

TP Alignement de génomes complets bactériens

Question 1. Comparaison et alignement de trois paires de génomes avec le logiciel Mummer

Trois lots de génomes complets vont sont fournis dans les répertoires STREPTO, HELICO et ACTO

- **STREPTO** : 2 génomes de l'espèce *Streptococcus thermophilus* : la souche CNRZ1066 (CP000024) la souche LMD9 (CP000419)
- **HELICO** : 2 génomes de l'espèce *Helicobacter Pylori* : la souche 26695 (AE000511) et la souche J99 (AE001439)
- **LACTO** : 2 génomes de 2 espèces du genre *Lactobacillus* : *Lactobacillus johnsonii* souche NCC533 (AE017198) et *Lactobacillus plantarum* souche WCFSI (AL935263). =>

1.1. Comparaison des génomes 2 à 2 en calculant les matches maximaux mais sans produire alignement.

- 1) lot STREPTO, génomes CP000023 et CP000419,
- 2) lot HELICO, génomes AE000511 et AE001439,
- 3) lot LACTO, génomes AE017198 et AL9352763.

Utiliser le programme « mummer » avec l'option permettant de détecter les MEM nucléiques de longueur minimale 20 pb. Visualiser ensuite les résultats graphiquement sous forme de dotplot avec la commande mummerplot. Donner des noms de fichiers résultats explicites.

Tester plusieurs tailles de MEMs (par exemple 50 et 20 nt) : qu'observez vous ? Comparer les résultats pour les 3 paires de séquences.

Indiquer si les génomes analysés sont colinéaires ou présentent des réarrangements.

Pouvez-vous avoir une idée de leur degré de divergence ?

Calculer la couverture des MEMs pour chaque paire de génomes alignés.

1.2. Réaliser les alignements des trois paires de génomes.

Utiliser le programme *nucmer* pour Strepter et Helico et *promer* pour Lacto.

Promer est un outil d'alignement adapté aux génomes divergents : toutes les étapes de recherche de matches et d'alignements se font sur les traductions protéiques dans les 6 phases de lecture des génomes, ce qui lui donne une meilleure sensibilité. Utiliser *promer* sur les génomes du lot LACTO d'abord avec les paramètres par défaut puis en baissant la taille des ancres (mais utiliser alors les MUMs pour que ce soit plus rapide), visualiser graphiquement et commenter les résultats.

Quelle est le pourcentage d'identité moyen des trois alignements de génome ?

Quelle est le nombre d'inversions détectées dans l'alignement HELICO ? A quoi correspondent elles ?

Question 2. Alignements multiples de 9 génomes de *Salmonella enterica*

On vous fourni un lot de 9 génomes complets (non drafts) proches de la sous espèce *Salmonella enterica subsp. Enterica* (voir fichier *Salmonella_dataset.csv* dans le répertoire SALMO).

Aligner les 9 génomes avec **mauve** et **parsnp** et visualiser les résultats.

Pour visualiser le résultat de l'alignement mauve graphiquement vous utiliserez l'application Java Mauve en local sur vos postes de travail (transfert du fichier résultat .mauve de genotoul vers votre poste et ensuite lancement de Mauve et dans menu 'File' cliquer sur 'Open Alignment').

Pour visualiser le résultat de l'alignement parsnp graphiquement vous utiliserez l'outil *gingr* fourni dans le package Harvest en local sur vos postes de travail (transfert du fichier résultat parsnp.gingr de genotoul vers votre poste et ensuite lancer la commande :
`/Applications/Gingr1.2.app/Contents/MacOS/gingr parsnp.ggr`).

2.1. Analyser la structure de l'alignement avec le viewer de Mauve.

Combien y'a-t-il de LCB dans l'alignement ? Y'a-t-il des réarrangements dans l'alignement ? Décrire la structure de l'alignement. Que se passe-t-il si on fait varier le poids des LCB ?

2.2. Regarder l'arbre phylogénétique produit par parsnp en utilisant le viewer gingr

L'arbre est-il cohérent avec les sérotypes ? Indiquer le nombre de SNPs détectés par génome.

2.3. Comparer la taille de l'alignement, le %moyen d'identité et le nombre de gaps des 2 alignements Mauve et ParSNP

Pour cela utiliser la commande *harvesttools* pour produire les alignements au format multifasta et ensuite la commande *infoalign* pour générer les stats d'alignement.

Commenter les différences entre Mauve et parsnp.

Question3. Alignement de 8 génomes complets et drafts du groupe A de *Wolbachia*

On vous fourni un lot de 8 génomes complets (4 complets et 4 drafts) du groupe A de l'espèce d'endosymbiont *Wolbachia*. (voir fichier *Wolbachia_dataset.csv* dans le répertoire WOLBA).

Comparer les génomes deux à deux avec *mummer* (prendre Wmel comme référence).
Interpréter

Aligner ensuite les génomes avec un outil de votre choix parmi *parsnp*, *ProgressiveMauve* et *Mugsy*.

Analyser et interpréter les résultats obtenus.