

Master Bioinformatique, biochimie structurale et génomique

| M1 Bioinformatique, Biochimie Structurale et Génomique

Semestre1 :

- **Introduction à la Génomique 1 (20h de CM - 10h de TD)**

Panorama transversal de la génomique. Les contours de la génomique seront définis et l'accent est mis sur l'aspect transversal de cette discipline. L'approche génomique sera illustrée aussi bien chez les procaryotes que chez les eucaryotes. Cette unité sera principalement composée de cours portant sur :

- sur le séquençage et l'annotation des séquences
- l'évolution des génomes (micro et macro évolution)
- l'expression des gènes
- les réseaux d'interaction protéique
- l'influence des variations du génome et du transcriptome sur les phénotypes au niveau de l'organisme.

Une série de TD viendra compléter cet enseignement théorique et permettra au travers d'études de cas d'illustrer les points cruciaux du cours.

- **UE Introduction à la Génomique 2 (30h de TD)**

Les différents grands thèmes de la génomique abordés dans l'unité d'introduction à la génomique 1 seront approfondis au travers de l'analyse d'un ou plusieurs articles en binôme de recherche scientifique récents en langue anglaise.

Ce module permettra aux étudiants d'approfondir leurs connaissances en fonction de leur projet de formation personnel.

Au cours de cette unité d'enseignement, les étudiants seront acteurs : chaque binôme réalisera une analyse approfondie de son ou ses articles, et effectuera une présentation orale. Le travail de préparation se fera en étroite interaction avec un enseignant spécialiste.

L'enseignant fournira des informations et/ou du matériel complémentaires (articles ou revues) utiles à la compréhension de l'article et au replacement la thématique dans le contexte scientifique actuel.

A l'issue des présentations, chaque binôme rédigera un rapport sur un des thèmes présenté par un des autres binômes.

- **Bioinformatique 1 (2h de CM - 24h de TP)**

Cette unité vise à approfondir les notions de base sur l'analyse in silico de séquences biologiques, en abordant notamment de manière plus complète :

1. la recherche et la caractérisation de domaines protéiques ;
2. la reconstruction phylogénétique ;
3. l'analyse et la prédiction de structures protéiques 3D ;
4. l'intégration de données génomiques hétérogènes pour la prédiction de fonction.

L'accent est mis avant tout sur la **pratique** ; l'objectif est de fournir aux étudiants la maîtrise des outils bioinformatiques permettant une analyse approfondie de séquences biologiques protéiques autour des quatre points évoqués ci-dessus. Elle s'organise donc sous forme de série de 4 TP (4x4 heures) abordant ces thèmes, au cours desquels les étudiants apprennent à manipuler et comparer différents outils, et à interpréter de manière critique les résultats. Chaque TP débute par une introduction théorique aux notions abordées durant la séance et se poursuit par une partie pratique qui aborde à la fois des exercices d'application simples, mais également la résolution de problèmes plus complets d'une manière semi-autonome.

L'évaluation se fait pour moitié sur des rapports de TP à rendre après chaque séance, et pour moitié sur un examen sur table.

- **Bioinformatique 2 (14h de CM - 8h de TD)**

Cette unité vise à comprendre l'intérêt et les principes de plusieurs méthodes bioinformatiques reposant sur l'analyse de séquences biologiques :

1. alignements multiples de séquences
2. prédiction de la structure secondaire
3. prédiction et modélisation de la structure des protéines
4. bases de la phylogénie moléculaire
5. cis-régulation

L'accent de cette demi-unité est de faire le lien entre l'utilité, la mise en place et fiabilité de diverses méthodes et applications bioinformatiques d'intérêt général.

Les étudiants sortiront avec des bases qui leur permettront de mieux comprendre l'utilité d'autres approches bioinformatiques.

Elle s'organise autour de cours théoriques complétés par 2 TD (2x4 heures) de consolidation des notions vues en cours.

Les séances de TD visent à illustrer les notions en cours.

L'évaluation se fait par un examen sur table.

- **Biochimie Structurale 1 (12h de CM - 8h de TD)**

Physicochimie – Structure et Fonction.

Comprendre le lien entre données expérimentales et concepts généraux. Apprendre à lire des articles de recherche.

A travers l'étude des relations structure-fonction d'une protéine les étudiants verront :

- l'importance des modèles pour la compréhension,
- l'utilisation de cycles thermodynamiques et l'épistasie,
- les limites des modèles.

- **Méthode d'analyse structurale 2 (18h de CM - 2h de TD)**

Faire découvrir les différentes techniques de biologie structurale.

- **TP de Biochimie structurale et de Génomique (Deux semaine de TP)**

Description non disponible.

- **Programmation structurée 1 (Durée non disponible)**

Initiation au langage de programmation Python.

Contrôle continu tout au long de l'UE par validation du travail (petits programmes) réalisé en TP. De plus un programme de traitement de données biologique sera demandé en fin d'UE.

- **Programmation structurée 2 (Durée non disponible)**

Perfectionnement au langage de programmation Python, interface graphique et module Biopython.

Même modalité de contrôle que précédemment.

Semestre2 :

- **Analyse d'article et présentation en anglais**

Description non disponible.

- **Mathématiques appliquées à la biologie**

Description non disponible.

- **Phylogénie et évolution**

Description non disponible.

- **Stage en laboratoire (7 semaines)**

Description non disponible.

Master 2 Recherche

Responsable

■ Dr. Hervé DARBON

Objectifs scientifiques et professionnels de la spécialité

- Acquérir les concepts spécialisés dans les domaines de recherche liés à la mention.
- Avoir, par un stage de 6 mois en laboratoire, une réelle approche du métier de chercheur (conception et suivi d'un projet scientifique, réalisation du plan d'expérience, synthèse des résultats sous forme d'un rapport de stage, présentation orale des résultats obtenus, recherche bibliographique associée au projet, suivi des séminaires de laboratoire, réalisation de posters scientifiques).

Semestre1 :

- **Conférences et projet bibliographique en anglais (6 crédits)**

Objectif(s) de l'UE ! Maîtrise de l'anglais scientifique écrit et oral, ainsi que des méthodes bibliographiques.

Contenu (liste des enseignements)

- Introduction et règles du jeu,
- choix des sujets,
- Guidance personnalisée des projets bibliographiques,
- Ecriture d'une mini revue en anglais sur la base de 3-4 articles principaux et (au plus) une douzaine de références au total,
- Présentation orale.
- Assistance de tous les étudiants aux exposés

Descriptif sous forme de texte : Etude bibliographique de la littérature internationale sur une thématique scientifique, qui donnera lieu à une synthèse écrite et une présentation orale.

- **Phylogénie, évolution et génomique environnementale (6 crédits)**

Objectif(s) de l'UE ! Spécialisation en analyses phylogénétiques et en génétique des populations afin d'analyser des problèmes évolutifs complexes.

Contenu (liste des enseignements)

Nous insisterons sur l'analyse phylogénétique, la modélisation de processus évolutifs complexes, et les tests d'hypothèses. Deux approches seront utilisées dans ce cours: la première est une connaissance poussée de la méthodologie (quelle méthode, pourquoi, comment), la deuxième partie permettra aux étudiants d'utiliser leur bagage théorique afin de démontrer les limites de certain concept évolutif (par exemple la notion d'espèce). Nous ferons un complément en génétique des populations, qui est à la base de la théorie de la microévolution. Le nombre de séquences moléculaires par individu ainsi que les données sur les génomes entiers se font de plus en plus nombreux,. On parle aujourd'hui de génomique des populations.

I. Approfondissement des méthodes :

- Evolution des génomes, brassage d'exon et évolution concertée.
- Duplication à grande échelle des génomes (ancienne et récente).
- Paralogon, sythénie conservée.
- Pression de sélection et shift fonctionnel (sélection par branche).
- Datation moléculaire
- Test de l'évolution parallèle et convergente.
- Inférence bayésienne et maximum de vraisemblance.

II. Identification de scénarios évolutifs :

Il s'agit d'un TP-TD sur la notion d'espèce dans le règne animal. Nous nous proposons d'étudier la divergence génétique entre deux espèces de Téléostéen et le brassage génétique possible entre celles-ci dans des conditions environnementales anthropisées.

Nous identifierons les espèces parentes et leurs hybrides à l'aide de séquences de l'ADN mitochondrial (cytochrome b) et de marqueurs microsatellites dont les régions flanquantes sont homologues avec une espèce modèle (le *Danio rerio*). Nous étudierons la diversité génétique des deux espèces à l'aide de Réseaux d'haplotypes. Nous ferons de la datation moléculaire des haplotypes, et estimerons les niveaux d'introgession des deux génomes. Les régions introgressées seront identifiées et cartographiées sur le génome du danio.

Une partie humide est proposée aux étudiants afin qu'ils génotypent eux meme les individus « purs » et « hybrides ». La partie in silico permettra aux étudiants d'analyser leurs propres données.

• **Génomique des organismes modèles (3 crédits)**

Objectif(s) de l'UE ! Comprendre les approches utilisant des organismes modèles en génomique.

Contenu :

- l'analyse fonctionnelle de l'ensemble des gènes chez différents systèmes modèles animaux.
- compréhension des mécanismes génétiques et moléculaires à la base des maladies humaines grâce à l'utilisation de modèles animaux.

Seront en particulier abordées dans ce module les notions d'organismes modèles en physiologie animale (modèle animaux eucaryotes : souris, nématode, drosophile, levure), modèles animaux transgéniques, techniques de transgénése, stratégies à grande échelle: inactivation génique, expression ectopique de gène, recherche de gènes modificateurs, crible de molécule chimique , « drug discovery »...

- **Génomique fonctionnelle (6 crédits)**

Objectif(s) de l'UE ! Maîtriser les bases conceptuelles des approches visant à la connaissance de la fonction des gènes au niveau des cellules et des organismes.

Contenu (liste des enseignements)

- approche transcriptome : historique et base théoriques
- technologies, les procédures d'analyse et outils d'interprétation
- application des technologies des biopuces à la santé humaine (oncologie et maladies infectieuses)
- annotation des génomes : utilisation des concepts de l'évolution biologique et ontologies
- génomique fonctionnelle chez les animaux d'élevage et chez les plantes
- clonage positionnel à l'échelle du génome de gènes de prédisposition à une maladie
- approche combinée génétique et transcriptomique
- caractérisation des éléments cis-régulateurs de l'expression des gènes
- analyse du protéome : méthodes, retombées scientifiques et cliniques
- bioinformatique fonctionnelle des protéines et analyse structurale des réseaux d'interactions
- séquençage à très haut débit : analyse de la variation génétique, de l'expression des gènes et de leur régulation.

Remarque : Le génome, transcriptome, le protéome, et les réseaux seront abordés. L'idée est notamment de montrer comment on peut chercher à expliquer les variations phénotypiques au niveau de la cellule, des tissus, des organismes et des populations sur la base d'une recherche ciblée sur les variations des séquences d'ADN, du taux d'ARN ou de la quantité des protéines, et sur une meilleure connaissance des réseaux (ADN-miARN, ADN-protéines, protéines-protéines).

- **Atelier biophysique (3 crédits)**

Objectif(s) de l'UE ! Approche pratique des outils biophysiques du chercheur en biologie structurale.

- **Atelier polymorphisme 1 (Avec Phylogénie, évolution et génomique environnementale)**

Objectif(s) de l'UE ! maîtrise pratique de méthodes d'analyse de polymorphismes

Contenu :

- panorama des différentes méthodes d'analyse de polymorphismes
- amplification des fragments choisis
- recherche des polymorphismes connus par PCR-RFLP et inconnus par PCR-DHPLC et PCR-SSCP
- analyse du polymorphisme des microsatellites par l'analyse de la taille des produits PCR
- identification des polymorphismes par séquençage (séquenceur automatique à capillaires ou séquenceur à très haut débit)
- méthodes d'analyses de polymorphismes exploitées par la police scientifique.

- **Atelier polymorphisme 2 (3 credits)**

Objectif(s) de l'UE ! Maîtrise pratique de méthodes d'analyses de données de polymorphismes.

Contenu :

- bioinformatique et base de données des polymorphismes :
- faire le point sur les polymorphismes connus pour un gène
- exploiter la séquence pour en déduire des amorces pour la PCR
- vérifier la spécificité des séquences identifiées par alignement multiple
- analyse de données issues des séquenceurs automatiques à capillaires ou des séquenceurs de nouvelle génération à très haut débit.
- analyse des déséquilibres de liaisons
- analyse d'association des polymorphismes avec un phénotype

- **Analyse statistique de données génomiques 1 (3 credits)**

Objectif(s) de l'UE ! L'objectif de l'atelier est d'acquérir les connaissances théoriques et pratiques sur les différentes méthodes d'analyses statistiques mises en place dans l'étude des données génomiques.

Contenu :

Présentation et application des différentes méthodes de normalisation des données génomiques.

Présentation et application des différentes méthodes d'analyses statistiques appliquées aux données génomiques (recherche de gènes discriminant, méthode de classification).

Présentation d'outils informatiques permettant l'analyse de données de biopuce. Initiation au langage R et aux bibliothèques Bioconductor.

Durant ce module les étudiants vont acquérir les connaissances théoriques et pratiques sur les différentes méthodes d'analyses statistiques utilisées dans l'étude de données génomiques (notamment dans le contexte de l'analyse par biopuces).

Ils vont apprendre dans un premier temps les différentes méthodes de normalisation des données génomiques (centrage et réduction, MAD, lowess, quantile, RMA, MAS5...).

Dans un deuxième temps les étudiants vont apprendre à manipuler les méthodes statistiques de recherche de gènes discriminant (SNR, t.test, SAM...) et de classification (k-means, clustering-hiérarchique...).

Certaines méthodes sont tout d'abord présentées à travers l'utilisation d'un tableur avant que les étudiants n'apprennent à maîtriser le langage R et les bibliothèques Bioconductor.

Les logiciels de classification courants sont présentés (TMEV, cluster, treeview...) et de nombreux exemples pris dans la littérature ou les bases de données publiques (GEO).

Les outils de connexion à des bases de données publiques via Bioconductor (notamment BiomaRt) sont présentés et illustrés par des exemples.

Semestre2 : Stage de laboratoire (6 mois)