

TUTORIAL Alignement de génomes complets bactériens

1. Connexion

Pour se connecter utiliser le login/mot de passe correspondant à votre numéro de machine. Ouvrir ensuite un terminal et se connecter sur **genotoul** : `ssh -XY genotoul.toulouse.inra.fr`. Avec les comptes « fleur » (anemone, arome, aster, bleuet, camelia, capucine, chardon, clematite, cobee, coquelicot, cyclamen, dahlia, digitale, geranium, gerbera, glaieul, hortensia, iris, et jacinthe).

Ensuite travailler dans un répertoire du /work et chacun lance ses alignements sur un nœud différent du cluster en utilisant la commande :

```
qlogin -l hostname=node001 (node002, node003, etc...)
```

2. Outils d'alignements utilisés et docs

Les logiciels à utiliser sont : mummer, yass, les outils mauve (mauve 2.3.1, progressiveMauve, MauveContigMover), mugsy, les outils de la suite Harvest et l'outil infoalign de Emboss..

Les documentations sont accessibles en ligne :

- **Mummer** : <http://mummer.sourceforge.net/manual/>

- **Mauve et ProgressiveMauve** : <http://darlinglab.org/mauve/mauve.html>

- **Mugsy** <http://mugsy.sourceforge.net>

- les outils de la suite **Harvest** qui inclut 3 outil : **parsnp**, **Gingr**, **HarvestTools**, doc :

<https://harvest.readthedocs.io/en/latest/index.html>

- la commande **infoalign** de la suite **Emboss** pour le calcul des stats (nombre identité, couverture) des alignements produits, doc :

<http://emboss.sourceforge.net/apps/release/6.2/emboss/apps/infoalign.html>

3. Commandes à utiliser pour le TP

- **Lancer un calcul de MEM de taille 20 sur une paire de génomes avec mummer**

```
mummer -maxmatch -b -c -l 20 <seq1.fsa> <seq2.fsa> > seq1-seq2.mem20
```

-maxmatch : Compute all maximal matches regardless of their uniqueness (=> Les MEMs et pas les MUMs)

-b : Compute both forward and reverse complement matches

-c : Report the query position of a reverse complement match relative to the forward strand of the query sequence

-l int : Minimum match length (default 20)

- **Visualiser le résultat de calcul des MEM sous forme de dotplot**

```
mummerplot seq1-seq2.mem20
```

- **Lancer un alignement nucmer ou promer**

```
nucmer --maxmatch -prefix=fichier seq1.fsa seq2.fsa
```

```
promer --maxmatch -prefix=fichier seq1.fsa seq2.fsa
```

Génère un alignement au format « .delta ».

-> Pour visualiser les alignements :

```
mummerplot fichier.delta
```

-> Pour afficher un résumé de l'information sur chaque alignement :

```
show-coords -c -d -k -r fichier.delta > fichier.coords
```

-c : Include percent coverage information in the output

-d : Display the alignment direction in the additional

FRM columns (default for promer)

-k : Knockout (do not display) alignments that overlap

another alignment in a different frame by more than 50% of their length, AND have a smaller percent similarity or are less than 75% of the size of the other alignment (promer only)

-r : Sort output lines by reference IDs and coordinates

-> Pour visualiser les alignements :

show-aligns fichier.delta > fichier.aligns

- **Lancer un alignement mauve**

mauveAligner [options] <seq1 filename> <sml1 filename> ... <seqN filename> <smlN filename>

ou

mauveAligner [options] <Multi-FastA file>

- **Lancer un alignement parsnp**

parsnp -x -g genome.gbk -d rep-fasta

Options

- x : permet le filtrage de la recombinaison

-g : fichier genbank

- d : répertoire contenant les génomes à aligner au format fasta

Résultats : fichiers de sortie .gingr .xmfa .tree dans un répertoire [./P_CURRDATE_CURRTIME]

- **Lancer un alignement ProgressiveMauve**

progressiveMauve --output=threeway.xmfa genome_1.gbk genome_2.gbk genome_3.gbk

Résultat : fichier XMFA

- **Lancer un alignement Mugs**

mugsy [-p output prefix] multifasta_genome1.fsa multifasta_genome2.fsa ... multifasta_genomeN.fsa

Résultat : fichier MAF

- **Conversions de format**

- **Fomat 'MAF' vers format 'multifasta'**

/usr/local/bioinfo/src/HarvestTools/harvesttools-Linux64-v1.2/harvesttools -a fichier.maf -M fichier.fsa

- **Fomat 'XMFA' vers format 'multifasta'**

/usr/local/bioinfo/src/HarvestTools/harvesttools-Linux64-v1.2/harvesttools -i fichier.xmfa -M fichier.fsa

- **Statistiques de l'alignement**

infoalign fichier.msa