

Compréhension entre les liens statistiques et biologiques dans le contexte des réseaux de gènes

Contexte :

Que cela soit dans les domaines de la médecine et de l'agronomie, des données omiques (génomique, transcriptomique, métabolomique, etc.) sont générées pour répondre à certaines questions biologiques. Ces données peuvent être utilisées pour inférer des réseaux biologiques. Un réseau biologique est constitué de nœuds représentant des molécules telles que des gènes, métabolites..., et des arêtes représentant les connexions statistiques ou biologiques entre ces molécules. Différentes méthodes statistiques ont été développées pour inférer des réseaux en identifiant des liens directs (par ex. corrélations partielles) (Altenbuchinger et al., 2020) ou indirects (par ex. corrélation de Pearson) (Langfelder and Horvath, 2008). Malgré l'utilisation de ces approches dans de nombreuses études visant à étudier des mécanismes biologiques d'intérêt, l'adéquation entre les liens statistiques et biologiques reste encore à explorer (Serin et al., 2016). Il s'avère donc essentiel d'étudier les apports de chaque méthode ainsi que leur complémentarité en lien avec la question biologique d'intérêt.

Objectif du Stage :

Comparaison des réseaux de gènes estimés avec différentes méthodes statistiques et des processus biologiques identifiés dans ces réseaux en lien avec l'adaptation des plantes à leur environnement.

Pour atteindre cet objectif, le candidat aura les missions suivantes :

- Revue bibliographique portant sur la comparaison des structures des réseaux de gènes (He et al., 2019) et sur leur validation biologique
- Identifier des méthodes de sélection des gènes adaptées à la question biologique
- Mettre en œuvre avec le logiciel R les méthodes d'inférence de réseaux de gène (WGCNA, GGM, etc.)
- Validation biologique des réseaux inférés
- Rédiger un R Markdown destiné aux biologistes
- Aide pour l'intégration des analyses à l'interface de visualisation R Shiny développée dans l'UMR AGAP

Mots clés : Réseaux statistiques, graphes, génomique

Profil recherché :

Avoir suivi ou suivre une formation en biostatistique / sciences des données

Compétences recherchées :

- Bonne maîtrise des modèles linéaires et des modèles linéaires généralisés
- Programmation R
- Intérêt pour la génomique
- Une connaissance basique de git serait appréciée.

Durée du stage : 4 à 6 mois entre janvier et aout 2024

Lieu du stage : Montpellier, CIRAD Lavalette, UMR AGAP

Indemnité mensuelle : 611 euros

Modalités de candidature : CV et lettre de motivation à envoyer par email Marie Denis (marie.denis@cirad.fr) et Bénédicte Favreau (benedicte.favreau@cirad.fr)

Altenbuchinger, M., Weihs, A., Quackenbush, J., Grabe, H. J., & Zacharias, H. U. (2020). Gaussian and Mixed Graphical Models as (multi-) omics data analysis tools. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1863(6), 194418.

He, H., Cao, S., Zhang, J. G., Shen, H., Wang, Y. P., & Deng, H. W. (2019). A statistical test for differential network analysis based on inference of Gaussian graphical model. *Scientific reports*, 9(1), 10863.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9(1), 1-13.

SERIN, E. A. R., NIJVEEN, H., HILHORST, H. W. M. & LIGTERINK, W. 2016. Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science*, 7, 444.