
Réseaux d'associations au sein de communautés bactériennes : Inférence de réseaux et distribution en loi de puissance de l'abondance des espèces

Arnaud Cougoul*¹

¹Unité d'Epidémiologie Animale (EpiA) – Institut national de la recherche agronomique (INRA) : UR0346 – INRA Auvergne-Rhône-Alpes, Route de Theix, 63122 Saint Genès Champanelle, France

Résumé

Les organismes (animaux, homme, plantes) sont dans leur majorité infectés simultanément par plusieurs parasites. Les prévalences des infections multiples peuvent atteindre 30% voire 80% dans certaines populations humaines. Ce multiparasitisme peut altérer le diagnostic, les traitements et l'épidémiologie des maladies infectieuses. Les épidémiologistes sont demandeurs d'outils pour analyser ces communautés de parasites et identifier de potentielles interactions biologiques impliquant des pathogènes d'importance médicale ou vétérinaire. Or deux caractéristiques rendent ces analyses complexes, d'une part les prévalences de chaque parasite dans les populations peuvent être très faibles, d'autre part, les co-occurrences de parasites peuvent résulter de facteurs de risque ou d'exposition communs (par exemples, les similarités d'environnement, de climat ou de susceptibilité physiologique). L'objectif de notre travail était de développer une méthode, basée sur les réseaux, permettant de prendre en compte ces deux caractéristiques pour améliorer la détection d'interactions au sein de communautés de parasites. Pour cela, nous avons travaillé sur des données de composition du microbiote de 267 tiques *Ixodes ricinus* femelles obtenues par séquençage de la région 16s du génome (JF Cosson et al). Les tiques ont été récoltées dans les Ardennes sur 9 sites différents. Les microbiotes de ces tiques sont diversifiés avec un total de 629 espèces de bactéries observées. Ces communautés de bactéries peuvent être représentées par un graphe biparti hôtes-parasites où chaque tique récoltée est reliée aux parasites dont elle est porteuse. Le but était de détecter les associations entre ces bactéries et de prendre en compte une structure dans la population hôtes, à savoir sur ces données, le lieu de récolte de la tique. Lors de l'inférence du réseau d'associations nous avons identifié certaines particularités dans les données : (i) la distribution de l'abondance des espèces suit une loi de Pareto (loi de puissance), (ii) il y a une majorité d'espèces rares et très peu d'espèces ubiquistes, (iii) plus de la moitié des espèces bactériennes ont une fréquence observée inférieure à 1 %. Cette distribution étalée, qui est souvent observée dans les réseaux écologiques, pose des problèmes méthodologiques car le nombre d'occurrences des bactéries est si faible que les interactions sont difficilement identifiables et peu robustes. De plus, il s'avère impossible de détecter une relation antagoniste sur plus de 90 % des paires de bactéries de notre étude car un déficit de co-occurrences de bactéries est indétectable quand les prévalences sont faibles. Afin d'améliorer la robustesse de la détection d'interactions, nous avons procédé à un seuillage sur les valeurs de prévalences. Une mise en parallèle avec les résultats obtenus sur un gradient du paramètre de pénalisation d'un modèle graphique gaussien a permis de

*Intervenant

caractériser l'effet du seuillage. L'effet du lieu de récolte de la tique sur l'abondance des bactéries a pu être pris en compte de deux façons : (i) les différents sites ont été vus comme des points du graphe (comme des bactéries) et les dépendances sites-bactéries ont été observées ; (ii) l'effet a été éliminé au préalable à l'aide d'une régression puis les résidus ont été utilisés. Pour conclure, les fréquences observées des bactéries ont une distribution sous la forme d'une loi de puissance ce qui modifie la distribution des corrélations entre bactéries. Nous avons pu mettre en avant des associations tout en tenant compte de l'effet lieu. La méthode présentée a permis de choisir un seuillage des prévalences approprié à nos données.