

TP1

Résumé de la séance

Au cours de cette séance, les séquences nucléotidiques des gènes codant les protéines GP de plusieurs espèces de virus Ebola sont récupérées à partir de la base de données publique Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>). Les séquences sont ensuite réunies dans une unique fichier qui sera utilisé lors des prochaines séances (alignement multiple et un arbre phylogénétique).

Accès aux documents

Les documents pédagogiques en relation avec cette séance de travail sont accessibles sur la plateforme de cours eCampus⁴.

- [Fichier PDF](#) de l'article Suzuki et Gojobori (1997). Il est demandé de lire cet article, utile pour une bonne compréhension et interprétation des analyses réalisées.
- [Rappel de la terminologie](#) des espèces.

Travail à réaliser

Partie 1

Cette première étape a pour objectif d'extraire depuis la base de données publique Genbank les séquences des gènes de glycoprotéines (GP) utilisées par Suzuki et Gojobori (1997) pour leur analyse phylogénétique des espèces d'Ebolavirus. **La structure d'un gène de GP et de ses différents cadres de lecture a été présentée en cours.** Les numéros d'accès des séquences sont listés dans la **Table 1** de l'article.

Votre travail consiste à récupérer les séquences codantes (nucléotidiques) des GP sur le site Genbank ("nucleotide") et d'enregistrer ces séquences dans un fichier texte au format FASTA. Notez ici qu'il ne faut pas récupérer les séquences de "Marburg", celles-ci sont trop éloignées des séquences "Ebola" (celles que nous étudions). Pour vous aider dans ce travail, vous pouvez suivre la procédure présentée dans cette [vidéo de démonstration](#) ("Comment récupérer une séquence depuis la banque de données Genbank", vidéo réalisée par Daniel Gautheret).

Lors de la recherche des séquences dans la banque de données Genbank, les fiches descriptives obtenues comprennent : 1) la séquence ADN du gènes (en bas)

⁴ En cas de problème d'accès à eCampus, le lien suivant peut être utilisé : https://drive.google.com/drive/folders/1Qflp6LlbiOFhwGUUn_jvKufS7TRSbTIQ?usp=sharing

et 2) les annotations des cadres de lecture (ou CDS=coding sequence) associés. Remarquez que la plupart des fiches présentent deux traductions différentes du même gène : GP et SGP. Nous souhaitons récupérer la séquence "GP". Pour cela, cliquez sur le lien "CDS" correspondant, puis en bas à droite sur l'onglet "FASTA". La séquence CDS seule est ainsi isolée. Sauvegardez cette séquence avec le lien : « send to » + « file » + "format FASTA" + "Create File". Manuellement, créez enfin un unique fichier au format MULTI-FASTA à partir des fichiers FASTA précédemment sauvegardés.

Vérifier la justesse de votre travail en complétant le questionnaire suivant : <https://forms.gle/bWzXoCLyqJEDKsoc6>.

Etape 2

Il s'agit à cette étape de reproduire l'arbre phylogénétique obtenu par Suzuki et Gojobori (1997). Pour cela le site Phylogeny.fr est utilisé (<https://ngphylogeny.fr/>). Réalisez un alignement multiple des séquences collectées à l'étape précédente, ainsi qu'un arbre phylogénétique. Vous utilisez les valeurs par défaut des différents paramètres (analyse en mode "one click"). Observez l'alignement multiple et l'arbre associé. Retrouvez-vous la topologie de l'arbre présenté Figure 1, dans la publication scientifique de 1997 ? Sauvegardez l'alignement au format FASTA.

Etape 3

La souche d'ebolavirus responsable de l'épidémie de 2014 tire son nom du fleuve Makona en Guinée et Sierra Leone. Nous voulons déterminer l'origine de cette souche en positionnant la séquence de son gène codant la GP au sein de l'arbre précédent.

Recherchez sur le site de Genbank/nucleotide (voir Étape 1) deux séquences d'ebolavirus de type « Makona ». Récupérez la séquence du gène (CDS) de la GP, ajoutez là au fichier FASTA et refaites un alignement multiple et arbre en utilisant Phylogeny.fr. La structure de l'arbre est-elle toujours cohérente avec l'article Suzuki 1997 ? Quelle hypothèse peut-on faire sur la souche Makona 2014 ? Sauvegardez l'alignement multiple en format FASTA.

Vérifier la justesse de votre travail en complétant le questionnaire suivant : <https://forms.gle/R6Udg4GLF1uRV9S29>.

Étape 4 (facultative)

Au cours de la prochaine séance, la console UNIX sera utilisée. Pour préparer cette séance, l'enseignant en salle vous montre : 1) comment ouvrir une console, 2) comment créer un dossier (commande "mkdir"), 3) comment se déplacer dans d'un dossier à un autre (commandes "pwd" et "cd") et 4) comment copier/déplacer un fichier (commandes "mv" et "cp").

Ressources complémentaires

- Le virus et la maladie Ebola : [Page internet de l'Institut Pasteur](#).
- La base de données Genbank : [Vidéo "GenBank Tutorial | How To Use GenBank Database"](#) (5 minutes).
- Les formats de fichiers FASTA et MULTI-FASTA : [Page internet Wikipedia](#).