Projet Annotation – Annotation d’un fragment génomique

Annoter l'un des deux fragments génomiques ci-joint en utilisant les différentes approches vues en cours. Les deux brins direct et complémentaire seront analysés.

Dans un premier temps, dans votre rapport expliquez la problématique, décrivez votre démarche (pourquoi telle méthode, que vous permet-elle de faire), présentez et analysez vos résultats. Il faudra donc réaliser un tableau de synthèse comme fait en TP et expliquez vos choix. Les résultats doivent inclure la prédiction fonctionnelle des produits des gènes identifiés sur le fragment génomique.

Intégrer ensuite les résultats dans Artemis qui vous permettra de produire une fiche en format EMBL de votre fragment, donc dans un format qui permetttrait son dépôt dans cette base de données. Dans le TP Utilisation d'Artemis, vous trouverez le lien pour télécharger le logiciel ainsi que son guide d'utilisation. Vous avez donc un travail d’autoformation à réaliser pour apprendre à utiliser cette palteforme.

Pour intégrer les résultats, il faudra réaliser plusieurs parsers pour mettre les résultats des programmes dans le format GFF compris par Artemis que vous trouverez décrit dans TP Utilisation Artemis et dans le document description des parsers : des résultats de GenMark, des résultats de GenMark.hmm et les résultats de scan\_for\_matches concernant la prédiction des RBS et donc du codon initiateur, des promoteurs et des terminateurs de transcription. Un seul programme doit pouvoir permettre de parser les différents résultats de scan\_for\_matches.

Comme les programmes à réaliser sont simples, pour vous habituer a utiliser des langages différents, il serait souhaitable que ces parsers soient écrits en Perl en utilisant le petit mémo qui est mis en ligne. La partie la plus importante étant la création de l’algorithme, structure logique de votre programme avant son implémentation dans un langage de programmation. Cependant, l’utilisation du langage Python est acceptée.

La prédiction fonctionnelle sera réalisée ensuite à partir des produits protéiques potentiels déduits de votre annotation. Cette annotation fonctionnelle comprendra la détection de domaines fonctionnels. Votre annotation finale devra être fournie dans un fichier au format EMBL que vous pouvez obtenir en sortie d'Artemis ensuivant le modèle qui vous est fourni en exemple.

Même si la sortie de GenMark.hmm est actuellement disponible au format GFF, je vous demanderai de bien vouloir réaliser le programme qui transforme la sortie LST en format GFF, ceci pour vous entraîner à la programmation.

Je vous demande aussi de me fournir les algorithmes de chacun de vos programmes (cf. ci-dessous rapport)

Envoyer par courrier électronique :

* un rapport écrit qui doit être réalisé individuellement. Ce rapport doit justifier votre démarche et votre raisonnement, présenter les résultats obtenus à chaque étape et leur synthèse. Il devra inclure une image de l'interface d'Artemis. Il devra intégrer les algorithmes des différents parsers (attention, pas le code, mais la structure logique du programme, ce que vous avez du voir avec Jérôme Farinas)
* Les fichiers de vos résultats qui ont été utilisé comme entrée de vos programmes (GenMark, scan\_for\_matches)
* Les fichiers de vos programmes (pas dans le rapport mais dans des fichiers séparés pour que je puisse les tester)
* Le fichier de la séquence annotée en format EMBL

Précision pour le parser des résultats de scan\_for\_matches (voir la documentation en ligne sur le site silico)