**CHAPITRE 4 : STRUCTURE ET EVOLUTION DES GENOMES VEGETAUX**

1. **Généralités sur le génome des plantes**

Les génomes des plantes contiennent plusieurs classes de gènes qui sont absents ou sous-représentés chez les animaux, par exemple les gènes codant pour :

* Les enzymes de biosynthèse de la paroi cellulaire
* Les protéines de transport des nutriments
* Les protéines spécifiques de la photosynthèse : antennes, transports des électrons, RuBisCo….
* Les produits impliqués dans la turgescence et les réponses à différents stress.
* Des gènes de résistance aux pathogènes, très polymorphes et disséminés dans tout le génome.

Les classes de gènes communes aux plantes et aux animaux sont celles impliquées dans les mécanismes les plus généraux de la biologie d’une cellule :

* la communication entre cellule
* la régulation de transcription
* la transduction des signaux

1. **La taille des génomes végétaux**

La taille du génome haploïde, correspondant à la moitié du stock chromosomique, varie de **100** millions de paires de bases (Mb) environ chez *Arabidopsis thaliana* à **123 000** Mb chez *Fritillaria assyriaca* (la fritillaire). Ainsi, ces deux espèces de plantes à fleurs (angiospermes dicotylédones) sont constituées des mêmes organes (racines, tiges, feuilles et fleurs) mais ont des dimensions de génomes qui diffèrent de 1 230 fois.

Il est vrai que de très nombreuses espèces végétales sont polyploïdes, c'est-à-dire avec plusieurs stocks chromosomiques par noyau. Le blé tendre, par exemple, est un hexaploïde ; son noyau de 42 chromosomes résulte de l'association de 3 génomes d'espèces proches ayant chacune 14 chromosomes. Mais la polyploïdie ne suffit pas à expliquer cette énorme variation de taille puisque ce blé a une taille de génome « haploïde » de **16 000 Mb**, alors que le riz (de la même famille des graminées) a un génome de *450 Mb*. C'est d'ailleurs la raison pour laquelle, outre son importance agronomique majeure, le riz est devenu l'autre espèce modèle des généticiens des végétaux, représentant les monocotylédones. La tomate (génome de **950 Mb**) et le maïs (**2 500 Mb**) sont aussi des espèces très étudiées par les généticiens.

Comme dans les autres espèces, les génomes nucléaires des végétaux sont organisés en chromosomes, dont le nombre varie de 10 chez l'arabette à plus de 100 chez la canne à sucre. Les chromosomes sont constitués d'une longue double hélice d'ADN entourant des nucléosomes. La chaîne nucléoprotéique constituant ces nucléosomes est plus ou moins condensée selon les phases du cycle cellulaire. Cette organisation en chromosomes permet la réplication puis la ségrégation des gènes à la mitose et à la méiose. Les centromères qui s'attachent au fuseau équatorial lors de la division cellulaire sont essentiels à la régularité de la ségrégation. Les télomères, à chaque extrémité des chromosomes, sont indispensables à la fidélité de la réplication.

1. **Nature de séquences constituantes les génomes**

Un génome de 100 Mb environ suffit pour le fonctionnement et l'organisation d'une plante à fleur évoluée, ce qui a donné lieu à de nombreuses spéculations sur la fonction de l'ADN apparemment en excès chez les espèces à grand génome. Dans les années 1970, les études de cinétique de réassociation de l'ADN ont permis de mieux décrire la complexité des génomes végétaux. On a pu ainsi distinguer les séquences **répétées** (qui se réassocient rapidement car elles existent en un très grand nombre d'exemplaires) et les séquences dites **uniques**, qui se réassocient très lentement.

1. **Les séquences répétées**

Cet d'ADN répété ne code vraisemblablement pas pour des protéines et son rôle est encore mal connu, il pourrait être impliqué dans des phénomènes de régulation ou d'organisation physique des chromosomes. La proportion d'ADN répété est d'autant plus importante que la taille du génome est grande. Ainsi les espèces à grand génome comme le blé ont plus de 80 % de séquences répétées, alors que l'arabette en a autour de 20 %. On distingue deux types de séquences répétées :

1. **Séquences répétées en tandem :** où les séquences identiques sont regroupées et se suivent le long de la molécule.

Parmi les séquences répétées en tandem se trouve **l'ADN ribosomique**, qui code pour les ARN des ribosomes où s'effectue la synthèse protéique. Plusieurs milliers de copies de cette séquence, dont seule une partie est effectivement transcrite, sont concentrées dans une ou quelques régions chromosomiques et peuvent couvrir plusieurs dizaines de kilobases (milliers de paires de bases). L'autre grand groupe est celui des **ADN satellites**, ainsi appelés d'après leur position par rapport à la plus grande partie de l'ADN après centrifugation dans un gradient de densité. Ces ADN satellites sont constitués d'unités répétées en tandem dont le motif élémentaire a de 150 à 350 paires de bases. On trouve ces familles de séquences dans les régions télomériques et dans les régions centromériques.

On trouve aussi dans les génomes des végétaux des séquences **VNTR** (pour Variable Number of Tandem Repeats), mini- ou microsatellites, comme dans les génomes des espèces animales. Ces séquences sont réparties de façon aléatoire le long des chromosomes et sont constituées d'un nombre variable de répétitions d'un motif élémentaire plus ou moins long. Les microsatellites sont des répétitions en tandem d'une courte séquence de 2 à 5 nucléotides. Le polymorphisme important quant au nombre de motifs répétés, constaté entre individus de la même espèce, fait de ces locus microsatellites d'excellents marqueurs pour l'établissement des cartes génétiques.

1. **Séquences répétées dispersées** : Parmi le groupe des séquences répétées dispersées dans l'ensemble du génome, la classe la plus abondante est représentée par des séquences caractéristiques des éléments transposables, gènes sauteurs capables de se déplacer (transposition) et de se multiplier au sein du génome. Les séquences de ces transposons, ou de leurs dérivés inactivés, peuvent représenter un pourcentage significatif du génome.
2. **Les séquences « uniques »**

Se trouvant en simple copie ou en faible nombre de copies dans un génome donné, elles représentent la fraction accessible à l'analyse génétique classique. Il s'agit des séquences codant pour les gènes exprimés et dont les mutations vont se manifester par une caractéristique particulière, visible à l'œil nu ou accessible à l'analyse biochimique. Cette matière de base des généticiens d'hier est aussi celle de la génétique moléculaire. Le très grand nombre de marqueurs nucléiques aujourd'hui disponibles, par exemple les RFLPs (polymorphismes de la longueur des fragments de restriction), font également fréquemment appel à cette fraction du génome qui est a priori la plus intéressante puisque codant pour des protéines intervenant dans le métabolisme.

L'utilisation de la transcriptase inverse, enzyme qui permet la synthèse d'ADN à partir d'une séquence d'ARN, a servi à isoler les ADN complémentaires (ADN-c) des ARN messagers dans différentes espèces, différents organes, différentes conditions environnementales. Ces ADN-c peuvent être utilisés comme sondes et fournissent ainsi des marqueurs génétiques permettant de différencier pour ces gènes, connus ou inconnus, les formes alléliques, polymorphisme indispensable à l'établissement des cartes génétiques.

1. **Mécanismes à l’origine d’une variabilité du génome**

L’analyse de la séquence des génomes a montré une grande variabilité de séquences inter espèces. L’étude de la génétique des populations a aussi montré une grande variabilité entre individus d’une même espèce, ceci est en partie due au mécanisme de variabilité le plus connu : la mutation ponctuelle d’un nucléotide (incluant l’insertion, la délétion ou la substitution). Cependant, il existe des génomes dont le taux de mutation est très faible et dont la principale source de polymorphisme provient de l’arrangement diffèrent de leur gènes le long des chromosomes.

**3-1 Les éléments transposables**

Ceux sont des séquences d'ADN capable de se déplacer et de se multiplier de manière autonome dans un génome, par un mécanisme appelé transposition. Les transposons ont été identifiés par Barbara McClintock qui a étudié des mutations instables induites par ces éléments chez le maïs.

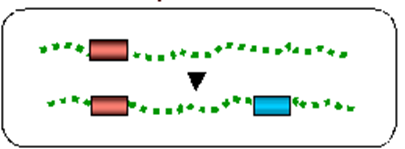
Ils présents chez tous les organismes vivants et sont un des constituants les plus importants des génomes eucaryotes, exemples : 45% du génome de l'homme et plus de 70% chez le maïs. Les éléments transposables peuvent expliquer d'importantes différences dans les tailles du génome d'organismes. Il y’a deux types de transposon :

1. **Transposon de classe I ou rétrotransposon** : ils transposent via un intermédiaire ARN selon un mode réplicatif (copier - coller).Leur ARN messager est rétro-transcrit en ADNc qui est intégré dans le génome. Les rétrotransposons sont divisés en 2 groupes :

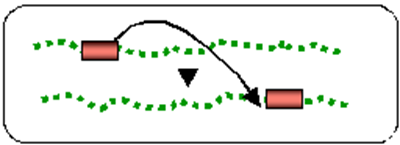
**1-les rétrotransposons à LTR :** (Longue séquence Terminale Répétée - "Long Termial Repeat") ou type rétroviral.

**2-les rétrotransposons sans LTR :** (les LINE - "Long Interspersed Nuclear Elements" - 1 à 7 kpb) et les SINE ("Short Interspersed Nuclear Elements" - 100 à 500 pb).

**Les séquences Alu :** Famille de séquences répétées issue d'un processus de rétrotransposition. Les séquences Alu contiennent un site de reconnaissance pour l'enzyme de restriction Alu I. Longueur environ 300 pb avec un nombre de répétitions d’environ 1 million de fois chez l'homme (11% du génome). La famille Alu est un exemple de SINE ("Short Interspersed Nuclear Elements". La plupart sont des rétroséquences issues d'une réverse transcription de l'ARN.



1. **Transposon de classe II** : ils transposent selon un mode conservatif (couper-coller).Ils sont caractérisés par la présence à leurs extrémités de séquences terminales inversement répétées (TIR). Ils sont de taille variable (100 à 20 000 pb).



**Tableau N°1** : nombre d’éléments transposables chez différentes espèces

|  |  |  |  |
| --- | --- | --- | --- |
| Eléments transposables | *Arabidopsis thaliana* | *Homo sapiens* | *Saccharomyces cerevisiae* |
| nombre de copies classe I - LTR | 1594 | 443.000 | 331 |
| nombre de copies classe I - non LTR | 515 | 2.426.000 | 0 |
| nombre de copies classe II | 2203 | 294.000 | 0 |
| **Les valeurs ci-dessus ne sont qu'indicatives. Elles évoluent en effet au fur et à mesure que les génomes sont re-séquencés, analysés structuralement et enfin ré-annotés.** | | | |

**3-2 la polyploïdie**

La polyploïdie est un processus important qui touche l’évolution de la plupart des espèces, on distingue deux types de polyploïdies : les autopolyploïdies qui résultent de dédoublement du stock chromosomique, et les allopolyploïdies qui proviennent d’une hybridation interspécifique suivie d’un dédoublement chromosomique.

D’un point de vue du mécanisme cytologique, la polyploïdie peut provenir des divisions méiotiques et mitotiques anormales, conduisant à la formation des gamètes diploïdes aboutissant à la formation de zygotes tétraploïdes. L’intérêt agronomique de la manipulation du nombre de chromosomes :

- Augmentation de la taille des plantes

- Contrôle de la fertilité/stérilité

- Facilite la recherche et la sélection de nouveaux variants (mutations). Donc amélioration des espèces ; supposons une plante constituée de 3 chromosomes (A B et C) :

|  |  |  |  |
| --- | --- | --- | --- |
| **Ploïdies** | **Désignation** | **Formule** | **Nb chromosomes** |
| **Monoploïdie** | n | A B C | 3 |
| **Diploïdie** | 2n | AA BB CC | 6 |
| **Triploïdie** | 3n | AAA BBB CCC | 9 |
| **Tétraploïdie** | 4n | AAAAA BBBB CCCC | 12 |
| **Monosomies** | 2n-1 | A BB CC  AA B CC  AA BB C | 5 |
| **Trisomies** | 2n+1 | AAA BB CC  AA BBB CC  AA BB CCC | 7 |

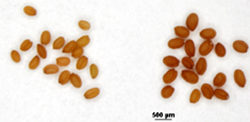
Les plantes polyploïdes peuvent être obtenues soit :

- par doublement du jeu de chromosomes (autopolyploïdes)

- par fécondation d'espèces proches (allopolyploïdes)

**1- Autopolyploïdes :** doublement

Espèce à 2n chromosomes Même espèce à 4n chromosomes Les plantes polyploïdes ressemblent aux plantes diploïdes mais sont en général plus grandes.

**Raisin**  **Arabidopsis thaliana Écotype**

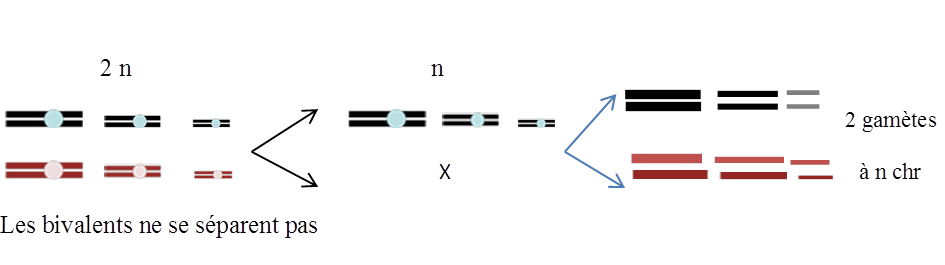
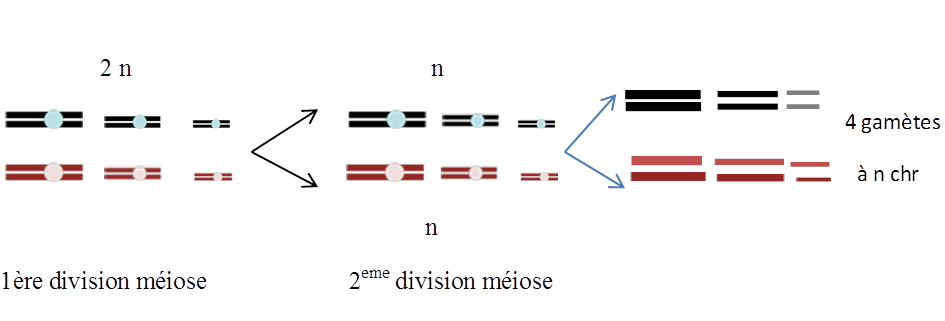
**Landsberg erecta (Ler)**



**Arabidopsis thaliana**

**Écotype Columbia (Col-0)**

**Doublement naturel des chromosomes:** Le doublement du nombre de chromosomes peut arriver spontanément par "accident".



**Doublement provoqué des chromosomes :** Le doublement du nombre de chromosomes peut être provoqué artificiellement.



Plante traité

Par la colchicine



On obtient des graines à 4 n

**2- allopolyploïde**

- Une alloploïdie n'est possible qu'entre 2 espèces évolutivement proches, donc inter fécondent.

- Si les 2 espèces qui fusionnent possèdent le même nombre de chromosomes, on parle d'amphidiploïde.

Espèce A Espèce B croisement Nouvelle espèce

2n chromosomes 2n chromosomes 4n chromosomes

**Exemple sur l’évolution : le génome du blé**

**Introduction**

La domestication et la culture du blé (blé tendre et blé dur) a été un élément fondateur des premières civilisations humaines dans le croissant fertile. En plus de son intérêt comme une des principales céréales apportant l’énergie dans l’alimentation, le blé est aussi la première source de protéines dans les pays en voie de développement.

Les réserves mondiales de blé sont en nette diminution. Par conséquent, il est important d’améliorer la rusticité et la qualité nutritionnelle du blé et cela dans une perspective d’agriculture durable, plus économe en intrant et plus respectueuse de l’environnement. La France, considérée comme le grenier de l’Europe pour le blé tendre, est un des principaux producteurs du blé.

Outre son importance économique nationale et internationale, le blé présente de nombreuses particularités scientifiques qui en font un modèle intéressant pour l’étude de l’organisation et l’évolution des génomes des plantes (monocotylédones), dans différents contextes de polyploïdie ainsi que sous la pression de la domestication et de la sélection.

**Evolution dynamique du génome du blé**

Le génome du blé semble avoir été envahi  par les éléments transposables (principalement retro éléments) qui constituent la majeure partie de ce génome (80-90%). Par ailleurs, les gènes ne semblent pas être distribués de façon homogène sur le génome du blé, mais plutôt regroupés en îlots ou régions dites riches en gènes.

De nombreuses études et analyses tendent à montrer un processus d’évolution dynamique du génome du blé impliquant des mécanismes conflictuels d’expansion et de réduction du génome. Les mécanismes d’expansion des génomes sont souvent dus à l’invasion du génome par des transposons et des rétrotransposons ainsi que des duplications de grands fragments génomiques. Les mécanismes qui conduisent à la réduction des tailles des génomes ou élimination des séquences sont souvent dus à des événements de recombinaisons inégales intra ou inter rétrotransposons et de délétion des séquences répétées. Contrairement aux petits génomes, comme celui d’*Arabidopsis*, de la vigne ou du riz, celui du blé semble présenter une évolution dynamique particulière. Son étude fine doit permettre de mieux comprendre les conséquences de cette stratégie évolutive particulière.

**Modèle d’étude des mécanismes de polyploïdisation**

Un des principaux résultats des programmes de séquençage des génomes des eucaryotes fut la découverte que les espèces considérées comme typiquement diploïdes (levure, *Arabidopsis thaliana* et même l’homme) sont en fait des anciens polyploïdes qui ont eu au cours de leur évolution de nombreux cycles de polyploïdisation et doublement chromosomique.

La polyploïdie joue un rôle très important dans l’évolution des eucaryotes et constitue ainsi un mécanisme important de diversification et génération de variabilité génétique. La majorité des plantes, y compris les plantes cultivées, sont des polyploïdes soit relativement récents (comme le colza, le blé, le cotonnier, la pomme de terre, la luzerne), soit anciens, retenant encore des « vestiges » d’événements de polyploïdisation plus ancestraux (le maïs, le soja, le chou).

Les raisons du « succès » de la polyploïdie chez les plantes ne sont pas claires. Pour s’établir comme des nouvelles espèces, les polyploïdes nouvellement formés doivent en effet supporter des biais d’appariement et de ségrégation chromosomique.

Le blé hexaploïde (*Triticum aestivum*) est issu de deux événements de polyploïdisation relativement récents entre trois espèces diploïdes bien identifiées. Le premier événement, impliquant *Triticum monococcum* et (putativement) *Aegilops speltoides*, a eu lieu il y a environ 0,5 million d’années et a conduit à l’apparition du blé dur (*Triticum durum*). Le deuxième événement de polyploïdisation a eu lieu il y a environ 9 000 ans entre le blé dur (tétraploïde) et un troisième diploïde (*Triticum tauschii*).

Le blé peut constituer un excellent modèle d’étude de la polyploïdie et de son effet sur l’évolution des gènes et des séquences chez les plantes. En effet, en plus des polyploïdes naturels, il est possible de « synthétiser » des nouveaux blés (*Triticum*) ayant différents niveaux de polyploïdie par croisement entre différentes Triticées diploïdes.

De nombreuses études conduites ces dernières années sur des polyploïdes naturels et synthétiques du blé ont porté sur la caractérisation de l’effet direct de la polyploïdie sur l’évolution et la régulation de l’expression des gènes. Des réarrangements dans les séquences non codantes ont été observés dans les polyploïdes synthétiques de Triticées et supposés comme contribuant au processus de stabilisation génomique.

D’autres études plus récentes ont montré l’effet de la polyploïdie sur la régulation de l’expression de plusieurs gènes chez des tétraploïdes synthétiques de Triticées. Cette régulation de l’expression est en partie due à des phénomènes épigénétiques et de silencing mais aussi à des excisions de gènes et séquences. Ceci suggère que des pertes de gènes peuvent être le résultat d’une évolution ou divergence à l’échelle de temps des génomes mais aussi d’une évolution rapide par élimination de séquences suite à la polyploïdisation.

L’effet de la polyploïdie sur l’évolution des génomes peut aussi être apprécié par analyse comparative de séquences génomiques d’espèces phylogénétiquement proches et ayant différents niveaux de polyploïdie. Un exemple assez illustratif a été récemment décrit autour d’un locus codant pour des gluténines de faible poids moléculaire (FPM) chez les Triticées. Le génome Am de *Triticum monococcum* contient 3 copies dupliquées en tandem au niveau de ce locus, alors que le génome Ad du tétraploïde *T. durum* contient une seule copie. Plus étonnant, en dehors des séquences des gènes de gluténines-FPM, le reste des régions comparées ne montrent pas d’homologie.

**Evolution sous pression de domestication et sélection**

Les blés ont subi une domestication avec différentes pressions de sélection, les faisant passer de graminée sauvage à l’espèce cultivée avec le niveau de production actuel. Un séquençage partiel permettrait de comprendre comment se manifestent ces pressions de sélection au niveau du génome et en liaison avec le processus général d’adaptation à la récolte, d’adaptation aux grandes régions géographiques, aux stress environnementaux, aux diverses utilisations de la récolte.

L’analyse de séquences de régions représentatives du génome devrait permettre de recueillir des informations scientifiques permettant d’apprécier de façon statistique les effets respectifs de la polyploïdisation, de la domestication et de la sélection, et ainsi d’augmenter les connaissances sur l’organisation du génome du blé et les inférences sur l’expression des gènes.