

ANT 3475 – Bioinformatique évolutive humaine

Trimestre: Hiver 2026
Horaire: Vendredi 8:30 – 11:30
Salle: Pavillon Lionel-Groulx C-3115

Professeur: Joseph Orkin
Bureau: Pavillon Lionel-Groulx 3072
Courriel: joseph.orkin@umontreal.ca
Disponibilité: Sur rendez-vous

Description du cours

Ce cours est une introduction pratique à la bioinformatique pour l'étude de l'évolution humaine et non-humaine. Les étudiant.e.s travailleront avec des données génomiques réelles et acquerront une expertise dans l'analyse de la structure des populations, de la phylogénétique, de la diversité génomique et des microbiomes. Le cours se concentrera principalement sur la familiarisation avec l'interface de calcul en ligne de commande utilisant UNIX et R pour manipuler des données volumineuses.

Ce cours est structuré comme un atelier d'analyse de données. Par conséquent, la présence en classe est obligatoire et la majeure partie de chaque cours sera consacrée à l'utilisation de tutoriels sur les données bioinformatiques. La notation des travaux dirigés en classe sera basée sur la participation active aux séries de problèmes et aux autres supports de travaux dirigés.

Chaque semaine, les étudiants auront un devoir à faire à la maison qui passera en revue les concepts de codage que nous avons appris en classe. Les concepts que nous étudions s'appuient les uns sur les autres, c'est pourquoi les devoirs doivent être rendus la semaine suivant leur attribution. Cependant, veuillez tenir compte des dates limites indiquées dans le horaire.

Évaluation

Participation aux ateliers en classe	20%
Problèmes hebdomadaires	40%
Examen intra	20%
Examen final	20%

Manuels scolaires

Nous travaillerons avec des données accessibles au public et des tutoriels en ligne, de sorte qu'aucun manuel n'est nécessaire pour ce cours. Cependant, je joins une liste de lectures complémentaires pour chaque cours. Ces textes et articles ont pour but de vous fournir des ressources pour le codage et des exemples du type d'analyses que nous pouvons effectuer en bio-informatique évolutive humaine. Ces textes ne feront pas l'objet d'examens, mais je vous encourage vivement à les lire avant chaque cours afin de mieux comprendre ce que nous allons faire.

Asare P and van Tonder A. (2023). Introduction to bacterial genomics.
https://avantonder.github.io/Ghana_course/

Fontsere, C. et al. 2019. Admixture in Mammals and How to Understand Its Functional Implications. *Bioessays*, 41(12):e1900123.

Janiak et al. 2018. Evolution of Acidic Mammalian Chitinase Genes (*CHIA*) Is Related to Body Mass and Insectivory in Primates. *Mol. Biol. Evol.* 35(3):607-622.

Kuderna, L. et al. 2023. A Global Catalog of Whole-Genome Diversity from 233 Primate Species, " *Science*, 380(6648):906-913.

Orkin et al. 2019. Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. *ISMEJ*. 13(1):183-196.

Novembre, J. et al. 2008. Genes Mirror Geography Within Europe. *Nature*, 456: 98–101.

Schotts, William E (2019). The Linux Command Line: A complete introduction. 2nd Edition. No starch press. San Francisco. <https://linuxcommand.org/tlcl.php>

Horaire et organisation des séances (sous réserve de modifications):

Cours	Date	Bloc	Thème	Lectures	Date limite
1	9 jan	Codage sous UNIX et R	Ligne de commande Unix	Schotts Ch. 1-10	
2	16 jan		Variables Bash	Schotts Ch. 11-13	PS1
3	23 jan		Expressions régulières	Schotts Ch. 19	PS2
4	30 jan		Boucles Bash	Schotts Ch. 24-29, 33	PS3
5	6 fev		Calcul de haute performance	Asare and van Tonder 2025, 12.1	PS4
6	13 fev		R et al visualisation		PS5
7	20 fev		Examen intra		
8	27 fev		Microbiomes	Orkin et al. 2019	
	6 mar		Pas de cours		
9	13 mar	Analyse des données	Sélection naturelle	Janiak et al. 2028	PS 6
10	20 mar		Phylogénétique	Asare and van Tonder 2025, 10.1	PS 7
11	27 mar		Introgression	Fontsere et al. 2019	PS 8
12	3 avr		Diversité génomique	Kuderna et al. 2023	PS 9
13	10 avr		Structure de la population	Novembre et al. 2008	PS 10
14	17 avr		Examen final		PS 11

Autres considérations :

Nous travaillerons sur une grappe d'ordinateurs à haute performance basée sur le cloud, de Calcul Québec. Cela nous permettra d'acquérir de l'expérience avec des systèmes informatiques avancés et de garder le matériel de cours bien organisé. La bionformatique et la biologie computationnelle utilisent presque exclusivement des systèmes de ce type, qui sont basés sur les systèmes d'exploitation Linux (par exemple, Ubuntu, Red Hat) et Unix (par exemple, Mac OSX). Heureusement, ce n'est pas un problème avec le système basé sur l'informatique en nuage.

Règles relatives à l'IA

Il est parfois approprié d'utiliser l'IA en bio-informatique, mais dans ce cours, ce n'est pas toujours le cas. L'IA est formidable pour déboguer votre code. Il suffit souvent de coller votre code dans l'invite et de demander pourquoi il ne fonctionne pas pour obtenir immédiatement une réponse. Parfois, cette réponse est erronée, ce qui peut avoir des conséquences désastreuses si vous ne savez pas que vous avez introduit une erreur à la ligne 35 d'un programme de 1000 lignes. En termes simples, si vous voulez comprendre comment analyser des données, vous devez apprendre à analyser des données. Cela signifie que vous devez fournir des efforts pour apprendre un nouveau langage.

En termes pratiques pour notre cours :

- 1) L'utilisation de toute IA est interdite pendant les examens. Si je vous vois l'utiliser, vous obtiendrez un zéro à l'examen.
- 2) Pendant les ateliers, vous pouvez poser des questions à l'IA si vous êtes vraiment bloqué. Cependant, si vous le faites régulièrement, vous n'apprendrez rien et vous risquez d'obtenir de mauvais résultats aux examens. Vous devez essayer de trouver la solution par vous-même. Utilisez l'IA comme un tuteur, pas comme une béquille.
- 3) Gardez à l'esprit que je peux savoir si vous utilisez l'IA pour résoudre vos problèmes. Il existe souvent plusieurs façons de répondre à une question de codage et si je vois un code étrange que je ne vous ai pas enseigné, je saurai qu'il provient de l'IA.