Les outils de la Bioinformatique au service d’un projet

de recherche en Sciences de la Vie

**(BIOINFORMATICS TRAINING PROGRAM)**



# Objectifs et Compétences

Cette formation intensive vise l'acquisition de compétences de base en programmation appliquée à la biologie, ainsi qu’une maîtrise pratique des principaux outils bioinformatiques, dont ceux conçus pour le traitement des données à haut débit (microarrays, Next-Generation Sequencing).

Elle se compose de deux modules (de base et avancé) d’une durée de 15 jours chacun, le module avancé n’étant accessible qu’aux stagiaires ayant réussi le module de base.

Au terme de cette formation, vos nouvelles compétences vous permettront d’exercer comme bioinformaticien ‘junior’ au sein d’instituts de recherche ou de poursuivre, convenablement outillés, une thèse de doctorat dans le secteur des Sciences de la Vie.

La formation sera axée sur la pratique, chaque stagiaire étant équipé d’un ordinateur portable sous Bio-Linux 8. Les formateurs proposeront en particulier des pas-à-pas à suivre individuellement. Ces séances seront régulièrement interrompues pour aborder succinctement en groupe complet les points plus théoriques. Des exercices dirigés et des devoirs à domicile compléteront le dispositif d’apprentissage en vue d’ancrer la matière vue. La progression de chacun sera assurée par des approches différenciées.

# Description de la formation

**Module de base :**

1. Introduction, présentation des objectifs et remise à niveau des candidats

2. Rappels de statistiques

L’objectif de ce module est la remise à niveau en matière de statistique de base.

- Statistiques descriptives : fréquence, moyenne, écart-type, médiane, erreur type, les graphiques

- Calcul de probabilité

- Notion de distribution : distribution normale et binomiale

- Les tests d’hypothèses : test d’1 et 2 moyennes, test d’1 ou 2 variances

- L’ANOVA à 1 facteur

- La régression simple

Les notions seront appliquées sur des exercices « papier » ou en utilisant le logiciel JMP.



3. Introduction à Linux et à la ligne de commande

Découvrir UNIX au travers d'un Linux conçu pour les biologistes

Installation, configuration et prise en main de Bio-Linux 8

(1 séance)

Introduction à la ligne de commande (2 séances)

Apprivoiser le Terminal

Utilisation des commandes usuelles (navigation dans l’arborescence, manipulation des fichiers et des répertoires, lecture et édition de fichiers…)

Du one-liner au programme (2 séances)

Introduction à la gestion des permissions, des processus

Création de programmes ‘scripts’

4. Modern Perl – Concepts de base

S’initier à la programmation en étudiant des programmes simples mais utiles

Apprendre à tirer parti des modules et de l'expressivité du langage Perl

Variables (scalaires, tableaux indicés, tables de hachage)

 (1 séance)

Valeurs, opérateurs, expressions et structures de contrôle

(2 séances)

Gestion des entrées/sorties et best of des modules CPAN

(1 séance)

Expressions régulières et one-liners (1 séance)

5. R – Introduction à l’analyse de données

Découvrir R en révisant les statistiques univariées de base

Aborder l'utilisation de R par l’exploration et la manipulation de jeux de données

Environnement et session sous RStudio, objets et opérateurs de base (1 séance)

Objets adaptés aux jeux de données, fonctions, import/export (1 séance)

Packages, conditions, boucles, graphiques, statistique descriptive (1 séance)

6. Applications bioinformatiques – Premiers contacts

S'initier à l'analyse de données de séquençage à haut-débit (Illumina, 454)

Métagénétique 16S (1 séance)

Assemblage de novo de génomes bactériens (2 séances)



**Module avancé :**

1. Introduction, présentation des objectifs

2. Modern Perl – Concepts avancés

Approfondir sa maîtrise de la programmation via des programmes complexes

Apprendre les bonnes pratiques, dont modularisation et documentation du code

Fonctions (1 séance)

Références et structures de données multidimensionnelles

(2 séances)

Modules, tests unitaires, programmes externes et templating

(1 séance)

Perl idiomatique (1 séance)

Introduction à la programmation orientée objet (1 séance)

3. R – Perfectionnement

Découvrir des fonctions R avancées en révisant les statistiques multivariées

S’initier au traitement de données en génomique et en écologie numérique

Statistiques avancées sous R (2 séances)

Statistiques multivariées, ANOVA

Distributions de probabilités, tests statistiques

Graphiques avancés (dont 3D), accompagnement au projet personnel (1 séance)

Production de rapports avec knitr et markdown (1 séance)

Analyse en Composantes Principales, dendrogrammes et clustering (2 séances)

Application à la génomique (microarrays) et à l’écologie numérique (2 séances)

. Introduction aux bases de données et au langage SQL

S'initier à la conception logique de bases de données

Apprendre à interroger efficacement une base de données

Aspects théoriques de la conception des bases de données

(3 séances)

Le modèle Entité-Association

Du modèle E-A au modèle relationnel et normalisation

De la théorie à la pratique... (1 séance)

Concevoir une base SQLite à l’aide du logiciel DB-Main

Introduction au langage SQL (2 séances)

Requêtes élémentaires et fonctions SQL

Jointures et requêtes complexes



5. Applications bioinformatiques : les classiques

Se perfectionner dans l'analyse de données de séquences ou de génomes complets

Aborder l'exploitation de ces données au travers de cas concrets

Introduction aux clusters de calcul intensif (1 séance)

Identification de SNPs entre souches bactériennes (1 séance)

RNA-seq chez un eucaryote et expression différentielle

(2 séances)

Cartographie d'une mutation dans un génome eucaryote

(2 séances)

Analyse phylogénétique d'une famille multigénique (2 séances)

6. Débriefing et conclusions

# Durée et horaire :

**Module de base :**

15 jours de formation (8h45-17h) du 4/11/2019 au 18/12/2019 (les lundis et vendredis ainsi que quelques mercredis).

**Module avancé :** 15 jours de formation (8h45-17h) janvier – mars 2020 (les lundis et vendredis ainsi que quelques mercredis). Ce module n’est accessible qu’après réalisation du Module de base.

Pour les demandeurs d'emploi, la réalisation d'un stage de fin de formation en entreprise (40 jours ouvrables) est possible à l'issue de la formation complète (module de base + module avancé).

Intéressé(e) ?

Pour vous inscrire à la séance d’information et de tests, appelez le 0800 93 947 ou davantage de précisions appelez Anne Huberty au 04/366.45.78 ou par mail : anne.huberty@forem.be