
Projet Bioinformatique

Master SD CD 2019_2020

SALAS VEGA Dominique - MOREL Louis - BONNETAIN Baptiste - LÉNON Lucas -
MONET Anaïs (SDIA TPS)

**Rapport : Big data en Bioinformatique - Statistiques sur les
trinuéotides dans les gènes**

Instructions et librairies permettant l'exécution du programme avec Eclipse

Nous utilisons les librairies Java de base (qui s'ajoutent automatiquement à la création d'un projet dans Eclipse) et Jxl qui permet de traiter les fichiers excel.

Informations sur l'arborescence locale et fichiers nécessaires

L'arborescence se met à jour et nécessite un dossier Results pour fonctionner. Le dossier Results contient 4 autres dossiers (Eucaryotes, Procaryotes...) contenant les différents types d'être vivant. Le niveau inférieur de l'arborescence contient les sous types des êtres vivants. Enfin nous avons les fichiers excel (un par être vivant).

Lors de la création de l'arborescence par l'interface graphique, les fichiers .DS_Store apparaissent. Ces fichiers sont invisibles dans l'arborescence locale. Nous n'avons pas réussi à les enlever. La mise à jour automatique de l'arborescence lors de la création d'un nouveau fichier ne fonctionne pas. Le code couleur non plus car nous ne faisons pas la mise à jour des fichiers, nous les créons seulement.

Nous n'avons pas eu le temps de télécharger tous les fichiers, notre code travaille seulement sur les fichiers non encore téléchargés.

Classe contenant main

La classe contenant le main est *interface.java* .

Remarques :

Nous n'utilisons pas l'API de GenBank pour accéder aux données. A défaut de temps et de compréhension vis à vis des API à utiliser, nous avons jugé adéquat l'utilisation de FTP. Nous nous sommes rendu compte plus tard qu'il fallait utiliser e-utilities.

Nous ne faisons pas de distinction entre les différents types de génome par faute de temps.

Le parsing fonctionne que pour les Eucaryotes car nous pensions que c'était identique pour le reste. En testant à la fin, nous nous sommes rendu compte que non et donc nous n'avons pas refait le parsing pour le reste. Il y a quelques fichiers dans les autres dossiers car notre parsing en détecte quelques uns.

La barre de chargement ne fonctionne pas.

Lorsque nous cliquons sur le bouton **démarrer** de l'interface, il y a une nouvelle fenêtre qui s'ouvre. Nous nous sommes rendu compte de cela au dernier moment en lançant le programme final sous windows. Nous avons testé que sous MacOS et il n'y avait pas ce problème. Nous n'avons pas eu le temps de gérer ce problème. Une fois cliqué sur le bouton, l'interface sur laquelle nous avons cliqué sur démarrer tourne sauf les logs qui se mettent à jour sur la deuxième interface.