

GROUPE	20	DIALLO, Abdoulaye Baniré	diallo.abdoulaye@uqam.ca	(514) 987-3000 3914	PK-4720
Mardi, de 18h00 à 21h00					

DESCRIPTION Étude des algorithmes, techniques et ressources logicielles appliquées aux structures en biologie moléculaire (arbres de phylogénie, structures tridimensionnelles, agglomérats moléculaires).
Construction des arbres de phylogénie. Prédiction de la structure secondaire de l'ARN et algorithmes de repliement. Comparaison des structures secondaires de l'ARN. Structures des protéines (secondaire, tertiaire, quaternaire). Classification des protéines. Interactions entre protéines. Visualisation des protéines. Simulation des voies de régulation. Ordinateurs biologiques.

OBJECTIFS Ce cours est structuré autour de deux aspects également importants :

- un ensemble de problèmes bioinformatiques mettant en jeu des structures non linéaires (i.e. qui ne sont pas uniquement des séquences), qui serviront de fil directeur principal du cours (phylogénies, structures d'ARN et de protéines, puces à ADN, réseaux, dynamique de la régulation) ; en particulier, nous survolerons, pour chacun de ces points au moins un logiciel de référence ;
- un ensemble de méthodes algorithmiques **générales**, mais introduites dans le cadre des problèmes bioinformatiques évoqués au point précédent : programmation dynamique, clustering, évaluation statistiques, recherche de motifs structurés, méthodes d'apprentissage (HMM, réseaux de neurones), théorie des graphes.

ÉVALUATION	Description sommaire	Date	Pondération
	Participation		10%
	Projet donnant lieu à un rapport		25%
	Remise de la description de projet : 5%		
	Remise du rapport projet : 20%		
	Soutenance		20%
	Présentation article		20%
	Examen oral		25%

CONTENU **Avertissement.** Le plan sommaire qui suit présente l'architecture générale de ce cours et il se peut que ce plan soit trop ambitieux.

- ❑ **Chapitre 1 (4 séances). Introduction à l'inférence de phylogénies:**
 - Méthodes de distances (algorithmes de clustering). Parcimonie maximum (programmation dynamique, exploration d'un espace de solution par branch-and-bound et heuristiques), incohérence de la parcimonie. Méthodes probabilistes (maximum de vraisemblance, méthodes bayésiennes, exploration probabiliste d'un espace de solution, modèles probabilistes d'évolution). Évaluation statistique d'une phylogénie (Consensus, bootstrap, jackknife, test de permutation, test du ratio de vraisemblance).
- ❑ **Chapitre 2 (2 séances) Structures secondaires de l'ARN.**
 - Rappels sur la structure des ARN. Méthodes de repliement : minimum d'énergie (programmation dynamique) et analyse de covariation. Recherche de motifs de structure secondaire (algorithmes recherche de motifs).
- ❑ **Chapitre 3 (1 séance) Modèles de Markov Cachés.**
 - Introduction à la notion d'homologie distance et de modélisation probabiliste d'une famille de séquences. Description des Modèles de Markov Cachés et des problèmes algorithmiques liés (théorie des graphes, programmation dynamique et apprentissage statistique). Applications 1 : hmmer et profiles, PFAM, détection de signaux peptides et ancrés GPI. Survol des problèmes pratiques liés à l'utilisation des HMM. Applications 2 : modélisation des structures secondaires d'ARN par grammaires stochastiques (HMM et grammaires).
- ❑ **Chapitre 4 (2 séances). Détection et analyse de séquences fonctionnelles.**
 - Rappels sur les différents types de séquences fonctionnelles et la structure des génomes. Détection de séquences fonctionnelles par mesure de la conservation (empreinte phylogénétique) Applications : détection d'ARN non codants, détection de régions régulatrices, reconstruction de séquences génomiques ancestrales, prédiction de gènes.
- ❑ **Chapitre 5 (1 séance). Bioinformatique des micropuces.**
 - Survol des différentes problématiques bioinformatiques liées à une expérience de micropuces. Mise au point d'un ensemble de sondes (recherche de motifs, repliement d'ADN). Analyse d'images (méthodes standard de segmentation). Traitement statistique : normalisation, mesure de l'expression d'un gène. Analyse des résultats d'une expérience : clustering, time series.

- ❑ **Chapitre 6 (2 séances). Analyse et prédiction de la structure des protéines.**
 - Rappels sur la structure des protéines. Classification structurale des protéines (CATH, SCOP, ...). Outils de visualisation de structures de protéines. Alignement structural de protéines (programmation dynamique). Prédiction de la structure des protéines (apprentissage machine, réseaux de neurones).
- ❑ **Chapitre 7 (1 séance). Réseaux.**
 - Présentation de différents types de réseaux (réseaux de régulation, réseaux d'interaction de protéines, réseaux métaboliques) et des ressources existantes. Présentation de modèles informatiques de réseaux (réseaux booléens, réseaux probabilistes, équations différentielles, réseaux de petri) et des algorithmes d'inférence et d'analyse correspondants.
- ❑ **Présentations des étudiants (2 séances).**

 RÉFÉRENCES

- U C <http://www.bioinfo.uqam.ca/bif7001>
Site web du cours.
- V C A. Baxevanis et F. Ouellette – *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Second Edition*. – Wiley-Interscience. 2001
Bonne introduction à de nombreux sujets, traités sous un angle pratique. 2 exemplaires sont disponibles à la bibliothèque (QP620.B57.2001) dont un est mis en réserve pour le cours BIF7001.
- V C A. Baxevanis et F. Ouellette – *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Third Edition: Student Companion Site*.
- V C Mount, D. – *Bioinformatics. Sequence and genome analysis, Second edition*. – Cold Spring Harbor Laboratory Press. 2004.
Un livre très complet, mélangeant concepts théoriques et pratiques. Un exemplaire de la première édition est disponible à la bibliothèque (QH441.2.M68.2001) mis en réserve pour le cours BIF7001.
- V C Durbin, R., Eddy, S., Krogh, A. et Mitchison, G. – *Biological sequences analysis: Probabilistic models of proteins and nucleic acids*. – Cambridge University Press. 1998.
Une bonne introduction aux aspects probabilistes de la bioinformatique. Un exemplaire est disponible à la bibliothèque (QP620B576.1998) mis en réserve pour le cours BIF7001.
- V C Ewens, W. et Grant, G. – *Statistical Methods in Bioinformatics. An introduction* – Springer. 2001.
Un livre plutôt théorique sur les aspects statistiques de la bioinformatique. Un exemplaire est disponible à la bibliothèque : R858.E94.2001. Une seconde édition a été publiée en 2004.
- V C Jones, N.C. et Pevzner, P.A. – *An Introduction to Bioinformatics Algorithms* – MIT Press. 2004.
Un bon livre pour ceux qui sont intéressés au lien entre les techniques algorithmiques de base et les problèmes bioinformatiques. Non disponible à la bibliothèque
- V C Neapolitan, R. et Naimipour, K. – *Foundations of Algorithms Using C++ Pseudo-code* – Jones and Bartlett Publishers. 1998
Une introduction claire aux principales techniques algorithmiques. Un exemplaire est disponible à la bibliothèque : QA9.58.N42.2004.
- V C Baldi, P. et Brunak, S. – *Bioinformatics - The Machine Learning Approach* – The MIT Press. 2001.
Excellente introduction aux techniques d'apprentissage-machine en bioinformatique. Un exemplaire est disponible à la bibliothèque (QH507.B35.2001) mis en réserve pour le cours BIF7001.

A : article – C : comptes rendus – L : logiciel – N : notes – R : revue –
S : standard – U : uri – V : volume

C : complémentaire – O : obligatoire – R : recommandé