

Etude de la robustesse aux informations de structure et d'apparentement de nouvelles mesures de déséquilibre de liaison

Thématique : statistique génétique

Lieu : INRA, UR875 Unité de Biométrie et Intelligence Artificielle, Chemin de Borde Rouge, BP 52627, F-31320 Castanet-Tolosan

Projet : DLVITIS

Encadrant : Brigitte Mangin (brigitte.mangin@toulouse.inra.fr), Christine Cierco-Ayrolles

Contexte

Le déséquilibre de liaison - l'association non-indépendante des allèles à différents loci dans un génome - est un outil important pour les généticiens. Le déséquilibre de liaison est largement utilisé pour étudier l'histoire évolutive des espèces ou à des fins de cartographie génétique et de sélection. Il existe plusieurs mesures pour capturer différentes caractéristiques du déséquilibre de liaison. La plus couramment utilisée, pour des loci bialléliques, est la mesure r^2 , le carré du coefficient de corrélation entre les variables nominales indiquant la présence de chacun des deux allèles.

Deux nouvelles mesures de déséquilibre de liaison ont été proposées par l'équipe dans l'objectif de corriger le biais de la mesure r^2 dû à la structure de l'échantillon ou à l'apparentement des individus. Ces deux nouvelles mesures utilisent des informations supplémentaires que sont la structure de la population et la matrice d'apparentement.

Ces nouvelles mesures ont été validées sur des données simulées et appliquées à des données réelles collectées sur la vigne. Les résultats montrent clairement que les informations fournies par la structure de la population et la matrice d'apparentement permettent une correction du biais de l'estimateur de la mesure r^2 .

Objectif du stage

Les informations supplémentaires nécessaires à la correction du biais sont rarement connues a-priori et sont généralement estimées à partir de données génomiques. Des méthodes nombreuses, s'appuyant sur des modélisations différentes, permettent à partir de ces données génomiques d'estimer la structure d'un échantillon et la matrice d'apparentement des individus. L'objet du stage est d'étudier la robustesse des nouvelles mesures aux choix de la modélisation et à l'incertitude découlant de l'estimation.

Travail à réaliser et compétences requises

Le stagiaire devra se familiariser avec les méthodes de l'estimation de l'apparentement et de la structure. Il mettra en oeuvre ces méthodes dans le logiciel R, lorsqu'elles se basent sur des méthodes de réduction de dimension (telles l'ACP, l'ACM,...) ou utilisera des logiciels dédiés lorsqu'elles se basent sur des modèles plus sophistiqués. Il appliquera sur les données de la vigne ces différentes méthodes d'estimation et analysera les résultats dans R.

Indemnisation : environ 400 Euros par mois