



Ingénieur-e biologiste en traitement de données

Corps : Ingénieur d'études

Nature du concours : Concours externe

BAP (branche d'activité professionnelle) : A

Emploi type : A2A41 – Ingénieur-e biologiste en traitement de données

Nombre de poste(s) offert(s) : 1

Localisation du poste : EPHE – CEFE (Centre d'Ecologie Fonctionnelle et Evolutive) - Campus CNRS, 1919 route de Mende - 34090 Montpellier.

ENVIRONNEMENT ET CONTEXTE DE TRAVAIL

L'École Pratique des Hautes Études (EPHE) est un grand établissement d'enseignement supérieur et de recherche, membre de l'Université Paris Sciences et Lettres (PSL), et membre fondateur du Campus Condorcet.

Le CEFE est un des plus importants laboratoires de recherche en écologie et évolution en France. C'est un laboratoire leader sur les sujets liés à la transition écologique, les sciences de la conservation et d'une très grande visibilité internationale pour la recherche fondamentale en écologie et évolution. L'EPHE en est tutelle principale au même titre que le CNRS, l'Université de Montpellier et l'IRD.

Le poste est rattaché à la plateforme GEMEX (Génomique, Écologie Moléculaire et Évolution expérimentale) du laboratoire. Il/Elle rejoindra une unité où la génomique est en forte croissance et y bénéficiera d'un excellent environnement de travail. Au CEFE, il/elle sera la personne référente en bio-informatique pour les projets du laboratoire impliquant des données génomiques. Il/Elle sera en charge de la gestion des données générées par le séquenceur Promethion présent sur la plateforme. Il/Elle pourra s'appuyer sur des moyens importants de calculs (machines GPU, serveurs de calcul locaux et clusters externes). Il/Elle animera le groupe bio-informatique du CEFE qui réunit plusieurs ingénieurs affectés à des projets ainsi que d'autres personnels impliqués de l'unité. Il/Elle assurera également le lien entre la partie biologie moléculaire de la plateforme GEMEX et le pôle numérique de l'unité. L'ingénieur.e en bio-informatique interviendra à la fois sur les projets scientifiques en interne et coordonnera les interactions avec les plateformes régionales/nationales de calcul (en particulier meso@LR, clusters de calcul de l'IFB). Il/Elle travaillera notamment en coordination avec les ingénieurs de la plateforme MBB (Montpellier Bioinformatique Biodiversité) et des autres plateformes de bio-informatique de la communauté.

MISSIONS

L'ingénieur-e en bio-informatique organise et réalise la gestion et le traitement bio-informatique de données de séquençage en génomique écologique, évolutive et environnementale de l'unité. Il/Elle accompagne et forme les utilisateurs de données de séquençage au sein de l'unité.

ACTIVITÉS PRINCIPALES

- Réaliser le traitement polyvalent des données de séquençage sur courts et longs reads.
- Gérer des données produites par le Promethion de GEMEX et support bio-informatique sur ce séquenceur.
- Assister, conseiller et former les utilisatrices et utilisateurs aux diverses techniques de traitement des données de séquençage dans différents environnements (clusters de calcul, cloud, Linux, Mac et Windows).
- Déboguer les scripts de bio-informatique des utilisatrices et utilisateurs.
- Organiser la mise en forme et le stockage des données et assurer la maintenance des bases de données créées.
- Appliquer et faire appliquer les principes FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable) de la science ouverte et reproductible, en lien avec l'ingénierie de recherche Science Ouverte du pôle numérique du CEFE.
- Assurer une veille technologique et scientifique et participer à des réseaux professionnels (par exemple plateformes de bio-informatique externes).

COMPÉTENCES REQUISES

Connaissances

- Connaissances et savoir-faire approfondis et polyvalents en bio-informatique : traitement de différents types de données de séquençage (génomomes entiers et réduits, transcriptomes, amplicons, ADN environnemental) par approches variées (exemples utilisés actuellement au CEFE : assemblages et annotations de génome, appel de variants, analyse d'expression, épigénétique, métabarcoding, métagénomique).
- Connaissances en langage (Bash, Python, R, conda, slurm) et environnements informatiques et gestionnaire de workflows.
- Connaissances en génétique et génomique.
- Intérêt pour la biologie évolutive, l'écologie, et le suivi de la biodiversité.
- Anglais avancé (niveau B2).

Savoirs-faire

- Utiliser des *pipelines* déjà existants en ligne de commande et les modifier si nécessaire pour le traitement de données de type NGS dans le cadre des projets menés par la plateforme GEMEX et les équipes.
- Traiter d'importants jeux de données en travaillant sur un cluster informatique.
- Interpréter les résultats des analyses dans leur contexte scientifique.
- Rédiger en anglais les parties des articles scientifiques en lien avec les données analysées.

Facultatif : analyser statistiquement les données obtenues à l'aide d'un langage de programmation (R, Python, Julia...) et/ou à partir de programmes existants pour des analyses plus spécialisées de génétique des populations ou phylogénie.

Savoirs-être

- Aptitude à travailler en équipe
- Capacité à restituer ses résultats à des publics divers, des plus expérimentés aux débutants
- Ecoute et pédagogie
- Esprit d'initiative et autonomie.
- Rigueur et sens de l'organisation.