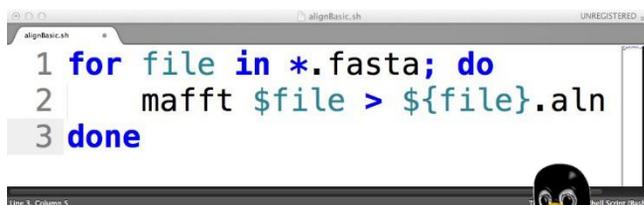


LINUX POUR L'ANALYSE DE DONNEES GENOMIQUES

Du 30 septembre à 9h au 4 octobre 2024 à 17h 35 h

Le système d'exploitation Linux est disponible sur de nombreux ordinateurs. Il constitue notamment le cœur du système d'exploitation des ordinateurs MAC, c'est également le système d'exploitation de la majorité des serveurs/clusters de calculs. Même s'il dispose d'une interface graphique conviviale, sa force repose essentiellement sur la possibilité d'utiliser un ensemble de commandes textuelles très puissantes qu'il est possible d'enchaîner pour réaliser automatiquement des tâches complexes et/ou répétitives.

Ceci permet de traiter plus facilement des données stockées sous formes de nombreux (des milliers) ou de très volumineux (>Gb) fichiers. De plus, des dizaines d'outils liés à l'analyse de données génomiques (notamment pour le traitement de données NGS) ne sont disponibles que pour Linux.



```
1 for file in *.fasta; do
2     mafft $file > ${file}.aln
3 done
```



OBJECTIFS

Connaître les commandes de base du système Linux et savoir les utiliser pour traiter des données génomiques.

A la fin du module vous serez capable d'écrire des scripts simples en « bash » (le langage de base de Linux)

Ceci constitue également une introduction à l'écriture de scripts en Perl ou Python.

Mots clefs : Génomique – Linux – programmation – alignement de séquences – phylogénie – bioinformatique – Cluster de calculs – Séquençage de nouvelle génération (NGS) – polymorphisme moléculaire – Blast

METHODES ET SUPPORTS PEDAGOGIQUES

Cours introductif

Mises en pratique en salle informatique sur des cas concrets. Accès temporaire à un cluster de calcul.

L'accent est mis sur la mise en pratique autonome sur des cas concrets, afin qu'à la fin de la formation les stagiaires soient autonomes et disposent d'une série de scripts utiles qu'ils ont développés eux-mêmes et qu'ils pourront réutiliser en les adaptant à leurs besoins.

public

Professionnels du secteur public ou privé confrontés à la nécessité d'automatiser des analyses génomiques du fait de l'augmentation des quantités de données.

TARIF

410 € nets de taxes/jour/participant soit 2 050€ pour les 5 jours (non dissociables)
425 € nets de taxes/jour/participant soit 2 125€ pour les 5 jours (non dissociables)

Agents Cirad, Inrae, et autofinancement : nous contacter

Déjeuners pris en charge par l'Institut Agro Montpellier

Encadrement pédagogique

Isabel Martin Grande, ingénieur pédagogique numérique, l'Institut Agro Montpellier

Intervenants :

Vincent Ranwez, professeur, Montpellier SupAgro

Muriel Tavaud, maître de conférences, Montpellier SupAgro

Alexandre Dehne Garcia, ingénieur INRAE

Responsable pédagogique :

vincent.ranwez@supagro.fr

Pré-requis

Connaissances de bases en génomique. Les exemples seront traités essentiellement du point de vue de l'automatisation des tâches (l'intérêt et les fondements de ces analyses ne seront que brièvement évoqués)

Aucune compétence en informatique n'est pré-requis, mais un goût pour l'outil informatique et pour la logique sont essentiels.

Contenus

- Manipulation de nombreux fichiers fasta et fastq (les renommer, en déplacer un sous-ensemble, les trier, compter le nombre de séquences qu'ils contiennent, etc.)
- Collecte des résultats contenus dans plusieurs fichiers pour obtenir un fichier de synthèse analysable facilement à l'aide de tableurs (e.g. Excel) ou de R
- Enchaînement des étapes nécessaires à une analyse phylogénétique s'appuyant sur plusieurs gènes
- Enchaînement des étapes nécessaires à l'identification de polymorphisme à partir de données NGS



Date limite d'inscription
09/09/2024

Contact inscriptions

Sarah Hémici

☎ 04 99 61 23 56

sarah.hemici@supagro.fr