

Chapitre1 Introduction et Notions Préliminaires

La génomique est l'étude exhaustive et multidisciplinaire des génomes, tant en ce que concerne le nombre et la disposition précise des gènes sur les chromosomes, que leur séquence et leur fonction. Il s'agit donc du déterminisme des grands processus cellulaires et de leur intégration au niveau des organismes entiers. Les données qu'elle fournit vont également éclairer de façon décisive les principes qui sous-tendent leur développement, leur adaptation au milieu et leur filiation évolutive.

L'objectif premier de la plupart des projets de génomique consistait à déterminer la séquence d'ADN, soit du génome, soit d'un grand nombre de transcrits

Cette discipline a connu un essor avec notamment le développement des techniques de séquençage des génomes. La génomique fut très médiatisée à la fin du 20^{ème} siècle avec la compétition entre différentes équipes scientifiques pour la publication de la première carte du génome humain.

La génomique se divise en deux branches : la génomique structurale, qui se charge du séquençage du génome entier, et la génomique fonctionnelle, qui vise à déterminer la fonction et l'expression des gènes séquencés en caractérisant le transcriptome et le protéome.

Le génome est l'ensemble du matériel génétique d'un individu ou d'une espèce codé dans son ADN (à l'exception de certains virus dont le génome est porté par des molécules d'ARN). Il contient en particulier toutes les séquences codantes (transcrites en ARN messagers, et traduites en protéines) et non-codantes (non transcrites, ou transcrites en ARN, mais non traduites).

Chez les virus, le génome est contenu soit dans une (ou plusieurs) molécule(s) d'ADN ou d'ARN, à simple ou double brin.

Chez les procaryotes (bactéries et archées), le génome est généralement contenu dans une molécule d'ADN circulaire. Peut aussi exister un génome extrachromosomique, contenu dans des plasmides et des épisomes.

Chez les eucaryotes, on distingue :

- le génome nucléaire, contenu dans le noyau qui caractérise les eucaryotes. C'est de ce génome dont on parle en général quand on parle du génome d'un eucaryote (animal, plante, champignon, etc.) ;
- les génomes non-nucléaires, contenus dans des organites :

- le génome mitochondrial, contenu dans les mitochondries, chez la quasi totalité des eucaryotes ;

- le génome chloroplastique, contenu dans les chloroplastes, chez les algues et les plantes supérieures.

Chez quelques eucaryotes (par exemple la levure) sont aussi présents des plasmides (de taille réduite).

Chez l'homme en particulier (organisme eucaryote), le génome nucléaire est réparti sur 46 chromosomes, soit 22 paires d'autosomes et deux gonosomes (XX chez la femme, XY chez l'homme). Il ne faut pas confondre le génome et le caryotype, qui caractérise les chromosomes.

La taille du génome se mesure en nombre de nucléotides, ou *bases*. La plupart du temps, on parle de pb (pour *paire de bases*, puisque la majorité des génomes est constituée de doubles brins d'ADN ou bien d'ARN). On emploie souvent les multiples kb (pour kilo-base) ou Mb (méga-base), qui valent respectivement 1 000 et 1 000 000 bases. La taille du génome peut aussi être exprimée en pg (picogrammes), ce qui correspond à la masse d'ADN (haploïde) par cellule. 1 pg représente environ 1 000 Mpb.

La taille du génome peut varier de quelques kilo-bases chez les virus à plusieurs centaines de milliers de Mb chez certains eucaryotes. La quantité d'ADN, contrairement à ce qui a été longtemps supposé, n'est pas proportionnelle à la complexité d'un organisme ; ainsi, l'amibe *Amoeba dubia*, un organisme unicellulaire, a un génome environ 200 fois plus grand que *Homo sapiens*. Ce constat est fréquemment appelé paradoxe de la valeur C.

Le *Paris japonica*, du genre *Paris* a le génome le plus vaste décrit : il comporte près de 150 milliards de paires de base, soit près de 50 fois la taille du génome humain

L'annotation d'un génome consiste à traiter l'information brute contenue dans la séquence dans le but :

1. de prédire, le contenu en gènes, la position des gènes à l'intérieur d'un génome (le début, la fin, et chez les eucaryotes, les introns et les exons), ainsi que leur organisation (gènes uniques ou en opéron, avec des séquences promotrices, des terminateurs, des sites de fixation ribosomiques "RBS" ...). Dans ce cas, on parle d'annotation structurale.
2. de prédire la fonction potentielle de ces gènes (leur attacher une étiquette, portant leur nom probable, leur fonction probable, leurs interactions probables). Dans ce cas on parle d'annotation fonctionnelle.

Les 2 types d'annotations

L'annotation peut être automatique c'est-à-dire s'appuyer uniquement sur des algorithmes recherchant des similarités (de séquence, de structure, de motifs, ...), permettant de prédire (en fait deviner) la fonction d'un gène. Elle aboutit au transfert « automatique » de l'information figurant dans l'étiquette d'un gène « similaire » d'un génome déjà annoté au génome en cours d'annotation. L'annotation automatique est parfois complétée par :

Une annotation manuelle par des experts qui valident ou invalident la prédiction en fonction de leurs connaissances ou de résultats expérimentaux. Celle-ci peut ainsi éviter le transfert automatique d'erreurs et donc leur propagation, ce qui peut devenir le grand problème que devra affronter la génomique, compte tenu de l'afflux massif de données issues en particulier, des nouvelles techniques de séquençage (voir notion de pyroséquençage dans les prochains chapitres).
