

Labo 3 – CRISPR

Professeur : Xavier Brochet

Assistant : Thibault Schowing

Modalités

Après la superbe introduction sur les systèmes de défense des bactéries en milieu fromager, suivez les instructions données dans ce document et répondez aux questions dans un fichier nommé **BBC_TP3_Nom_Prenom.pdf** que vous rendrez au format pdf **sur Teams** (Ou Cyberlearn selon la mode de l'année).

Objectifs

En tant que producteur de fromage, vous utilisez des cultures de bactéries très précieuses car elles vous permettent de garantir les saveurs qui ont rendu votre fromage célèbre. Vous voulez savoir si vos bactéries ont des défenses adaptatives et contre quels bacteriophages connus.

Étape 1 : télécharger le génome

Télécharger le génome de « votre » bactérie directement depuis <https://www.ncbi.nlm.nih.gov/>. Sélectionner la base de donnée « Assembly » dans le menu déroulant puis chercher la référence suivante : **GCF.000011845.1**. Vous pouvez aussi directement accéder à la base de données RefSeq et saisir la référence dans la barre de recherche ici : <https://www.ncbi.nlm.nih.gov/refseq/>.

Selon le chemin choisi dans les méandres du site du NCBI ou même la version du site, vous aurez plusieurs options pour télécharger le génome. Vous pouvez utiliser le menu « Send to » ou vous aurez directement un bouton « Download assembly » ou « Download ».

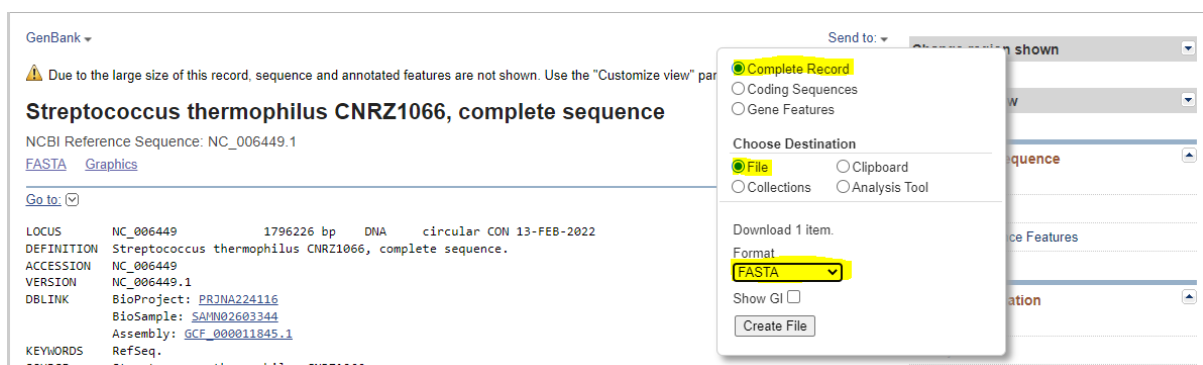


Figure 1 Enregistrement du génome GCF.000011825.1 - menu "Send to"

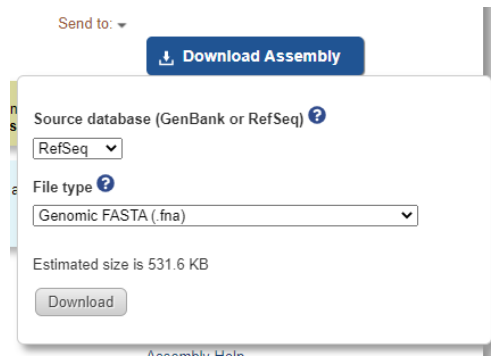


Figure 2 Variante avec bouton "Download Assembly"

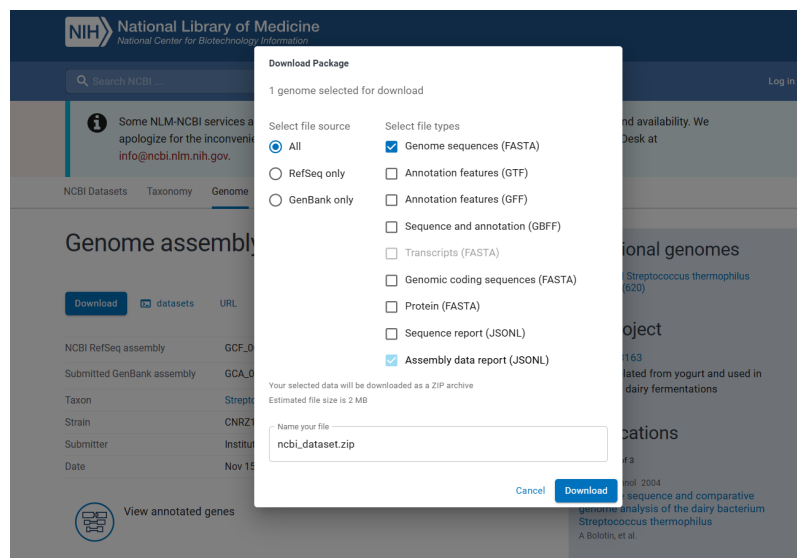


Figure 3 Variante 2025 avec simplement l'option "download"

Une fois que vous avez téléchargé et extrait le génome, vérifiez que la taille du fichier *.fna* est d'environ 1'777 KB.

Avant de continuer, répondez aux questions suivantes :

1. Que sont les bases de données « Assembly » et « RefSeq » ? Expliquez rapidement avec vos propres mots.
2. Si vous devez télécharger plusieurs centaines de génomes, il vous faudra automatiser (script) le processus. Recherchez rapidement ce que le NCBI propose comme solution et écrivez une courte description.
3. L'assemblage d'un génome peut être composé de plusieurs séquences. Combien de séquences notre fichier contient-il ? Donner le(s) identifiant(s). (note : notepad++ suffit)

Étape 2 : Analyser le génome (CRISPR)

Nous voulons savoir si notre bactérie possède un système de défense adaptatif CRISPR. Commencez par répondre aux questions suivantes (petit rappel de biologie dans le contexte) :

1. Que signifie l'acronyme CRISPR-Cas ?
2. Quels sont les éléments principaux qui composent le système CRISPR ?
3. Décrivez de manière concise le fonctionnement de CRISPR Cas en général.

Nous allons maintenant utiliser l'outil CRISPRCasFinder, qui permet de trouver dans un génome les différents éléments composant le système CRISPR. Rendez-vous sur la page <https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index> puis observez la page et les différentes options qui vous sont proposées pour obtenir les éléments CRISPR Cas d'un génome.

1. Quelles sont ces options ?
2. Dans le cas où vous auriez plus de demandes que ce que ne l'autorise la version web, quelles seraient vos options pour analyser vos génomes avec CRISPRCasFinder ?

Pour un résultat rapide et consistant, vous allez utiliser la référence du génome et rechercher dans la base de données CRISPRCasDB (premier champ sur la page d'accueil de CRISPRCasFinder).

1. Décrivez les résultats obtenus.
 - a. Quels éléments obtenez-vous ?
 - b. Que contiennent ces éléments ?
 - c. Cliquez sur « CP000024_1 » de l'élément « CRISPR ». Quelles informations trouvez-vous dans l'onglet « Détails » (*omettez les éléments à 0 ou NA*) ? Et dans l'onglet « Fasta » ?

Étape 3 : BLAST

Nous voulons maintenant analyser les différents spacers trouvés pour découvrir s'ils correspondent à un virus connu ou à un autre organisme. Pour cela, nous devons chercher des séquences similaires dans une base de données. Nous allons utiliser BLASTn (BLAST nucleotide – nucleotide) disponible à l'adresse suivante :

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BLASTSearch

Copiez les séquences (des spacers uniquement) au format fasta dans la zone de texte prévue à cet effet, vérifiez que vous allez bien chercher dans « Nucleotide collection (nr/nt) » et que la recherche est optimisée pour les séquences très similaire (megablast). Ce sont normalement les paramètres par défaut. Exécutez le BLAST et attendez les quelques seconds nécessaires pour que les résultats s'affichent.

1. Que signifie le nt et nr de la base de données dans laquelle nous avons cherché ?
2. Combien de séquences produisent un alignement significatif **pour le premier spacer** ?
Décrivez les informations que vous obtenez sur la page de résultats ainsi que sur les séquences correspondantes (quelques phrases, résumé des métriques). Ajoutez une capture d'écran des résultats (pour le premier phage).
3. Pour la première séquence (le premier spacer), à quelles espèces appartiennent les séquences correspondantes trouvées ?

Étape 4 : Un peu de parsing !

Dans la ligne « Results for » vous pouvez choisir pour laquelle des séquences vous voulez les résultats. Voir les résultats un par un peut être utile mais dans notre cas, il serait plus simple d'avoir tous les résultats d'un coup afin de pouvoir les analyser manuellement. Les résultats sont disponibles dans une variété de formats et contiennent plus ou moins d'informations.

1. Téléchargez les résultats au format JSON en cliquant sur « Single-file JSON »
2. En utilisant la méthode de votre choix (Python, SED, AWK, Notepad++, Bash, Assembleur, ...) extrayez les informations qui concernent les phages du fichier JSON. Pour cela testez simplement si le mot « phage » est contenu dans le champ « title » de la description. Certaines descriptions contiennent deux éléments, vous pouvez ignorer le second car ils sont souvent identiques.

A combien de phages nos spacers correspondent-ils ?

A combien de phages uniques correspondent-ils ?

Bravo, vous avez identifié des phages contre lesquels votre bactérie a une défense adaptative !

Résumez les étapes de ce labo en quelques phrases. Résumez les étapes nécessaires si vous voudriez en savoir plus sur le génome de votre bactérie. Écrivez un court paragraphe sur l'annotation de génomes et les étapes que vous pensez devoir accomplir pour la réaliser. Explorez rapidement les données déjà présentes sur la page du NCBI où vous avez téléchargé le génome.