



Vendredi 14 décembre 2018, 14:30

Salle de réunion

SOUTENANCE DE THÈSE

ANALYSE DE LA DIFFÉRENCIATION GÉNÉTIQUE À L'ÈRE DES NOUVELLES TECHNOLOGIES DE SÉQUENÇAGE

par

Valentin Hivert

Thèse, CBGP

📍 L'avancée des technologies de séquençage et de génotypage à haut-débit permet la comparaison de patrons de polymorphisme à un très grand nombre de marqueurs génétiques. L'analyse de la différenciation des populations à une échelle génomique rend ainsi possible la recherche de régions génomiques impliquées dans l'adaptation locale des organismes à leur environnement. Dans cette thèse, nous avons suivi deux approches complémentaires pour caractériser la différenciation génétique à partir de données de génotypage à haut-débit.

📍 Dans un premier temps, nous avons développé un estimateur non-biaisé du paramètre F_{ST} pour des données de génotypage d'individus en mélange (Pool-seq). La construction de cet estimateur, dans un contexte d'analyse de variance, a nécessité de bien prendre en compte les différentes étapes de l'échantillonnage : des gènes dans le mélange d'individus et des lectures de séquençage parmi les gènes. Nous montrons qu'il surpasse les estimateurs utilisés jusqu'à présent.

📍 Dans un deuxième temps, nous avons développé une méthode d'analyse de la différenciation génétique à l'échelle du génome, dans le cadre d'un modèle bayésien hiérarchique, pour distinguer l'effet de la démographie de celui de la sélection. Pour cela, nous avons implémenté plusieurs extensions au modèle SelEstim, pour exploiter l'information de déséquilibre de liaison entre les marqueurs. Une première stratégie a consisté à analyser des données multialléliques, obtenues par le regroupement local de marqueurs SNPs en blocs d'haplotypes. Une stratégie alternative a consisté à intégrer un modèle de lissage prenant en compte la dépendance spatiale entre marqueurs adjacents. Cette approche repose sur l'analyse de données bialléliques, ce qui la rend applicable à la fois à des données de génotypage individuel et à des données Pool-seq. Nous discutons, sur la base de l'analyse de jeux de données simulées, des mérites relatifs de ces différentes approches.