**CHAPITRE 3 : Cartographie fine et étude physique des génomes**

**INTRODUCTION**

Différentes méthodes d’analyse du génome existent. Elles se classent 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 (les centiMorgans) 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.

1. **Les différents niveaux de résolution de la structure physique du génome :**

Ils permettent de décrire la structure physique du génome nucléaire des plantes et les renseignements qu’elles ont apporté ; ces méthodes recouvrent :

- la cytogénétique,

- les méthodes de fractionnement et de cartographies de restriction

- l’organisation de la chromatine.

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 et 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.

Les différences de taille dans le génome des plantes sont devenues évidentes le jour où l’on a su mesurer la quantité d’ADN par noyau en utilisant soit :

- Des méthodes d’extraction suivies de dosage,

- Des méthodes de cytophotométrie (Evaluation 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 variation des quantités d’ADN par noyaux est sans rapport avec la complexité génétique (nombre de gènes).

La quantité d’ADN d’un génome est appelée la C value. La teneur en ADN peut varier, au sein d’une même plante, selon le tissu, ou l’organe examiné. Les quantités d’ADN par noyaux sont très variables, (de quelques centaines de millions de pb à quelques dizaines de milliards de pb.

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

|  |  |
| --- | --- |
| Espèce végétale | Taille du génome (million de pb) |
| Arabidopsis thaliana | 130 |
| Riz | 430 |
| Medicago truncatula | 650 |
| Brassica oleracea(chou) | 700 |
| Sorghon | 800 |
| Tomate | 1000 |
| Colza | 1200 |
| Maïs | 2500 |
| Orge | 52000 |
| Pois | 5000 |
| Blé | 16000 |
| Tulipe | 3000 |

Le génome d’une cellule peut en effet subir, au cours du développement, des phénomènes d’endo duplication (Cas des cellules de l’albumen de maïs, des trichomes d’Arabidopsis et des cellules provasculaires (3 ou 4 cycles d’endoréplication). Ainsi que les cellules du suspenseur de l’embryon chez les légumineuses.

De nombreuses espèces végétales sont polyploïdes (auto- ou allopolyploïdes) peuvent avoir des variations d’un facteur de 2 à 8 dans la teneur en ADN (Colza et blé).

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

**2- méthodes de fractionnement du génome : Séparation physique du génome**

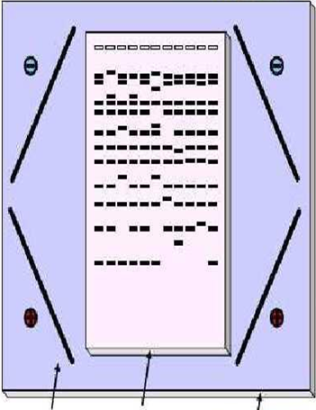
La plupart des êtres vivants contiennent des génomes fragmentés : chromosomes et plasmides bactériens (procaryotes); chromosomes nucléaires, génomes chloroplastiques et mitochondriaux chez les eucaryotes.

La première étape d’étude d’un génome peut-être de le résoudre en ses composants. Parmi les techniques utilisées pour analyser de façon globale la structure des génomes et les fractionner en leurs composants, on trouvera :

1. **Méthodes physiques de séparation des génomes entiers ou fragmentés**

**1-a Séparation en fonction de la taille**

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**

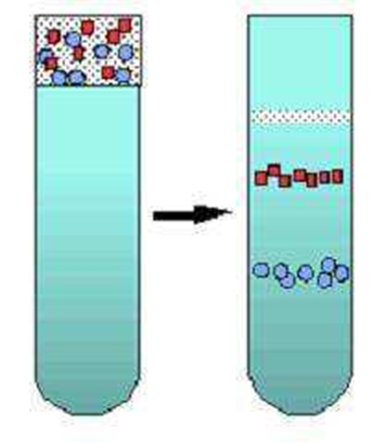
Les plus petits chromosomes d’eucaryotes supérieurs ; dont la taille dépassent la dizaine de Mb ne sont pas résolus par cette technique, de ce fait n’entrent pas dans la maille du gel.

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= GGCCNNNNNGGCC, 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.

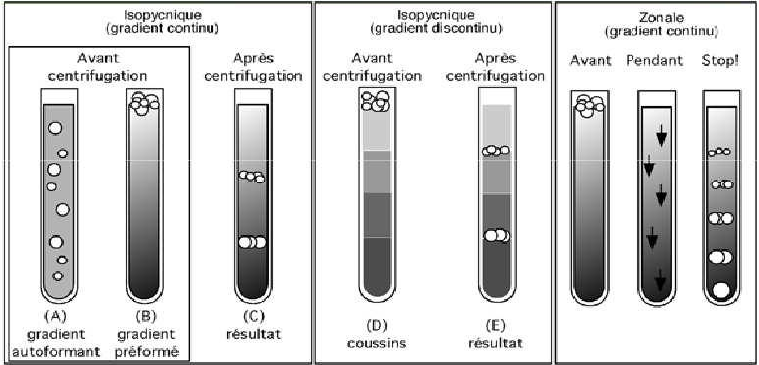


**1-b Séparation en fonction de la forme**

***-* 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 g/ml à 7.5 mol/l.

Cette possibilité d'atteindre des densités si élevées en solution aqueuse est son principal avantage.



Différents types de gradients permettant la séparation de particules en fonction de leur densité.

**1-c Séparation en fonction de la composition en bases**

***-* Gradient de densité (CsCl)**

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

***-* 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 déflecteur est activé et oriente la particule vers le collecteur.

**1-d Séparation en fonction simple ou double brin**

Les acides nucléiques simple brin s’accrochent spontanément à la nitrocellulose. L'ADN double brin s'adsorbe à l'hydroxyapatite (Ca10(PO4)6(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.

1. **analyse de l’organisation physique de l’ADN par des méthodes issues du clonage du gène**

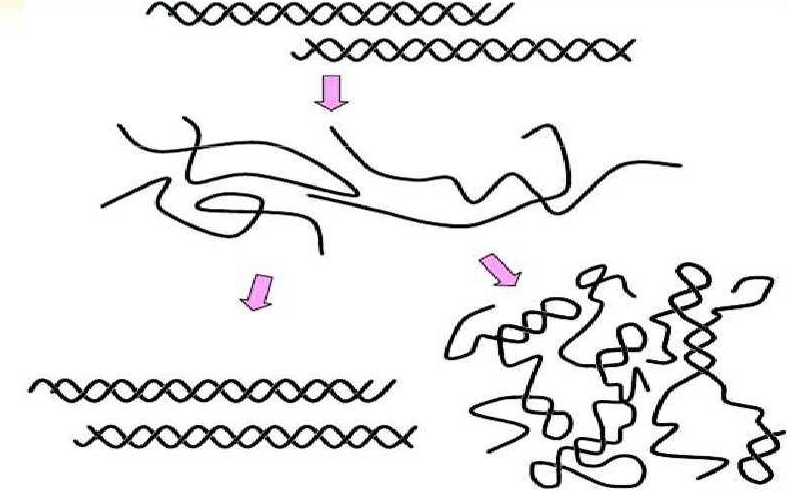
**2-1 Fragmentation spécifique des génomes par des enzymes de restriction**

L’isolement sous forme purifiée de courtes séquences d’ADN directement analysables peut s’effectuer par les endonucléases sites-spécifiques, (enzymes de restriction) ou endonucléases de restriction. Si on utilise une quantité d’enzyme minimale inférieure à la quantité ainsi définie (ou un temps d’incubation inférieur), tous les sites ne seront pas coupés en une heure (Selon les molécules, 0, 1,2 ,… n sites sont coupés). Ces conditions de digestion partielle sont recherchées pour générer une série de fragments chevauchants, lors de la construction de banques génomiques.

**2-2 Repérage de séquences spécifiques par hybridation**

**1- Dénaturation ou fusion de l’ADN** ; On peut rompre les liaisons hydrogènes entre bases appariées d’une molécule d’ADN bicaténaire en chauffant la molécule ou en manipulant les conditions de milieu : solutions alcalines à pH 12,5 avec NaOH, solutions concentrées d’urée ((NH2)2CO) ou de formamide (HCONH2). Les molécules simple brin résultantes sont stables tant que les conditions dénaturantes sont maintenues. Les simples brins absorbent beaucoup plus les UV que les doubles brins. On appelle hyperchromicité cet effet qui permet de suivre facilement la cinétique de dénaturation d’un acide nucléique maintenues.

La dénaturation de l’ADN peut être suivie par l’augmentation de densité optique et est définie par la iTm. Si on reste “longtemps” autour du Tm du double brin, les simples brins complémentaires déroulés pourront se rencontrer par hasard, former des appariements transitoires dans les régions (déplaçant d’éventuelles zones monocaténaires en double brin) et refermer la double hélice comme une fermeture à glissière. On dit qu’il y a eu renaturation de l’ADN.



Refroidissement rapide

Refroidissement progressif

Dénaturation (chauffage)

Ce Tm est établi dans des conditions bien définies de force ionique (en masquant les charges des phosphates, les ions Na+ stabilisent la double hélice; si leur concentration s’élève 10 fois, le Tm augmente de 16,6°C) et éventuellement d’agents dénaturants (1 % de formamide abaisse le Tm de 4°C environ). Deux conséquences importantes :

- si la concentration des séquences complémentaires est forte, l’appariement initial se fera très vite (probabilité de rencontre dans une réaction bi-moléculaire) : dans un génome complexe, les séquences répétées se dénatureront bien plus vite que les séquences uniques : on peut ainsi mesurer la fraction répétitive d’un génome.

- si on ajoute un excès d’une séquence particulière, complémentaire d’une séquence unique d’un génome complexe, la séquence unique du génome s’hybridera plus vite avec la séquence ajoutée qu’elle ne retrouvera son partenaire initial (PRINCIPE DES SONDES).

**2-Hybridation sur support : techniques de Southern et de northern**

Si la séquence ajoutée est marquée (une sonde), elle permettra de repérer aisément les hybrides. Ce principe est à la base de deux techniques principales de repérage / dosage de séquences particulières :

**- l’hybridation d’empreintes sur support (après transferts dits de Southern ou Northern**) : les acides nucléiques séparés sur un gel d’agarose sont transférés et fixés sur une membrane avant hybridation à une sonde. On parle de Southern blot quand on a transféré un ADN, de Northern blot quand on a transféré des ARN (analyse des ARNm).

- **l’hybridation en tache sur support (dot blot en anglais)** : une goutte de l’ADN à tester est dénaturée et fixée sur le support, puis hybridée à la sonde.

**- l’hybridation sur colonies ou plages de lyse phagiques** : permet de repérer parmi plusieurs candidats ceux qui portent une séquence particulière ; très utilisé pour le criblage des banques.

- **l’hybridation en tâche (sur ADN ou ARN dénaturés) :** permet de déterminer la présence d’une séquence particulière dans une préparation (Une goutte de la préparation d'acide nucléique en solution est appliquée à la membrane)

- **l’hybridation sur tissus, sur coupe cellulaire :** permet de détecter où et quand s’exprime un gène donné en repérant les ARN transcrits grâce à une sonde génique radioactive

- **l’hybridation sur chromosomes en métaphase**: permet de positionner les gènes sur un chromosome ou de « peindre » des chromosomes utilisant un ensemble de séquences connues pour être présentes sur un chromosome donné : elles s’hybrideront tout au long de leur chromosome cible.

**Sources de sondes :** Toute séquence nucléique peut servir de sonde :

un ARN purifié ou son ADNc peut servir à rechercher les séquences génomiques qui le codent, un fragment de restriction peut servir à rechercher les ARNm correspondants etc.Une classe de sondes qui s’est beaucoup développée récemment est représentée par des oligonucléotides synthétiques.

=> En pratique, la taille des sondes varie de quelques dizaines de nucléotides à quelques kb.

=> La découverte de nouvelles enzymes de restriction, la possibilité d’effectuer des doubles ou triples digestions et la capacité de repérer les séquences en fonction de leur taille et de leur hybridation allait permettre de dresser des cartes de petites régions de génome, balisées par les sites de restriction.

Le clonage aléatoire de petits fragments d’ADN génomique de plantes, suivi de l’utilisation de fragments clonés comme sonde pour évaluer la distribution des séquences correspondantes dans l’ADN génomique a révélé différents cas de figure :

* Les fragments révélés sont uniques ou en nombre très limité (révélés par hybridation des sondes fortement radioactives et des temps d’exposition prolongés) ;
* Les fragments sont en nombre limité, mais révélé rapidement et sans utiliser de sondes particulièrement chaudes ;
* Les fragments sont en nombre important, mais distincts et distribués selon un profil en échelle dont chaque barreau correspond à une taille multiple de celle du plus petit fragment
* Certaines sondes révèlent des profils d’hybridation mal définis, qui ne sont pas organisées de façon régulière par rapport à des sites de restriction (effet des éléments transposables et leurs dérivés)

Un génome peut être complètement cloné dans quelques dizaines de milliers de clones BAC ou YAC alors que plusieurs dizaines de millions de phages ou quelques centaines de milliers de cosmides étaient nécessaires avec des fragments de 20 ou 30 kb.

1. **La carte physique**

La carte physique est l'ordonnancement 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.

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

1. **La carte génétique**

La carte génétique est l'ordonnancement de marqueurs grâce à l'analyse statistique de leur ségrégation au cours des générations. 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 1ère fois par Morgan et Sturtevant en 1913). Unité de mesure de la distance sur la carte génétique. 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 évènement 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).

1. **Marqueur génétique et/ou moléculaire**

Cette notion de marqueur génétique a été introduite en 1980. Il s’agit d’une séquence d'ADN repérable spécifiquement. 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).

* 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.

* 1. **Microsatellites**

Une séquence d’AND dite microsatellite ou Simple Sequence Rrepeats SSR ou Shorte Tendem 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.

* 1. **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.

* 1. **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.

1. **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.

1. **La cartographie de liaison**

La carte de liaison est l'ordonnancement 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.