Les molécules du vivant (Biologie) - 2025 - Prof. Professeur 1

# 📚 Course Content

## 1. Introduction

* Ce cours porte sur l’architecture du génome humain et présente les grandes notions nécessaires à sa compréhension, en lien avec la biologie moléculaire et l’évolution. Une attention particulière sera portée à la diversité et à la complexité qui caractérisent notre patrimoine génétique.
* Le génome humain sert de référence fondamentale pour l’identification des variations génétiques et l’étude de leur impact biomédical. Il s’intègre avec d’autres référentiels, tels que le méthylome, le protéome et le transcriptome, pour mieux comprendre le fonctionnement cellulaire global.

### Notions fondamentales

* L’évolution biologique est un processus fondamental qui façonne la diversité des formes de vie. Elle repose sur des concepts centraux comme l’adaptation, la sélection naturelle, le temps, et la notion d’ancêtre commun universel, appelé LUCA (Last Universal Common Ancestor).
* Chez les eucaryotes, la classification repose sur l’étude de différents organismes modèles, choisis pour des raisons économiques, techniques et génétiques (partage de séquences d’ADN et d’éléments transmis de génération en génération).
* L’homologie désigne un trait partagé par plusieurs espèces, hérité d’un ancêtre commun. Il ne faut pas confondre homologie (héritage évolutif partagé) et analogie (similitude fonctionnelle sans héritage commun, par exemple les ailes des oiseaux et des insectes).
* Deux grands types de gènes homologues existent : les orthologues (issus d’une spéciation, présents dans plusieurs espèces) et les paralogues (issus d’une duplication au sein d’une même espèce).
* L’analyse des séquences d’ADN ou de protéines par alignement permet de juger de leur proximité évolutive. On évalue l’identité stricte et la similarité (tenant compte de similitudes physicochimiques). Un pourcentage d’identité significatif (>20 %) suggère une possible homologie.
* Il est essentiel de noter qu’un pourcentage faible de similarité/identité n’exclut pas l’homologie, notamment dans le cas de régions peu complexes ou d’évolution convergente. À l’inverse, des séquences proches peuvent n’être apparentées que par hasard.
* L’étude comparative des génomes, appuyée par ces notions de similarité, joue un rôle central pour appréhender les mécanismes évolutifs et l’organisation des gènes chez l’humain et les autres organismes.
* Le dogme central de la biologie moléculaire explique le flux d’information génétique de l’ADN vers l’ARN, puis la protéine. Ce schéma justifie l’importance d’étudier les gènes sous leurs différentes formes et fonctions.

## 2. Le génome humain

* Le génome humain constitue l’ensemble du matériel génétique portant l’information nécessaire au développement, à la reproduction et au bon fonctionnement de l’organisme. Il manifeste une organisation complexe et une composition diversifiée.
* Les progrès réalisés dans son séquençage et son exploration contribuent à la compréhension des bases moléculaires des maladies, de la diversité humaine et de l’évolution.

### Projet de séquençage

* Le séquençage du génome humain a représenté un projet scientifique colossal, débuté en 1990 et étalé sur environ 15 ans, mobilisant des centres du monde entier. Ce projet a nécessité un budget de près de trois milliards de dollars.
* L’objectif principal était la lecture complète de la séquence de l’euchromatine humaine. Des méthodes de cartographie génétique (basée sur les fréquences de recombinaison) et physique (mesure en kilobases ou mégabases) ont été mises en œuvre.
* Le séquençage a véritablement démarré en 1999, avec la collaboration de 20 centres répartis dans 6 pays. Les premières versions de la séquence ont été rendues publiques en 2000 (« draft » à 90 %) et finalisées à 99 % en 2003.
* L’évolution du projet est marquée par des versions successives du génome de référence, identifiées par des numéros de version (e.g., hg38, hg19, etc.), qui corrigent et complètent progressivement la séquence, notamment certains chromosomes ou régions complexes.
* Malgré une version considérée comme complète en 2022, des ajustements ponctuels restent à réaliser, notamment pour certaines régions du chromosome Y. Les mises à jour actuelles améliorent la qualité et la représentation de la diversité.
* Des outils de visualisation en ligne (par exemple le UCSC Genome Browser) permettent aujourd’hui d’explorer la séquence, d’y localiser des gènes, des annotations fonctionnelles ou des zones conservées, facilitant la recherche biomédicale.
* L’utilisation de ces bases de données, associées à une exploration interactive, offre une vue détaillée de la structure, du contenu et de la variabilité du génome humain. L’accès à ces données est crucial pour la recherche et la médecine personnalisée.

### Architecture globale

#### Génome mitochondrial

* Le génome mitochondrial humain forme une petite molécule d’ADN circulaire de 16,6 kb, dont la séquence complète a été établie en 1981. Il est organisé de façon très compacte et compte 37 gènes, dont 13 codent pour des protéines impliquées dans la phosphorylation oxydative, et 24 pour des ARN de transfert (ARNt) et ribosomaux (ARNr).
* Le génome mitochondrial présente quelques différences dans l’utilisation du code génétique universel, notamment au niveau de certains codons stop, et une densité en gènes bien plus élevée que dans le génome nucléaire.
* Cette structure rappelle fortement celle de bactéries, confortant l’hypothèse d’une origine endosymbiotique (alpha-protéobactérie). L’héritage mitochondrial se fait exclusivement par la mère et il existe plus de 1000 copies par cellule.
* Le phénomène d’hétéroplasmie, défini par la cohabitation de plusieurs types de génomes mitochondriaux dans une même cellule, varie selon les tissus et au cours de l’âge, avec des implications potentielles sur certaines maladies.
* Comparé à d’autres organismes, le génome mitochondrial de l’humain diffère par sa taille, son organisation (circulaire ou linéaire), le nombre de copies par cellule et le contenu en gènes, illustrant l’adaptabilité évolutive de ce compartiment.

#### Génome nucléaire

* Le génome nucléaire humain est constitué de 3,1 milliards de paires de bases (haploïde), portées par 22 paires de chromosomes homologues (autosomes) et une paire de chromosomes sexuels (gonosomes : X et Y). La séquence de ce génome a été obtenue sous forme de brouillon en 2001 et grandement améliorée jusqu’en 2022.
* Il se divise conceptuellement en hétérochromatine (0,2 Gb, zones condensées, peu ou pas transcrites, riches en éléments répétés) et euchromatine (2,9 Gb, partie riche en gènes et transcrite activement). La compréhension de cette distinction est clé pour l’étude de la régulation génétique.
* La nomenclature chromosomique actuelle dérive d’accords internationaux (Nomenclature de Paris, 1971), permettant de standardiser la description des structures chromosomiques.
* À l’échelle fine, le génome nucléaire présente une complexité organisationnelle : introns, exons, sens de transcription, et zones hautement conservées (environ 5 % du génome).
* Des outils interactifs tels que le Genome Browser facilitent l’exploration des séquences, le repérage des gènes, des protéines associées et des indices de conservation évolutive.

### 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 simple brin lors de la transcription. On distingue les gènes codants (pour des protéines, via ARNm ; environ 20 000 pour 1,1 % du génome humain) et de très nombreux gènes non-codants (>40 000, dont les ARNt, ARNr, miARN, snARN, lncARN, etc.).
* Les gènes diffèrent notablement par leur taille, le nombre d’exons et la complexité de leur structure. Par exemple, certains (comme la dystrophine) s’étendent sur plusieurs mégabases et contiennent des centaines d’exons.
* La répartition des gènes sur le génome se fait de manière non regroupée (territoire génique indépendant) ou en clusters (groupes de gènes homologues issus de duplications, comme les histones ou les récepteurs olfactifs). Des pseudogènes (copies défectueuses d’anciens gènes) sont fréquemment associés à ces clusters.
* Les pseudogènes représentent environ 10 % de l’ensemble des gènes codants. On distingue les pseudogènes processés (issus de rétrotranspositions), non processés (issus de duplications) et unitaires (issus de la perte de fonction d’un gène unique). Bien que longtemps assimilés à des séquences sans fonction, 15 % d’entre eux sont transcrits.
* Les éléments répétés constituent une très grande part du génome humain. Ils regroupent des éléments en tandem (satellites, mini- et microsatellites, participant aux empreintes génétiques et aux études de variabilité individuelle) et des éléments dispersés (40-45 % du génome), principalement des rétrotransposons LINE, SINE (dont Alu), rétrovirus, et des transposons.
* Les éléments répétés jouent un rôle dans l’évolution et la plasticité du génome. Ils participent à la génération de variations génétiques, à la structuration du génome, mais aussi à certains mécanismes pathologiques.
* Bien qu’en majorité inactifs ou fossiles chez l’humain, quelques éléments (comme certains LINE-1) restent actifs et peuvent induire des mutations récentes. La diversité de ces éléments montre l’importance des événements de mobilité et de recombinaison au cours de l’évolution.
* Des outils bioinformatiques permettent de visualiser précisément la répartition de ces gènes, pseudogènes et éléments répétés à l’échelle du génome, contribuant à une meilleure compréhension de sa dynamique et de son organisation.

## 3. Comparaison aux autres génomes

* Le génome humain se compare à une incroyable diversité d’autres génomes, tant par sa taille, le nombre de gènes et l’organisation des chromosomes.
* La capacité codante n’est pas proportionnelle à la taille totale du génome ou au nombre de chromosomes. Par exemple, certains organismes comme les protozoaires ont de petits génomes très denses en gènes, alors que certains eucaryotes supérieurs montrent de très grands génomes avec de nombreux éléments non codants.
* La distribution du nombre de gènes varie énormément selon les domaines du vivant et n’est pas un reflet direct de la complexité biologique. Certains organismes possèdent en effet bien plus de gènes que l’humain, sans pour autant être plus complexes.
* Il existe une relation linéaire entre la taille du génome et le nombre de gènes chez les procaryotes et certains eucaryotes simples (génomes compacts). Mais cette corrélation disparaît chez les eucaryotes supérieurs à cause de la grande proportion de séquences non codantes et d’éléments répétés.
* La comparaison entre génomes nucléaires de différentes espèces montre des variations énormes : la drosophile (Drosophila melanogaster) a un génome de 122 Mb avec ~14 000 gènes, la levure (Saccharomyces cerevisiae) 12,2 Mb pour ~5 800 gènes, certains végétaux (Oryza sativa japonica) 430 Mb pour ~40 000 gènes et la plante Paris japonica atteint même ~150 Gb.
* Chez l’humain, la densité en gènes, la proportion de séquences répétées, le nombre d’introns ou la part du génome codant, diffèrent nettement de la mitochondrie ou des autres espèces, illustrant une importante diversité adaptative.

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

* Le génome humain montre une immense variabilité, entretenue par différents types de variations. Parmi celles-ci, de petites variations ponctuelles (SNV/SNP), des indels, des variations structurales et la répétition de séquences sont nombreuses et peuvent avoir des conséquences diverses.
* Une mutation en biologie désigne toute modification héritée de l’information contenue dans le génome. Dans un contexte médical, on parle de mutation lorsque cette modification est responsable d’un phénotype, notamment pathologique.
* Le terme polymorphisme désigne une variation bénigne, présente à une fréquence élevée dans la population (≥1%). On distingue aussi les variations rares ou fréquentes, bénignes ou pathogènes, selon qu’elles affectent ou non la santé.
* Une nomenclature précise est utilisée pour décrire et localiser ces variations (génomique g., cDNA c., protéique p.), ce qui facilite leur identification et leur interprétation en génétique médicale.
* Les variations peuvent toucher une base (SNV, polymorphisme de substitution), quelques bases ou de grands segments du génome (variations de structure, CNV, équilibrée ou non). Les effets de ces variations dépendent de leur nature et de leur localisation (exon, intron, intergénique).
* Les substitutions et indels sont les sources les plus fréquentes de variation entre deux génomes. Ils peuvent être silencieux, faux-sens, non-sens, ou provoquer des cadres de lecture décalés et des anomalies d’épissage.
* L’expansion de triplets répétés (ex : (CGG) dans l’X-Fragile ou (CAG) dans la maladie de Huntington) est à l’origine de maladies dynamiques avec instabilité au cours de la réplication.
* Les variations structurales concernent de grands segments d’ADN, provoquées par des remaniements chromosomiques, équilibrés ou déséquilibrés (CNV). Elles représentent la principale source de variation en nombre de bases affectées.
* D’autres mécanismes, plus rares, incluent la conversion génique, l’empreinte parentale (modification épigénétique/méthylation), et des types spécifiques de mutation selon les gènes impliqués.
* Les conséquences fonctionnelles sont multiples : perte de fonction, gain de fonction ou effet dominant négatif selon l’impact sur la protéine codée. La variabilité génétique contribue à la diversité humaine, mais aussi à l’apparition de maladies génétiques.

## 5. La variabilité du génome humain

* La variabilité du génome humain est fondamentale pour expliquer la diversité interindividuelle et l’adaptation aux conditions environnementales. Elle résulte de nombreuses sources de variation, tant ponctuelles que structurales.
* Certaines variations sont rares, d’autres plus fréquentes à cause d’effets fondateurs ou de pressions sélectives. Plus d’1,1 milliard de variations sont déjà décrites dans la base dbSNP, couvrant environ 34 % du génome humain, et plus de 7 millions de variations structurales sont référencées dans la DGV, soit près de 80 % du génome.
* La connaissance et la documentation de cette variabilité permettent d’identifier les liens entre génotype et phénotype, de déceler l’origine génétique de certaines maladies, ou d’adapter les prises en charge médicales en fonction des profils individuels.

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

* L’évolution des génomes désigne l’ensemble des mécanismes contribuant à la modification, l’ajout, la perte ou la réorganisation du matériel génétique au cours du temps. Ces processus favorisent l’adaptation aux conditions de vie, l’émergence de nouvelles espèces et la diversité biologique.
* Les principales forces à l’œuvre sont la sélection naturelle, l’adaptation, et le passage du temps. Les variations apportées peuvent conférer un avantage sélectif à certains individus.
* Les événements de duplication d’exons, présents dans environ 10 % des gènes humains, ajoutent des domaines protéiques répétés et résultent souvent de recombinaisons ou de crossing-over inégaux.
* Le brassage des exons, ou exon shuffling, peut se produire via différents mécanismes, tels que la recombinaison non-allélique ou l’action de rétrotransposons (notamment les LINEs). Ce phénomène permet l’apparition de nouveaux gènes à fonctions mixtes.
* La duplication de gènes entiers, ou même de l’ensemble du génome, a joué un rôle considérable dans l’évolution et la complexification des organismes, ouvrant la voie à l’apparition de nouvelles fonctions ou à la spécialisation.
* Ces mécanismes moléculaires, combinés à une diversification par mutation et réarrangement, rendent compte de la dynamique et de la richesse des architectures génétiques observées dans le vivant.

# 📄 Slide Mapping

|  |  |
| --- | --- |
| Slides | Section Path |
| 68-73 | Comparaison aux autres génomes |
| 105-110 | Evolution des génomes: notions essentielles, mécanismes |
| 1, 3, 111 | Introduction |
| 4-11 | Introduction → Notions fondamentales |
| 103-104 | La variabilité du génome humain |
| 12-13 | Le génome humain |
| 31-33 | Le génome humain → Architecture globale → Génome mitochondrial |
| 14-17, 24-30, 34-35 | Le génome humain → Architecture globale → Génome nucléaire |
| 36-67 | 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: 11