

INF4500: Bioinformatique
Session : Automne 2013
Université du Québec à Montréal
Professeur : Abdoulaye Baniré Diallo
diallo.abdoulaye@uqam.ca

Devoir 1 : Assemblage et comparaison de séquences
(énoncé de 2 pages)
Date de remise : 1^{er} novembre

Introduction :

Ce devoir vous permet de pratiquer l'utilisation de quelques méthodes de comparaison de séquences et d'assemblage.

Pour toutes les questions, il faudra fournir :

- Le canevas utilisé pour trouver la solution
- Les résultats trouvés
- Votre description des résultats.

Pour les données volumineuses, vous pourrez les mettre dans un dossier compressé et les mettre dans votre répertoire public sur le serveur *bioinfo* et m'envoyer par courriel l'adresse url. **Votre rapport ne devra pas dépasser 10 pages** (sans compter les annexes).

Question 1 (25%):

- a) Donnez une description sommaire de ce locus chez **l'humain**. Celle-ci devra minimalement contenir les informations suivantes : sa location précise, sa taille, son rôle et toutes les autres descriptions que vous jugerez pertinente.
- b) Utilisez *MapViewer* du *NCBI* pour déterminer les locus voisins du **ALS2**. Comparez leur description à celle du **ALS2**. Existe t-il une relation entre ces loci? Si oui, décrivez la, Sinon commentez les différences.
- c) Utilisez *MapViewer* du *NCBI* pour donner une description des différentes annotations assignées au locus **ALS2**.
- d) Analysez le fichier *genbank* du *NCBI* correspondant à ce locus. Que pouvez vous dire des régions codantes, des STS et CDS?
- e) Est ce que gène a des homologues chez d'autres mammifères? Si oui lesquels? Ont –ils les mêmes roles dans ses organismes.

Question 2 (25%) :

Utilisez le programme CAP3 (plusieurs versions disponibles sur Internet, mais utilisez de préférences Mobyle) pour assembler les reads du SARS. Vous obtiendrez comme résultat plusieurs contigs. Concentrez votre analyse sur 5 contigs que vous choisirez

- a) Que pouvez vous dire sur la qualité des contigs choisis?

Attention : les cas de plagiat seront soumis aux règlements de l'UQAM
copyright@Abdoulaye Baniré Diallo

- b)** Blastez vos contigs.
- c)** Quelles sont les fonctions prédites par les différents contigs choisis?
- d)** Quel(s) sont le(s) génome(s) dans lesquels vous trouvez des séquences fortement similaires? Que pouvez vous en déduire?

Question 3 (25%) :

Vous disposez d'un vecteur fictif, nommé pANNE, fourni par Madame Bergeron. Ce vecteur contient 6627bp et provient d'une combinaison de plusieurs séquences de vecteurs employées en laboratoire. Vous trouverez une copie de ce vecteur dans le fichier *pANNE.txt*. Identifiez la composition précise de ce vecteur, en indiquant les différentes régions, le sens des séquences, les chevauchements, etc. Vous pouvez présenter vos résultats de façon imagée (soyez créatif).

Question 4 (25%) :

Utilisez blast pour trouver des orthologues du FOXP4. Recupérez le CDS de 8 orthologues.

- a)** Alignez ces CDS en utilisant ClustalW, dialign et Mavid.
- b)** Discutez la performance en terme de facilité d'utilisation et de rapidité des différents programmes?
- c)** Analysez et discutez la qualité de l'alignement donné par chaque méthode?
- d)** Quel est votre meilleur alignement? Justifiez votre choix.