

TITRE: Développements de modèles informatiques et statistiques spécifiques aux séquences génomiques virales.

EQUIPE/THEME: MabioVis / Modélisation et Comparaison de séquences et structures biologiques

DIRECTEURS: Macha Nikolski et Patricia Thébault

COURRIELS: [macha@labri.fr](mailto:macha@labri.fr) <<mailto:macha@labri.fr>>, [patricia.thebault@labri.fr](mailto:patricia.thebault@labri.fr) <<mailto:patricia.thebault@labri.fr>>

MOTS-CLES: métagénomique virale, NGS, bioinformatique, méthode d'apprentissage, inférence de motifs, comptage OTU

DIRECTEURS HABILITES: Macha Nikolski

DESCRIPTION du SUJET:

La métagénomique, basée sur l'analyse de séquences d'ADN (l'ensemble des génomes est appelé le métagénome) offre une alternative intéressante en contournant la nécessité de la culture bactérienne (étapes de clonages), pour amplifier spécifiquement ou de façon aléatoire un fragment d'ADN isolé. Ainsi, l'avènement ces dernières années des technologies de séquençage à haut débit offre de nouvelles perspectives pour l'étude d'échantillons écologiques correspondant à différentes communautés. En lien direct avec la production de ces nouvelles données et nouvelles questions, leur analyse in silico nécessite de revisiter les algorithmes traditionnels mais également de proposer de nouvelles approches bioinformatiques adaptées. Nous envisageons dans ce contexte plusieurs niveaux de développements bioinformatiques pour proposer de nouveaux outils dédiés à la caractérisation et analyse de séquences virales issus de différents échantillons de métagénomique. Un premier volet s'intéressera à l'inférence de paramètres permettant de différencier des espèces proches ou de relier des espèces éloignées de virus et un second pour le développement de méthodes de clustering ou de barcoding adaptés à l'analyse de la biodiversité d'un échantillon virale. Un certain nombre d'approches ont été proposé pour la métagénomique bactérienne, basée sur l'analyse du 16S, pour l'inférence de signature et/ou analyse de biodiversité. En comparaison avec les bactéries, les génomes viraux contiennent en moyenne une douzaine de gènes et offrent moins de marqueurs phylogénétiques et rendent la question plus ardue. On s'intéressera donc à adapter des méthodes existantes ou en développer de nouvelles dans ce cadre de calcul distribué. Les développements informatiques seront appliquées aux données produites en métagénomique virale par l'UMR BFP autour de ces deux axes.