elle utilise une population de cas et de témoins non apparentés pour étudier les relations entre les marqueurs génétiques et les traits quantitatifs. Les marqueurs génétiques sont comparés à la présence ou à l'absence de la caractéristique étudiée pour déterminer s'il y a une association significative entre eux.

L'analyse d'association est utile pour l'identification des QTL car elle permet l'identification de variations génétiques associées à des traits quantitatifs dans des populations de taille importante. Cependant, cette technique est limitée par la nécessité d'utiliser des marqueurs génétiques qui sont en forte LD (déséquilibre de liaison) avec les QTL, ce qui peut rendre difficile l'identification des variations génétiques réelles qui sont responsables des QTL.

2.2.2. Étude d'association pangénomique

L'étude d'association pangénomique (GWAS) est une technique qui utilise des marqueurs génétiques couvrant l'ensemble du génome pour identifier les variations génétiques associées à des traits quantitatifs. Cette technique est similaire à l'analyse d'association, mais utilise un grand nombre de marqueurs génétiques pour maximiser la couverture du génome.

Les GWAS ont permis l'identification de QTL associés à de nombreux traits quantitatifs dans diverses espèces, notamment chez les humains, les plantes et les animaux. Cette technique permet également d'identifier les gènes et les voies moléculaires impliqués dans les traits quantitatifs, ce qui peut fournir des informations importantes sur les mécanismes moléculaires sous-jacents.

Cependant, l'étude d'association pangénomique est également limitée par le coût élevé des analyses de génotypage et la nécessité d'utiliser des populations de grande taille pour identifier des associations significatives avec une puissance statistique suffisante. De plus, cette technique ne permet pas de cartographier les QTL de manière précise et de déterminer les effets d'interaction entre les différents QTL.

2.3. Utilisation de marqueurs moléculaires pour la cartographie de QTL

L'utilisation de marqueurs moléculaires pour la cartographie de QTL (Quantitative Trait Loci) est une technique largement utilisée pour la détection de QTL dans les espèces animales et végétales. Cette technique est basée sur la construction d'une carte génétique qui relie les QTL aux marqueurs moléculaires.

La construction d'une carte génétique se fait en deux étapes principales : le génotypage et la cartographie. Le génotypage consiste à déterminer les génotypes pour les marqueurs moléculaires dans une population de lignées recombinantes. La cartographie consiste ensuite à placer les marqueurs moléculaires sur une carte génétique et à les lier aux QTL.

Il existe deux types de cartes génétiques : les cartes génétiques basées sur des marqueurs (ou cartes de liaison) et les cartes physiques (ou cartes de position). Les cartes de liaison sont construites à partir de marqueurs moléculaires qui sont étroitement liés aux QTL, tandis que les cartes physiques sont construites à partir de marqueurs moléculaires qui sont associés à des positions physiques connues dans le génome (Fig. 23).

Une fois la carte génétique établie, les QTL peuvent être cartographiés en utilisant des méthodes statistiques telles que l'analyse de régression et l'analyse de variance. Ces méthodes permettent de déterminer la contribution de chaque QTL à la variation du trait quantitatif étudié.

L'utilisation de marqueurs moléculaires pour la cartographie de QTL présente plusieurs avantages par rapport aux méthodes classiques basées sur la sélection phénotypique. Tout d'abord, cette technique permet de localiser les QTL de manière précise sur la carte génétique, ce qui facilite leur caractérisation moléculaire et leur clonage. De plus, elle permet de cartographier des QTL qui sont difficiles à détecter par des méthodes classiques de sélection phénotypique, tels que les QTL à effet mineur.

Cependant, l'utilisation de marqueurs moléculaires pour la cartographie de QTL présente également des limites, notamment la nécessité de construire une carte génétique précise et la complexité des interactions entre les QTL, qui peuvent être difficiles à détecter. De plus, cette technique ne permet pas d'identifier les mécanismes moléculaires sous-jacents aux QTL, ce qui nécessite des études supplémentaires.

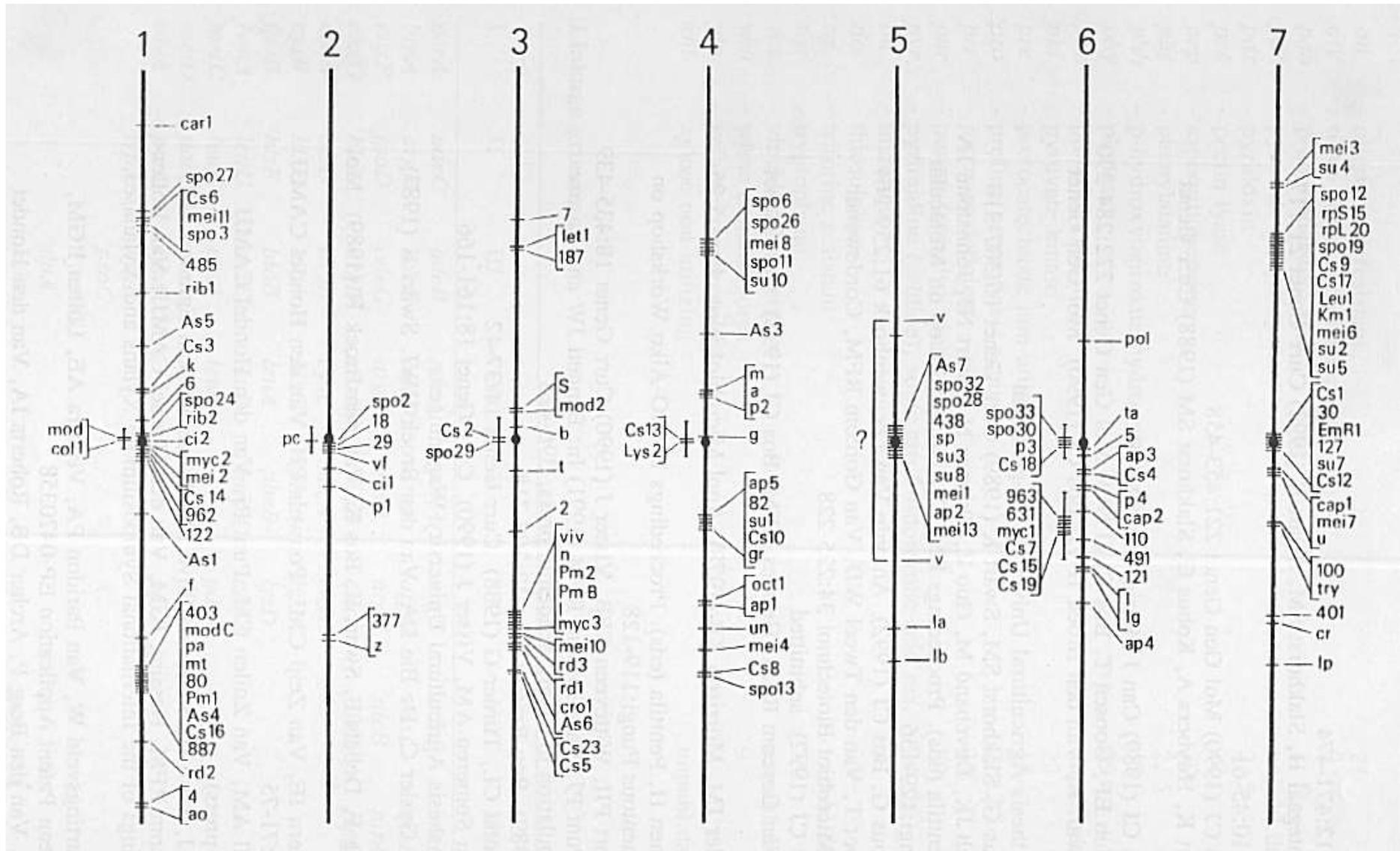


Figure 23 : Exemple d'une carte génétique : Carte génétique de *Podospora anserina*.

2.4. Techniques de géotypage haut débit pour la cartographie de QTL

Les techniques de géotypage haut débit sont des outils de plus en plus utilisés pour la cartographie de QTL (Quantitative Trait Loci). Ces techniques permettent de géotyper simultanément des milliers de marqueurs moléculaires, ce qui permet de construire des cartes génétiques avec une résolution élevée et de détecter des QTL avec une précision accrue.

Les deux principales techniques de géotypage haut débit utilisées pour la cartographie de QTL sont le séquençage de nouvelle génération (NGS) et les puces à ADN (ou microarray).

2.4.1. Séquençage de nouvelle génération (NGS)

Il permet de séquencer rapidement de grandes quantités d'ADN à un coût abordable. Cette technique peut être utilisée pour le géotypage de marqueurs moléculaires tels que les SNP (Single Nucleotide Polymorphisms) et les SSR (Simple Sequence Repeats). Les données de séquençage peuvent ensuite être analysées pour détecter des QTL en utilisant des méthodes statistiques telles que l'analyse de régression.

La commercialisation depuis 2005 des technologies de NGS a révolutionné au cours de ces dernières années la dimension des analyses génétiques par un changement majeur d'échelle des capacités de séquençage.

Ayant trouvé rapidement de nombreuses applications dans le domaine de la recherche, en particulier pour l'identification de nouveaux gènes impliqués dans des maladies monogéniques, le NGS a progressivement été validé pour des applications en diagnostic génétique.

Le NGS repose sur la génération massive de données de séquences obtenues par des cycles successifs d'incorporation de nucléotides, et ainsi l'émission de signaux qui sont ensuite convertis en information de séquence. Différentes technologies existent actuellement, notamment basées sur un séquençage en parallèle de millions de molécules d'ADN, avec une augmentation toujours croissante des capacités de séquençage associée à une diminution progressive des coûts, et de nouvelles approches sont en développement (en particulier le séquençage direct de molécules d'ADN uniques). De manière schématique, le processus de NGS est constitué de multiples étapes de génération et d'analyse des données, avec la prise en compte de critères de qualité de séquençage (en particulier l'analyse de la « couverture » et de la « profondeur de lecture » de la séquence d'intérêt), qui sont présentées de manière synthétique dans la Figure 24 en association avec des termes d'usage courant associés.

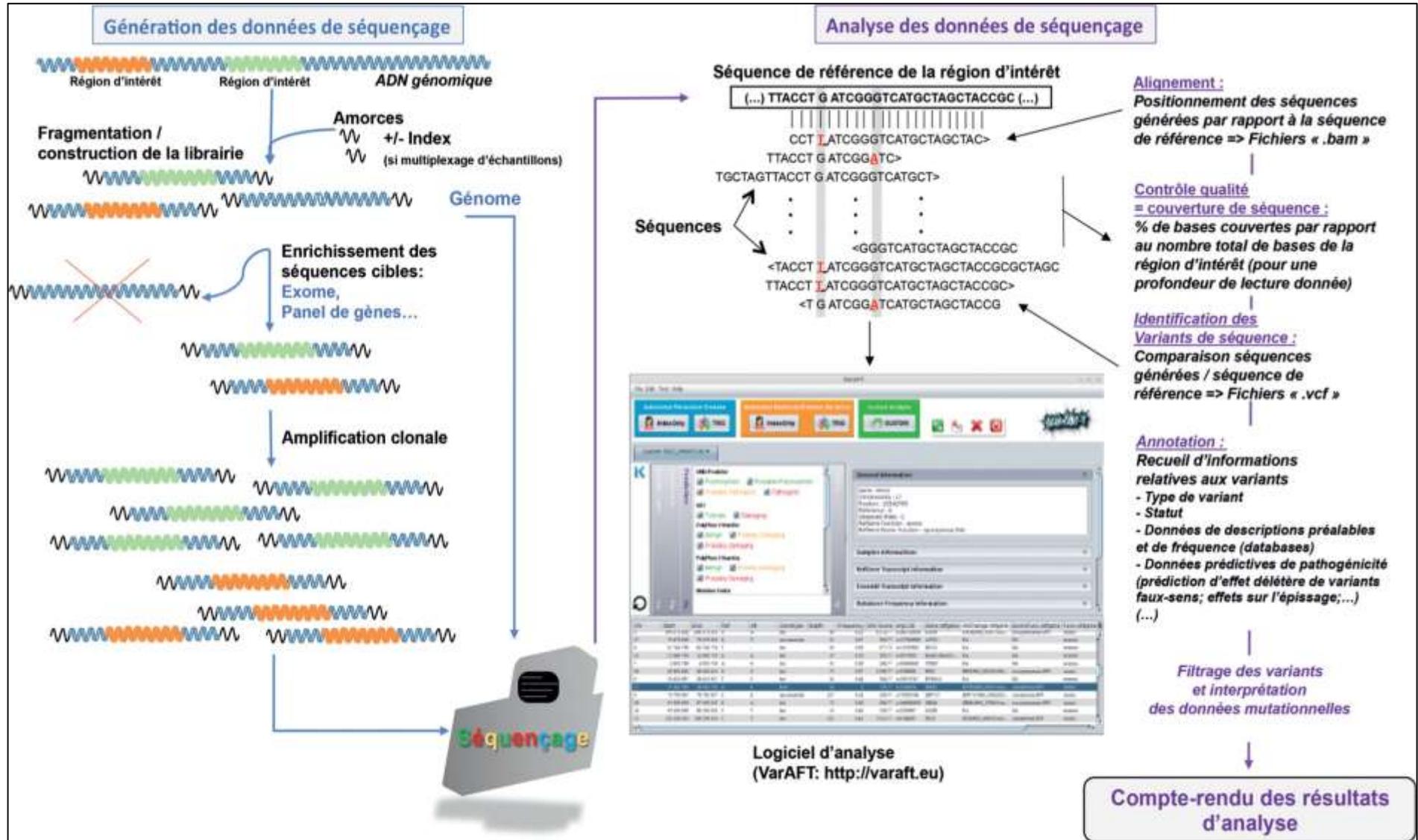


Figure 24 : Principe de la méthode de séquençage de nouvelle génération (NGS)

2.4.2. Les puces à ADN

Les puces à ADN permettent de génotyper simultanément des milliers de marqueurs moléculaires, tels que les SNP et les SSR, à un coût abordable. Ces puces sont disponibles pour de nombreuses espèces animales et végétales, ce qui les rend largement utilisées pour la cartographie de QTL. Les données de puces à ADN peuvent être analysées pour détecter des QTL en utilisant des méthodes statistiques telles que l'analyse d'association pangénomique.

Le principe d'une puce à ADN réside dans la reconnaissance, c'est-à-dire l'hybridation, entre deux molécules d'ADN simple brin complémentaires (Fig. 25). L'échantillon (ADN ou ARN), marqué de manière fluorescente, est mis en contact avec la puce portant plusieurs milliers de sondes qui sont des fragments d'ADN ou des oligonucléotides de séquence connue. Après lavage du matériel fixé de manière non spécifique, le signal est quantifié au niveau de chaque sonde. Sa valeur dépendra de la concentration en molécules marquées complémentaires de la sonde dans l'échantillon et du degré de complémentarité (le pourcentage d'identité) avec la sonde.

Actuellement, l'utilisation la plus courante des puces à ADN concerne la quantification des ARN messagers d'un échantillon (transcriptome) afin de comparer les profils de transcription de deux échantillons obtenus dans des conditions de croissance différentes.

Dans le cadre de l'identification bactérienne et du typage, utilisation des puces d'ADN est différente : ces puces permettent la détection d'une séquence d'ADN dans un mélange, d'identifier un polymorphisme de séquence (SNP) et de re-séquencer un fragment d'ADN (Fig. 26). Les puces d'ADN servent de :

a) Outils de détection

Des puces à ADN sont utilisées pour détecter un produit de PCR et remplacent ainsi la migration sur gel en validant la spécificité de l'amplification par sa complémentarité avec la sonde. Pour l'analyse d'un échantillon simple, cette procédure est peu compétitive par rapport à la PCR en temps réel ou à l'électrophorèse capillaire. En revanche, dans le cas d'un échantillon complexe, les puces à ADN permettent l'analyse simultanée d'un grand nombre de fragments. Des puces permettant d'identifier les bactéries présentes dans un échantillon sur la base de l'analyse des séquences d'ARN ribosomique 16S sont en développement dans plusieurs institutions (Fig. 26A).

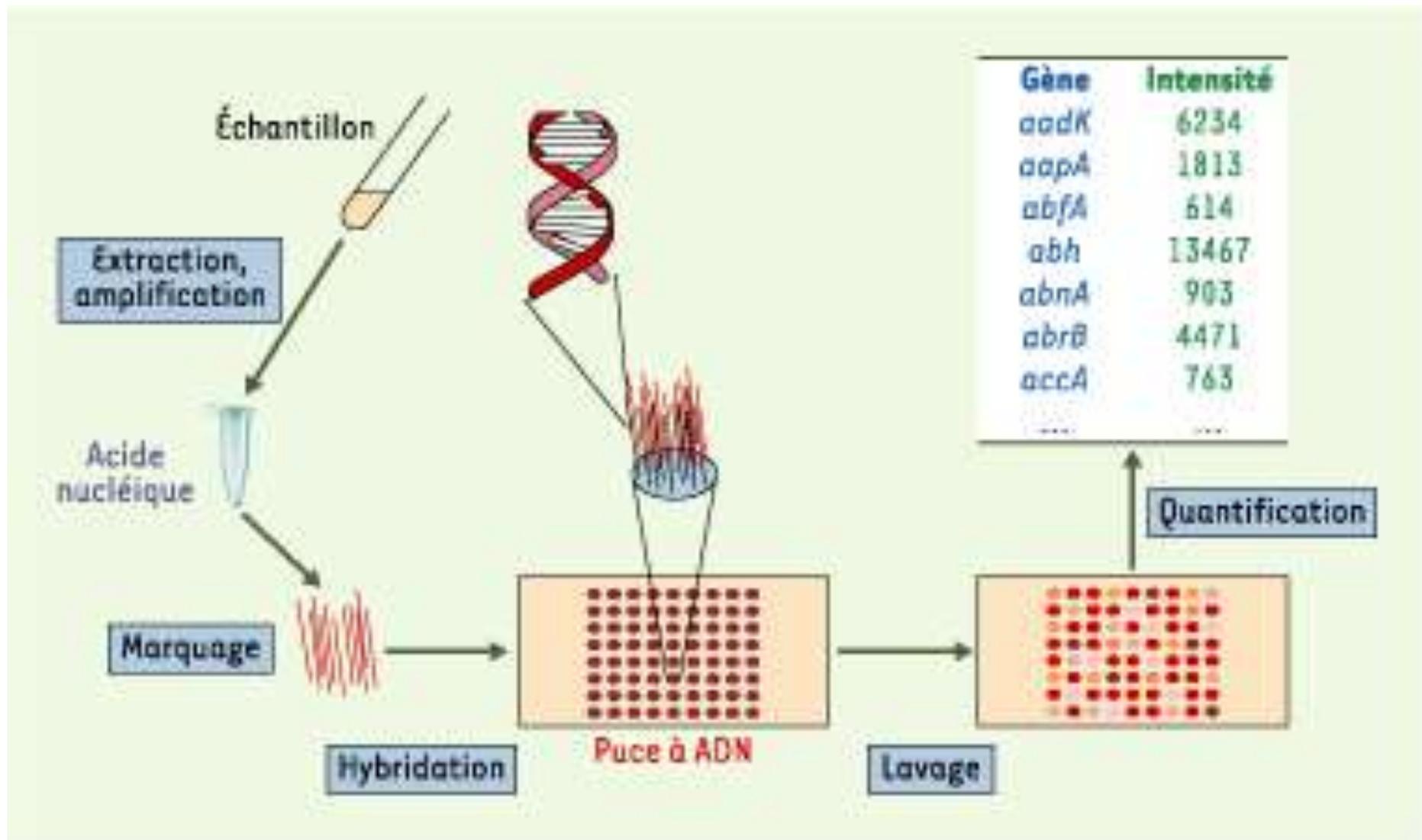


Figure 25 : Analyse d'acide nucléique par puce à ADN.

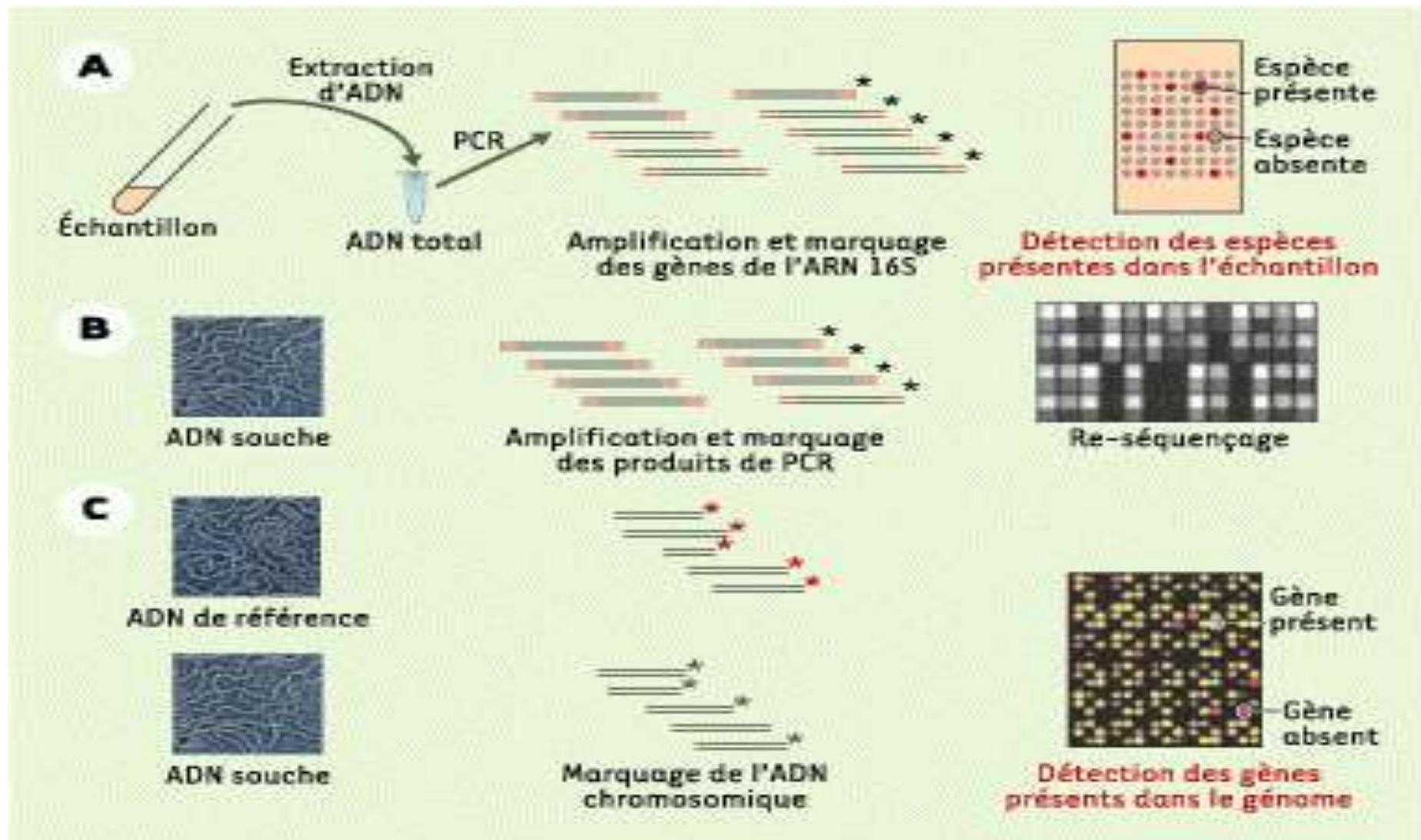


Figure 26 : Trois applications des puces à ADN dans l'analyse microbiologique.

(A : Analyse d'une population bactérienne mixte, B : Puce Affymetrix de re-séquençage, C : Détection de régions chromosomiques spécifiques d'un isolat).

L'ADN total est extrait à partir d'un échantillon et l'ensemble des ADN codant pour les ARN 16S sont amplifiés, en utilisant des amorces universelles, et hybridés sur des puces portant des sondes spécifiques de l'ARN 16S des espèces recherchées. Cet outil est développé pour la surveillance du bioterrorisme avec un ensemble de sondes spécifiques des agents pathogènes potentiellement utilisés.

b) Pour le re-séquençage

La connaissance de la séquence complète d'un génome est le niveau ultime de typage, mais, dans le contexte actuel, le séquençage de chaque isolat n'est pas envisageable. Des puces oligonucléotides sont donc développées pour obtenir des informations partielles sur les génomes bactériens (Fig. 26B). Des puces sont utilisées pour le typage par MLST de *Staphylococcus aureus*. Les sept locus analysés sont amplifiés par PCR et hybridés à la puce au lieu d'être séquencés un par un. L'amélioration de la sensibilité de la technique doit permettre d'utiliser l'ADN génomique total pour détecter un ensemble de positions polymorphes réparties sur le génome. Il serait alors possible d'avoir une vision globale du génome et de pointer sur des mutations particulières, comme celles entraînant la résistance à un antibiotique.

c) Pour la caractérisation génomique

Les puces à ADN sont aussi utilisées pour caractériser la partie variable du génome d'un clone. Ces puces « biodiversité » portent des sondes correspondant à des gènes qui ne sont pas présents dans tous les isolats d'une espèce (Fig. 26C). Elles sont établies à partir de la comparaison des séquences de plusieurs génomes. Par une seule expérience d'hybridation, ces puces permettent d'établir une véritable empreinte digitale correspondant aux gènes présents ou absents dans un clone. Ainsi, à la différence de la majorité des méthodes de typage, les puces à ADN apportent une information fonctionnelle sur la nature des gènes qui différencient deux isolats. Ces résultats peuvent être comparés aux données phénotypiques sur les souches, notamment en relation avec leur virulence.

2.4.3. Avantages et limites des techniques de génotypage haut débit pour la cartographie de QTL

Les techniques de génotypage haut débit pour la cartographie de QTL présentent plusieurs avantages par rapport aux techniques classiques basées sur des marqueurs moléculaires limités. Tout d'abord, ces techniques permettent une couverture plus complète du génome, ce qui permet de détecter des QTL qui seraient difficiles à détecter avec des marqueurs

moléculaires limités. De plus, ces techniques permettent une analyse à haute résolution de la structure du génome et des variations génétiques, ce qui facilite l'identification de QTL associés à des traits complexes.

Cependant, les techniques de génotypage haut débit pour la cartographie de QTL présentent également des limites, notamment la nécessité de disposer de ressources informatiques importantes pour l'analyse des données et la nécessité de vérifier les résultats avec des techniques de validation indépendantes. De plus, ces techniques ne permettent pas d'identifier les mécanismes moléculaires sous-jacents aux QTL, ce qui nécessite des études supplémentaires.

3. Analyse statistique des QTL

3.1. Modèles de régression et d'analyse de variance

L'analyse statistique des QTL (Quantitative Trait Loci) repose sur l'utilisation de modèles de régression et d'analyse de variance (ANOVA) pour identifier les régions du génome qui sont associées à des variations phénotypiques.

Les modèles de régression sont largement utilisés pour l'analyse des QTL. Ces modèles permettent d'analyser la relation entre un trait phénotypique et les marqueurs moléculaires qui sont associés à ce trait. Le modèle de régression le plus couramment utilisé est le modèle linéaire simple, qui peut être écrit comme suit :

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Où : Y est le trait phénotypique, X est le génotype du marqueur moléculaire, β_0 et β_1 sont les coefficients de régression et ε est l'erreur résiduelle.

L'analyse de variance (ANOVA) est une autre méthode couramment utilisée pour l'analyse des QTL. L'ANOVA permet de décomposer la variation phénotypique totale en composantes de variation associées aux effets génétiques et environnementaux. Cette méthode peut être appliquée à des données provenant d'un croisement biparental, dans lequel des individus sont générés en croisant deux lignées parentales différentes.

L'ANOVA pour les QTL peut être effectuée en utilisant un modèle à un ou plusieurs QTL. Le modèle à un QTL peut être écrit comme suit :

$$Y_{ij} = \mu + G_i + E_j + \varepsilon_{ij}$$

Où : Y_{ij} est la valeur phénotypique de l'individu i du groupe j , μ est la moyenne générale, G_i est l'effet génétique de l'individu i , E_j est l'effet environnemental du groupe j et ε_{ij} est l'erreur résiduelle.

Le modèle à plusieurs QTL peut être écrit comme suit :

$$Y_{ij} = \mu + \sum \beta_k Z_{ik} + E_j + \varepsilon_{ij}$$

Où : Z_{ik} est la valeur du marqueur moléculaire k chez l'individu i , β_k est l'effet du marqueur moléculaire k et le symbole \sum indique la somme sur tous les marqueurs moléculaires. Ce modèle permet de détecter plusieurs QTL simultanément.

Les modèles de régression et d'ANOVA pour les QTL présentent des avantages et des inconvénients. Les modèles de régression sont simples et faciles à utiliser, mais ils ne tiennent pas compte des effets environnementaux et ne permettent pas de détecter plusieurs QTL simultanément. Les modèles d'ANOVA permettent de tenir compte des effets environnementaux et de détecter plusieurs QTL simultanément, mais ils nécessitent un grand nombre d'échantillons et peuvent être plus complexes à mettre en œuvre.

L'analyse statistique des QTL repose sur l'utilisation de modèles de régression et d'ANOVA pour identifier les régions du génome qui sont associées à des variations phénotypiques. Le choix du modèle dépend de la question de recherche et des données disponibles.

3.2. Modèles mixtes linéaires pour la détection de QTL

Les modèles mixtes linéaires sont une méthode avancée d'analyse statistique pour la détection de QTL (Quantitative Trait Loci). Ces modèles sont particulièrement utiles pour tenir compte des effets environnementaux, des effets de parenté et des corrélations entre les marqueurs moléculaires. Les modèles mixtes linéaires combinent une composante fixe, qui décrit les effets principaux des marqueurs moléculaires, et une composante aléatoire, qui décrit les effets aléatoires de l'environnement et de la parenté.

Le modèle mixte linéaire peut être écrit comme suit :

$$Y = X\beta + Zu + e$$

Où : Y est le vecteur des valeurs phénotypiques, X est la matrice de conception pour les effets fixes des marqueurs moléculaires, β est le vecteur des coefficients pour les effets fixes, Z est la matrice de conception pour les effets aléatoires de la parenté et de l'environnement, u est le vecteur des effets aléatoires et e est le vecteur des erreurs résiduelles.

Les modèles mixtes linéaires permettent d'estimer les effets génétiques des marqueurs moléculaires tout en prenant en compte les effets environnementaux et de parenté. Ces modèles sont particulièrement utiles pour les données provenant de croisements de lignées apparentées, dans lesquelles les individus partagent une proportion variable de leur ADN.

Une méthode couramment utilisée pour l'analyse des QTL à l'aide de modèles mixtes linéaires est la méthode EMMA (Efficient Mixed-Model Association). Cette méthode permet de tenir compte des corrélations entre les marqueurs moléculaires et les effets environnementaux, ce qui peut améliorer la précision de la détection des QTL. Elle est également efficace pour les données de génotypage à haut débit, qui peuvent comporter des milliers ou des millions de marqueurs moléculaires.

Les modèles mixtes linéaires sont une méthode avancée pour l'analyse statistique des QTL. Ces modèles permettent de tenir compte des effets environnementaux, de parenté et de corrélations entre les marqueurs moléculaires, ce qui peut améliorer la précision de la détection des QTL. La méthode EMMA est une méthode couramment utilisée pour l'analyse des QTL à l'aide de modèles mixtes linéaires.

3.3. Correction de l'effet de population et de l'effet environnemental

Lors de la détection de QTL (Quantitative Trait Loci), il est important de prendre en compte les effets environnementaux et de population pour éviter les faux positifs et améliorer la précision de la détection. Les effets environnementaux peuvent inclure des variations dues aux conditions de culture, aux variations climatiques et aux différences dans les pratiques agricoles, tandis que les effets de population peuvent être dus à la structure de la population ou à la parenté.

La correction de l'effet de population peut être réalisée en utilisant différentes approches, telles que l'ajustement pour la structure de la population en utilisant des méthodes de sous-échantillonnage, l'ajout de Co-variables dans les modèles d'analyse statistique, ou encore l'utilisation de modèles mixtes linéaires qui prennent en compte les effets de parenté.

La correction de l'effet environnemental peut également être réalisée en utilisant différentes approches. Par exemple, des blocs de parcelles peuvent être définis pour contrôler les variations environnementales et minimiser l'impact des erreurs de mesure, ou encore des Co-variables environnementales peuvent être ajoutées aux modèles d'analyse statistique.

Il existe également des méthodes pour corriger à la fois les effets environnementaux et de population. L'une de ces méthodes est la méthode de contrôle de qualité (Quality Control, QC), qui utilise des marqueurs génétiques pour détecter les erreurs expérimentales, les erreurs de génotypage et les individus mal étiquetés, afin d'éliminer les données bruyantes et améliorer la précision de l'analyse.

La correction de l'effet de population et de l'effet environnemental est essentielle pour la détection précise des QTL. Les approches pour la correction de l'effet de population et de l'effet environnemental incluent l'ajustement pour la structure de la population, l'ajout de Co-variables dans les modèles d'analyse statistique, l'utilisation de modèles mixtes linéaires et l'utilisation de méthodes de contrôle de qualité. La combinaison de ces approches peut améliorer la précision de l'analyse et permettre une meilleure compréhension des mécanismes sous-jacents aux traits quantitatifs étudiés.

3.4. Évaluation de la signification statistique des QTL

Une fois que des QTL (Quantitative Trait Loci) ont été détectés, il est important de déterminer leur signification statistique pour évaluer leur pertinence biologique. Différentes approches sont utilisées pour évaluer la signification statistique des QTL.

L'une des approches les plus courantes consiste à utiliser des seuils de significativité basés sur des tests statistiques tels que le test du rapport de vraisemblance (Likelihood Ratio Test, LRT) ou le test de Wald. Ces tests permettent de comparer la qualité de l'ajustement entre un modèle avec le QTL et un modèle sans le QTL, et de déterminer si la différence dans l'ajustement est significative.

Les seuils de significativité sont souvent établis à l'aide de simulations, en utilisant des données simulées pour estimer la distribution nulle de la statistique de test sous l'hypothèse qu'il n'y a pas de QTL. Cette approche permet de contrôler le taux de faux positifs et d'établir des seuils de significativité appropriés pour chaque étude.

Une autre approche pour évaluer la signification statistique des QTL consiste à utiliser des intervalles de confiance (IC) pour estimer la précision de la position et de l'effet du QTL. Les IC permettent de déterminer si les positions et les effets estimés sont suffisamment précis pour être considérés comme significatifs.

Enfin, il est également possible d'utiliser des approches de permutation pour évaluer la signification statistique des QTL. Cette approche consiste à permuter les génotypes des

individus tout en maintenant les phénotypes et les positions des marqueurs génétiques, afin d'estimer la distribution nulle de la statistique de test. Les QTL détectés dans les données réelles sont alors comparés à la distribution nulle pour déterminer leur signification statistique.

L'évaluation de la signification statistique des QTL est essentielle pour déterminer leur pertinence biologique. Différentes approches sont utilisées pour évaluer la signification statistique des QTL, notamment l'utilisation de seuils de significativité basés sur des tests statistiques, l'utilisation d'intervalles de confiance et l'utilisation d'approches de permutation. L'utilisation combinée de ces approches peut améliorer la fiabilité de la détection des QTL et permettre une meilleure compréhension des mécanismes sous-jacents aux traits quantitatifs étudiés.

4. Applications des QTL dans la recherche agricole

4.1. Sélection assistée par marqueurs pour la sélection de caractères complexes

Les QTL (Quantitative Trait Loci) ont de nombreuses applications dans la recherche agricole, notamment pour la sélection assistée par marqueurs (Marker-Assisted Selection, MAS). Cette technique permet d'identifier et de sélectionner des individus porteurs de marqueurs moléculaires associés à des caractères d'intérêt, tels que la résistance aux maladies, la qualité du grain ou la croissance.

L'utilisation de marqueurs moléculaires permet d'identifier rapidement les individus porteurs des marqueurs associés aux caractères d'intérêt, sans avoir à effectuer des tests de phénotypage coûteux et chronophages. Cette technique permet également de sélectionner des individus porteurs des marqueurs pour des caractères complexes, qui sont généralement difficiles à sélectionner par des méthodes conventionnelles.

La sélection assistée par marqueurs est utilisée dans de nombreuses cultures pour améliorer la qualité et le rendement des cultures. Par exemple, dans la culture du riz, des QTL ont été identifiés pour la résistance aux maladies, la tolérance à la salinité et la qualité du grain. En utilisant des marqueurs moléculaires associés à ces QTL, les sélectionneurs peuvent identifier rapidement les individus porteurs des marqueurs pour ces caractères, ce qui permet d'accélérer le processus de sélection.

Dans la sélection des animaux d'élevage, la sélection assistée par marqueurs est utilisée pour améliorer les caractéristiques de production, telles que la croissance, la qualité de la viande et la résistance aux maladies. Les QTL ont été identifiés pour ces caractères dans de nombreuses espèces animales, telles que les bovins, les porcs et les poulets. En utilisant des marqueurs

moléculaires associés à ces QTL, les sélectionneurs peuvent identifier rapidement les animaux porteurs des marqueurs pour ces caractères, ce qui permet d'accélérer le processus de sélection et d'améliorer la productivité et la rentabilité de l'élevage.

Exemple de l'introgression du gène B_t chez le maïs

On réalise une série de backcross entre une lignée élite et une lignée transformée génétiquement. Cette dernière est caractérisée par une insertion unique du gène B_t (transgène de résistance à la pyrale) sur le chromosome 1.

Au cours des rétrocroisements, on peut sélectionner les individus porteurs du gène B_t , et ayant recombiné le plus petit fragment de la lignée donneuse autour du gène B_t . En effet, grâce aux marqueurs moléculaires, on sélectionne les individus ayant pour les marqueurs proches du gène, le génotype de la lignée élite (Fig. 27).

De plus, il est également possible d'accélérer le retour vers le parent élite grâce aux marqueurs moléculaires répartis sur l'ensemble du génome. A chaque rétrocroisement, seront choisis les individus ayant le plus de fragments issus du parent élite récurrent.

À la quatrième génération, on obtient une lignée quasi isogénique de la lignée élite, c'est-à-dire identique à la lignée élite de départ, mais ayant intégré le gène B_t . Il y a donc bien gain de temps, donc d'efficacité.

4.2. Cartographie de QTL pour la résistance aux maladies et aux ravageurs

La cartographie de QTL (Quantitative Trait Loci) pour la résistance aux maladies et aux ravageurs est une application importante des QTL dans la recherche agricole. Cette technique permet d'identifier les régions du génome associées à la résistance aux maladies et aux ravageurs, ce qui permet de développer des variétés de plantes résistantes à ces facteurs de stress biotiques.

La résistance aux maladies et aux ravageurs est une caractéristique importante pour la production agricole, car elle permet de réduire les pertes de récolte et d'améliorer la qualité des cultures. Les maladies et les ravageurs peuvent causer des dommages importants aux cultures, ce qui peut entraîner une baisse de la productivité et de la rentabilité. La cartographie de QTL pour la résistance aux maladies et aux ravageurs permet d'identifier les régions du génome qui contrôlent ces caractéristiques, ce qui permet de développer des variétés de plantes résistantes à ces facteurs de stress.

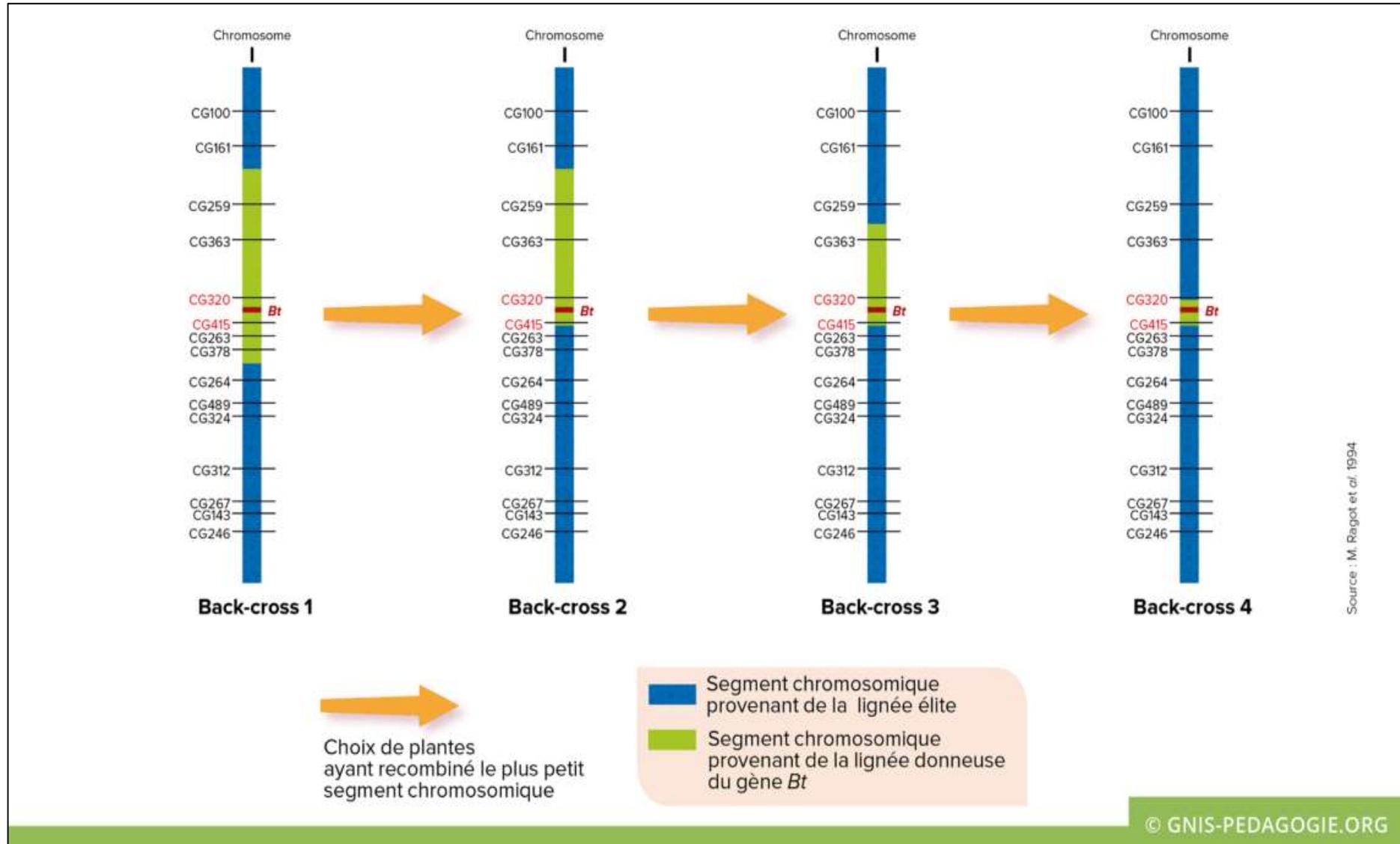


Figure 27 : Sélection assistée par marqueurs moléculaire, Exemple de l'introgession du gène *Bt* chez le maïs

Par exemple, des QTL ont été identifiés pour la résistance à la rouille des feuilles du blé et à la mosaïque du maïs. Ces QTL ont été utilisés pour développer des variétés de blé et de maïs résistantes à ces maladies, ce qui a permis de réduire les pertes de récolte et d'améliorer la qualité des cultures. De même, des QTL ont été identifiés pour la résistance à la pourriture des racines chez les tomates et les poivrons, ce qui a permis de développer des variétés résistantes à cette maladie.

En utilisant la cartographie de QTL, il est également possible de comprendre les mécanismes moléculaires qui sous-tendent la résistance aux maladies et aux ravageurs. Par exemple, des QTL ont été identifiés pour la résistance à la rouille des feuilles du blé, qui sont associés à des gènes de résistance à la maladie. Cette connaissance peut être utilisée pour développer des stratégies de lutte contre les maladies et les ravageurs, qui ciblent spécifiquement ces gènes de résistance.

La cartographie de QTL pour la résistance aux maladies et aux ravageurs est une application importante des QTL dans la recherche agricole. Cette technique permet d'identifier les régions du génome associées à la résistance aux maladies et aux ravageurs, ce qui permet de développer des variétés de plantes résistantes à ces facteurs de stress biotiques. Cela permet d'améliorer la productivité et la rentabilité des cultures, tout en réduisant les pertes de récolte et en améliorant la qualité des cultures.

4.3. Cartographie de QTL pour la qualité des cultures et des animaux

La cartographie de QTL (Quantitative Trait Loci) est également largement utilisée dans la recherche agricole pour améliorer la qualité des cultures et des animaux. Les QTL associés à des caractéristiques de qualité telles que la teneur en protéines, la teneur en lipides, la teneur en sucre, la taille des fruits et la qualité de la viande ont été identifiés dans diverses espèces.

En utilisant la cartographie de QTL, les scientifiques peuvent identifier les gènes qui contrôlent les caractéristiques de qualité et les régions du génome qui sont associées à ces gènes. Cette information peut être utilisée pour améliorer les variétés de cultures et les races d'animaux pour améliorer leur qualité et leur valeur économique.

Par exemple, des QTL ont été identifiés pour la teneur en protéines dans le soja, ce qui a permis de développer des variétés de soja à haute teneur en protéines pour une utilisation dans l'alimentation animale et humaine. De même, des QTL ont été identifiés pour la teneur en huile dans le maïs, ce qui a permis de développer des variétés de maïs à haute teneur en huile pour une utilisation dans l'industrie alimentaire.

En ce qui concerne les animaux, des QTL ont été identifiés pour la qualité de la viande bovine, notamment la tendreté, la jutosité et la saveur. Ces QTL ont été utilisés pour développer des races bovines qui produisent de la viande de qualité supérieure pour le marché des produits alimentaires haut de gamme.

En outre, la cartographie de QTL a également été utilisée pour améliorer la qualité des fruits et des légumes. Des QTL ont été identifiés pour la taille, la couleur et la saveur des fruits et légumes, ce qui a permis de développer des variétés de fruits et de légumes de qualité supérieure pour les consommateurs.

La cartographie de QTL pour la qualité des cultures et des animaux est une application importante des QTL dans la recherche agricole. Cette technique permet d'identifier les régions du génome associées aux caractéristiques de qualité, ce qui permet de développer des variétés de cultures et des races d'animaux qui répondent aux besoins des consommateurs et de l'industrie alimentaire. Cela permet d'améliorer la valeur économique des cultures et des animaux, tout en répondant aux besoins des consommateurs en matière de qualité des produits alimentaires.

4.4. Amélioration de la qualité nutritionnelle des aliments

La cartographie de QTL (Quantitative Trait Loci) est également utilisée dans la recherche agricole pour améliorer la qualité nutritionnelle des aliments. En identifiant les régions du génome qui sont associées à la teneur en nutriments, les scientifiques peuvent développer des variétés de cultures qui ont une teneur élevée en nutriments essentiels pour la santé humaine.

Par exemple, des QTL ont été identifiés pour la teneur en acides gras oméga-3 dans le saumon, ce qui a permis de développer des variétés de saumon à haute teneur en acides gras oméga-3 pour répondre à la demande des consommateurs pour des aliments plus sains. De même, des QTL ont été identifiés pour la teneur en vitamine C dans les agrumes, ce qui a permis de développer des variétés d'agrumes à haute teneur en vitamine C pour répondre à la demande des consommateurs pour des aliments riches en nutriments.

En outre, la cartographie de QTL a également été utilisée pour améliorer la teneur en nutriments essentiels tels que le fer et le zinc dans les cultures de base, telles que le riz, le blé et le maïs. Cela est particulièrement important pour les régions du monde où la malnutrition est répandue et où les cultures de base constituent une grande partie de l'alimentation quotidienne.

La cartographie de QTL pour l'amélioration de la qualité nutritionnelle des aliments est une application importante des QTL dans la recherche agricole. Cette technique permet d'identifier les régions du génome associées à la teneur en nutriments, ce qui permet de développer des variétés de cultures qui ont une teneur élevée en nutriments essentiels pour la santé humaine. Cela contribue à améliorer la qualité de l'alimentation et à lutter contre la malnutrition dans les régions du monde où elle est répandue.

5. Applications des QTL en médecine et en écologie

5.1. Cartographie de QTL pour les maladies humaines

La cartographie de QTL (Quantitative Trait Loci) est une technique utilisée en génétique pour identifier les régions du génome qui sont associées à des traits quantitatifs, tels que la taille, le poids ou la pression artérielle. Cette technique est également utilisée en médecine pour identifier les régions du génome qui sont associées à des maladies humaines.

La cartographie de QTL est particulièrement utile pour les maladies complexes, telles que le diabète, les maladies cardiovasculaires et les maladies auto-immunes, qui sont influencées par de nombreux facteurs génétiques et environnementaux. En identifiant les régions du génome qui sont associées à ces maladies, les scientifiques peuvent mieux comprendre les mécanismes génétiques qui sous-tendent ces maladies et développer de nouveaux traitements plus ciblés.

La cartographie de QTL est généralement réalisée en analysant les données génétiques provenant de familles touchées par une maladie spécifique. Les scientifiques examinent les différences génétiques entre les membres de la famille atteints de la maladie et les membres de la famille qui ne sont pas atteints de la maladie pour identifier les régions du génome qui sont associées à la maladie.

Une fois les régions du génome associées à la maladie identifiées, les scientifiques peuvent effectuer des études plus approfondies pour comprendre les mécanismes génétiques qui sous-tendent la maladie. Cela peut inclure l'identification des gènes spécifiques qui sont impliqués dans la maladie, ainsi que des mutations génétiques qui augmentent le risque de développer la maladie.

En utilisant ces informations, les scientifiques peuvent développer des tests génétiques pour prédire le risque de développer la maladie, ainsi que des thérapies génétiques et des médicaments plus ciblés. La cartographie de QTL est donc une technique importante pour la compréhension et le traitement des maladies complexes chez l'homme.

5.2. Identification des QTL impliqués dans la résistance aux maladies chez les animaux sauvages et domestiques

La cartographie de QTL est également utilisée dans la recherche sur les animaux pour identifier les régions du génome qui sont associées à des traits d'intérêt, tels que la résistance aux maladies chez les animaux sauvages et domestiques.

La résistance aux maladies est un trait important dans l'agriculture, car elle peut avoir un impact significatif sur la santé et la production des animaux. La cartographie de QTL peut être utilisée pour identifier les régions du génome qui sont associées à la résistance aux maladies chez les animaux, ce qui peut aider les éleveurs à sélectionner des animaux qui sont plus résistants aux maladies.

La cartographie de QTL pour la résistance aux maladies chez les animaux peut être réalisée de différentes manières. Dans certains cas, les données génétiques sont collectées à partir d'animaux sauvages ou domestiques qui sont naturellement résistants à une maladie particulière. Les différences génétiques entre ces animaux résistants et ceux qui sont sensibles à la maladie sont ensuite analysées pour identifier les régions du génome qui sont associées à la résistance.

Dans d'autres cas, des croisements entre des animaux résistants et sensibles peuvent être utilisés pour produire une population de descendants qui présentent une variabilité génétique pour la résistance à la maladie. Les données génétiques de cette population peuvent ensuite être utilisées pour identifier les régions du génome qui sont associées à la résistance.

Une fois les QTL associés à la résistance aux maladies identifiés, les éleveurs peuvent utiliser cette information pour sélectionner des animaux résistants à la maladie pour la reproduction. Cela peut aider à améliorer la résistance aux maladies dans les populations animales, ce qui peut avoir des avantages pour la santé et la production animale.

La cartographie de QTL est une technique importante dans la recherche sur les animaux pour identifier les régions du génome qui sont associées à des traits d'intérêt, tels que la résistance aux maladies. Cette information peut être utilisée pour sélectionner des animaux résistants pour la reproduction, ce qui peut aider à améliorer la résistance aux maladies dans les populations animales.

5.3. Identification des QTL impliqués dans l'adaptation aux changements environnementaux

La cartographie de QTL est également utilisée pour identifier les régions du génome qui sont impliquées dans l'adaptation aux changements environnementaux chez les plantes et les animaux. Les changements environnementaux tels que le changement climatique, la pollution et la dégradation des habitats peuvent avoir un impact important sur la survie et la reproduction des espèces.

La capacité d'une espèce à s'adapter à ces changements dépend de sa variabilité génétique et de sa capacité à évoluer rapidement. La cartographie de QTL peut être utilisée pour identifier les régions du génome qui sont associées à des traits d'adaptation, tels que la tolérance aux températures extrêmes, la résistance aux maladies, la capacité à utiliser différents types de nutriments et la capacité à s'adapter à des habitats différents.

La cartographie de QTL pour l'adaptation aux changements environnementaux peut être réalisée en utilisant des populations d'espèces qui présentent une grande variabilité génétique et une grande diversité phénotypique. Les données génétiques et phénotypiques de ces populations peuvent ensuite être analysées pour identifier les régions du génome qui sont associées à des traits d'adaptation.

Une fois les QTL associés à l'adaptation identifiés, il est possible d'utiliser ces informations pour sélectionner des individus qui présentent des traits d'adaptation désirables pour la reproduction. Cela peut aider à améliorer l'adaptation des populations d'espèces aux changements environnementaux, ce qui peut avoir un impact positif sur leur survie et leur reproduction à long terme.

La cartographie de QTL est une technique importante pour identifier les régions du génome qui sont impliquées dans l'adaptation aux changements environnementaux chez les plantes et les animaux. Cette information peut être utilisée pour sélectionner des individus avec des traits d'adaptation désirables pour la reproduction, ce qui peut aider à améliorer l'adaptation des populations d'espèces aux changements environnementaux.

6. Techniques avancées en génétique quantitative

6.1. Analyse d'interaction génétique

L'analyse d'interaction génétique est une technique avancée en génétique quantitative qui cherche à identifier les effets combinés de plusieurs gènes sur un trait donné. Cette

technique est particulièrement importante pour comprendre les processus complexes qui sous-tendent les traits quantitatifs et peut être utilisée pour améliorer la précision de la prédiction des phénotypes.

L'interaction génétique peut être définie comme une situation où l'effet d'un gène sur un trait dépend de l'allèle présent à un autre locus. Par exemple, deux gènes qui ont un effet positif sur la production de lait peuvent avoir un effet négatif lorsque les deux sont présents dans la même lignée. Cette interaction peut être causée par des mécanismes moléculaires tels que la modulation de l'expression génique ou des interactions entre protéines.

Il existe plusieurs méthodes pour l'analyse d'interaction génétique, notamment les modèles additifs, les modèles multiplicatifs et les modèles non additifs. Les modèles additifs considèrent que chaque allèle contribue de manière indépendante à l'expression du trait. Les modèles multiplicatifs prennent en compte les effets combinés des allèles présents sur le trait. Les modèles non additifs considèrent que les effets des allèles dépendent de leur interaction avec d'autres allèles.

L'analyse d'interaction génétique peut être réalisée en utilisant des populations d'espèces qui présentent une grande variabilité génétique et une grande diversité phénotypique. Les données génétiques et phénotypiques de ces populations peuvent ensuite être analysées pour identifier les interactions génétiques qui sont associées à des traits quantitatifs.

Les résultats de l'analyse d'interaction génétique peuvent être utilisés pour comprendre les mécanismes moléculaires qui sous-tendent les traits quantitatifs et pour améliorer la précision de la prédiction des phénotypes. Par exemple, les interactions génétiques peuvent être utilisées pour prédire la réponse d'une population à des traitements spécifiques, tels que l'administration de médicaments ou de suppléments alimentaires.

L'analyse d'interaction génétique est une technique avancée en génétique quantitative qui permet d'identifier les effets combinés de plusieurs gènes sur un trait donné. Cette technique est particulièrement importante pour comprendre les processus complexes qui sous-tendent les traits quantitatifs et peut être utilisée pour améliorer la précision de la prédiction des phénotypes.

6.2. Modèles de régression bayésienne

Les modèles de régression bayésienne sont une technique avancée d'analyse statistique qui est de plus en plus utilisée en génétique quantitative pour l'analyse de QTL. Cette approche

permet de prendre en compte l'incertitude dans les estimations des paramètres et des effets génétiques, ainsi que la structure hiérarchique des données.

Le modèle de régression bayésienne suppose que les paramètres sont distribués selon une distribution de probabilité a priori et que la distribution a posteriori est obtenue en combinant cette information avec les données observées. L'utilisation de l'inférence bayésienne permet ainsi de mettre à jour les distributions de probabilité des paramètres à mesure que de nouvelles données sont collectées.

Cette approche est particulièrement utile pour l'analyse de données complexes, telles que les données d'expression génique, où la dimension des données est élevée et où les effets génétiques sont souvent faibles. Les modèles de régression bayésienne permettent également de prendre en compte l'effet de nombreux gènes en même temps, ainsi que l'effet des interactions entre ces gènes.

Cependant, cette approche peut être plus complexe à mettre en œuvre et nécessiter des ressources informatiques importantes. De plus, les résultats peuvent être sensibles aux choix des distributions a priori et des paramètres du modèle. Par conséquent, une interprétation soignée des résultats est essentielle pour éviter des erreurs d'interprétation.

6.3. Méthodes de sélection assistée par marqueurs avancées

Les méthodes de sélection assistée par marqueurs (MAS) sont des techniques avancées en génétique quantitative qui permettent de sélectionner les individus ayant les meilleures caractéristiques génétiques pour un trait d'intérêt, en se basant sur des marqueurs moléculaires spécifiques.

La sélection traditionnelle implique l'évaluation des performances phénotypiques des individus, ce qui est souvent coûteux, prend du temps et peut être imprécis. La sélection assistée par marqueurs permet de surmonter ces obstacles en utilisant des marqueurs moléculaires pour prédire la valeur génétique des individus pour un trait d'intérêt, sans avoir à évaluer directement le phénotype.

Les méthodes de sélection assistée par marqueurs avancées, telles que la sélection génomique ou la sélection génomique large, utilisent des informations génomiques provenant de l'ensemble du génome plutôt que de se concentrer sur un petit nombre de marqueurs. Ces méthodes sont basées sur l'utilisation de puissants outils statistiques pour prédire la valeur génétique des individus en utilisant l'ensemble des informations génomiques disponibles.

La sélection génomique utilise des marqueurs moléculaires pour prédire la valeur génétique des individus, tandis que la sélection génomique large utilise l'ensemble du génome pour prédire la valeur génétique. Ces méthodes ont été largement utilisées dans l'amélioration des plantes et des animaux, où elles ont permis d'accélérer considérablement le processus de sélection et d'obtenir des gains génétiques importants.

Cependant, il est important de noter que l'utilisation de la sélection assistée par marqueurs avancées nécessite une connaissance approfondie de la génétique du trait d'intérêt, une quantité importante de données génomiques et une expertise en bio-informatique et en statistiques. Par conséquent, ces méthodes ne sont pas toujours accessibles aux petits producteurs ou aux éleveurs qui ne disposent pas de ressources suffisantes.

7. Développements futurs dans le domaine des QTL

7.1. Évolution des technologies de séquençage haut débit

Les QTL (quantitative trait loci) sont des régions du génome qui sont associées à des variations phénotypiques quantitatives pour des traits d'intérêt. Les avancées technologiques dans le domaine du séquençage haut débit ont révolutionné la façon dont les QTL sont identifiés et caractérisés, et ont conduit à de nouveaux développements futurs passionnants dans ce domaine.

L'une des tendances clés dans le domaine des QTL est l'utilisation de technologies de séquençage haut débit pour cartographier les QTL à l'échelle du génome entier. Cette approche permet d'identifier des QTL qui ne peuvent pas être détectés par les méthodes de cartographie traditionnelles, ce qui peut améliorer la précision de la sélection des individus dans les programmes d'amélioration des plantes et des animaux. De plus, le séquençage haut débit permet également d'identifier des variants génétiques rares qui pourraient avoir des effets phénotypiques importants.

Une autre tendance importante dans le domaine des QTL est l'utilisation de méthodes de séquençage haut débit pour comprendre les mécanismes moléculaires sous-jacents aux variations phénotypiques. Les données de séquençage permettent de relier les variants génétiques à des changements dans la structure ou la fonction des protéines, ce qui peut aider à comprendre les voies métaboliques et les réseaux de régulation impliqués dans la variation phénotypique.

Enfin, les avancées dans le domaine du séquençage haut débit ont également conduit à l'émergence de nouvelles approches pour l'analyse des QTL, telles que l'utilisation de réseaux

de régulation génétique pour comprendre les interactions entre les gènes et les voies métaboliques impliqués dans la variation phénotypique.

Dans l'ensemble, les développements futurs dans le domaine des QTL seront probablement liés à l'évolution continue des technologies de séquençage haut débit, ainsi qu'à une meilleure compréhension des mécanismes moléculaires sous-jacents aux variations phénotypiques. Ces avancées devraient contribuer à accélérer le développement de nouvelles variétés de plantes et de nouvelles races animales avec des caractéristiques améliorées et plus résilientes.

7.2. Nouvelles approches pour l'analyse de données à haut débit

Les développements futurs dans le domaine des QTL seront également liés à l'évolution de nouvelles approches pour l'analyse de données à haut débit. Les technologies de séquençage haut débit produisent des quantités massives de données génomiques, ce qui nécessite des outils avancés pour l'analyse et l'interprétation des données.

Une des approches en développement est l'intégration de données multi-omiques, qui consiste à combiner des données génomiques avec d'autres types de données omiques, tels que des données transcriptomiques, protéomiques et métabolomiques. Cette approche permet de comprendre les mécanismes moléculaires sous-jacents à la variation phénotypique à travers différentes échelles biologiques, allant du niveau de l'ADN aux niveaux protéiques et métaboliques.

Une autre approche importante est l'utilisation de l'apprentissage automatique (machine learning) pour l'analyse de données génomiques à haut débit. Cette approche permet d'identifier des modèles complexes dans les données et de prédire des relations entre les gènes et les traits phénotypiques. L'apprentissage automatique peut également être utilisé pour l'identification de variants génétiques rares et pour la classification de sous-populations génétiques.

Enfin, les nouvelles approches pour l'analyse de données à haut débit incluent également des outils pour la visualisation interactive des données et pour la création de bases de données intégrées qui permettent de stocker et d'interroger des données génomiques et phénotypiques à grande échelle.

Les développements futurs dans le domaine des QTL seront largement dépendants de l'évolution de nouvelles approches pour l'analyse de données à haut débit. Ces avancées devraient contribuer à une meilleure compréhension des mécanismes moléculaires sous-jacents

aux variations phénotypiques et à l'identification de nouveaux traits d'intérêt pour l'amélioration des plantes et des animaux.

7.3. Perspectives pour l'application des QTL dans la sélection assistée par marqueurs et la biologie de synthèse

Les QTL ont des implications importantes dans le domaine de la sélection assistée par marqueurs, qui est une méthode de sélection d'individus pour des traits d'intérêt en utilisant des marqueurs génétiques associés à ces traits. Les avancées dans la cartographie et la caractérisation des QTL, ainsi que dans l'analyse de données à haut débit, ont permis d'améliorer la précision de la sélection assistée par marqueurs. De plus, l'utilisation de techniques de biologie de synthèse pour la modification des génomes offre des possibilités pour la création de nouveaux QTL artificiels.

Dans le domaine de la sélection assistée par marqueurs, les développements futurs devraient se concentrer sur l'identification de marqueurs génétiques à haute résolution, ainsi que sur la mise en œuvre de méthodes de sélection assistée par marqueurs basées sur des modèles statistiques plus sophistiqués. Ces méthodes devraient permettre d'optimiser la sélection pour des traits complexes et multigéniques.

En ce qui concerne la biologie de synthèse, les développements futurs pourraient permettre de créer des QTL artificiels pour des traits qui n'existent pas naturellement dans la population d'intérêt. Les techniques de biologie de synthèse peuvent être utilisées pour synthétiser des gènes et des réseaux de régulation pour contrôler des voies métaboliques spécifiques, ce qui pourrait permettre de modifier la production de métabolites et la réponse à des stress biotiques et abiotiques.

Enfin, les QTL peuvent également être utilisés dans la conception de systèmes biologiques artificiels, tels que des réseaux de régulation génétique modulaires pour la production de molécules à des fins médicales et industrielles. Les QTL peuvent également être utilisés pour concevoir des systèmes de biosenseurs pour la détection de molécules spécifiques dans l'environnement.

Les perspectives pour l'application des QTL dans la sélection assistée par marqueurs et la biologie de synthèse sont prometteuses. Les développements futurs dans ces domaines devraient permettre d'élargir l'éventail des traits phénotypiques disponibles pour la sélection et la création de systèmes biologiques artificiels. Cependant, ces avancées soulèvent également

des questions éthiques et réglementaires importantes, qui devront être prises en compte pour garantir une utilisation responsable de ces technologies.

Références bibliographiques

Références bibliographiques

- Aguet, F., Alasoo, K., Li, Y.I., Battle, A., Im H.K., Montgomery, S.B. Lappalainen, T. (2023). Molecular quantitative trait loci. *Nature Reviews Methods Primers*, 3 : 4.
- Azouzi, B. (2006). L'outil statistique en expérimentation. Ed. Office des publications universitaires, Alger. 163 p.
- Baril, C., Bergonzini J.C. (1994). De la Génétique des Populations à la Génétique Quantitative. Ed. CIRAD, Forêt. 75 p.
- Barton, N. H., Turelli, M. (1989). Evolutionary Quantitative Genetics: How Little Do We Know? *Annual Review of Genetics*, 23(1): 337–370.
- Barton, N.H., Keightley, P.D. (2002). Multifactorial genetics: Understanding quantitative genetic variation. *Nature Reviews Genetics*, 3: 11–21.
- Becker, W.A. (1967). Program in Genetics: Manual of procedures in quantitative genetics. Ed. Pullman, Washington State University. 130 p.
- Broman, K. W., Speed, T. P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 641–656.
- Brun, J.M. (1992). Les bases de la génétique quantitative : Définition et mesure des paramètres du croisement. In. *Eléments de Génétique Quantitative et application aux populations animales*. Ed. INRA, Production animale. Pp : 101-105.
- Bulmer, M. G. (1985). The mathematical theory of quantitative genetics. Ed. Clarendon Press, Oxford, UK. 255 p.
- Caballero, A. (2020). Quantitative Genetics. Ed. Cambridge University Press. 325 p.
- Charlesworth, B. (1990). Optimization Models, Quantitative Genetics, And Mutation. *Evolution*, 44(3): 520–538.
- Charlesworth, D., Charlesworth, B. (1995). Quantitative Genetics in Plants: The Effect of the Breeding System on Genetic Variability. *Evolution*, 49(5): 911–920.
- Choo, T. M., Reinbergs, E., & Jui, P. Y. (1988). Comparison of F₂ and F_∞ diallel analyses in barley. *Genome*, 30(6), 865–869.

- Churchill, G.A., Doerge, R.W. (1998). Mapping Quantitative Trait Loci in Experimental Populations. In. Paterson, A.H. (eds.), *Molecular Dissection of Complex Traits*. CRC Press. 11 p.
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(1-2): 169–196.
- Cristobal, M.S., Robert-Granié, C., Foulley, J.L. (2002). Hétéroscédasticité et modèles linéaires mixtes : théorie et applications en génétique quantitative. *Journal de la société française de statistique*, 143 (1-2) : 155-165.
- Cruz, C. D. (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum, Agronomy*, 35(3): 271-276.
- Cyplik, A., Bocianowski J. (2022). Analytical and numerical comparisons of two methods of estimation of additive×additive×additive interaction of QTL effects. *Journal of Applied Genetics*, 63: 213–221.
- De Jong, G. (1990). Quantitative Genetics of reaction norms. *Journal of Evolutionary Biology*, 3(5-6), 447–468.
- Doebeli, M. (1996). Quantitative Genetics and Population Dynamics. *Evolution*, 50(2), 532–546.
- Doerge, R.W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics*, 3: 43–52.
- Doerge, R.W., Zeng Z.B., Weir B.S. (1997). Statistical Issues in the Search for Genes Affecting Quantitative Traits in Experimental Populations. *Statistical Science*, 12 (3): 195-219.
- Doumith, M., Cazalet, C., Simoes, N., Frangeul, L., Jacquet, C., Kunst, F., Martin, P., Cossart, P., Glaser, P., Buchrieser, C. (2004). New Aspects Regarding Evolution and Virulence of *Listeria monocytogenes* Revealed by Comparative Genomics and DNA Arrays. *Infection and Immunity*, 72(2), 1072–1083.
- Ducrocq, V. (1992). Les bases de la génétique quantitative : Du modèle génétique au modèle statistique. In. *Eléments de Génétique Quantitative et application aux populations animales*. Ed. INRA, Production animale. Pp : 75-81.

- Erickson, D. (2005). Quantitative trait loci: Mapping the future of QTL's. *Heredity*, 95: 417–418.
- ERNER, L.M., 1954. Genetic Homeostasis. Ed. Oliver and Boyd, Edinburgh. 134 p.
- Falconer, D.S. (1983). Problems on quantitative genetics. Ed. Longman, London, UK. 104 p.
- Falissard, B. (2005). Comprendre et utiliser les statistiques dans la science de la vie. Ed. Masson, Paris. 372 p.
- Frankham, R. (1999). Quantitative genetics in conservation biology. *Genetical Research*, 74(3), 237–244.
- Frost, J., 2020. Hypothesis Testing: An intuitive guide for making data driven decisions. Ed. Jim Frost. 371 p.
- Gallais, A. (1990). Théorie de la sélection en amélioration des plantes. Ed. Masson, Paris. 588 p.
- Gallais, A. (2009). Hétérosis et variétés hybrides en amélioration des plantes. Ed. Quae, France. 356 p.
- Garcia-Oliveira, A. L., Tan, L., Fu, Y., Sun, C. (2009). Genetic Identification of Quantitative Trait Loci for Contents of Mineral Nutrients in Rice Grain. *Journal of Integrative Plant Biology*, 51(1), 84–92.
- Gibson, G., Weir, B. (2005). The quantitative genetics of transcription. *Trends in Genetics*, 21(11), 616–623.
- Glaser, P. (2005). Les puces à ADN vont-elles révolutionner l'identification des bactéries ? *Médecine/sciences*, 21(5), 539–544.
- Goffinet, B., Gerber, S. (2000). Quantitative Trait Loci: A Meta-analysis. *Genetics*, 155(1): 463–473.
- Gomulkiewicz, R., Kirkpatrick, M. (1992). Quantitative Genetics and the Evolution of Reaction Norms. *Evolution*, 46(2): 390–411.
- Gorokhova, S., Biancalana, V., Lévy, N., Laporte, J., Bartoli, M., Krahn, M. (2015). Clinical massively parallel sequencing for the diagnosis of myopathies. *Revue Neurologique*, 171: 558–571.

- Grant, I., Beversdorf, W.D. (1985). Heterosis and combining ability estimates in spring-planted oilseed rape (*Brassica napus* L.). *Canadian Journal of Genetics and Cytology*, 27(4), 472–478.
- Hackett, C. A., Weller, J. I. (1995). Genetic Mapping of Quantitative Trait Loci for Traits with Ordinal Distributions. *Biometrics*, 51(4): 1252.
- Hallauer, A.R., Carena, M.J., Miranda Filho, J.B. (2010). *Quantitative Genetics in Maize Breeding*. Ed. Springer Science & Business Media, Iowa State University. 662 p.
- Hill, W. G. (2009). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537): 73–85.
- <https://www.semae-pedagogie.org/>
- Jacquard, A. (1970). Génétique des populations et raisonnement probabiliste. *Journal de la société statistique de Paris*, 111 : 223-229.
- Jacquard, A. (1970). *Structure génétique des populations*. Ed. Masson, France. 388 p.
- Jiang, C., Zeng Z.B. (1995). Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. *Genetics*, 140: 1111 – 1127.
- Johannes, F., Colot, V., Jansen, R.C. (2008). Epigenome dynamics: a quantitative genetics perspective. *Nature Reviews Genetics*, 9: 883–890.
- Kao, C.H., Zeng, Z.B., Teasdale R.D. (1999). Multiple Interval Mapping for Quantitative Trait Loci. *Genetics*, 152: 1203–1216.
- Kearsey, M.J. Farquhar A.G.L. (1998). QTL analysis in plants; where are we now? *Heredity*, 80:137–142.
- Knapp, S. J., Bridges, W. C., & Birkes, D. (1990). Mapping quantitative trait loci using molecular marker linkage maps. *Theoretical and Applied Genetics*, 79(5).
- Krahn, M., Arveiler, B. (2016). Le séquençage de nouvelle génération : principe, applications en diagnostic et perspectives. In : *Livre national d'enseignement - Génétique Médicale - DFGSM2/3*. Ed. Issy-les-Moulineaux : Elsevier-Masson.
- Kumar, J., Gupta, D. S., Gupta, S., Dubey, S., Gupta, P., Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Reports*, 36(8): 1187–1213.

- Lamkey K.R., Edwards J.W. (1999). Quantitative Genetics of Heterosis. In. Coors, J.G., Pandey, S., Lamkey, K.R., Edwards, J.W. (eds), The Genetics and Exploitation of Heterosis in Crops. *ACSESS Publications*. Pp: 31- 48.
- Le Roy, P. (1992). Les bases de la génétique quantitative : Les méthodes de mise en évidence des gènes majeurs. In. *Eléments de Génétique Quantitative et application aux populations animales*. Ed. INRA, Production animale. Pp : 93-99.
- Lefort-Buson, M., Dattée, Y., Lavoisier, O. (1982). Genetic study of some agronomic characters in winter oilseed rape (*Brassica napus* L.) II. - Genetic parameters. *Agronomie*, 2 (4): 323-332.
- Lellouch, J., Lazar, P. (1974). *Méthodes statistiques en expérimentation biologique*. Ed. Flammarion, France. 283 p.
- Louis, O. (2002). *Éléments de génétique quantitative*. Ed. Quae, France. 184 p.
- Malécot, G., Malecot, G. (1955). *La Génétique de population : Principes et applications*. *Population* (French Edition), 10(2) : 239.
- Mallar, J. (1992). Les bases de la génétique quantitative : Population et Variabilité. In. *Eléments de Génétique Quantitative et application aux populations animales*. Ed. INRA, Production animale. Pp : 55-60.
- Mallol, C., Varro, L. (). Conditions d'existence d'une loi de Hardy-Weiberg périodique dans des populations sélectivement neutres. In. Costa, R., Grishkov, A., Guzzo, H.V., Peresi, L.A. (eds). *Nonassociative algebra and its application*. Marcel Dekker Inc., New York. Pp: 439-446.
- McIntyre, L.M., Coffman, C.J., Doerge, R.W. (2001). Detection and localization of a single binary trait locus in experimental populations. *Genetics Research*, 78(01): 79-92.
- Miles, C., Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nature Education*, 1(1): 208.
- Moll, R. H., Stuber, C. W. (1974). Quantitative Genetics—Empirical Results Relevant to Plant Breeding. *Advances in Agronomy*: 277–313.
- Morgenthaler, S. (2008). *Génétique statistique*. Ed. Springer-Verlag France. 156 p.
- Motulsky, H. (2018). *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*. Ed. Oxford University Press. 605 p.

- Motulsky, H. (2019). Biostatistique : Une approche intuitive. Ed. De Boeck Supérieur s.a., France. 577 p.
- Mullin, T. J., Park, Y. S. (1992). Estimating genetic gains from alternative breeding strategies for clonal forestry. *Canadian Journal of Forest Research*, 22(1), 14–23.
- Namkoong, G. (1979). Introduction to Quantitative Genetics in Forestry. *Technical Bulletin*, 1588. 342 p.
- Posthuma, D., Beem, A. L., de Geus, E. J. C., van Baal, G. C. M., von Hjelmberg, J. B., Iachine, I., Boomsma, D. I. (2003). Theory and Practice in Quantitative Genetics. *Twin Research*, 6(05), 361–376.
- Poujardieu, B., Mallard, J. (1992). Les bases de la génétique quantitative : Les méthodes d'estimation de l'héritabilité et des corrélations génétiques. In. *Eléments de Génétique Quantitative et application aux populations animales*. Ed. INRA, Production animale. Pp : 87-92.
- Roff, D. A. (2007). A Centennial Celebration For Quantitative Genetics. *Evolution*, 61(5): 1017–1032.
- Roff, D.A. (1997). *Evolutionary Quantitative Genetics*. Ed. Springer Science & Business Media. 492 p.
- Rousson, V. (2013). *Statistique appliquée aux sciences de la vie*. Springer-Verlag France. 327 p.
- Rouyre, N., Gilles, B. (1995). *Génétique quantitative : aspects généraux et application à l'amélioration des plantes*. Ed. CIRAD-Forêt, France. 31 p.
- Rumbaugh M.D., Caddel J.L., Rowe D.E. (1988). Breeding and Quantitative Genetics. In. Hanson, A.A., Barnes, D.K., Hill, R.R., Rumbaugh, M.D., Caddel, J.L., Rowe, D.E., *Alfalfa and Alfalfa Improvement. Agronomy Monograph*, 29: 777-808.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences*, 93(20): 10614–10619.
- Seaton, G., Haley, C. S., Knott, S. A., Kearsey, M., Visscher, P. M. (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics*, 18(2): 339–340.

- Sehgal, D., Singh, R., Rajpal, V. R. (2016). Quantitative Trait Loci Mapping in Plants: Concepts and Approaches. *Sustainable Development and Biodiversity*, 31–59.
- Shi, X.M. (2020). eQTL Analysis: Methods and Protocols. Ed. eQTL Analysis Methods and Protocols. Ed. Humana Press, Springer Protocols, USA. 252 p.
- Southern, E. M. (1996). DNA chips: analysing sequence by hybridization to oligonucleotides on a large scale. *Trends in Genetics*, 12(3), 110–115.
- Stuber, C.W., Edwards, M. D., Wendel, J.F. (1987). Molecular Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. II. Factors Influencing Yield and its Component Traits¹. *Crop Science*, 27(4) : 639.
- Tourab, D. (2006). Initiation à la biostatistique. Ed. Amazon Fulfillment, Poland. 220p.
- Trad, S., Allignet, J., Frangeul, L., Davi, M., Vergassola, M., Couve, E., Morvan, A., Kechrid, A., Buchrieser, K., Glaser, P., El Solh, N. (2004). DNA Macroarray for Identification and Typing of *Staphylococcus aureus* Isolates. *Journal of Clinical Microbiology*, 42(5), 2054–2064.
- Troesch, A., Nguyen, H., Miyada, C.G., Desvarenne, S., Gingeras, T.R., Kaplan, P.M., Cros, P., Mabilat, C. (1999). Mycobacterium Species Identification and Rifampin Resistance Testing with High-Density DNA Probe Arrays. *Journal of Clinical Microbiology*, 37(1), 49–55.
- Van Buijtenen, J.P. (2001). Genomics and quantitative genetics. *Canadian Journal of Forest Research*, 31(4), 617–622.
- Van Leeuwen, W.B., Jay, C., Snijders, S., Durin, N., Lacroix, B., Verbrugh, H.A., Enright, M.C., Troesch, A., van Belkum, A. (2003). Multilocus Sequence Typing of *Staphylococcus aureus* with DNA Array Technology. *Journal of Clinical Microbiology*, 41(7): 3323–3326.
- Vilain, M. (1999). Méthodes expérimentales en agronomie : Pratique et analyse. Ed. Technique et Documentation, Paris. 337 p.
- Wada, Y., Akita, T., Awata, T., Furukawa, T., Sugai, N., Ishii, K., Ito, Y., Kobayashi, E., Kusumoto, H., Matsumoto, T., Mikawa, S., Miyake, M., Murase, A., Shimanuki, S., Sugiyama, T., Uchida, Y., Yanai, S., Yasue, H. (2000). Quantitative trait loci (QTL) analysis in a Meishan x Gottingen cross population. *Animal Genetics*, 31(6), 376–384.

- Walsh, B. (2001). Quantitative Genetics. *Encyclopedia of Life Sciences*: 1-7.
- William, R. Meredith, J. (1984). Quantitative Genetics. In. Kohel, R.J., Lewis, C.F., Meredith, W.R. (eds), Cotton, *Agronomy Monograph*, 24: 131-150.
- Wricke, G., Weber, E. (1986). Quantitative Genetics and Selection in Plant Breeding. Ed. Walter de Gruyter, Berlin. 407 p.
- Xu, Y., Li, P., Yang, Z., Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal*, 5(2): 175–184.
- Yeri, S.B., Kumari, V., Sharma, R., Punia S.S. (2022). Quantitative Trait Locus (QTL) Mapping in Crop Improvement. In. Kumar, N. (eds), *Biotechnology and Crop Improvement*. CRC Press. 10 p.
- Yun, S. J., Gyenis, L., Hayes, P. M., Matus, I., Smith, K. P., Steffenson, B. J., & Muehlbauer, G. J. (2005). Quantitative Trait Loci for Multiple Disease Resistance in Wild Barley. *Crop Science*, 45(6), 2563.
- Zeng, Z.B. (1994). Precision Mapping of Quantitative Trait Loci. *Genetics*, 136: 1457-1468.