

Au sein du laboratoire de biologie moléculaire on trouve des:

- **Produits** : amorces, enzymes (de restriction, taq-polymérase, ARNase, DNAase...), chloroforme, agarose, **BET**,
- **Consommables** : tubes eppendorf, gants, plaques PCR, embouts pour micropipettes,
- **Appareillages** : vortex, genogrinder, thermocycleur, machine à glace, bain marie, spectrophotomètre, hotte chimique, électrophorèse, centrifugeuse (réfrigérée), micropipettes, autoclave, ...

Au sein du laboratoire de biologie moléculaire il faut:

- **Hygiène et sécurité** : toute activité dans un laboratoire peut nécessiter l'utilisation de produits chimiques potentiellement dangereux, pour soi et pour l'environnement. Il est donc important de suivre les consignes d'hygiène et de sécurité des biens et des personnes.
- **Il faut impérativement travailler avec blouse et avec des gants.**
- Certains solvants nécessitent d'être préparés et manipulés sous des hottes chimiques.
- Dans un laboratoire de biologie moléculaire, le produit le plus dangereux est le **BET (bromure d'éthidium)**, classé parmi les CMR.
- Les appareils utilisés peuvent aussi être une source de danger: la centrifugeuse qui tourne à haute vitesse, les autoclaves, les électrophorèses, les lampes UV...
- **Travailler sans gaspillage des produits et manipuler soigneusement les appareils.**

I. Analyse du polymorphisme de l'ADN

I.1.Extraction de l'ADN

L'extraction de **haute qualité et de grande quantité d'ADN** génomique à partir de tissus est au cœur de nombreux tests moléculaires et, plus récemment, l'application de séquençage pour caractériser la diversité des plantes, la récupération de l'ADN peut être considérée comme un outil fondamental de la science de la plante.

Les étapes de base de l'extraction de l'ADN sont les suivantes :

1. La collecte et le stockage de tissus végétaux
2. La lyse des cellules végétales
3. La solubilisation des lipides et des protéines avec des détergents
4. La séparation de l'ADN des autres molécules
5. La purification de l'ADN séparé
6. La suspension dans un tampon approprié

Les méthodes d'extraction d'ADN végétal :

- Solutions préparées
- Kits commerciaux (Qiagen, MB Biomedical, Fisher Scientific...)

I.2. Dosage et analyse de la pureté de l'ADN extrait

Le dosage de la quantité de l'ADN extrait peut se faire par 4 méthodes :

Électrophorèse sur gel d'agarose, le spectrophotomètre, le Qubit et la qPCR

I.2.1. L'électrophorèse est un outil important dans l'analyse qualitative et quantitative des acides nucléiques (ADN, ARN). Son principe repose sur la séparation des molécules. Les critères de séparation utilisés sont la taille (poids moléculaire) et la charge électrique.

Les fragments de l'ADN sont séparés en fonction du poids moléculaire. Le gel utilisé est le gel d'agarose (généralement à 1%): par utilisation de marqueur de taille « Low DNA » qui nous permet, selon la migration de connaître la taille de notre ADN.

Après la migration d'électrophorèse, le gel est soumis à une lumière sous ultraviolet afin d'observer les bandes d'ADN fluorescentes. Le marqueur d'acide nucléique contenu dans le gel devient fluorescent avec une couleur rouge orangée.

L'estimation de la taille des fragments est faite à l'aide du marqueur de taille moléculaire (Low DNA Mass Ladder) utilisé simultanément dans un autre puits lors de la migration.

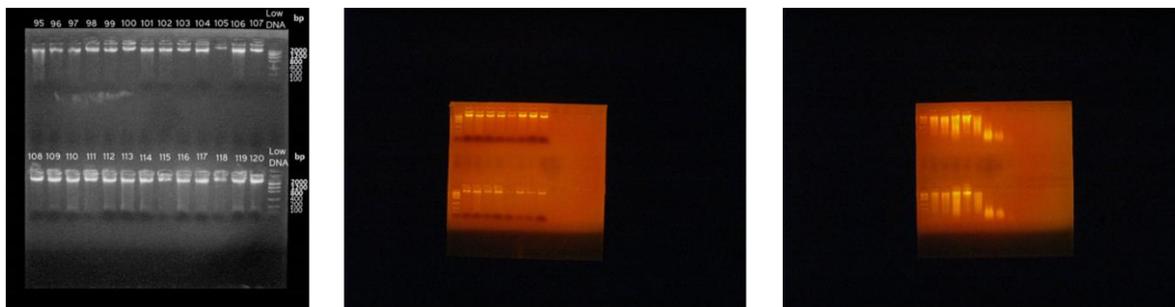


Fig. 1 : Gels d'agarose après migration et après autoradiographie sous UV

I.2.2. Spectrophotomètre (nano-drop)

Les acides nucléiques absorbent dans l'UV à $\lambda_{\text{max}} = 260 \text{ nm}$. En mesurant l'absorbance d'une solution d'acide nucléique pure à 260 nm, on obtient sa concentration à l'aide de l'équation de la loi Beer-Lambert. En pratique, c'est une équation modifiée qui est employée :

$$c_m = (A \cdot e) / l$$

c_m = concentration massique d'acide nucléique en ng/ μ l

A = absorbance à 260 nm

e = coefficient d'extinction massique ou la constante en ng.cm/ μ l

l = trajet optique en cm (généralement 1 cm).

1 unité de DO = 50 ng/ μ l

Des coefficients d'extinction acceptés par la communauté scientifique sont listés ci-après: ADN double brin = ~50 ng.cm/ μ l; ADN simple brin = ~33 ng.cm/ μ l; ARN = ~40 ng.cm/ μ l

➤ **Analyse de la pureté par spectrophotométrie**

Le ratio A260 nm/280 nm: Ce ratio explique la pureté des acides nucléiques (évaluer la contamination de protéines). La pureté est considérée comme acceptable lorsque le ratio A260 nm/A280 nm est compris entre 1,8 – 2,0 pour l'ADN et entre 2,0 – 2,2 pour l'ARN.

Le ratio A260 nm/A230 nm: Un ratio A260/A230 compris entre 2,0 – 2,2 indique une bonne pureté des acides nucléiques. Lorsque ce ratio est significativement plus faible que 2,0 – 2,2, cela révèle la présence de contaminants absorbant à 230 nm dans la solution (le phénol, les carbohydrates, les peptides, l'acide humique, l'urée, l'EDTA, les polysaccharides ...).

I.2.3. Le Qubit :

Son principe est la fluorométrie basée sur la mesure de la fluorescence des fluorophores spécifiques des acides nucléiques. Mais des réactifs et standards dédiés sont nécessaires pour réaliser ces mesures sur le Qubit

Contrairement à la spectrophotométrie, la fluorimétrie ne permet pas d'analyser la pureté et ni d'identifier des contaminants dans une solution d'acides nucléiques.

I.2.4. La qPCR: la PCR quantitative ou PCR en temps réel

Son principe est de suivre la quantité d'ADN cycle après cycle grâce à un marqueur fluorescent dont l'intensité d'émission est proportionnelle à la quantité de produits.

Le thermocycleur est branché à un ordinateur qui, avec un programme dédié va mesurer l'intensité de fluorescence cycle après cycle et il va tracer une courbe.

NB: Les auteurs recommandent d'utiliser en parallèle le NanoDrop et le Qubit (moins coûteux) ou le NanoDrop et la qPCR (plus coûteux), car ni le Qubit ni la qPCR ne permet la détection de contaminants présents dans les ADN purifiés. Les ratios A260/A230 et A260/A280 obtenus par le NanoDrop sont de bons indicateurs de pureté.

II. Méthodes d'analyse du polymorphisme de l'ADN: marqueurs moléculaires

Le mot de **polymorphisme** signifie « plusieurs formes ».

Le polymorphisme génétique est l'existence, dans une population, de plusieurs états alternatifs de l'ADN, ou **allèles**, en une position définie du génome, ou *locus*.

Au sein du génome, différents phénomènes plus ou moins complexes sont responsables du polymorphisme:

1. Des mutations dues à un remplacement d'une ou plusieurs bases par d'autres sans le changement de la longueur de l'ADN (substitution)
2. Des insertions ou des délétions correspondant à l'ajout ou à la disparition de bases dans la séquence d'ADN (changement de la longueur de l'ADN) (pour le RFLP: perte ou gain d'un site de restriction)
3. Des réarrangements chromosomiques dus à des recombinaisons génétiques ou à l'insertion de transposons. Les recombinaisons génétiques sont la conséquence d'une cassure de l'ADN suivie d'un échange de matériel entre deux chromosomes.

Pour étudier le polymorphisme, des marqueurs morphologiques, biochimiques et moléculaires sont utilisés.

Les marqueurs morphologiques (phénotypiques): Caractères morphologiques ou phénotypiques (couleur des poils, capacité à pousser en milieu carencé, présence ou absence de taches foliaires, etc...). Ce sont des caractères faciles à mesurer mais peu polymorphes, rarement co-dominants (généralement dominants), sont influencés par l'environnement, sont dépendants souvent du stade de développement de la plante et souvent ne sont pas reproductible sous différents environnements.

Un marqueur moléculaire est un locus **polymorphe** qui renseigne sur le génotype de l'individu qui le porte. Ils permettent à la fois un diagnostic extrêmement **fin** de la variabilité et la mise en place de stratégies très rapides de création et sélection variétale.

Ces marqueurs présentent également différents avantages comparés aux autres marqueurs:

- très nombreux et très polymorphes
- neutres vis à vis de la sélection,
- couvrent le génome entier,
- indépendants de la partie de la plante prélevée, du stade de développement et des fluctuations environnementales.

Un marqueur génétique peut être défini comme **un gène ou une séquence polymorphe** d'ADN, facilement détectable grâce à son emplacement connu sur le chromosome. Ces marqueurs peuvent être utilisés pour différencier les individus, les espèces, les genres, les variétés...

Un marqueur moléculaire, devrait idéalement présenter les caractéristiques suivantes:

1. Être **polymorphe**, c'est à dire posséder plus d'un allèle dans la population étudiée (variable entre individus).
2. Être **co-dominant**, ce qui signifie qu'un hétérozygote peut être différencié des homozygotes c'est-à-dire absence d'interaction intra locus.
3. Non **épistatique**, c'est-à-dire que son génotype peut être lu à partir de son phénotype quel que soit le génotype des autres locus (absence d'interaction inter locus).
4. Être **neutre**, c'est-à-dire que son polymorphisme est sans effet sur la variabilité phénotypique.
5. Être **insensible au milieu**, c'est-à-dire le génotype peut être inféré à partir du phénotype quel que soit le milieu.
6. **Reproductible** d'une expérience à une autre

Les marqueurs moléculaires peuvent se diviser en deux grands groupes, selon leur principe:

- Les marqueurs basés sur l'hybridation moléculaire: RFLP
- Les marqueurs basés sur l'amplification enzymatique *in vitro* (la PCR): SSR, AFLP, RAPD

1. Les marqueurs basés sur l'hybridation moléculaire: RFLP pour Restriction Fragment Length Polymorphism ou Polymorphisme de longueur des fragments de restriction

Le principe de base de cette méthode repose sur la comparaison de profils de coupure par les enzymes de restriction suite à l'existence d'un polymorphisme dans la séquence d'une molécule d'ADN par rapport à une autre. L'ADN des individus à comparer est donc digéré par une ou plusieurs enzymes de restriction. Les produits de restriction sont ensuite séparés par électrophorèse sur gel d'agarose ou d'acrylamide en présence de marqueurs de taille ou de poids.

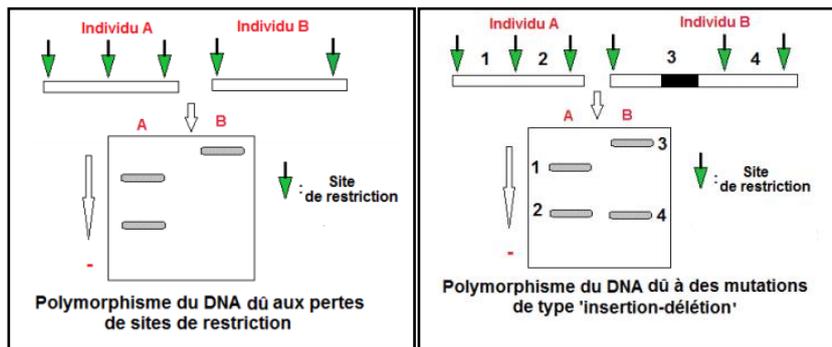
Le polymorphisme observé est utilisé comme critère d'identification

1. Les causes du polymorphisme dans les RFLP

Les causes du polymorphisme sont :

- perte ou gain d'un site de restriction
- mutation de type insertion-délétion

Si deux individus diffèrent par un ou plusieurs sites ou, même, par la distance séparant deux sites identiques consécutifs, il se crée une différence dans la longueur des fragments générés par l'enzyme de restriction.



Après électrophorèse, une hybridation est nécessaire (Southern blotting).

Southern Blotting: Hybridation moléculaire ADN/ADN (Blotting = Transfert)

- Elle consiste à détecter **spécifiquement** des fragments d'ADN transférés sur filtre par leur hybridation à **une sonde** (petit morceau d'ADN synthétisé dont la séquence correspond au moins en partie à celle d'un des brins de l'ADN).
- En marquant cette sonde par un atome radioactif (P32) ou par une molécule fluorescente liée à un des nucléotides de la sonde, on peut détecter la sonde et le morceau de l'ADN complémentaire qui lui est hybridé.
- Les ADN cibles sont dénaturés puis transférés par capillarité sur une membrane de nitrocellulose ou de nylon.
- Les fragments séparés sont hybridés avec un ADN sonde.
- Une fois transféré sur une membrane de nylon ou de nitrocellulose et hybridé par un ADN sonde, l'ADN cible digéré est visualisé par autoradiographie.
- Après mise en contact de la membrane avec la sonde marquée dans des conditions favorables à l'hybridation, des lavages sont effectués afin de faire apparaître après

autoradiographie des bandes marquées correspondant à des fragments d'ADN complémentaires de la sonde.

2. Les marqueurs moléculaires basés sur la PCR

Le développement de la technique PCR (*Polymerase Chain Reaction*) offre l'avantage d'analyser les marqueurs moléculaires en un temps court tout en utilisant des concentrations faibles d'ADN. La réaction de polymérisation en chaîne a aussi l'avantage de ne pas dégrader définitivement l'ADN, contrairement à la RFLP.

Une PCR classique se déroule dans un petit tube lui même placé dans un appareil programmable appelé thermocycleur. Ce dernier se contente de placer le tube aux températures voulues pendant les durées programmées et de recommencer en effectuant des cycles. Un cycle reproduit **trois températures différentes** pendant des durées différentes. La durée d'un cycle est de l'ordre de la minute. Après n cycle, on obtient 2^n molécules d'ADN identiques.

Avant la réaction, tous les « acteurs » de la PCR sont introduits dans le même tube. Il s'agit de:

- l'ADN à amplifier,
- des oligonucléotides (ou amorces), spécifiques du segment d'ADN voulu (on parle de couple d'amorces)
- de l'ADN polymérase (ici la Taq polymérase)
- Les ions Mg^{2+} pour maintenir stable le pH du milieu
- et enfin du mélange des quatre désoxyribonucléotides constitutifs de l'ADN.

Les étapes de la PCR:

- **Dénaturation** : l'ADN à amplifier est dénaturé à $95^{\circ}C$, les deux brins se séparent et servent de matrice.
- **Hybridation des amorces** : on baisse la température ($50-60^{\circ}C$) pour permettre à l'amorce de s'hybrider spécifiquement aux extrémités de la séquence cible à amplifier.
- **Elongation ou polymérisation à $72^{\circ}C$** : à cette température optimale de l'activité de l'enzyme Taq polymérase, l'amorce est allongée par complémentarité au brin matrice, ce qui produit deux nouvelles molécules d'ADN.

Il faut un couple d'amorces : **Forward et Reverse**

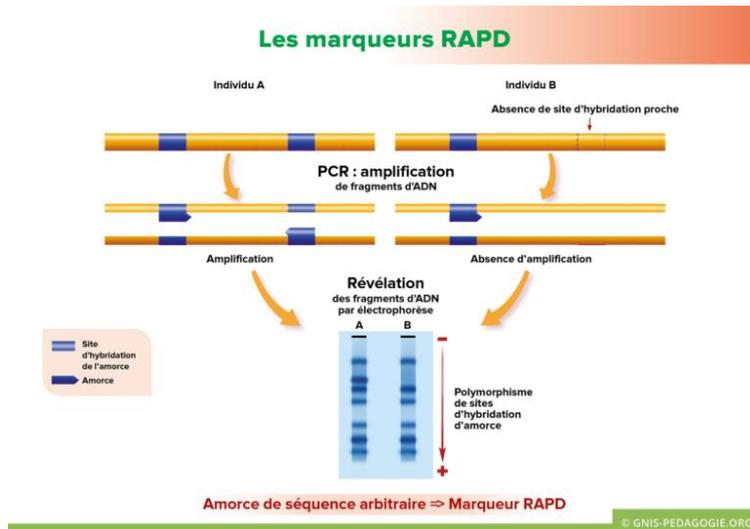
Chaque amorce doit :

- Contenir 17 à 28 nucléotides (environ 20)
- % GC: plus de 50 % et moins de 68 %
- Extrémité 3': de préférence se termine par GC/CG/CC/GG
- Des séquences qui ne soient pas complémentaires entre elles
- Eviter les palindromes
- Equilibre des bases : 20-30 % G; 20-30 % C; 20-25 % A; 20-25 % T.
- T_h (température d'annealing) comprise entre 55 et $65^{\circ}C$

$$T_h = T_m - 5 \quad / \quad T_m = 4(G+C) + 2(A+T)$$

2.1.RAPD Random Amplified Polymorphism DNA: Polymorphisme d'amplification aléatoire de l'ADN

Cette technique consiste à réaliser une PCR en utilisant une amorce courte d'une dizaine de nucléotides, de séquence **arbitraire**. Cette amorce va s'hybrider **au hasard** dans le génome. C'est une technique qui ne nécessite pas obligatoirement de connaître la séquence de l'ADN cible. Les amorces utilisées seules ou par couples. Lorsqu'une seule amorce aléatoire est utilisée, l'amplification a lieu une fois, celle ci se fixe sur 2 sites complémentaires peu distants (3000 pb au max)



Le polymorphisme révélé est un polymorphisme de sites d'hybridation d'amorce. Les amorces constituent donc les marqueurs. Pour l'ensemble du génome, une dizaine de fragments sont amplifiés en moyenne puis séparés par électrophorèse. Cette méthode est rapide et d'une faible technicité. Toutefois, sa reproductibilité est difficile à obtenir.

2.2.AFLP (Amplified Fragment Length Polymorphism) Polymorphisme d'amplification de fragment d'ADN

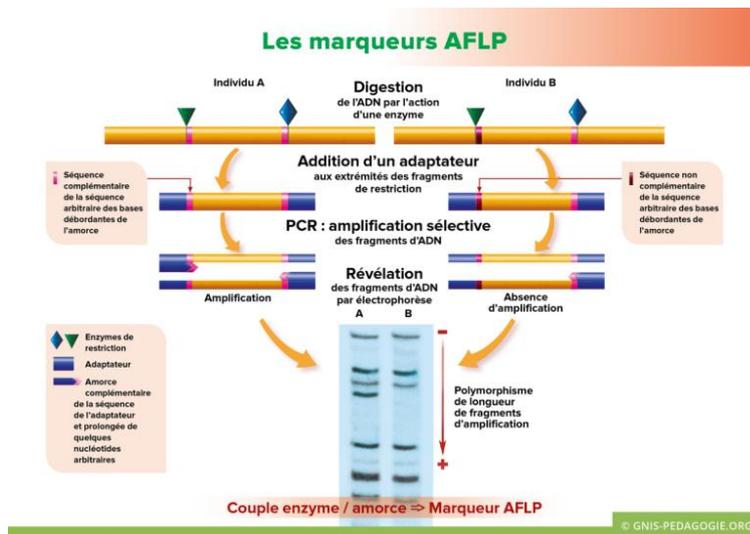
La technique AFLP (Amplified Fragment Length Polymorphism) est fondée sur la mise en évidence conjointe de polymorphisme de sites de restriction et d'hybridation d'amorces arbitraires.

Cette technique utilise à la fois les enzymes de restriction et l'amplification PCR. On distingue les étapes suivantes :

- L'ADN génomique est coupé par deux enzymes de restriction.
- Des adaptateurs de **séquences connues** et spécifiques des enzymes de restriction utilisées, sont ajoutés aux extrémités des fragments de restriction générant ainsi une matrice pour l'amplification.
- Une première amplification (pré-amplification) est réalisée à l'aide d'amorces de séquences complémentaires à la séquence des adaptateurs et des sites de restriction.
- La deuxième amplification (amplification sélective) utilise des amorces identiques aux premières mais **prolongées** à l'extrémité 3' de quelques nucléotides arbitraires (de 1 à 3 nucléotides).

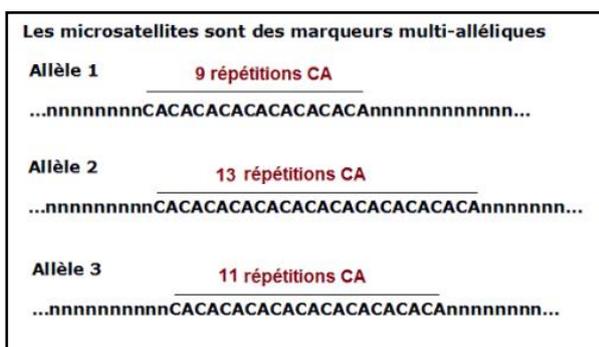
- Ainsi, seuls sont amplifiés les fragments possédant les bases complémentaires de ces bases arbitraires. Ces amorces sélectives permettent de réduire le nombre de fragments amplifiés à une centaine.
- Ces fragments sont séparés par électrophorèse sur gel d'acrylamide dénaturant puis visualisés par coloration au nitrate d'argent ou révélés grâce à un marquage radioactif ou fluorescent réalisé lors de l'amplification sélective.

C'est la combinaison enzyme de restriction/amorce qui permet de révéler le polymorphisme entre les individus.



2.3. Marqueurs de type 'microsatellites' (SSR : Simple Sequence Repeats)

C'est le **Polymorphisme de nombre d'unités de répétition**. Sur le génome, il existe des séquences constituées d'unités répétées de 1 à 4 nucléotides. Ce sont les microsatellites. Les plus courants sont (A)_n, (TC)_n, (TAT)_n et (GATA)_n, les valeurs de n pouvant aller de quelques unités à plusieurs dizaines. On parle de séquences répétées en tandem. L'intérêt de ces microsatellites réside dans leur polymorphisme. Celui-ci repose sur la variation du nombre d'unités de répétition.



Les polymorphismes sont identifiés en construisant des amorces PCR dans la région adjacente de la région microsatellite. C'est la paire d'amorces spécifique des bordures droite et gauche du microsatellite qui constitue le marqueur. Il faut connaître, synthétiser et tester les amorces bordant le microsatellite. Etant donné qu'il n'existe que de faibles différences de tailles sur les microsatellites, ces marqueurs sont souvent séparés par électrophorèse sur gel de polyacrylamide (au lieu de l'agarose). Ces marqueurs ont été utilisés pour des études de la diversité génétique. Ainsi que dans l'élaboration des cartes génétiques des plantes.

Les SSR constituent de bons marqueurs moléculaires (reproductibles, codominants et aisés d'utilisation).

II. La sélection assistée par marqueurs (SAM) :

La SAM est basée sur l'utilisation **de marqueurs moléculaires**, de petits segments d'ADN situés près **d'un gène d'intérêt** (résistance à une maladie, par exemple) dans l'ADN de la plante. Les marqueurs sélectionnés et utilisés au laboratoire permettent de détecter précisément la présence ou l'absence du (ou des) gène(s) d'intérêt recherché(s) dans les plants en cours de sélection.

La SAM présente également un intérêt important dans les programmes d'**introgression** d'un gène d'intérêt dans une variété élite par **retrocroisements**.

Les marqueurs moléculaires sont souvent utilisés pour conduire cette introgression car ils permettent de réduire les temps de sélection.

Avec l'aide des marqueurs moléculaires, 4 rétrocroisements suffisent pour arriver au même résultat car on peut à chaque génération choisir les plantes ayant recombiné le plus petit segment chromosomique.

Quels sont les avantages de la sélection assistée par marqueurs ?

- **Rapide** (quelques jours d'analyses au laboratoire suffisent pour établir un diagnostic de la présence ou de l'absence des gènes d'intérêt dans les plants en cours de sélection). La SAM permet de trier rapidement un grand nombre de plantes à un stade précoce du processus de sélection afin de recommencer rapidement un nouveau **cycle de sélection**.
- **Précise** (détection fiable de la présence ou de l'absence du gène recherché) ;
- **Non destructive** car l'analyse des plants est réalisée sur une très faible quantité de matériel végétal.
- **Non influencée par des facteurs environnementaux**.
- Cette méthodologie de sélection permet de cumuler le plus vite possible le maximum de gènes favorables dans un même génotype et ainsi de réduire considérablement le temps nécessaire à l'introduction de nouvelles variétés.
- La Sélection Assistée par marqueurs peut être couplée à toute technique de sélection de plante comme l'amélioration conventionnelle, la mutagenèse, la transgénèse.

La SAM présente cependant quelques **limites** :

- Elle n'est **pas compétitive en termes de coût et de temps** par rapport aux méthodes de sélection traditionnelles, **lorsque le phénotype peut être déterminé facilement** (hauteurs des plantes, précocité, résistance à certaines maladies dont l'observation est facile...) ;
- Elle est souvent **inefficace** pour la sélection de **caractères agronomiques à déterminisme génétique complexe** (comme le rendement, par exemple), gouvernés par un grand nombre de gènes ou de QTL (Quantitative Trait Loci).

III. Introduction à l'analyse des QTLs et cartographie

Les caractères d'intérêts agronomiques en grande partie sont des caractères quantitatifs dits polygéniques qui sont plus complexes en comparaison avec les caractères mendéliens.

Un Qtl (Quantitative Trait Locus) ou Locus de caractère quantitatif peut être défini comme une région chromosomique contenant un ou plusieurs gènes affectant un caractère **quantitatif**. Les QTL sont étudiés tout particulièrement chez les plantes : les marqueurs génétiquement liés à des QTL permettent de sélectionner, parmi un grand nombre d'individus ceux qui sont les plus performants.

Les marqueurs permettent ainsi, l'élaboration de **cartes génétiques** où chaque chromosome est représenté par un ensemble de marqueurs moléculaires dont l'ordre et l'espacement sont déterminés en comparant les individus de la descendance d'un croisement. On parle de la **cartographie**.

La cartographie génétique, basée sur l'analyse de ségrégation, consiste à positionner de façon relative des locus, utilise les propriétés de la recombinaison à la méiose. Le principe est de localiser un Qtl.

NB : Les cartes génétiques reposent sur l'évaluation de la distance relative séparant des caractères héréditaires ou marqueurs génétiques. La distance génétique entre deux gènes se définit par la fréquence d'apparition d'événements de recombinaisons entre ces gènes d'une génération à l'autre : 1% de recombinaison correspondant à une distance de 1 centi-Morgan

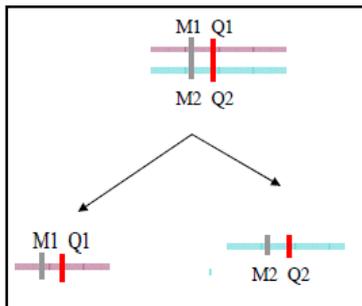
En premier, comment sont détectés les Qtls ?

La recherche de gènes a été rendue possible grâce au développement des cartes génétiques, car la ségrégation de marqueurs moléculaires dans la population est mise en relation avec la variabilité des caractères pour rechercher des zones du génome portant les gènes affectant ces caractères quantitatifs (QTL).

On peut déduire la présence d'un *QTL* quand sont observées des différences phénotypiques significatives entre des individus ayant hérité différents allèles, liées à des marqueurs, de leurs parents.

En effet, puisque les allèles des QTL ne sont pas identifiables (leur position est à *priori* inconnue), l'information disponible devrait venir de marqueurs moléculaires proches (dont les localisations sont connues à partir de la carte génétique).

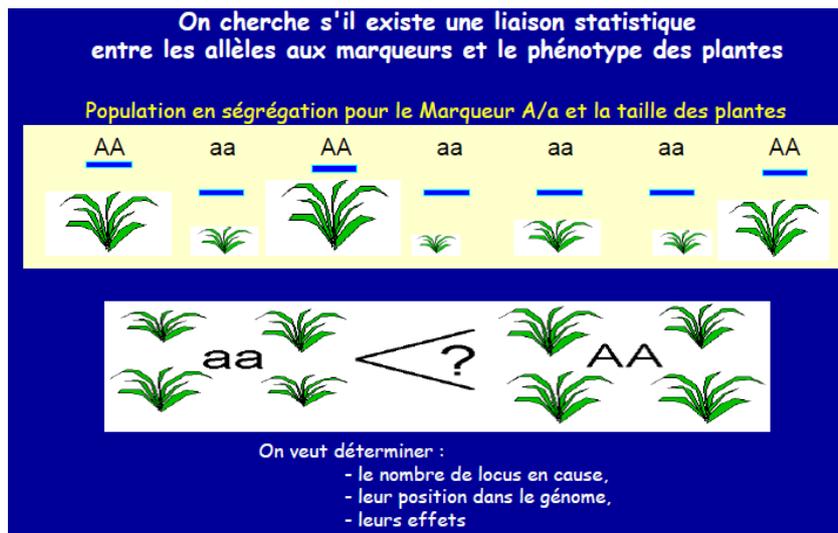
L'idée est donc de tester si, pour un groupe de marqueurs qui sont dans une zone spécifique du génome, la distribution des phénotypes est la même quel que soit leur génotype.



Si on considère un locus marqueur M très proche du locus du gène quantitatif Q , la probabilité de recombinaison entre les deux locus est très faible: ils seront transmis ensemble lors de la méiose. Ainsi, l'allèle $Q1$ sera transmis avec l'allèle $M1$. En d'autres termes, M « marque » la transmission sans recombinaison du segment qui l'entoure et de tous les gènes qu'il contient.

Pour un marqueur donné, le principe de la détection de QTL est d'observer dans la descendance d'un parent hétérozygote pour un marqueur donné ($M1/M2$), s'il existe une différence entre les moyennes des phénotypes selon l'allèle marqueur $M1$ ou $M2$ transmis. Si cette différence existe, elle suggère la ségrégation des allèles en un QTL ($Q1$ et $Q2$) lié à ce marqueur.

L'idée est que, si cette différence existe, elle s'explique par la ségrégation des allèles, $Q1$ ou $Q2$, en un QTL génétiquement lié au marqueur.



Modèles et outils de recherche des QTL

Le principe de détection des QTL se base sur la recherche systématique de déséquilibres de liaisons entre les locus marqueurs et locus contrôlant les caractères quantitatifs.

Des marqueurs dont la ségrégation des allèles dans la descendance fait ressortir une liaison statistique significative avec la valeur phénotypique d'un caractère donné chez les individus.

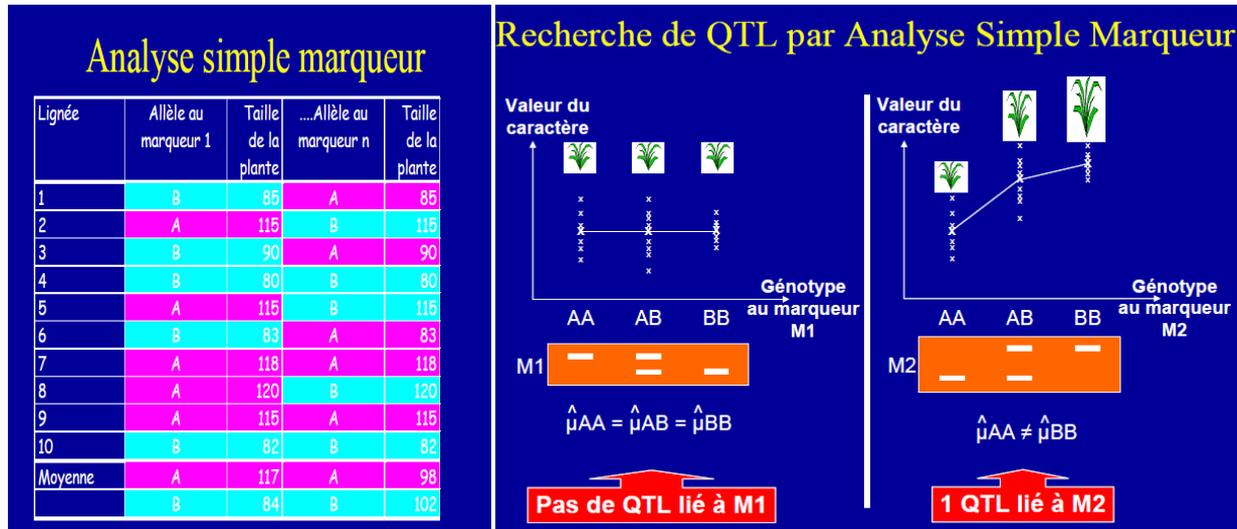
Pour cela, plusieurs méthodes sont utilisées:

1. Méthode de détection marqueur par marqueur (« single marker »)

C'est une méthode qui a été initiée en 1976. L'analyse est réalisée en modèle linéaire qui permet de rechercher par analyse de variance ou régression multiple, une relation entre le génotype de chaque locus marqueur et la valeur phénotypique du caractère quantitatif considéré.

En effet, pour chaque marqueur, il est possible de classer les individus selon l'allèle qu'ils portent (classes génotypiques) et calculer par la suite les moyennes pour chaque groupe.

Les différences significatives, si elles existent, entre ces moyennes seront déterminées suite à une analyse de variance. L'explication de l'effet du génotype au marqueur passe par la présence d'un QTL polymorphe génétiquement lié à ce marqueur.



2. Méthodes de cartographie d'intervalle (« interval mapping »)

Elle se base sur l'hypothèse qu'il y a un QTL au plus dans l'intervalle qui sépare deux marqueurs, liés avec un taux de recombinaison r . Les distances entre marqueurs sont connues et la position du QTL par rapport aux marqueurs est alors déterminée.

On considère qu'un QTL existe entre deux marqueurs co-dominants (M et N), et on se place dans le cas d'un croisement de deux lignées pures l'une $M1Q1N1 / M1Q1N1$ et l'autre $M2Q2N2 / M2Q2N2$. Un F1 où tous les individus sont $M1Q1N1 / M2Q2N2$ est produit.

Cette méthode prend en considération les fréquences des recombinaisons entre le QTL et les deux marqueurs.

Comme dans l'approche précédente, la moyenne de chaque classe de marqueur dans les individus F2 peut être dérivée, et ces contrastes entre moyennes de chaque classe de marqueur peuvent être établis pour démontrer ou non la présence d'un QTL entre les deux loci marqueurs.

Il y'a deux types de cartographie par intervalle: cartographie **d'intervalle simple (Simple Interval Mapping, SIM)** et **cartographie d'intervalle composite (Composite Interval Mapping, CIM)**.

La cartographie des QTL nécessite donc 4 éléments :

- 1) des marqueurs moléculaires polymorphes couvrant, si possible, tout le génome pour construire une carte génétique et pour établir des liaisons avec des QTL,
- 2) Des mesures d'un caractère phénotypique recherché,
- 3) une population d'étude en ségrégation pour le caractère recherché et les marqueurs moléculaires,
- 4) des outils statistiques pour trouver des associations entre les états alléliques liées aux marqueurs et les valeurs du caractère phénotypique.

Une carte génétique est élaborée par l'ordonnement de marqueurs grâce à l'analyse statistique de leur ségrégation au cours des générations. Elle permet ainsi de situer dans l'ordre les positions relatives des marqueurs les uns par rapport aux autres. Sa construction s'appuie sur l'estimation du

taux de recombinaison entre tous les locus pris deux à deux. L'intérêt des cartes génétiques est de positionner sur les chromosomes des locus d'intérêt agronomique, voire identifier les gènes

Les principaux critères de l'élaboration d'une carte génétique sont :

1. Développer une population (de 50 à plusieurs milliers d'individus) de cartographie appropriée, souvent issue d'un croisement entre deux lignées pures (F2, lignées recombinantes consanguines (RILs), des lignées consanguines de rétrocroisement ou back-cross, lignées haploïdes doublées HD,...) ;
2. choisir un type de marqueurs moléculaires adéquat ;
3. rechercher le polymorphisme chez les deux parents avec les marqueurs moléculaires choisis puis faire le génotypage de la population de cartographie avec ces marqueurs ;
4. faire des analyses de liaisons qui se basent sur le calcul des fréquences de recombinaisons entre marqueurs, établir les groupes de liaison, ensuite estimer l'ordre et les distances des

IV. Séquençage des macromolécules

Le séquençage est un procédé utilisé pour déterminer l'ordre (la séquence) des acides aminés d'une protéine ou des bases dans les acides nucléiques (ADN et ARN).

Les techniques de séquençage peuvent être divisées selon leur apparition en :

- Première génération (méthode Sanger et méthode Gilbert)
- Deuxième génération (NGS: Next generation sequencing): Pyroséquençage 454 de Roche, Solexa d'Illumina et Solid d'Applied Biosystems .
- Troisième génération (NNGS: Next Next generation sequencing): HeliScope d'Helicos, Technologies des nanopores , Plateforme de Pacific Biosciences et Starlight .

1. La technique Sanger

C'est une méthode très utilisée qui donne de bons résultats mais elle est lente. Elle est basée sur l'interruption de la synthèse enzymatique d'un brin d'ADN complémentaire: arrêt d'élongation.

L'ADN à séquencer est cloné (de nombreuses copies simple brin sont produites).

Une courte amorce d'oligonucléotides (**marquée**) est ajoutée à l'ADN. Le point de fixation de l'amorce sert de point de départ pour la synthèse du brin complémentaire.

La polymérase est ajoutée avec :

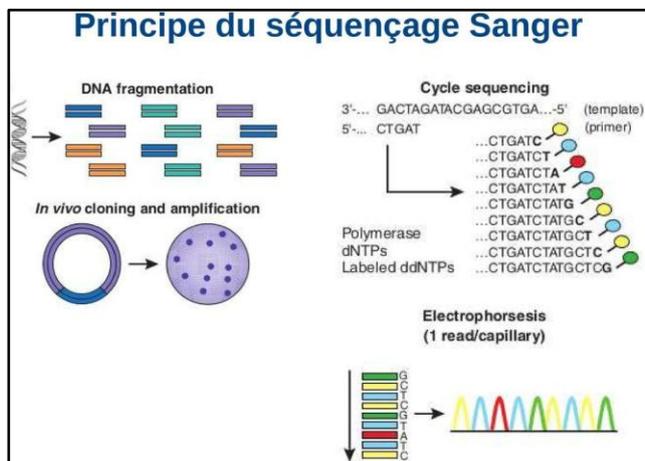
- Les 4 nucléotides normaux : d-ATP, d-CTP, d-GTP et d-TTP (dont un est marqué au P32, ou au S35) ;
- une faible concentration de 4 nucléotides analogues (des didésoxynucléotides: ddNTP) dans des incubations séparées. Les ddNTP sont identiques aux nucléotides normaux sauf que les groupes hydroxyles (OH) des riboses sont remplacés par des hydrogènes (H). La polymérase ne peut pas distinguer ce substrat des nucléotides normaux.

L'intégration d'un didésoxynucléotide dans le brin synthétisé entraîne **l'arrêt de l'élongation en raison de l'absence du groupement OH, nécessaire à l'extension.**

Les 4 incubations contiennent donc un mélange de molécules partiellement synthétisées d'ADN double brin marqué. La longueur des fragments d'ADN varie en fonction du point d'intégration du didésoxynucléotide. Comme cette intégration est aléatoire, l'ensemble des molécules dans un mélange représente l'ensemble des positions pour une base particulière. Les 4 mélanges sont analysés simultanément sur un gel d'électrophorèse qui contient un composé qui entraîne la dénaturation de l'ADN et le processus est mené sous un voltage fort pour éviter la réassociation des brins. Les bandes sont révélées par autoradiographie et la séquence est lue directement sur le gel.

Une adaptation de cette technique consiste à marquer les didésoxynucléotides (Chaque didésoxyribonucléotide est marqué par un fluorophore spécifique), ce qui permet de n'utiliser qu'un seul mélange au lieu des quatre nécessaires dans le cas d'un marquage des amorces. Les fragments d'ADN synthétisés portent ce fluorophore terminal. On les appelle des terminateurs.

Dans ce cas on parle d'un **séquençage par arrêt de synthèse** à l'aide d'un terminateur marqué.

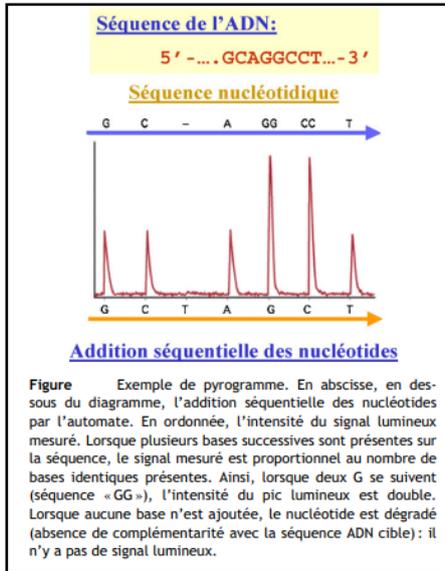


2. Le Pyroséquençage

Cette technique a été publiée pour la première fois en 1998 alors que le principe a été décrit en 1985. Elle permet d'analyser la synthèse d'ADN cible en temps réel. **On parle de séquençage par synthèse d'ADN.** Son principe consiste à hybrider une amorce à l'ADN cible (amplifié par PCR), puis à ajouter séquentiellement et dans l'ordre une base à partir de l'extrémité 3 de l'amorce. Chaque base est marquée par un fluorophore différent dont le signal est mesuré par bioluminescence à condition que la base complémentaire de la cible soit incorporée.

La séquence est déduite en fonction de l'ordre d'incorporation des nucléotides sur l'ADN complémentaire de la cible néosynthétisée.

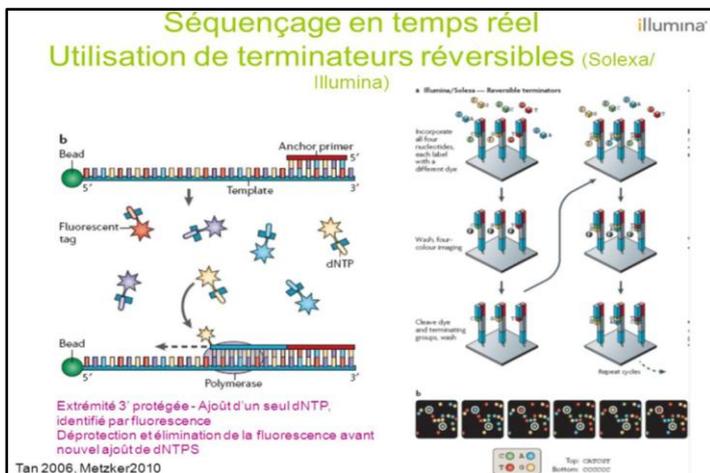
Une caméra CCD mesure le signal de bioluminescence produit. Le pyroséquençage, du fait de la faible longueur des bases pouvant être lues, a eu longtemps une indication limitée dans le séquençage. L'arrivée d'automates, tel que le 454 permettant le séquençage massif parallèle, a permis le développement de cette technique à une grande échelle.



3. Séquençage en temps réel: Utilisation de terminateurs réversibles (Solexa/ Illumina)

Cette technique se déroule selon les étapes suivantes :

- Génération d'une banque d'ADN double brin à partir de l'échantillon à analyser par fractionnement aléatoire en morceaux de 200 pb.
- Ajout d'adaptateurs spécifiques aux extrémités.
- Dénaturation de l'ADN en simple brin.
- Fixation de l'extrémité des simples brins aléatoirement à la surface du « flowcell ».
- PCR « bridge » en phase solide. Création d'un double brin. Dénaturation et création de groupes (clusters) denses où les fragments sont amplifiés.
- Le premier cycle de séquençage commence en ajoutant les 4 terminateurs réversibles marqués, les amorces et l'ADN polymérase.
- Après excitation par un laser, la fluorescence émise par chaque cluster est récupérée et la première base est lue.
- Le cycle suivant continue en ajoutant les 4 terminateurs réversibles marqués. • Après excitation l'image est acquise de la même façon et la deuxième base est lue.
- Les cycles de séquences sont répétés pour lire chaque base les unes après les autres.
- La lecture est effectuée à chaque position sur toutes les séquences en parallèle.
- Les erreurs majeures de séquences proviennent d'erreur de séquençage (99%)



Comparaison de quelques méthodes

	Sanger	454/ Roche GS FLX Titanium	Illumina/ Solexa Genetic Analyser	Life:APG SOLiD	SMRT/ Pacific Biosciences
Clonage	Oui	Non	Non	Non	Non
Technique	ddNTPs	Pyro- séquençage	Termineurs réversibles	Ligation	Polymérisation en direct Molécule unique
Lecture	800 bases/ lecture 384 lectures/ run 270 kb/run/3j	350 à 600 bases/lecture 800.000 lectures/run 0.45 Gb /run/ 8h	75 ou 100 bases/lecture plus M lectures/run 14 Gb/run/ 4j	75 bases/ lecture 100 à 200 M lectures/run 30 Gb/run/ 7j	1000 à 3000 bases/lecture et plus à venir 75000 lectures/ run XX Gb / run / 30 min,
Avantages	Fiabilité et taille des fragments	Taille des fragments (seq rep.) et rapidité	Plateforme la plus utilisée Rapidité	Encodage/ 2 bases limite erreurs	Grand potentiel pour de gros fragments
Incon- vénients	Temps et coût	Coût réactifs Erreurs / homo- polymères Indels	Longueur des lectures	Un peu plus long	Grand taux d'erreurs

Stratégie de séquençage d'une protéine

Les étapes impliquées dans le séquençage d'une protéine sont:

- Détermination du *nombre de chaînes polypeptidiques (sous-unités)* dans la protéine
- Coupure des *liens disulfures (inter- ou intramoléculaires)*
- Détermination de la *composition des acides aminés de chaque chaîne* polypeptidique
- Les chaînes polypeptiques sont souvent trop longues pour être directement séquencées; elles doivent avant être coupées en des plus *petits fragments peptidiques par des réactions spécifiques*.
- Détermination de la séquence de chacun des fragments peptidiques par la *dégradation d'Edman*
- Éluclidation de la séquence de chacune des chaînes polypeptidiques par *recouvrement des séquences des différents fragments*
- Détermination de la structure de la protéine entière, incluant les liens disulfures entre les sous-unités