**CHAPITRE I : CONCEPTE DE LA GENOMIQUE**

1. **Introduction**

L’information héréditaire définissant les caractéristiques biologiques d’un organisme eucaryote est portée par ses chromosomes et est codée dans l’ADN. Cet ADN se compose d’un enchaînement de quatre types de molécules appelées nucléotides: l’adénine (A), la guanine (G), la cytosine (C) et la thymine (T). Comme une série de lettres forme un mot, certaines séquences de nucléotides ont une signification fonctionnelle et constituent les **gènes**. La traduction des ADN codants en protéines passe par une phase de transcription de l’ADN en ARN. Les protéines ainsi synthétisées remplissent des rôles variés de structure, de transport, de communication, de défense, etc. D’autres gènes spécifient la synthèse de divers ARN impliqués dans la machinerie de traduction ou la régulation de l’expression génique. L’ensemble du patrimoine génétique d’un individu constitue son **génome**. Toutes les cellules d’un organisme possèdent le même génome, mais cette information s’exprime différemment selon les tissus, sous l’effet de divers mécanismes de régulation. L’ADN peut subir des changements (mutations) sous l’effet, par exemple, d’agents extérieurs (UV, radioactivité, produits chimiques…). Les mutations ponctuelles modifiant un seul nucléotide sont appelées **polymorphisme SNP** (Single-Nucleotide Polymorphism). Les mutations créent des versions différentes d’un même gène (allèles) et peuvent parfois en modifier la fonction. Les mutations apparues dans les cellules germinales sont transmises à la descendance, contrairement aux mutations somatiques. L’ensemble des allèles constitue le **génotype**. L’interaction génotype et environnement conditionne les caractéristiques de l’organisme (**phénotype**).

1. **Les débuts de la génomique**

Un bref regard en arrière souligne la vitesse avec laquelle notre connaissance des génomes s’est développée, ce qui peut laisser penser que les progrès futurs seront tout aussi spectaculaires.

* **1953** : Watson et Crick découvrent la structure de l’ADN
* **1956** : F. Sanger élabore le protocole de séquençage des protéines : Technique délicate à réaliser qui nécessite un matériel cher.
* **1977** : F. Sanger élabore le protocole de séquençage de l’ADN : Technique plus abordable et moins coûteuse.
* **Années 80** : • Technique de séquençage de l’ADN peu répandue,
* Détermination de la séquence d’un seul gène,
* Etude expérimentale et informatique de la fonction de la protéine codée par ce gène.

**Développement de la génomique**

* Automatisation des techniques expérimentales
* Mise au point de séquenceurs automatiques (le 1er en 1987). C'est le boom de la génomique qui a induit le développement de la bioinformatique (création de banques de données...)
* Création de laboratoires spécialisés en séquençage
* **1990** : lancement officiel du programme international du séquençage du génome humain
* **1991** : apparition de la première "puce à ADN" construite par la Société AFFYMETRIX
* **1995** : Séquençage du 1er génome procaryote Haemophilus influenzae (1,83 Mb)
* **1996** : Séquence du 1er génome eucaryote Saccharomyces cerevisiae (12 Mb)
* **1997** : séquençage du génome de la bactérie Escherichia coli
* **1998** : séquençage du génome du nématode Caenorhabditis elegans
* **2000** : séquençage des génomes de la drosophile, de la souris et de la plante Arabidopsis
* **2001 :** premières données sur le séquençage complet du génome humain
* **2005** : 1740 projets génomes existent parmi lesquels les projets de génomes modèles.

De l’ensemble de ces avancées, on peut retenir quelques grandes idées parfois surprenantes :

L’homme partage **40%** de ses gènes avec les plantes, plus de **80%** avec les mammifères et **98,5%** avec les chimpanzés, et du point de vue génétique, les êtres humains sont à **99,9%** identiques. Mais les faibles différences **(0,1%)** autorisent la détermination des empreintes génétiques à partir de l’ADN. En effet, deux êtres humains diffèrent par 3 millions de bases nucléotidiques sur 3 milliards au total. Ces empreintes individuelles ont de multiples applications comme la recherche de paternité, la recherche en criminologie chez l’homme, la traçabilité des produits carnés chez les bovins. Environ la moitié du génome humain est constitué de longues séquences répétitives sans fonction apparente. Seulement **5** à **10%** du génome correspond à des gènes qui peuvent être exprimés. L’étude du transcriptome ne s’adresse donc qu’à une toute petite partie du génome, mais certainement la plus importante car fonctionnelle.

1. **Apparition de la génomique**

**La génétique directe** : C’est l’étude génétique qui commence par l’observation d’une variation visible dans la morphologie ou la physiologie – c’est-à-dire elle commence par la recherche d’une différence génétique qui provoque une différence phénotypique.

**La génétique inverse :** Avec le développement de notre connaissance de l’ADN et de la façon dont il code l’information de la séquence d’acides aminés des protéines ainsi que le contrôle cellulaire de la production de celles-ci, une autre forme d’étude génétique devient possible : L’étude basée sur l’ADN. Au contraire, le programme de génétique inverse débute par des changements génétiques connus et par la recherche des modifications qui en résultent dans l’organisme.

Une analyse par génétique inverse est basée sur l’utilisation d’une séquence normale d’ADN. En insérant une mutation dans l’ADN (ou en le comparant avec l’ADN d’autre génomes), on peut analyser la fonction de la séquence d’ADN.

À l’aide d’une séquence normale d’ADN, on peut lire la séquence d’acides aminés de la protéine synthétisée (correspondance code d’ADN/ AA codés).

L’ADN peut ensuite être modifié de façon spécifique, aboutit à un blocage total de la production de la protéine ou à un changement de la protéine, (modification ou suppression de son activité métabolique).

Les conséquences d’une telle mutation complète sur le développement du phénotype peuvent ensuite être suivies de la même façon que les effets mutationnels dans le programme de génétique directe.

L’un des avantages du programme inverse est qu’il permet la création sur mesure d’un grand nombre de mutations spécifiques.

Au lieu de cela, on utilise l’information déjà disponible sur les divergences des génomes des différentes espèces apparentées.

Les génomes complets de nombreux organismes ont été séquencés grâce aux applications des techniques appropriées, ce qui a donné lieu à une nouvelle discipline au sein de la génétique, appelée **génomique**

1. **Définition**

La génomique est l’étude de l’ensemble des gènes d’une espèce (génome), de leur disposition sur les chromosomes (établissement de cartes génétiques et physiques), de la description fine de leur structure et de leurs fonctions. Pour le moment, les entreprises de génomique s’attachent essentiellement au séquençage du génome et dans une moindre mesure à la recherche de la fonction des gènes séquencés.

La génomique est composée de trois volets complémentaires : la génomique structurale, la génomique fonctionnelle et la protéomique.

* **La génomique structurale** : analyse la structure des gènes et d’autres parties du génome, donc elle contribue à l’annotation des génomes et à l’identification des séquences informatives (les gènes avec ou sans introns codant des protéines ou des ARN fonctionnels, les séquences régulatrices, les séquences répétés, les éléments transposables, la cartographie physique du génome, le séquençage systématique du génome)
* L**a génomique fonctionnelle** : analyse la fonction des gènes et autres parties du génome (elle inclue l’analyse du transcriptome (ARNm) elle contribue aussi à l’annotation des génomes et à l’identification des séquences informatives- la mutagénèse intentionnelle, le séquençage partiel de gènes exprimés (Expressed Sequence Tags ou EST), la transformation génétique, l'évaluation fine du matériel transformé)
* **La protéomique** : c’est l’analyse du protéome (protéine). Elle contribue aussi à l’annotation des génomes et à l’identification des séquences informatives. Ces trois approches sont complémentaires.
1. **Applications de la génomique**
2. La génomique est utilisée dans le cas de maladies infectieuses (recherche de gènes de résistance aux antibiotiques, recherche et identification de gènes impliqués dans la pathogénicité)
3. La génomique est utilisée dans le cas de maladies génétiques :
* recherche de gènes responsables de susceptibilité génétique aux maladies
* recherche de modifications de l'ADN à l'origine du cancer
* recherche de modification des profils d'expression des gènes, à l'origine de certains disfonctionnements cellulaires.
1. La génomique est utilisée dans l'industrie :
* recherche de gènes impliqués dans les pathologies (cibles pharmaceutiques, marqueurs diagnostiques)
* recherche de gènes codant des molécules d'intérêt
* recherche de gènes responsables de résistances
* recherche de gènes permettant de comprendre les mécanismes clés (cancer, vieillissement, etc)
1. **Génomique Végétale :** Elle permet d’identifier ce que l’on peut appeler des **gènes d’intérêt** intervenant dans :
* la résistance aux pathogènes (virus, bactéries, champignons),
* la qualité (saveur des fruits et des légumes, composition en acides gras ou en vitamines, qualité boulangère...),
* mais également dans les performances agronomiques (précocité, architecture, économies d’intrants, rendement...).
* Exploitant la diversité existant entre tous les individus d’une même espèce et les formes sauvages apparentée.
* La génomique permet d'approfondir notre connaissance des principales voies métaboliques dans les plantes et de rationaliser la sélection de nouvelles variétés en combinant les caractères les plus « performants » dans des génotypes élites.
1. **Complexité des génomes des plantes :** La diversité du règne végétal reflète la grande complexité des génomes des plantes :
* Taille parfois considérable,
* Duplications partielles ou complètes,
* Addition de génomes issus d’espèces apparentées, fruits d’une évolution naturelle depuis des millions d’année mais aussi des différentes formes de sélection (domestication) pratiquées par l’homme au cours des siècles.

Tout ceci nécessite la mise en place de méthodes et d’outils spécifiques. Les chercheurs ont choisi de travailler sur des espèces dites **modèles**, à génome de taille relativement réduite, comme l’arabette (*Arabidopsis thaliana*) et le riz.

L’intégration des données obtenues sur ces espèces modèles et leur extrapolation à l’étude des espèces d’intérêt agronomique, comme le blé, le maïs ou le colza, imposent l’utilisation de techniques de biologie moléculaire à haut débit et le développement d’outils bioinformatiques puissants.

1. **Objectifs de la génomique végétale**
* Etude de l’organisation des gènes et des séquences régulatrices
* Etude de l’expression des gènes dans l’espace et le temps
* Etude de la structure et de la fonction des protéines et des ARN pour lesquelles les gènes codent
* Analyse d’un grand nombre d'organismes possibles
* Utiliser ces informations dans les études du métabolisme (métabolome), des interactions entre les macromolécules biologiques (interactome), dans le développement…etc.
* La génomique vise à caractériser la fonction des milliers de gènes grâce à des approches globales (transcriptome : étude globale de l’expression du génome ; protéome et métabolome), les réseaux d’interaction entre eux en réponse à divers stress abiotiques (exp sécheresse) ou biotiques (exp attaque par un champignon phytophatogène).
* L’analyse fonctionnelle du génome de plantes modèles (Arabidopsis, riz)
* Développement des outils pour la génomique végétale (transcriptomique, protéomique, métabolomique, bioinformatique, production de lignées mutantes, production de populations recombinantes)
1. **La bioinformatique :** c’est une science nouvelle, se fondant sur le séquençage massif et récent du génome de plusieurs organismes procaryotes et eucaryotes, dont celui des plantes supérieures (Arabidopsis, riz). Sa mission est double :

**1/ Annoter le génome des plantes :** L’un des enjeux majeur est de comprendre la structure des gènes (détermination des séquences promotrices gouvernant l’expression et la régulation des gènes, des séquences codantes dictant la synthèse des protéines impliquées dans l’activité cellulaire).Un autre enjeu de cette annotation est de comprendre la fonction des gènes ; il ressort en effet que près de 50 % des gènes prédits par le séquençage du génome d’Arabidopsis ont une fonction inconnue.

**2/ Mettre en place des bases de données :** regroupant des informations structurales sur les gènes et les génomes et leur expression dans des contextes physiologiques. Ces bases de données constituent l’un des outils majeurs, via la génomique comparative, pour repérer des gènes jouant un rôle dans le développement ou en réponse aux stress biotiques et abiotiques.

**Conclusion**

La génomique structurale, la génomique fonctionnelle, la transcriptomique, la protéomique et la bioinformatique sont des approches complémentaires.

Le matériel de base de la génomique est l'ensemble des séquences d'acides nucléiques et des séquences polypeptidiques obtenues par différentes méthodes de séquençage.

Ces séquences et d'autres types d'informations qui découlent de leur analyse sont stockés dans des bases de données. L'accès aux bases de données s'effectue via le Web et Internet.

L'analyse de l'ensemble de ces données nécessite des méthodes bioinformatiques.