

---

# Inférence et clustering de graphes pour l'exploration du microbiote intestinal humain

Hanna Julienne\*<sup>†1</sup>, Magali Berland<sup>1</sup>, Nicolas Pons<sup>1</sup>, Emmanuelle Le Chatelier<sup>1</sup>, and Dusko Ehrlich<sup>1</sup>

<sup>1</sup>MetaGénoPolis, INRA, Université Paris-Saclay (MGP) – Institut National de la Recherche Agronomique - INRA – 78350 Jouy en Josas, France, France

## Résumé

La métagénomique du microbiote intestinal humain est un champ de recherche actif et prometteur. Le séquençage massif nous donne maintenant accès à notre second génome qui joue un rôle crucial dans notre physiologie. L'étude du microbiote est donc nécessaire pour une meilleure compréhension de la santé humaine.

En effet, grâce à la reconstruction d'espèces métagénomiques sans génome de référence par la détection de gènes co-abondants [1], l'association significative d'espèces bactériennes a été mise en évidence avec des pathologies telles que la cirrhose, l'obésité et le diabète de type 2 [2-4].

Le format des données issues de la métagénomique quantitative pose en lui-même des challenges inédits au biostatisticien. Le nombre de variables observés (~15 millions de gènes, ~1000 espèces métagénomiques) pose une difficulté de synthèse majeure. A cela s'ajoute le fait que la taille des cohortes étudiées est relativement faible par rapport aux nombres de variables métagénomiques. Dans cette configuration, l'approche par tests d'hypothèses, bien que concluante dans de nombreuses études [2-4], peut se heurter à un problème de puissance. Par ailleurs, la détection d'espèces significativement enrichies chez les individus sains ou malades, bien que cruciale, nous éclaire peu sur la structuration de l'écosystème sain et sa perturbation chez le malade.

Dans ce contexte, nous proposons d'utiliser l'inférence de graphe sur les espèces métagénomiques afin de donner une image globale de leur organisation. La méthode choisie [5] pour l'inférence du graphe a été spécifiquement développée pour corriger les biais de composition dus aux abondances relatives et exclure les corrélations partielles.

Une fois le graphe inféré, il est possible d'utiliser sa structure pour identifier des communautés d'espèces bactériennes co-occurentes. Le nombre relativement restreint de communautés permet de synthétiser l'état du microbiote sur quelques variables. Nous montrerons sur une cohorte de patients cirrhotiques [4] que ces communautés forment un regroupement de haut niveau qui peut-être plus pertinent du point de vue clinique que le regroupement par genre ou par phylum.

Le modèle statistique sous-jacent au graphe permet de simuler l'effet de l'augmentation

---

\*Intervenant

<sup>†</sup>Auteur correspondant: hanna.julienne@jouy.inra.fr

d'une espèce bactérienne sur le reste de l'écosystème. Nous nous proposons d'identifier des espèces leviers qui pourraient être des pistes thérapeutiques prometteuses pour restaurer un écosystème intestinal sain.

Références :

Nielsen, H. B., Almeida, M. et al. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature biotechnology*, 32(8), 822-828.

Le Chatelier, E. et al. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature*, 500(7464), 541-546.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., ... & Peng, Y. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, 490(7418), 55-60.

Qin, N. et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature*, 513(7516), 59-64

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol*, 11(5), e1004226.