



Aide bioinformatique en Biotechnologie : Correction des erreurs des séquençages génomiques par HMM

HAMIDECHI Mohamed Abdelhafid

*Département de Biologie Appliquée, Faculté des Sciences de la Nature et de la Vie, Université des Frères Mentouri,
Constantine 1, Algérie*

Résumé

La BioInformatique, une filière émergente, assiste beaucoup de spécialités (biologie, les sciences médicales, la biotechnologie, etc.) afin de mieux explorer les données massives issues des expérimentations et surtout des assemblages génomiques post-séquençage.

Les deux approches pour assembler les génomes séquençés sont : *de novo* et le génome référence. L'assemblage *de novo* est un processus automatisé et assisté par un algorithme dérivant de la transformée de Burrow-Wheeler et d'un assembleur SGA basé sur la détection de chevauchements entre les différents reads issus du séquençage (1,2 milliard de reads ou courtes séquences) pour construire des contigs puis des scaffolds dans le bon ordre. Une deuxième approche, avec laquelle contribue la bioinformatique dans l'assemblage *de novo* est l'utilisation des graphes de DeBruijn.

De nombreux défis font face à l'assemblage des génomes parmi lesquels la gestion des lectures erronées nées de ces assemblages en utilisant les modèles cachés de Markov (HMMs : Hidden Markovian Models) basées sur le calcul des probabilités conditionnelles. Ces HMMs peuvent également être utilisés pour résoudre divers problèmes d'analyse de séquences, tels que les alignements de séquences par paires et multiples, l'annotation de gènes, la classification, la recherche de similarité, etc.

Mots-clés : Séquençage, Assemblage de novo, Graphes SGA, Graphes de De Bruijn, HMMs.