



Université Batna 2

Faculté des sciences de la nature et de la vie

Département de biologie des organismes

**Cours de**  
**Génomique et protéomique fonctionnelle**  
**M2 BMC-BPC**  
**Dr. LAANANI.I**

2022-2023

## Table des matières

<b>I- Etude de quelques projets du génome et les ressources d'internet .....</b>	<b>2</b>
<b>1. Le Projet Génome Humain :.....</b>	<b>2</b>
<b>1.1. Les objectifs de Projet Génome Humain :.....</b>	<b>2</b>
<b>1.2. Les caractéristiques du génome humain : .....</b>	<b>3</b>
<b>1.3. Les ressources de l'internet : .....</b>	<b>4</b>
<b>1.4. Les aspects éthiques, juridiques et sociaux du projet Génome humain :.....</b>	<b>4</b>
<b>2. Les projets génomes concernant des virus :.....</b>	<b>5</b>
<b>3. Les projets génomes concernant des procaryotes :.....</b>	<b>5</b>
<b>Les bactéries :.....</b>	<b>5</b>
<b>4. Les projets concernant les génomes des eucaryotes : .....</b>	<b>6</b>
<b><i>Saccharomyces cerevisiae</i> (<i>S. cerevisiae</i> levure, eucaryote unicellulaire): .....</b>	<b>6</b>
<b><i>Caenorhabditis elegans</i> (vers nématodes, métazoaire, eucaryote) :.....</b>	<b>6</b>
<b><i>Drosophila melanogaster</i> (La mouche ou mouche du vinaigre) :.....</b>	<b>7</b>
<b><i>Arabidopsis thaliana</i> (Plantes, une mauvaise herbe) : .....</b>	<b>7</b>

## I- Etude de quelques projets du génome et les ressources d'internet

Le rassemblement et l'analyse des informations génomiques représentent un travail à grande échelle, requérant la coordination de nombreux laboratoires. Pour cette raison les généticiens forment souvent des collaborations, appelées des projets génomes. L'un des plus grands et le plus connu est Le Projet Génome Humain.

Les premiers organismes dont l'ADN a été séquencé sont un virus de bactérie et de petites bactéries. En 1985, les scientifiques se lancent dans le séquençage de l'ADN d'organismes multicellulaires, tout d'abord celui du ver *Caenorhabditis elegans* (finalisé en 1998), puis celui d'une plante, *Arabidopsis thaliana* (achevé en 2000), avant d'envisager de déterminer la séquence de l'ADN humain, qui compte environ trois milliards de nucléotides répartis sur vingt-trois paires de chromosomes.

### 1. Le Projet Génome Humain :

Le Projet Génome Humain (*Human Genome Project* HGP) est un programme entrepris en 1990 dont la mission était d'établir le séquençage complet de l'ADN du génome humain. Le séquençage avait déjà été réussi pour *Escherichia Coli*, *Saccharomyces Cerevisiae* et *Drosophila Melanogaster*. Il impliquait plus de 20 laboratoires de 6 pays (USA, Grande -Bretagne, Japon, France, Allemagne et Chine), l'ensemble étant coordonné par une collaboration internationale. Cette entreprise de grande ampleur est le résultat d'une coopération scientifique internationale qui s'est étalée sur près de quinze ans. Ces groupes ont utilisé la méthode de séquençage ordonnée ou hiérarchique. La séquence brute du génome humain est livrée en février 2001 et la séquence complète en 2003. Son achèvement a été annoncé le 14 avril 2003.

Un projet concurrent, nommé *Celera Genomics*, lancé par des sociétés pharmaceutiques privées, disposant de beaucoup plus de moyens que HGP a aussi tenté de séquencer le génome humain. La société *Celera Genomics* s'est lancée dans la bataille en 1998 en utilisant une stratégie de séquençage aléatoire.

Le projet Génome humain comportait en outre l'établissement de la séquence des nucléotides des génomes d'autres organismes. Plusieurs d'entre eux ont été utilisés pour des expériences en laboratoire, ce qui fait que certains de leurs mécanismes biochimiques et certaines de leurs fonctions physiologiques sont très bien connus. Comme le génome de ces organismes présente de nombreuses ressemblances avec le génome humain, la connaissance de leurs séquences aidera à l'identification de gènes humains et à l'établissement de la fonction de ces derniers.

#### 1.1. Les objectifs de Projet Génome Humain :

On peut brièvement résumer les objectifs du Projet Génome Humain comme suit :

- La construction d'une séquence complète du génome ;
- L'identification de tous les gènes portés par l'ADN humain ;

- La création de cartes génétiques et physiques à haute résolution pour aider à la localisation de gènes associés à une maladie ;
- La compilation de banques contenant l'ensemble des données de polymorphisme, en particulier de polymorphisme de nucléotide unique (SNP) ;
- L'amélioration des performances du séquençage automatique ;
- L'élaboration des outils informatiques permettant de traiter, d'archiver et de communiquer l'ensemble des données produites.

### 1.2. Les caractéristiques du génome humain :

Le génome humain est l'ensemble de l'information génétique portée par l'ADN sur les 23 paires de chromosomes présent dans le noyau plus l'ADN mitochondrial (hérité de la mère uniquement). Il porte l'ensemble de l'information génétique humaine, estimée à 100 000 gènes avant le séquençage et qui s'est révélée contenir finalement de 25 000 à 30 000 gènes. Le génome humain est complexe mais pas plus qu'un autre génome. Il contient  $3,2 \times 10^9$  pb dont seulement 5 % codent pour des gènes et on estime qu'il contient 25 000 à 30 000 gènes codant apparemment pour un nombre très important de protéines. Plus de 40% des gènes identifiés n'ont pas de fonction moléculaire connue.

Il a été démontré que **95 à 98 %** d'ADN est non-codant (pour une protéine) et moins de 5 % (**2-5%**) d'ADN est codant pour une protéine. Certains chercheurs dénomment l'ADN non-codant de l'ADN poubelle (*JUNK DNA*). En fait, la nature faisant rarement les choses au hasard, cet ADN aurait des fonctions de régulations très importantes pour la cellule.

**Le chromosome 19** a la plus forte densité génique et **les chromosomes 13 et Y** ont la plus faible densité génique. Les gènes humains sont plus grands et contiennent des introns plus nombreux et plus grands que les gènes des génomes des invertébrés tels que celui de la drosophile. Le plus grand gène humain connu code **la dystrophine**, une protéine musculaire. Ce gène, associé quand il est muté, à la dystrophie musculaire (myopathie), a une longueur de 25 Mb.

Il y a parmi les 3,2 milliards de paires de bases du génome humain environ 7 millions de variantes existant à une fréquence  $> 5$  %, ce sont en majorité des polymorphismes ponctuels, c'est-à-dire des **SNPs**.

-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 99,5% identiques. Mais les faibles différences (0,5%) autorisent la détermination des empreintes génétiques à partir de l'ADN.

### 1.3. Les ressources de l'internet :

Une ressource clef de l'internet pour tous les projets de génomiques est le serveur Web géré par le Centre National d'Information en Biotechnologie (National Center for Biotechnology Information, NCBI), sous l'égide de l'institut de Santé des Etats-Unis et de la Bibliothèque Nationale de Médecine (<http://www.ncbi.nlm.nih.gov>). Des liens informatiques particuliers offerts par ce site, tel le moteur de recherche BLAST et la banque de données de séquences Genbank.

Le guide des ressources consacrées au génome humain (Human Genome Resources) du site NCBI fournit un point de départ pour faire de la recherche en génomique humaine. A partir d'idiogrammes de chromosomes comportant des liens hypertexte « cliquable », on a directement accès à des cartes physiques à haute résolution de n'importe quel segment du génome, avec la possibilité de télécharger en quelques secondes plusieurs centaines de kilobases de séquences achevées, correspondant à une région d'intérêt. Différents logiciels de visualisation du génome sont disponibles, chacun ayant ses avantages et ses caractéristiques (Figure 1).



**Figure1:** Les ressources disponible sur le site Web NCBI

En plus des sites associés au NCBI, un site consacré au génome humain particulièrement populaire est celui fourni par le « Projet Ensembl » du Sanger Centre et du European Bioinformatics Institute (EBI), il a été financé par le groupe Wellcome (<http://www.ensembl.org>).

### 1. 4. Les aspects éthiques, juridiques et sociaux du projet Génome humain :

#### L'existence de l'information génétique

- La connaissance de la prédisposition à certaines maladies et la capacité de mettre au point des thérapies « faites sur mesure » peuvent être d'une grande utilité pour le traitement des maladies.
- Les données de projets génome humain sont utilisées pour accentuer les **différences raciales** et justifier des mesures racistes, on craint que l'information issue du projet Génome

humain, de même que de projets complémentaires comme **le projet Diversité du génome humain**, n'exacerbe le racisme dans un monde où il est déjà beaucoup trop répandu.

- Le projet Génome humain a pour but de prévenir les incapacités et les maladies, d'augmenter la longévité, de réduire la mortalité infantile et d'accroître l'intelligence. Étant donné que nous savons déjà que des **facteurs environnementaux et sociaux** interviennent dans des maladies.
- Constitution d'une banque de données d'ADN pour permettre au système judiciaire de prouver l'innocence de personnes accusées à tort de certains crimes.
- L'existence des gènes peuvent pousser des personnes à commettre des actes criminels (des personnes avec **une propension génétique à la criminalité**).
- Un gène associé à un risque de maladie.

#### **Propriété et commercialisation :**

- Le 11 novembre 1997, l'UNESCO adoptait sa Déclaration universelle sur le génome humain et les droits de l'Homme. L'article 4 prévoit que « le génome humain en son état naturel ne peut donner lieu à des gains pécuniaires ».
- **L'octroi de brevets pour des séquences d'ADN.**

#### **Thérapie génique :**

- Remplacer un gène défectueux : cellules germinales ou somatiques
- Le système CRISPR CAS

## **2. Les projets génomes concernant des virus :**

Les premiers génomes à avoir été séquencés avec succès sont les virus parce qu'ils ont des codes génétiques plus courts. Le génome viral du bactériophage fX174 (qui ne comprend que 5 368 paires de bases) a été le premier organisme séquencé.

## **3. Les projets génomes concernant des procaryotes :**

### **Les bactéries :**

Les premiers génomes de cellules à avoir été séquencés, étaient ceux de procaryotes d'abord celui d'*Haemophilus influenzae*, rapidement suivi par celui de *Mycoplasma genitalium* puis en 1997 celui d'*Escherichia coli*.

En 1995, pour la première fois, la séquence complète du génome d'une cellule vivante a été déterminée. Il s'agissait d'*Haemophilus influenzae*, une bactérie responsable d'infections bronco-pulmonaires chez les jeunes enfants, dont le génome est composé d'un seul chromosome circulaire long de 1,8 Mb et comportant 1 727 gènes.

*Mycoplasma genitalium* est un mycoplasme endoparasitaire des cellules épithéliales du tractus urogénital humain, agent infectieux pathogène pour l'homme, responsable d'urétrites et

d'autres maladies sexuellement transmissibles, éventuellement en association avec un autre mycoplasme. 481 gènes sont prédits pour le génome de 0.6 Mb de *M. genitalium*. La longueur moyenne d'un gène dans chaque espèce est proche de 1,1 kb.

*Escherichia coli*, également appelée colibacille et abrégée en *E. coli*, est une bactérie intestinale (Gram négatif) des mammifères, très commune chez l'être humain. En effet, elle compose environ 80 % de notre flore intestinale aérobie. *E. coli* possède un génome de 4,6 Mb contenant 4 288 gènes avec une densité très proche d'un gène par kilobase. Le patrimoine génétique de la souche *E. coli* de laboratoire non pathogène a été entièrement séquencé en 1997. Son génome comprend 4,6 Mb codant environ 4 200 protéines.

#### 4. Les projets concernant les génomes des eucaryotes :

##### ***Saccharomyces cerevisiae* (*S. cerevisiae* levure, eucaryote unicellulaire):**

*Saccharomyces cerevisiae* est une levure unicellulaire, organisme eucaryote appartenant au règne des champignons et à la famille des saccharomycètes. Les levures sont des champignons unicellulaires.

La séquence complète du génome de *S. cerevisiae* (appelée également levure de boulanger ou levure de bière) est l'une des premières à avoir été obtenue au printemps 1996, suite à un effort international. C'est le premier génome eucaryote séquencé. Il s'agit de l'un des organismes les plus étudiés du point de vue expérimental (génétique et biologie moléculaire). *S. cerevisiae* possède un génome de 12 Mb avec 6200 gènes se répartissent sur 16 chromosomes. On pense que seulement environ 5.800 de ces gènes sont fonctionnel. On estime que l'Homme partage 23 % de ses gènes avec cette levure. L'ensemble des données génomiques concernant *Saccharomyces cerevisiae* sont rassemblées sur [Saccharomyces Genome Database](#). *S. cerevisiae* est un organisme modèle car elle est l'organisme eucaryote le plus simple. Les chercheurs peuvent donc essayer de comprendre les procédés des mécanismes biologiques basiques avant de le faire chez l'Homme. Elle présente l'avantage de se reproduire rapidement et peut également être transformée pour exprimer de nouveaux gènes, tester des molécules...

##### ***Caenorhabditis elegans* (vers nématodes, métazoaire, eucaryote) :**

*Caenorhabditis elegans* est un petit ver d'un millimètre de long environ, transparent et non-parasitaire. Le nématode *C. elegans* est le premier organisme multicellulaire dont le génome est entièrement séquencé.

En 1998, ce ver fut le premier animal dont le génome ait été totalement séquencé. Il possède un génome de 97 Mb réparti sur 6 chromosomes, avec un total de 19 099 gènes environ. La densité génique est beaucoup plus faible que chez la levure, avec une densité moyenne d'environ 1 gène pour 5 kb. La densité moyenne est d'un gène tous les 5 kb. L'ensemble des

informations génomiques concernant cet organisme sont compilées dans WormBase (<http://www.wormbase.org>).

***Drosophila melanogaster* (La mouche ou mouche du vinaigre) :**

La drosophile ou mouche du vinaigre (*Drosophila melanogaster*) mesure quelques millimètres de long et est connue pour pondre dans les fruits. Le génome des *Drosophila* contient 4 paires de chromosomes : une paire X/Y, et trois autosomes appelés 2, 3, et 4. Le génome contient environ 165 millions de bases et environ 13 000 gènes. Le génome a fini d'être séquencé et annoté en 2000. D'un point de vue génétique, les êtres humains et les drosophiles ont des similitudes significatives. Selon une étude de mars 2000 de l'Institut américain de recherche sur le génome humain, environ 60 % des gènes sont conservés entre les deux espèces. Selon une analyse de 2001, 77 % des gènes associés à des maladies humaines identifiées ont un homologue dans le génome de la drosophile. 50 % des protéines de cette mouche ont des analogues chez les mammifères. *Drosophila* est utilisée comme modèle génétique pour diverses maladies humaines dont la maladie de Parkinson.

***Arabidopsis thaliana* (Plantes, une mauvaise herbe) :**

L'arabette des demoiselles ou *Arabidopsis thaliana* est une espèce de plantes appartenant à la famille des Brassicacées. Cette petite plante annuelle, originaire d'Europe, d'Asie et du nord-ouest de l'Afrique, a un cycle de vie rapide, elle est résistante et peut s'autoféconder. Ces qualités en ont fait depuis 1998 un organisme modèle de référence, pour la recherche aussi bien biologique que génétique.

En 2000, le génome d'*Arabidopsis thaliana* a été le premier génome de plante à être totalement séquencé. C'est une petite plante qui est utilisé comme modèle génétique pour étudier l'organisation du génome et le développement des plantes florales. Son génome de 120 Mb est distribué sur 5 paires de chromosomes. Ce génome contient environ 30 000 gènes avec une densité de l'ordre de 1 gène pour 5 bases. La génomique des plantes a joué un rôle croissant dans le développement de cultures résistantes aux parasites, aux herbicides, et à la sécheresse et plus riches en nutriment. L'ensemble des ressources concernant *Arabidopsis thaliana* (séquences, lignées, graines, expression des gènes, etc.), sont compilées par The Arabidopsis Information Resource et l'European Arabidopsis Stock Centre.

Peu de temps après le séquençage du génome humain, ils ont achevé celui du riz, de l'abeille à miel, du poulet et de la vache. En 2009, ils ont parvenu à une ébauche des génomes des germes de soja, du cheval, du cochon, du sorgho et du maïs. En 2010, se sont ajoutés à la liste le blé, les pommes, les cultures de graines oléagineuses et de biodiesel. En 2011, ont été séquencés les

génomomes du colza, d'autres arbres fruitiers et des plantes modèles. En 2012, se sont ajoutés à la liste de génomes importants pour l'agriculture ceux du coton, des tomates et des bananes.

Les organismes dont le poids économique est important dans les applications biotechnologiques industrielles ont fait preuve d'une extension similaire quant au séquençage. Nombre de protistes, tels que les amibes et les parasites, les moules, les champignons et les algues, ont été séquencés. Le séquençage des agents pathogènes pour les êtres humains et les animaux a suscité un intérêt particulièrement marqué puisqu'on espérait qu'il permettrait de trouver des méthodes de diagnostic et de traitement de maladies. Le séquençage d'autres organismes présente un énorme intérêt pour l'industrie grâce aux produits qui pourraient en découler, comme l'huile à partir des algues et les antibiotiques à partir des moules et des champignons.

**Tableau 1** : Listes des génomes complètement séquencés (2011).

Nom d'espèce	Nom commun	Année de publication	Taille du génome Mb	Nombre de gènes	Distance moyenne entre gènes Kb	Fraction couverte par des gènes codants %	Fraction non-codante %	Fraction répétitive %	Fraction transcrite %	Remarques
<b>Bactérie</b>										
<i>Mycoplasma genitalium</i>	<i>Mycoplasma</i>	1995	0.6	481	1.2	90	10			Petit génome (Intracellulaire)
<i>Haemophilus influenzae</i>		1995	1.8	1 717	1.0	86	14			Premier génome bactérien séquencé
<i>Escherichia coli</i>	<i>Entérobactérie</i>	1997	4.6	4 289	1.1	87	13			
<b>Levures</b>										
<i>Saccharomyces cerevisiae</i>	<i>Levure du boulanger</i>	1996	12	6 286	1.9	72	28			Premier génome eucaryote
<b>Animaux</b>										
<i>Caenorhabditis elegans</i>	<i>Ver nématode</i>	1998	97	19 000	5	27	73			Premier génome de métazoaire
<i>Drosophila melanogaster</i>	<i>Mouche à vinaigre</i>	2000	165	16 000	10	15	85			
<i>Clona intestinalis</i>			174	14 180	12					
<i>Danio rerio</i>	<i>Poisson zèbre</i>		1 527	18 957	81					
<i>Xenopus laevis</i>	<i>Xénope (amphibien)</i>		1 511	18 023	84					
<i>Gallus gallus</i>	<i>Poule</i>		2 961	16 736	177					
<i>Ornithorynchus anatinus</i>	<i>Ornithorynque</i>		1 918	17 951	107					
<i>Mus musculus</i>	<i>Souris</i>	2002	3 421	23 493	146					
<i>Pan troglodytes</i>	<i>Chimpanzé</i>		2 929	20 829	141					
<i>Homo sapiens</i>	<i>Humain</i>	2001	3 200	21 528	149	2	98	46	28	Version "brouillon"
1000 génomes humains		> 2008								Projet annoncé en janvier 2008
<b>Plantes</b>										
<i>Arabidopsis thaliana</i>	<i>Arabette</i>	2001	120	27 000	4	30	70			Premier génome de plante
<i>Oryza sativa</i>	<i>Riz</i>		390	37 544	10					
<i>Zea mays</i>	<i>Maïs</i>		2 500	50 000	50			50		Nb de gènes approximatif
<i>Triticum aestivum</i>	<i>Blé</i>		16 000							Génome hexaploïde
<i>Lilium</i>	<i>Lys</i>		120 000							
<i>Psilotum nudum</i>			250 000							

**ANNEXE** : Liste des bases de données

<ul style="list-style-type: none"> <li>• <b>Base de Données ADN</b> <ul style="list-style-type: none"> <li>– GenBank, DDBJ, EMBL,...</li> </ul> </li> <li>• <b>Base de Données Protéines</b> <ul style="list-style-type: none"> <li>– PIR, Swiss-Prot, PRF, GenPept, TrEMBL, PDB,...</li> </ul> </li> <li>• <b>Base de Données EST</b> <ul style="list-style-type: none"> <li>– dbEST, DOTS, UniGene, GIs, STACK,...</li> </ul> </li> <li>• <b>Base de Données Structure</b> <ul style="list-style-type: none"> <li>– MMDB, PDB, Swiss-3DIMAGE,...</li> </ul> </li> <li>• <b>Base de Données voies métabol.</b> <ul style="list-style-type: none"> <li>– KEGG, BRITE, TRANSPATH,...</li> </ul> </li> <li>• <b>Base de Données intégrées</b> <ul style="list-style-type: none"> <li>– SRS</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• <b>Base de Données de Motifs</b> <ul style="list-style-type: none"> <li>– Prosite, Pfam, BLOCKS, TransFac, PRINTS, URLs,...</li> </ul> </li> <li>• <b>Base de Données sur les maladies</b> <ul style="list-style-type: none"> <li>– GeneCards, OMIM, OMIA,...</li> </ul> </li> <li>• <b>Base de Données taxonomique</b></li> <li>• <b>Base de données littérature scient.</b> <ul style="list-style-type: none"> <li>– PubMed, Medline,...</li> </ul> </li> <li>• <b>Base de données de brevets</b> <ul style="list-style-type: none"> <li>– Apipa, CA-STN, IPN, USPTO, EPO, Beilstein,...</li> </ul> </li> <li>• <b>Autres...</b> <ul style="list-style-type: none"> <li>– RNA databases, QTL...</li> </ul> </li> </ul>
---	---