

INF4500 Bioinformatique

Automne 2013 – Groupe 10

Université du Québec à Montréal

Cours : Pr. Abdoulaye Baniré Diallo (diallo.abdoulaye@uqam.ca)

Atelier : Mohamed Amine Remita (remita.mohamed_amine@uqam.ca)

Atelier # 2 : Biojava

Téléchargement : http://biojava.org/wiki/BioJava:Download_1.8 (Biojava 1.8)

Tutoriel : <http://biojava.org/wiki/BioJava:CookBookLegacy>
<http://biojava.org/wiki/BioJava:BioJavaXDocs>

Exercice 1

- En utilisant Biojava écrivez un programme qui
 - Télécharge à partir de la base de données NCBI Nucleotide le fichier Genbank correspondant au numéro d'accèsion NM_001123784 (l'extension des fichiers Genbank est gb);
 - Affiche sur la console les différentes informations associées avec cette séquence (description, taille, date de publication, organisme, keywords...);
 - Traduit cette séquence en une séquence protéique;
 - Enregistre la séquence de protéine dans un fichier Fasta.

Exercice 2

- Téléchargez à l'aide d'un programme le fichier Genbank du génome du HPV11 qui a un numéro d'accèsion FR872717;
- Listez les gènes, leurs positions et les numéros d'accèsion de leurs protéines;
- Extrayez la sous séquence nucléotidique du **gène** L1; et traduisez la en une séquence protéique selon le cadre de lecture 1;
- C'est quoi le numéro d'accèsion de la **protéine** L1; à partir de cet id téléchargez sa séquence;
- Comparez les deux séquences protéiques.

Exercice 3

Dans le fichier `taxid151340.acc.lst` on trouve une liste de 20 numéros d'accèsion correspondant à des VPH (papillomavirus humain).

- Téléchargez automatiquement les fichiers Genbank de ces génomes;
- Calculez et affichez la fréquence des nucléotides pour chaque séquence;
- Calculez la distribution des fréquences des nucléotides pour le lot de ces séquences.