

Bio-informatique avancée

Groupe 20

Mercredi, de 17h30 à 20h30 PK-4323 (cours)

Responsable(s) du cours

Nom du coordonnateur : DIALLO, Abdoulaye Banire**Nom de l'enseignant :** OUANGRAOUA, Aida**Local :** PK-4320**Téléphone :** (514) 987-3000 ##8214**Disponibilité :****Courriel :** ouangraoua.aida@uqam.ca**Site Web :**

Autre enseignant : DIALLO, Abdoulaye Banire

Description du cours

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 du cours

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.

Contenu du cours

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 (5 séances) - Introduction à l'inférence de phylogénies.
 - Méthodes de distances (algorithmes de clustering). Parcimonie (programmation dynamique, exploration d'un espace de solution par branch-and-bound et heuristiques), incohérence de la parcimonie.

- Chapitre 2 (2 séances) - Réarrangements de génomes.
 - Structure et organisation de génomes. Modèles de réarrangements. Méthodes de construction de blocs de synténies. Méthodes de reconstruction de génomes ancestraux.
- Chapitre 3 (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 4 (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 ancrs GPI. Survol des problèmes partiques liés à l'utilisation des HMM. Applications 2 : modélisation des structures secondaires d'ARN par grammaires schotastiques (HMM et grammaires).
- Chapitre 5 (1 séance) - 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 6 (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'un expérience : clustering, time series.
- Chapitre 7 (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 structurel de protéines (programmation dynamique). Prédiction de la structure des protéines (apprentissage machine, réseaux de neurones).
- Présentations des étudiants (1 séance).

Modalités d'évaluation

Description sommaire (date)	Pondération
Projet donnant lieu à un rapport	50%
Recherche bibliographique : 5% (5 février)	
Présentation orale : 20% (23 avril)	
Rapport écrit : 25% (30 avril)	
Devoir #1 (12 février)	25%
Devoir #2 (9 avril)	25%

Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements, consultez le site suivant :

<http://www.sciences.uqam.ca/etudiants/integrite-academique.html> et

<http://www.bibliotheques.uqam.ca/recherche/plagiat/index.html>

Politique d'absence aux examens

L'autorisation de reprendre un examen en cas d'absence est de caractère exceptionnel. Pour obtenir un tel privilège, l'étudiant-e doit avoir des motifs sérieux et bien justifiés.

Il est de la responsabilité de l'étudiant-e de ne pas s'inscrire à des cours qui sont en conflit d'horaire, tant en ce qui concerne les séances de cours ou d'exercices que les examens. **De tels conflits d'horaire ne constituent pas un motif justifiant une demande d'examen de reprise.**

Dans le cas d'une absence pour raison médicale, l'étudiant-e doit joindre un certificat médical original et signé par le médecin décrivant la raison de l'absence à l'examen. Les dates d'invalidité doivent être clairement indiquées sur le certificat. Une vérification de la validité du certificat pourrait être faite. Dans le cas d'une absence pour une raison non médicale, l'étudiant-e doit fournir les documents originaux expliquant et justifiant l'absence à l'examen – par exemple, lettre de la Cour en cas de participation à un jury, copie du certificat de décès en cas de décès d'un proche, etc. Toute demande incomplète sera refusée. Si la direction du programme d'études de l'étudiant-e

constate qu'un étudiant a un comportement récurrent d'absence aux examens, l'étudiant-e peut se voir refuser une reprise d'examen.

L'étudiant-e absent-e lors d'un examen doit, dans les cinq (5) jours ouvrables suivant la date de l'examen, présenter une demande de reprise en utilisant le formulaire prévu, disponible sur le site Web du département à l'adresse suivante : <http://info.uqam.ca/politiques/>

L'étudiant-e doit déposer le formulaire dûment complété au secrétariat de la direction de son programme d'études : PK-3150 pour les programmes de premier cycle, PK-4150 pour les programmes de cycles supérieurs. Pour plus de détails sur la politique d'absence aux examens du Département d'informatique, consultez le site web suivant : <http://info.uqam.ca/politiques>

Intégrité académique

PLAGIAT Règlement no 18 sur les infractions de nature académique. (extraits)

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes;
- l'utilisation totale ou partielle du texte d'autrui en la faisant passer pour sien ou sans indication de référence;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant;
- l'obtention par vol, manoeuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-crédation, un rapport de stage ou un rapport de recherche;
- Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18.

Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements, veuillez consulter les sites suivants : <http://www.sciences.uqam.ca/etudiants/integrite-academique.html> et <http://www.bibliotheques.uqam.ca/recherche/plagiat/index.html>

Médiagraphie

UC <http://moodle.uqam.ca>
Site web du cours.

UC J. Felsenstein. Inferring Phylogenies. voir site web : evolution.gs.washington.edu.

VC 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.

VC A. Baxevanis et F. Ouellette -- *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Third Edition: Student Companion Site.*

VC 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.

VC 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.

VC 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.

VC 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

VC 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.

VC 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
S: Standard - U : uri - V : volume

C : complémentaire - O : Obligatoire - R : recommandé