

### 1- Introduction

Différentes méthodes permettent d'analyser le génome, elles sont classées en trois grandes catégories:

- **l'analyse de la structure physique** : vise à décrire l'organisation de l'ADN en termes de distance physique, en termes de succession de différents types de séquences, et en termes de distribution et d'empaquetage de l'ADN du génome au sein d'un ensemble de chromosomes. La structure physique est mesurée par des distances exprimées en nanomètres (nm) ou en paires de bases (pb).
- **l'analyse de la structure génétique** : vise à comprendre comment les gènes sont distribués le long des chromosomes, comment ils ségrégent et se recombinent à chaque génération. L'analyse de la structure génétique conduit à établir des cartes génétiques des chromosomes sur lesquelles les distances sont exprimées en unités de recombinaison (centimorgan = cM) qui sont en fait le reflet de l'aptitude à recombiner.
- **l'analyse de l'expression du génome** : consiste à analyser les produits de transcription ou transcriptomes, ainsi que les produits de traduction ou protéomes. Le résultat de l'activité des enzymes représentées dans un protéome sur une variété de substrats produit tout un ensemble de molécules organiques qui constituent le métabolome.

### 2- Les différents niveaux de résolution de la structure physique du génome

Ils permettent de décrire la structure physique du génome nucléaire des plantes, ces méthodes recouvrent :

- la cytogénétique,
- l'organisation de la chromatine.
- les méthodes de fractionnement et de cartographies de restriction

#### 2-1 Etude cytogénétique

Le premier niveau d'étude d'un génome est l'observation microscopique. Elle est réalisée à des stades du cycle cellulaire où les chromosomes sont individualisés (plaque métaphasique des noyaux en mitose ou condensés (stades pachytènes ou diplotènes de la première division méiotique) on peut alors dénombrer les chromosomes, mesurer leur taille, observer leur morphologie, localiser la constriction centromérique ou l'organisateur nucléolaire. La collection de chacun des chromosomes d'une cellule constitue le caryogramme.

Les différentes régions stratégiques du chromosome peuvent alors être décrites: télomère à chaque extrémité, centromère dont la position va déterminer le type chromosomique. Ces premières observations révèlent une hétérogénéité structurale des chromosomes. Le nombre de

chromosomes chez les végétaux est une caractéristique spécifique pour chaque espèce, ce nombre est toujours paire puisque les plantes sont des organismes diploïdes ou polyploïdes.

On peut mesurer la quantité d'ADN par noyau en utilisant soit des méthodes d'extraction suivies de dosage, soit des méthodes de cytophotométrie (évaluation de la teneur en ADN d'un échantillon sur des noyaux isolés ou la cytométrie en flux) ou sur des préparations microscopiques, après coloration spécifique de l'ADN (réactif de Schiff, DAPI, iodure de propidium) et comparaison avec un standard interne. La quantité d'ADN d'un génome est appelée la C value. La quantité d'ADN par noyau est très variable d'une espèce à l'autre (Tableau I) et cette variation est sans rapport avec la complexité génétique (nombre de gènes).

**Tableau I :** Taille des génomes en millions (pb)

Espèce végétale	Taille du génome (million de pb)
<i>Arabidopsis thaliana</i>	130
<b>Riz</b>	430
<i>Medicago truncatula</i>	650
<i>Brassica oleracea</i> (chou)	700
<b>Sorghon</b>	800
<b>Tomate</b>	1000
<b>Colza</b>	1200
<b>Maïs</b>	2500
<b>Orge</b>	52000
<b>Pois</b>	5000
<b>Blé</b>	16000
<b>Tulipe</b>	30000

La cytogénétique permet aussi de révéler des altérations de la structure d'un génome (translocations, inversions, délétion ou duplications).

## 2-2 Organisation de la chromatine

La chromatine correspond à la substance nucléaire qui fixe les colorants basiques. Dans le noyau interphasique, elle représente un mélange de fibres irrégulières dont le diamètre est variable en fonction du degré d'empilement. L'unité structurale de la chromatine est le **nucléosome** et représente le premier niveau de compaction de l'ADN. Cette structure est ensuite régulièrement répétée pour former le **nucléofilament** qui peut, lui-même, adopter des niveaux

d'organisation plus compacts, le niveau de condensation le plus élevé étant atteint au sein du chromosome métaphasique.

A partir d'observations histologiques, il est apparu que la chromatine avait une structure hétérogène faite d'un enchevêtrement de fibres dont le diamètre varie non seulement au cours du cycle cellulaire mais aussi en fonction des régions chromosomiques observées. Au sein du noyau interphasique, la chromatine est organisée en territoires fonctionnels et divisée en euchromatine et hétérochromatine.

### 2-2-1 L'euchromatine (vraie chromatine)

L'euchromatine est la fraction de chromatine qui apparaît décondensée pendant l'interphase.

Il peut se présenter sous deux formes :

- Une forme active (transcriptionnelle) : Elle est alors constituée d'une fibre dont le diamètre est de 10 à 11 nm (diamètre du nucléosome).
- Une forme inactive : Elle s'enroule en un solénoïde sous la contrainte d'une histone (H1) qui va relier deux nucléosomes consécutifs. Son diamètre est de 30 nm. D'autres protéines, non histones interviennent pour replier la fibre de chromatine en boucles.

L'euchromatine est décondensée pendant l'interphase du noyau ; elle représente la majeure partie du génome et contient les gènes structuraux (gènes codant pour les protéines et les ARN non traduits). Elle est localisée à l'intérieur du nucléoplasme et représente la forme transcriptionnellement active de l'ADN.

### 2-2-2 L'hétérochromatine

L'hétérochromatine a été définie comme une structure qui ne change pas l'état de condensation au cours du cycle cellulaire et localisé principalement en périphérie du noyau et du nucléole on distingue :

- **L'hétérochromatine constitutive** qui contient peu de gènes, formée principalement de séquences répétées et dont les plus grandes régions sont situées à proximité des centromères et des télomères
- **L'hétérochromatine facultative** qui contient des régions codantes pouvant adopter les caractéristiques structurale et fonctionnelle de l'hétérochromatine, comme le chromosome X inactif chez la femelle des mammifères (Figure 1)

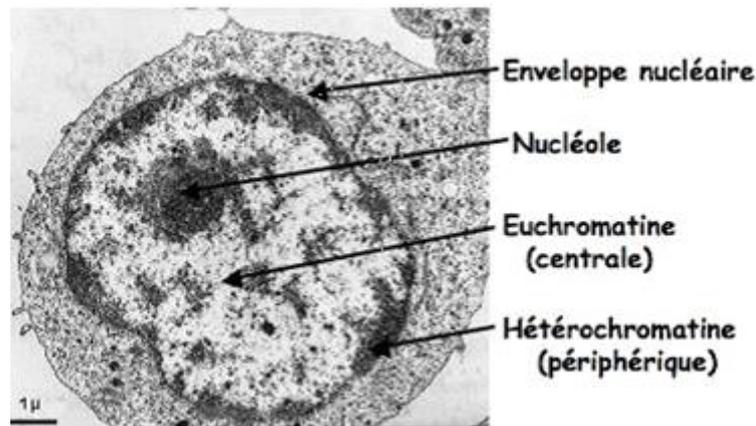


Figure 1 : localisation de la chromatine

### 2-3 Méthodes de fractionnement du génome : Séparation physique du génome

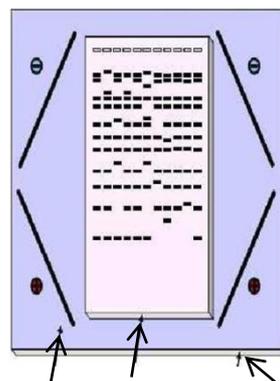
La plupart des êtres vivants contiennent des génomes fragmentés : chromosomes et plasmides bactériens chez les procaryotes et chromosomes nucléaires, génomes chloroplastiques et mitochondriaux chez les eucaryotes. La première étape d'étude d'un génome peut-être de le résoudre en ses composants. Parmi les méthodes physiques de séparation des génomes entiers ou fragmentés on trouvera :

#### 2- 3-1 Séparation en fonction de la taille

Cette séparation peut être s'effectuée par plusieurs méthodes :

- **Electrophorèse**

Les ADN de grande taille peuvent être séparés sur gel d'agarose soumis à un champ électrique alternatif (champ pulsé)



Électrode gel d'agarose plaque

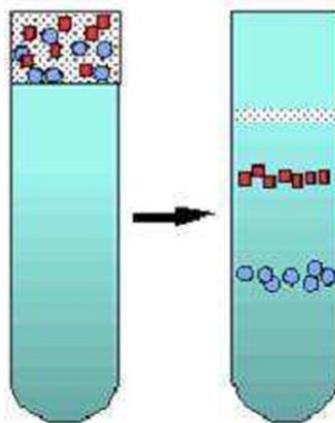
#### Figure N° 1 : Electrophorèse d'AND en champ pulsé

Les plus petits chromosomes d'eucaryotes supérieurs dont la taille dépasse la dizaine de Mb ne sont pas résolus par cette technique car ils n'entrent pas dans la maille du gel. Dans ce cas, il est cependant possible de fragmenter les chromosomes purifiés à l'aide d'endonucléases de restriction à sites rares (par ex. SfiI= GGCCNNNNGGCC, NotI=GCGGCCGC, ...) : les fragments générés sont alors séparables par cette technique.

- **-Gradients de sédimentation (saccharose)**

On a souvent recours un type de gradient: isopycnique utilisant un gradient discontinu de sucrose ; un coussin à haute teneur en sucrose (mettons 80%) se trouvera dans le fond du tube, avec un coussin de 50% par-dessus, puis un coussin de 25% puis un coussin de 10%.

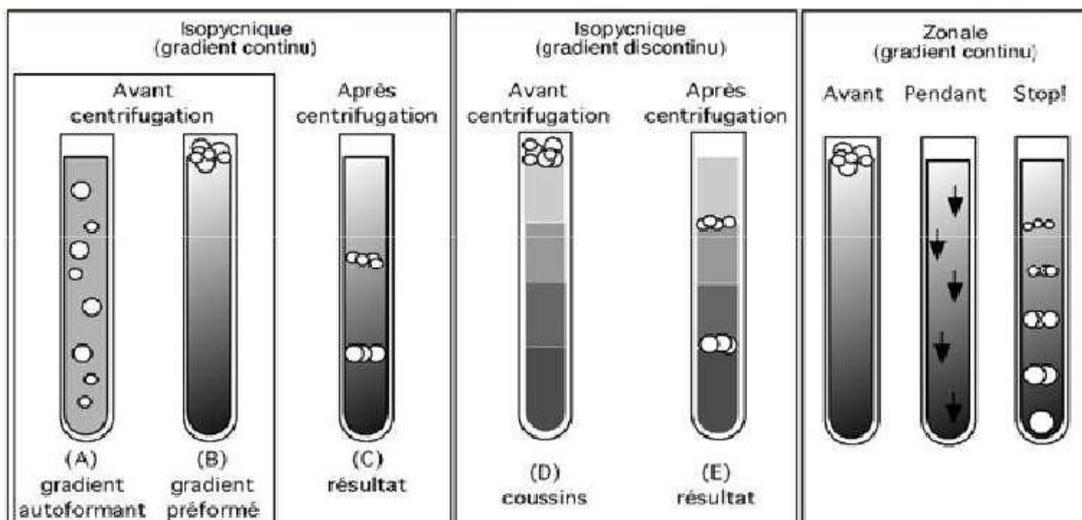
Les particules qui traversent un coussin pourraient être arrêtées à la surface du suivant.



**2-3-2 Séparation en fonction de la forme**

- **Gradients de densité (CsCl)**

Un autre produit couramment utilisé, particulièrement dans la séparation des acides nucléiques, est le chlorure de césium (CsCl). Ce sel peut atteindre une densité très élevée, de l'ordre de 1.9 g/ml à 7.5 g/ml. Cette possibilité d'atteindre des densités si élevées en solution aqueuse est son principal avantage.



**2-3-3 Séparation en fonction de la forme simple ou double brin**

Les acides nucléiques simple brin s'accrochent spontanément à la nitrocellulose. L'ADN double brin s'adsorbe à l'hydroxyapatite ( $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$ ), le simple brin ne s'y adsorbe pas.

Leur flexibilité étant très supérieure à celle d'une molécule double brin, ils ont tendance à se replier sur eux-mêmes pour former des régions en double hélice irrégulière, qui compactent la molécule. En gel d'acrylamide non-dénaturant, la vitesse de migration des polynucléotides simple brin dépend de leur longueur, mais aussi de la structure secondaire que prennent les molécules.

#### 2-3-4 Séparation en fonction de la composition en bases

- **Gradient de densité (CsCl)**

La densité des molécules d'ADN est fonction de leur composition en bases. Les ADN riches en GC sont plus denses que les ADN riches en AT. Il est possible ainsi de séparer les ADN mitochondriaux et nucléaires par centrifugation à l'équilibre sur un gradient de densité.

- **Trieurs de chromosomes**

Certains colorants ont une affinité pour l'ADN qui dépend de la concentration locale en bases particulières (séquences riches en AT par exemple). Les chromosomes, en solution, passent, un à la fois, devant un LASER à émission ultraviolette. Un détecteur mesure les intensités des émissions fluorescentes des colorants fixés à l'ADN (à des longueurs d'onde spécifiques de colorant). Chaque chromosome a un ratio et une intensité d'émission qui lui est spécifique.

Les gouttelettes de solution contenant le chromosome cherché peuvent alors être déviées par un champ électrique appliqué au moment de leur passage. Les deux signaux de fluorescence émis sont mesurés et, s'ils sont dans le bon rapport, un défecteur est activé et oriente la particule vers le collecteur.

### 3- La carte physique

La carte physique est l'ordonnement de fragments clonés chevauchants reconstituant la molécule d'ADN de départ. C'est à partir de cette carte que sera choisi l'ensemble minimal de fragments assurant la couverture complète du génome à séquencer. Les distances, mesurées en paires de bases (pb), entre les différents marqueurs sont dites absolues.

L'établissement d'une carte physique utilise les données des cartes de liaison et des informations de recouvrement partiel entre les fragments clonés permettant de définir des groupes de chevauchements. Selon les techniques de la cartographie physique, le niveau de résolution peut varier du chromosome entier à la paire de base. ; de ce fait il existe deux classes de cartographie physique:

- **La cartographie physique à faible résolution** dont la plus petite unité de carte pouvant être résolue varie habituellement entre une et quelques méga bases d'ADN.
- **La cartographie physique à haute résolution** dont la résolution est habituellement très grande, de quelques centaines de kilobases jusqu'à un seul nucléotide

#### 4- La carte génétique

C'est la détermination de la position d'un gène ou d'un marqueur génétique sur un chromosome en fonction du taux de recombinaison génétique. Cette fréquence de recombinaisons méiotiques est utilisée pour estimer les distances entre les marqueurs. Ces marqueurs doivent donc être polymorphes (présenter différents allèles identifiables) et peuvent être localisés dans un exon (variations phénotypiquement détectables) ou (plus couramment) dans un intron (variations détectables au niveau de l'ADN). Les distances se mesurent en centimorgans (cM).

- **Centimorgan (cM)**

Utilisé pour la première fois par Morgan et Sturtevant en 1913 lorsqu'ils ont réalisé la première carte génétique sur le chromosome X de la drosophile. La première plante cartographiée est le maïs en 1935. Lors de la méiose, des marqueurs situés sur un même chromosome peuvent être séparés s'il se produit un crossing-over dans la région qui les sépare. La probabilité qu'un tel événement se produise est proportionnelle à la distance qui sépare ces marqueurs. La fréquence de recombinaison reflète donc des distances entre marqueurs et l'unité de distance est le centiMorgan (cM) : 1 cM est égal à 1% de crossing-over (1 crossing-over pour 100 méioses).

Pour mesurer la distance D entre deux locus A et B on applique la loi suivante :

$$D (A - B) = \frac{\text{Nombre des individus recombinants}}{\text{Nombre total des descendants}}$$

**Remarque :** Notons que l'ordre est toujours le même sur une carte génétique et une carte physique. Seules les distances relatives changent.

#### 5- Marqueur génétique et/ou moléculaire

Cette notion de marqueur génétique a été introduite en 1980. Il s'agit de n'importe quelle séquence d'ADN pouvant être localisée et repérée spécifiquement sur un génome. En cartographie génétique, le marqueur est utilisé pour "baliser" le génome. En contrôle du transfert de gène, le marqueur est un gène associé au gène d'intérêt, codant une caractéristique détectable facilement et précocement, facilitant le repérage des cellules au sein desquelles la transgénèse a réussi. La détection d'un marqueur génétique peut s'effectuer par hybridation avec une sonde complémentaire, ou par son expression phénotypique. Les marqueurs peuvent être de différentes natures : STS, RFLP, microsatellites, SNP, EST, gènes ... Il existe aussi des marqueurs « anonymes » qui correspondent à des séquences non traduites (les variations alléliques de ces

marqueurs ne sont pas décelables au niveau phénotypique mais au niveau de la séquence).

### **5-1 RFLP (Restriction Fragments Length Polymorphism)**

Variation entre individus ou souches dans le profil d'ADN obtenu après coupure par diverses enzymes de restriction. Le polymorphisme de taille des fragments de restriction (RFLP) reflète directement des variations de séquence de segments précis d'ADN et est utilisé comme marqueur sur les cartes génétiques et physiques.

### **5-2 Microsatellites**

Une séquence d'AND dite microsatellite ou Simple Sequence Repeats SSR ou Short Tandem Repeats STR est formée par une répétition continue de motifs formé de 2 à 10 nucléotides Exp : le motif CAGT

Un microsatellite peut être présent à des milliers d'exemplaires dans le génome. Chez les végétaux supérieurs, il y'aurait en moyenne un microsatellite tous les 50 KB.

### **5-3 SNP (Single Nucleotide Polymorphism (= snips))**

Polymorphisme d'un seul nucléotide (polymorphisme nucléotidique). Variation stable de la séquence d'ADN génomique, portant sur une seule base, toutes les 100 à 300 bases environ du génome et affectant au moins 10% de la population. Beaucoup de SNP n'ont pas d'implications fonctionnelles mais ils définissent un locus unique dans le génome et sont polymorphes.

Les SNP se trouvant dans les régions codantes (SNPc) et dans les régions régulatrices des gènes seront particulièrement intéressants pour réaliser la cartographie.

### **5-4 STS (Sequenced Tag Site)**

C'est une courte séquence représentée de façon unique dans le génome. Il est facilement amplifiable par PCR et est archivé sous forme des amorces oligonucléotidiques qui le définissent.

### **5-5 QTL (Quantitative Trait Locus)**

Locus dont l'unité de fonction contribue, pour une part plus ou moins importante, à l'élaboration d'un caractère quantitatif (locus responsable de la variabilité de ce caractère). Plusieurs loci intervenant dans la réalisation d'un même caractère sont souvent groupés dans une même région chromosomique. Les QTL sont étudiés tout particulièrement chez les plantes : les marqueurs génétiquement liés à de tels QTL permettent de sélectionner, parmi un grand nombre de plantes, les individus les plus performants. Ils peuvent ainsi être utilisés pour construire un génotype idéal par croisements successifs ou pour améliorer l'évaluation de la valeur des individus.

## **6- Les cartes de restriction par digestion enzymatique**

Représentation graphique de la localisation des sites de restriction sur une molécule

d'ADN, après avoir soumis celle-ci à une digestion enzymatique.

Une carte de restriction identifie dans l'ADN une suite linéaire de sites séparés par une distance réelle (en nombre de paires de bases). Ces sites correspondent à des coupures de la séquence par des enzymes de restriction. Une enzyme de restriction est un outil biologique qui coupe, de manière reproductible, la molécule d'ADN en des sites correspondant à l'occurrence d'un motif de 4, 5 ou 6 bases spécifique à l'enzyme. On dit aussi que l'enzyme *digère* une séquence. Les tailles des fragments qui résultent de ces coupures sont ensuite mesurées.

## **6 La cartographie de liaison**

La carte de liaison est l'ordonnement de marqueurs (le plus souvent, de courts fragments d'ADN génomiques) définissant chacun un locus (localisation) unique. Ces cartes indiquent les positions relatives des marqueurs les uns par rapport aux autres. Deux marqueurs sont d'autant plus difficilement séparables qu'ils sont éloignés l'un de l'autre.