**CHAPITRE 8 : CARACTERISTIQUES STRUCTURALES DE GENOME D’*ARABIDOPSIS Thaliana***

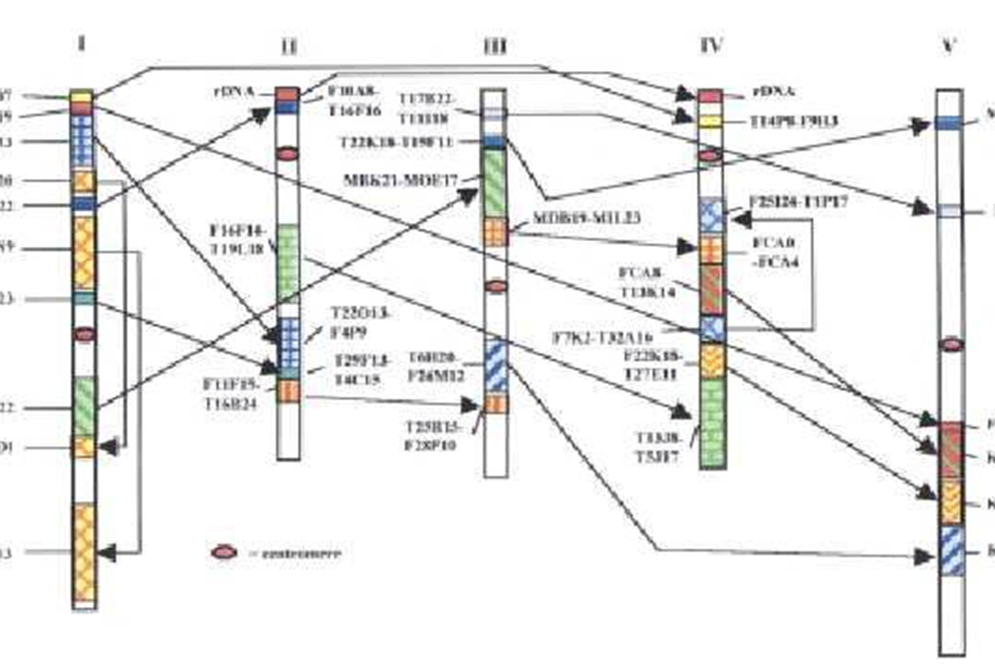
**1-Introduction**

|  |  |
| --- | --- |
| *Arabidopsis thaliana* a été découverte au 16eme siècle par Johannes Thal. Elle est aussi nommée arabette rameuse ou arabette des dames.    Plus de 750 accessions naturelles d'*Arabidopsis thaliana* ont été collectées dans le [monde](http://arabidopsis.org/images/Geo_dist_map.gif) et sont disponibles dans les 2 centres de semences : [ABRC](http://arabidopsis.org/servlets/Order?state=catalog) et le [NASC](http://arabidopsis.info/) ("*Nottingham Arabidopsis Stock Centre*").  **Quelques caractéristiques:**   * taille à l'âge adulte : 30-40 cm * racine pivotante, feuilles alternes, inflorescence en grappe * fleur à symétrie bilatérale comprenant 4 sépales, 4 pétales, 6 étamines et 2 carpelles * fruits de 3 à 5 mm offrant 30 à 60 graines. |  |
| La complexité et la grande taille des génomes des plantes cultivées rendent difficiles leur analyse génétique et moléculaire.Mais comme il y a des similitudes entre les génomes des différentes espèces (cultivées ou non), on a choisi une plante ayant le plus petit génome possible, pour l'analyser et utiliser les informations obtenues pour l'étude des génomes plus complexes et plus grands.  Le choix de la communauté scientifique internationale s'est porté sur *Arabidopsis thaliana* qui présente de nombreux avantages :   * un très petit génome donc plus facilement séquençable. Par comparaison, le riz et le maïs ont un génome respectivement 4 et 20 fois plus grands * facilité de culture liée à sa petite taille ([infrastructure légère pour la cultiver](http://www.wisc.edu/biotron/arabidopsis.html)) * reproduction principalement par autofécondation. * rapidité de développement : son [cycle de reproduction](http://www.arabidopsis.org/portals/education/growth.jsp) est court (6 semaines de la germination à la graine mature) * enfin, et c'est un point important : aucune importance économique donc pas de risque de"conflits" d'intérêts et une meilleur collaboration internationale.   **Tableau I.** Caractéristiques des génomes d’*Arabidopsis* et de végétaux économiquement importants.  **Espèce Taille du génome Nombre**   |  |  |  | | --- | --- | --- | | **EspèceEspèceesEspèces** | Taille du genome(MP) | Nombre de chromosomes | | *Arabidopsis thaliana* (arabette)  *Lycopersiconesculentum*(tomate)  *Zea mays* (maïs)  *Oryza sativa* (riz)  *Triticumaestivum*(blé)  *Sorghum bicolor* (sorgho)  *Hordeumvulgare*(orge) | 125  100  2500  430  16000  800  5200 | 5 (2n=2x=10)  12 (2n=2x=24)  10 (2n=2x=20)  12 (2n=2x=24)  7 (2n=6x=42)  10 (2n=2x=20)  7 (2n=2x=14) |  |  | | --- | | **2-Le génome d'*Arabidopsis thaliana*** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | * Les cartes physiques des 5 chromosomes ont été complètement déterminées en 1997. * Le génome d*'Arabidopsis thaliana* a été séquencé par une méthode ordonnée ("*top-down*") et publié en 2000 : Arabidopsis Genome Initiative (2000) "*Analysis of the genomesequence of the flowering plant Arabidopsis thaliana*". * Le génome d'*Arabidopsis thaliana* sert de modèle pour l'un des deux grands groupes de plantes à fleurs, les dicotylédones. * taille du génome : 114 500 000 paires de base (pb) + 10 Mb (organisateurs nucléolaires et centromères) = environ 125 Mb * la base de données [TAIR](http://www.arabidopsis.org/portals/genAnnotation/genome_snapshot.jsp) contient environ 26000 gènes codant des protéines - 924 pseudogènes- 4827 transposon. * En 2009, environ 40% des gènes ont une fonction qui n'est toujours pas connue  |  |  |  |  | | --- | --- | --- | --- | | chromosome | longueur (Mpb) | teneur en [G+C] (%) | Protéines codées | | 1 (métacentrique) | 30,0 | 35,7 | 8521 | | 2 (acrocentrique) | 20,0 | 35,8 | 5182 | | 3 (submétacentrique) | 23,0 | 36,3 | 6432 | | 4 (acrocentrique) | 19,0 | 36,2 | 4934 | | 5 (submétacentrique) | 27,0 | 35,9 | 7545 | | [**chloroplastique**](http://atensembl.arabidopsis.info/Arabidopsis_thaliana_MIPS/mapview) | 0,15 | 36,3 | 85 | | [**mitochondrial**](http://atensembl.arabidopsis.info/Arabidopsis_thaliana_MIPS/mapview) | 0,37 | 44,8 | 117 | | | |

* Nombre d’exons : 132 982
* taille moyenne des exons : 251pb
* Nombre d’introns : 107 484
* taille moyenne des introns : 170 pb
* Nombre moyen d’exons par gène : 5,2
* nombre d'ARNt : 631
* 79% des gènes contient des introns
* séquences type rétrovirus : 4,8%.
* séquences type transposons : 5,1%
* Fréquence des séquences répétées : environ 10%, ce qui est une fréquence faible par rapport aux autres plantes

**Remarque :** tous ces chiffres ne sont qu'indicatifs car ils évoluent au fur et à mesure que l'ensemble du génome est ré-analysé et donc ré-annoté.

Le degré de redondance (la duplication de gènes ou de régions identifiés sur un chromosome) est élevé : 24 régions d'une taille supérieure à 100 kb sont dupliquées, soit 65,6 Mb, soit 58% du génome.



**Fig 1 : Carte des régions dupliquées dans le génome d’Arabidopsis**

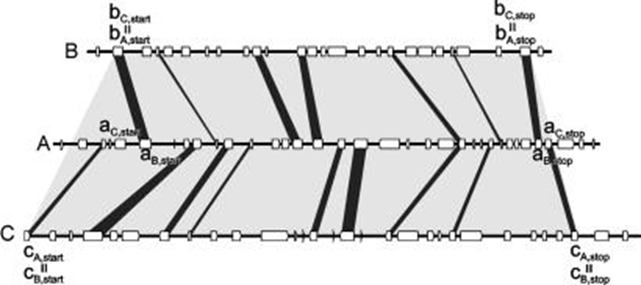
Cela suggère qu'*Arabidopsis thaliana* soit passé récemment par un état tétraploïde avec ensuite, une perte d'une partie des gènes dupliqués.

3 duplications :

* 221 millions d'années
* 162 millions d'années
* 75 millions d'années

Calcul : 6.1 mutations / millions d'années / site

Par ailleurs, un transfert du génome mitochondrial vers le noyau a certainement eut lieu récemment chez Arabidopsis thaliana puisque une région de 620 kb identique à 99% au génome mitochondrial a été identifiée sur le chromosome 2.



**FIG 2 : Duplication chez *Arabidopsis***

**3-Les génomes mitochondriaux des plantes**

Ils se distinguent de ceux des autres eucaryotes. Leur taille est plus importante et beaucoup plus variable (de 200 kb chez les brassicacées à 2500 kb chez les cucurbitacées, contre 16 à 20 kb chez les mammifères).La caractéristique de ces génomes est la présence de séquences répétées (impliquées dans des recombinaisons entre molécules d'ADN mitochondrial) variables en taille, en nombre de répétition et en orientation.Cependant, malgré leur grande variabilité en taille, les génomes mitochondriaux des plantes semblent avoir le même contenu en information génétique : de 100 à 120 gènes (inclus ceux qui codent pour les ARN de transfert et les ARN ribosomiques).Il existe un processus propre aux mitochondries végétales : l'édition des messagers. Le transcrit primaire peut subir des modifications post-transcriptionnelles qui changent spécifiquement certaines cytidines en uraciles. Les codons de l'ARN messager sont donc modifiés.

**4-Le génome chloroplastique**

Les chloroplastes possèdent plusieurs copies d'un génome circulaire mesurant de 120 à 217 kb (150 kb en moyenne) chez les plantes supérieures (pour Arabidopsis 154 Kpb avec 79 gènes).Le séquençage complet de l'ADN chloroplastique de plusieurs plantes très éloignées au sens évolutif (algues, bryophytes, gymnospermes, angiospermes) a révélé que chez les plantes terrestres, l'organisation de ce génome circulaire est remarquablement conservée.

Hormis pour certaines légumineuses et des conifères, le génome chloroplastique est caractérisé par la duplication d'une région contenant l'ADN ribosomique chloroplastique.

Les gènes de l'ADN chloroplastique se répartissent en deux catégories :

•l'une contient :

* Tous les ARN nécessaires à l'expression des gènes (30 ARN de transfert et 4 ARN ribosomiques)
* 21 protéines ribosomiques
* 4 sous-unités de l'ARN polymérase

•l'autre contient :

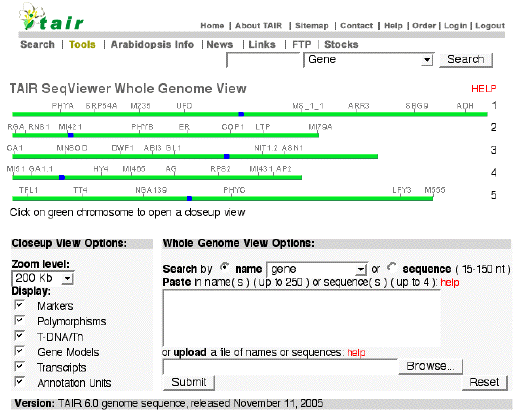
* 29 gènes impliqués dans la photosynthèse
* 11 gènes impliqués dans la photo respiration
* En particulier, le gène rbcL qui code pour la grande sous-unité de la ribulosebiphosphate carboxylase (RuBisCO), l'enzyme clé de la photosynthèse. La RuBisCO est constituée de sous-unités codées par le génome chloroplastique et par le génome nucléaire, ce qui démontre l'étroite collaboration entre les deux génomes.

L'ADN chloroplastique a un taux d'évolution assez lent. Ainsi, le séquençage de parties de l'ADN chloroplastique (comme le gène rbcL) permet de retracer l'évolution des grandes familles de plantes via des reconstructions phylogénétiques.

**4. Bases de données relatives à *Arabidopsis thaliana***

**4-1 TAIR :** The Arabidopsis Information Resource

Descriptions de l'ensemble des données de la base par catégories : TAIR DatabaseStatistics. L’image au-dessous montre la page d'accueil de "SequenceViewer", logiciel pour la recherche et la visualisation de séquences de gènes, de séquences transcrites, de marqueurs (et bien d'autres informations) sur les 5 chromosomes.



**Application**

Recherche du gène codant les deux types de sous-unités de la ribulose 1,5-bisphosphate carboxylase - oxygénase ou RuBisCO.

1. Aller au TAIR

2. Sélectionner "Gene" dans le menu déroulant "Search".

3-Taper "rubisco" (en haut à droite).

|  |  |  |  |
| --- | --- | --- | --- |
| Locus | Gène | Description | chromosome |
| Exemple de gènes codant la [RuBisCO](http://biochimej.univ-angers.fr/Page2/COURS/Zsuite/2Photosynthese/7CycleCALVIN/1CycleCALVIN.htm) | | | |
| ATCG00490 | ATCG00490.1 | chaîne lourde | [chloroplastique](http://atensembl.arabidopsis.info/Arabidopsis_thaliana_MIPS/mapview) |
| AT1G67090 | AT1G67090.1 (rbcS1A) et AT1G67090.2  (épissage alternatif) | chaîne légère 1A | 1 |
| AT5G38430 | AT5G38430.1 | chaîne légère 1B | 5 |
| AT5G38420 | AT5G38420.1 | chaîne légère 2B |
| AT5G38410 | AT5G38410.1 | chaîne légère 3B |
| AT2G07732 | AT2G07732.1 | pseudogéne : semblable à la chaîne lourde de la RuBisCO | 2 |
| Exemple de gènes ne codant pas la RuBisCO | | | |
| AT1G26230 | AT1G26230.1 | chaperonin | 1 |
| AT1G55490 | AT1G55490.1 | RuBisCO subunit binding-protein beta subunit | 1 |
| AT2G39730 | AT2G39730.1 | RuBisCOactivase | 2 |
| AT4G20130 | AT4G20130.1 | RuBisCO large subunit N-methyltransferase-related | 4 |
| AT5G58240 | AT5G58240.1 et AT5G58240.2  ([épissage alternatif](http://biochimej.univ-angers.fr/Page2/COURS/3CoursdeBiochSTRUCT/8Spliceosome/1Spliceosome.htm)) | bis(5'-adenosyl)-triphosphatase (putative) | 5 |

**4-2Flagdb++**

Le but de la base de données Flagdb++ est de développer une interface de travail basée sur l'intégration de données autour du génome d'Arabidopsis thaliana, afin d'aider les utilisateurs à comprendre la fonction biologique des gènes végétaux en les considérants dans un contexte large : une famille multigénique et/ou un réseau fonctionnel.

FLAGdb++ est développée de façon générique pour pouvoir s'appliquer à différents génomes et pour pouvoir stocker, organiser, visualiser et exploiter de nombreux types de données de génomique. L'intégration des données se fait par leur association aux séquences nucléiques, indépendamment les unes des autres.Elle concerne actuellement :

* l'annotation structurale et fonctionnelle du génome d'Arabidopsis thaliana
* les EST
* les étiquettes de mutants d'insertion
* les familles multigéniques
* les motifs protéiques
* les séquences répétées
* les oligonucléotides et sondes spécifiques pour des approches de génomique (puces à ADN, synténie, RNAi, criblage PCR...)

**4-3AraCyc (Arabidopsis thaliana Biochemical Pathways)**

Exemple de recherche de données sur le cycle de Calvin et la RuBisCO.

Aller à AraCyc. Taper "Calvin" dans la fenêtre "Search" en haut à droite.

Activer le lien vers les gènes codant la RuBisCO en passant la souris sur la flèche entre "D-ribulose 1,5-bisphosphate" et "3-phosphoglycérate".

**4-4 Le centre de ressources Arabidopsis thaliana pour la génomique (INRA Versailles)**

Ce centre de ressources propose une collection de plusieurs dizaines de milliers de mutants d'insertion indépendants (mutants pour environ 80% des gènes) obtenus par mutagenèse T-DNA (insertion d'une cassette d'ADN dans l'écotype WS).

* la méthode de transformation s'est appuyée sur les propriétés de transfert d’Agrobacterium tumefaciens.
* chaque mutant "knock-out" est observé et son phénotype donne des indications sur la fonction du gène inactivé.
* possibilité de rechercher un mutant dans un gène de séquence connue mais de fonction inconnue et d'observer son comportement.
* la plupart des mutants sont disponibles sous la forme d'une lignée d'autofécondation T3 (et quelques lignées en T2).