

Intitulé de l'UE	Informatique pour le séquençage
Section(s)	- (6 ECTS) Master en Sciences de l'Ingénieur industriel orientation Life data technologies / Cycle 2 Bloc 1

Responsable(s)	Heures	Période
Vincent BRANDERS	60	Quad 2

Activités d'apprentissage	Heures	Enseignant(s)
Exploitation des ressources bioinformatiques	30h	David COORNAERT
High-throughput sequencing algorithms	30h	Vincent BRANDERS

Prérequis	Corequis

Répartition des heures
Exploitation des ressources bioinformatiques : 10h de théorie, 20h d'exercices/laboratoires
High-throughput sequencing algorithms : 15h de théorie, 15h d'exercices/laboratoires

Langue d'enseignement
Exploitation des ressources bioinformatiques : Français
High-throughput sequencing algorithms : Français

Connaissances et compétences préalables
Séquençage nouvelle génération ¹ et services web bioinformatique
High-throughput sequencing algorithms :
<ul style="list-style-type: none"> • Bioinformatique de base : notions de biologie moléculaire, en particulier concernant l'ADN, les séquences nucléotidiques et les processus évolutifs. • Algorithmes : maîtrise des concepts d'algorithmes classiques (complexité, structures de données) et notions de programmation. • Statistiques : compréhension des probabilités et des statistiques pour l'analyse des données issues du séquençage à haut débit.

Objectifs par rapport au référentiel de compétences ARES
Cette UE contribue au développement des compétences suivantes

- Identifier, conceptualiser et résoudre des problèmes complexes
- Concevoir et gérer des projets de recherche appliquée
- S'engager dans une démarche de développement professionnel

Objectifs de développement durable

Aucun

Acquis d'apprentissage spécifiques

Capacité à définir le processus adéquat pour traiter des données NGS dans le cadre d'une recherche en particulier.
Aptitude à exploiter les ressources existantes pour les intégrer à un projet de recherche.

Contenu de l'AA Exploitation des ressources bioinformatiques

L'amplification du recours aux techniques NGS dans tous les pans de la recherche biologique aboutit à l'engorgement quotidien de ce type de données dans leur différentes phases de maturation ; données brutes, filtrées, nettoyées, assemblées, corrigées. Nous allons adresser ces questions en puisant abondamment au sein des banques afin de répéter les différentes étapes de finalisation allant des données brutes initiales pour finir par un résultat d'analyse complet : génome complet, analyse des variations dans le profil d'expression par exemple. Nous allons "rejouer" les travaux effectués lors des stages/TFEs des étudiants des années précédentes lorsque les dites données sont rendues publiques.

Contenu de l'AA High-throughput sequencing algorithms

- Rappels du séquençage à haut débit (NGS) et à ses applications.
- Algorithmes d'alignement de séquences (Needleman-Wunsch, Smith-Waterman, BLAST).
- Concepts de k-mers et leur utilisation dans l'assemblage de génomes (graphes de De Bruijn, filtres de Bloom).
- Assemblage de novo et problèmes liés aux séquences répétitives dans les génomes.
- Gestion des erreurs de séquençage, et méthodes de correction des erreurs (ex. : normalisation digitale, correction des erreurs par k-mers).
- Étude des structures de données pour les séquences et indexage : arbres de suffixes, tries, et index de suffixes.

Méthodes d'enseignement

Exploitation des ressources bioinformatiques : cours magistral, travaux de groupes, approche interactive, utilisation de logiciels

High-throughput sequencing algorithms : cours magistral, travaux de groupes

Supports

Exploitation des ressources bioinformatiques : notes de cours, activités sur eCampus

High-throughput sequencing algorithms : copies des présentations, activités sur eCampus

Ressources bibliographiques de l'AA High-throughput sequencing algorithms

Ukkonen, E. « On-line construction of suffix trees. » *Algorithmica*, 1995.

Margaret Dayhoff et al. « Atlas of protein sequence and structure », 1966

Pevzner, P. « Computational Molecular Biology: An Algorithmic Approach », MIT Press, 2000

Fedoroff, N. « Transposable elements, epigenetics, and genome evolution. » *Science*, 2012

Évaluations et pondérations

Évaluation	Note globale à l'UE
Langue(s) d'évaluation	Français
Méthode d'évaluation	Pondération 50% Algorithmes de séquençage haut débit et 50% Exploitation des ressources bioinformatiques

Examen oral avec exercices pratique pour Exploitation des ressources bioinformatiques.

High-throughput sequencing algorithms :

- Examen écrit

Report de note d'une année à l'autre pour l'AA réussie en cas d'échec à l'UE

Exploitation des ressources bioinformatiques : **non**

High-throughput sequencing algorithms : **non**

Année académique : **2024 - 2025**