Course Draft

# 📚 Course Content

## 1. Introduction

### Notions fondamentales

* L’arbre de la vie illustre les concepts majeurs de l’évolution, du passage du temps, de l’adaptation et de la sélection naturelle, ayant mené à la diversification des espèces depuis le dernier ancêtre commun universel (LUCA). Les eucaryotes se regroupent en fonction de besoins économiques, techniques et de leur facilité d’étude en laboratoire, parmi lesquels on compte 35 espèces modèles, dont 18 métazoaires.
* L’homologie biologique désigne des caractères partagés par plusieurs espèces issus d'un ancêtre commun, qu’ils soient anatomiques, moléculaires ou génétiques. L’orthologie fait référence à des gènes dans différentes espèces issus d’un gène ancestral suite à une spéciation, tandis que la paralogie décrit des gènes issus d’une duplication au sein d’une même espèce. La comparaison de séquences permet d’estimer leur identité ou similarité, même si une faible similarité ne signifie pas toujours absence d’homologie, et la similitude peut parfois survenir par hasard ou convergence évolutive.
* Pour déterminer l’homologie, il est nécessaire d’analyser l’alignement des séquences génomiques ou protéiques et de calculer le pourcentage d’identité ou de similarité. Un seuil significatif est généralement supérieur à 20%, mais il dépend de la famille de protéines. Certaines séquences similaires peuvent n’être homologues ni par lignée ni par convergence.

## 2. Le génome humain

* Le génome humain est une référence biologique fondamentale, au même titre que le méthylome, le protéome et le transcriptome. Il comprend le génome nucléaire et mitochondrial, la structure des chromosomes, et présente à la fois une organisation complexe et une forte variabilité.
* Le génome humain se compose d’un génome nucléaire, contenant 22 paires d’autosomes et une paire de chromosomes sexuels (X et Y), et d’un génome mitochondrial à molécule circulaire. Le génome nucléaire est organisé en 3,1 milliards de paires de bases, comportant des régions euchromatiniques riches en gènes et actives, ainsi que des régions hétérochromatiniques condensées et inactives. Le génome mitochondrial, plus compact, comprend 37 gènes, dont 13 codants pour des protéines et 24 non-codants, et son héritage est strictement maternel, avec plus de 1000 copies par cellule.
* Le séquençage du génome humain, débuté en 1990 avec un grand projet international, a duré environ 15 ans et mobilisé des ressources importantes. Les cartes génétiques étaient fondées sur les fréquences de recombinaison, tandis que les cartes physiques mesuraient la distance réelle entre les gènes. Le séquençage a atteint une version "draft" en 2000 (90%) et a été finalisé en 2003 (99%). Plusieurs versions du génome humain existent, la plus récente étant GRCh38 publiée en 2013, continuellement améliorées par des patchs et des mises à jour.
* Pour l’exploration du génome, l’UCSC Genome Browser permet de visualiser la structure chromosomique, les annotations de gènes et leur contexte, ainsi que la conservation des séquences. On peut y examiner le sens de transcription, la distinction entre exons et introns, le zoom jusqu’au niveau des acides aminés et des nucléotides, et constater que seulement 5% du génome est conservé au cours de l’évolution.
* Le génome mitochondrial humain a une taille de 16,6 kb et ressemble étroitement à celui de certaines bactéries, confirmant son origine endosymbiotique. Il comporte 37 gènes, a une organisation circulaire, est hérité de manière maternelle, et présente parfois de l’hétéroplasmie, c’est-à-dire la coexistence de copies différentes au sein d’une même cellule. La densité génique y est élevée et la plupart des gènes codent pour des éléments indispensables à la phosphorylation oxydative.
* En comparaison, le génome nucléaire humain contient majoritairement des régions euchromatiniques actives (2,9 Gb) et une minime partie hétérochromatinique (0,2 Gb). Les régions hétérochromatiniques sont constituées principalement de séquences répétées et sont transcrites de manière inactive. La structure chromosomique et l’activation transcriptionnelle diffèrent sensiblement entre ces deux entités génomiques.
* Le génome humain comporte différents types de gènes : codants pour des protéines (~20 000 gènes) et non-codants (>40 000 gènes dont ARNt, ARNr, miARN, snARN, etc.), ainsi que de nombreux pseudogènes issus de divers mécanismes. Les éléments répétés forment une grande part du génome, comme les satellites, minisatellites, microsatellites, ainsi que les transposons (LINE, SINE, rétrotransposons, etc.).
* Les gènes peuvent être isolés ou organisés en clusters, parfois avec des pseudogènes associés. Les familles géniques importantes comprennent celles des histones ou des récepteurs olfactifs. Les éléments répétitifs sont en grande partie inactifs mais jouent un rôle dans l’évolution génomique.

### Projet de séquençage

* Le Projet Génome Humain fut une initiative internationale majeure commencée en 1990, menée sur environ 15 ans avec un budget de près de 3 milliards de dollars. Ses objectifs comprenaient la cartographie génétique et physique du génome humain, ainsi que l’obtention de la séquence entière de l’euchromatine.
* La cartographie génétique s'est basée sur les fréquences de recombinaison alors que la cartographie physique mesurait les distances réelles entre les marqueurs génétiques. Le séquençage a démarré en 1999 dans plusieurs centres mondiaux et la version "draft" (90%) a vu le jour en 2000, suivie par la publication d'une version finalisée à 99% en 2003. La gestion et la mise à jour du génome humain ont conduit à de nombreuses versions successives, la version GRCh38 étant actuellement la référence internationale. Les versions sont régulièrement enrichies de nouveaux patchs et corrections et il existe aujourd’hui une séquence complète pour tous les chromosomes.
* Des ressources comme l’UCSC Genome Browser permettent de visualiser et d’explorer finement le contenu du génome, d’accéder aux annotations des gènes et d’identifier leurs fonctions potentielles par rapport à l’évolution des différentes versions de référence.

### Architecture globale

* L’architecture globale du génome humain se compose de génomes nucléaire et mitochondrial. Le génome nucléaire est diploïde, constitué de 3,1 Gb, et subdivisé en euchromatine active (2,9 Gb, riche en gènes) et hétérochromatine inerte (0,2 Gb, zones hautement répétées). Des versions successives du génome de référence retracent les progrès dans la cartographie et le séquençage depuis les années 2000.
* La visualisation de la structure génomique, notamment par les navigateurs comme l’UCSC Genome Browser, permet d’explorer les gènes, leur organisation, les exons, introns, acides aminés, et les annotations associées. Seulement 5% du génome humain est conservé à travers les espèces au cours de l’évolution, témoignant d’une grande variabilité des séquences.
* Les éléments constitutifs du génome incluent des gènes, des zones répétées, des introns et des exons, et une part significative du génome est formée d’éléments répétés, tant en tandem que dispersés. Ces dernières comprennent des satellites, minisatellites, microsatellites et des transposons, avec un impact marquant sur la structure et la fonction du génome.

#### Génome mitochondrial

* Le génome mitochondrial humain est compact (16,6 kb), entièrement séquencé dès 1981, et contient 37 gènes : 13 codants (indispensables à la phosphorylation oxydative) et 24 ARN non codants (22 ARNt, 2 ARNr). Il utilise un code génétique légèrement différent du code universel et il existe en de nombreuses copies par cellule.
* Son organisation rappelle celle d’une bactérie et il est hérité de façon strictement maternelle. On observe parfois une hétéroplasmie, reflet de la coexistence de différents types de mitochondries au sein d’une cellule, en quantité variable selon le tissu et l’âge.
* La comparaison des génomes mitochondriaux de différentes espèces, comme la drosophile, la levure ou l’arabette, révèle des différences de taille, de structure (circulaire ou parfois linéaire), du nombre de copies par cellule et du nombre/répartition de gènes codants ou ARN. Le génome mitochondrial humain est particulièrement dense en gènes et peu sujet aux séquences répétées.

#### Génome nucléaire

* Le génome nucléaire humain comporte environ 3,1 Gb (haploïde). Il fut séquencé sous forme de brouillon en 2001 et finalisé en 2022. Il se divise en hétérochromatine (0,2 Gb, zones condensées et répétées, transcription inactivée) et en euchromatine (2,9 Gb, zones riches en gènes et actives transcriptionnellement).
* Les chromosomes humains sont organisés différemment selon la phase cellulaire. L’essentiel de la fonction génétique active est localisée dans les espaces euchromatiniques, alors que l’hétérochromatine contient surtout des séquences répétitives.
* La complexité du génome nucléaire est renforcée par la présence de nombreux gènes codants et non codants, d’éléments répétés, et d’une architecture fine de l’organisation génique ainsi que de la régulation de l’expression. La distinction entre les deux entités majeures du génome (nucléaire et mitochondrial) est essentielle pour comprendre leur héritabilité et leurs fonctions respectives.

### Description des principaux éléments constituants

#### Gènes (codants et non-codants), pseudogènes et éléments répétés

* Un gène est un segment d’ADN utilisé comme matrice pour la synthèse d’un ARN par transcription ; il existe des ARN codants (messagers, pour la synthèse protéique) et non-codants (comme ARNt, ARNr, miARN, lncARN, etc.). Le nombre total de gènes dépasse aujourd’hui 60 000 (codants et non-codants), et continue d’être affiné grâce à l’amélioration de l’annotation.
* Les gènes varient notablement en taille, nombre d’exons/introns, et organisation ; certains sont groupés en clusters résultant de duplications, d’autres disséminés de façon isolée. Certaines familles, telles que celles des histones ou des récepteurs olfactifs, illustrent cette complexité et peuvent inclure de nombreux pseudogènes.
* Les pseudogènes sont des copies défectueuses de gènes fonctionnels : ils peuvent être processés (issus de la rétrotransposition : perte de la structure exon/intron), non processés (issus d’une duplication : conservation de la structure initiale mais accumulation de mutations), ou unitaires (qui ont perdu leur fonction sans duplication). Ils représentent près de 10% des gènes codants, nombre très élevé, et un rôle potentiel dans l’évolution. Une part significative du génome humain est aussi constituée d’éléments répétés, en tandem (microsatellites, minisatellites, satellites) ou dispersés (transposons, rétrotransposons LINE, SINE), la majorité étant inactifs mais constituant un mécanisme important du remodelage génétique.

## 3. Comparaison aux autres génomes

* La comparaison des génomes révèle que la capacité codante d’un organisme n’est pas proportionnelle à la taille de son génome ni au nombre de ses chromosomes. Le nombre de gènes varie grandement selon les espèces, sans lien direct avec la complexité biologique.
* Chez les eucaryotes, la taille du génome n’est pas corrélée au nombre de gènes, certains organismes simples ayant pourtant des génomes bien plus vastes que l’humain. Les génomes nucléaires de différentes espèces (drosophile, levure, riz, Paris japonica) illustrent la diversité de taille et de nombre de gènes.
* Le génome humain se distingue par une densité génique relativement faible (environ 1 gène tous les 120 kb pour le nucléaire, contre 1 gène/0,45 kb dans le mitochondrial), un taux de séquences codantes très faible (environ 1,1%) et une forte proportion de séquences répétées.

## 4. Les types de variations du génome humain et leurs conséquences

* Le génome humain est soumis à diverses variations : mutations de bases simples (SNV), insertions/délétions, variations de structure touchant de larges segments d’ADN (CNV, inversions). Ces modifications sont caractérisées selon leur impact : certaines sont bénignes (polymorphismes), d’autres pathogènes ou causales de pathologies.
* Le vocabulaire génétique distingue mutations, variations, polymorphismes, pathogénicité, et décrit précisément leur localisation et leur effet (exon, intron, intergénique). Les variations peuvent provoquer des pertes ou gains de fonction, produire un effet dominant négatif ou être neutres, avec des répercussions majeures sur la santé humaine.
* Ces variations touchent tous les niveaux du génome : substitution, indels, expansions de triplets, remaniements structuraux, et altérations affectant l’épissage, la lecture du cadre ou la composition des protéines. On estime à plusieurs millions le nombre de variations entre deux individus humains, expliquant, entre autres, la diversité génétique et la survenue des maladies.
* Les variations de structure représentent la plus grande source de variabilité en nombre de bases affectées, avec plusieurs milliers d’événements structuraux detectés entre deux génomes et à l’échelle de dizaines de millions de bases.
* Les conséquences des variations génomiques peuvent entraîner une perte de fonction totale ou partielle (non-sens, décalages du cadre de lecture), un gain de fonction, ou un effet dominant négatif où la protéine mutée interfère avec la fonction de l’allèle normal.

## 5. La variabilité du génome humain

* La variabilité du génome humain découle de multiples sources – mutations ponctuelles, variations de structure, expansions, etc. Certaines variations sont rares alors que d’autres sont fréquentes, parfois issues d’un effet fondateur ou de mécanismes mécaniques.
* À ce jour, plus de 1,1 milliard de variations ont été référencées (soit 34% du génome) et plus de 7 millions de variations structurelles décrites (soit près de 80% du génome), montrant l’ampleur de la diversité humaine. Cette variabilité contribue à l’adaptabilité et à la diversité phénotypique au sein de l’espèce humaine.

## 6. Evolution des génomes: notions essentielles, mécanismes

* L’évolution des génomes s’explique par des processus dynamiques d'ajout, de perte ou de remaniement du matériel génétique en réponse aux conditions environnementales, offrant parfois un avantage sélectif. Ceci aboutit à l’émergence de nouvelles espèces ou d’adaptations spécifiques.
* Les mécanismes principaux incluent les variations ponctuelles de séquence (substitutions, indels), les duplications d’exons (responsables de domaines répétés dans les protéines), les brassages d’exons (exon shuffling), la duplication de gènes ou de segments entiers, et même la duplication complète du génome. D’autres phénomènes, comme la conversion génique ou l’empreinte parentale, interviennent aussi dans cette dynamique évolutive.
* Au fil du temps, ces mécanismes modifient la structure et la diversité du génome, favorisant la création de nouvelles fonctions et l’adaptation des organismes à leur environnement. L'étude de ces processus permet de mieux comprendre la variabilité génétique et les mécanismes sous-jacents de l’évolution.

# 📄 Slide Mapping

|  |  |
| --- | --- |
| Slides | Section Path |
| 1, 111 | (unmapped) |
| 68-73 | Comparaison aux autres génomes |
| 105-110 | Evolution des génomes: notions essentielles, mécanismes |
| 3, 12 | Introduction |
| 4-11 | Introduction → Notions fondamentales |
| 103-104 | La variabilité du génome humain |
| 13 | Le génome humain |
| 14-17, 24-30, 67 | Le génome humain → Architecture globale |
| 31-33 | Le génome humain → Architecture globale → Génome mitochondrial |
| 34-35 | Le génome humain → Architecture globale → Génome nucléaire |
| 36 | Le génome humain → Description des principaux éléments constituants |
| 37-66 | Le génome humain → Description des principaux éléments constituants → Gènes (codants et non-codants), pseudogènes et éléments répétés |
| 18-23 | Le génome humain → Projet de séquençage |
| 74-102 | Les types de variations du génome humain et leurs conséquences |

**Summary:** Total slides mapped: 110, Unique sections: 14