Faculté des Sciences
Aix*Marseille Université

A envoyer au responsable du master BBSG :
Herve.darbon@afmb.univ-mrs.fr

FORMULAIRE STAGE Recherche-M2 BBSG
(période de stage : du 5 janvier 2017 au 3 juillet 2017)

Titre du stage : `Méthodes bioinformatiques pour la discrimination de contaminants dans les séquences NGS`

Laboratoire (intitulé, adresse, site web):
Institut Méditerranéen de Biodiversité et Ecologie Marine et Continentale (IMBE)
+ Technologies Avancées pour la Génomique et la Clinique (TAGC).

Equipe : Emese Meglécz (IMBE) + Jacques van Helden (TAGC)

`Maitre de stage :`

E-mail :
Emese Meglecz <emese.meglecz@imbe.fr>,
Jacques van Helden <Jacques.van-helden@univ-amu.fr>

Téléphone : : +33 4 13 55 11 97

Descriptif du stage :

La détection des éventuels contaminants dans les données de séquençage à haut débit (NGS) est cruciale avant toute analyse biologique des données. Les méthodes basées sur la recherche de similarité dans des bases de données de séquences sont très coûteuses en temps et en capacité de calcul quand il s'agit d'aligner des millions de fragments de lecture (reads). Le but de ce stage sera d'explorer la possibilité de détection des contaminants en utilisant la distribution de k-mers (sous-séquence de longueur k) des reads NGS.

Le stage consistera à développer des méthodes de classification de séquences sur base de comptage de k-mères, et à tester la capacité de ces méthodes à détecter les contaminants dans une librairie de reads NGS. Ces méthodes prendront en compte la nature particulière des données : comptages discrets (nombres naturels) avec des espérances faibles et des matrices de comptages relativement creuses. Le logiciel sera développé sous forme d'une librairie R.