

FICHE TYPE DE RECUEIL DES FAITS MARQUANTS 2016 DES DEPARTEMENTS/CENTRES

(Renseigner une fiche par fait marquant, classification des rubriques en annexe)

Titre du fait marquant :

Une méthode d'inférence de Combinaison de Traits Fonctionnel Agrégés pour caractériser le potentiel fonctionnel d'un écosystème microbien à partir de données de métagénomique. Application à la dégradation des fibres par le microbiote intestinal humain.

Catégorie : Etude

(Publication –en indiquant le doi) , Colloque, Partenariat, ESCO, Prospectives, Etudes, Brevets, Lancements/Inaugurations, Autres : précisez)

Contact : Béatrice Laroche

Unité : UR 1404 MaIAGE

Département : MIA

Centre INRA de Recherche : Jouy-en-Josas

Axe du document d'orientation 2010-2020 : Approches prédictives pour la biologie, Développement de systèmes alimentaires sains et durables

Axe du tripode : Alimentation

Domaine d'activité : Mathématique - modélisation - informatique

Méta-programme (si adapté) : MEM

Mots-clés (rubrique libre) : Digestion des fibres, microbiote intestinal, métagénomique, modélisation

Résumé (5 lignes) : Nous avons développé une approche computationnelle mêlant statistiques et modélisation pour caractériser une fonction écosystémique d'un écosystème microbien à partir de données de métagénomique shotgun. Nous avons appliqué cette approche à l'étude de la dégradation des fibres par le microbiote intestinal humain. Nous trouvons qu'elle est caractérisée par 4 combinaisons de traits fonctionnels agrégés, qui semblent différentiellement abondantes chez les malades de Crohn.

Contexte et enjeux :

Les équipes MaIAGE-DynEnVie (MIA) et MICALIS-Phylhom (MICA) ont collaboré dans le cadre d'un projet et d'un co-encadrement de thèse financés par le métagprogramme MEM pour mener à bien cette étude.

Les communautés de microorganismes sont ubiquitaires dans notre environnement et jouent un rôle très important dans de nombreux domaines allant des biotechnologies à l'agriculture, l'alimentation et la santé humaine et animale. La description et la compréhension des

fonctions de ces écosystèmes complexes, pour pouvoir à terme les modifier, constitue un enjeu de recherche majeur.

Les communautés microbiennes sont souvent caractérisées par une très forte diversité taxinomique et une forte variabilité temporelle. Dans ce contexte, l'étude de caractéristiques des microorganismes reliées à leur capacités fonctionnelles (ou traits fonctionnels) constitue une approche séduisante. En effet, dans de nombreux cas, sous l'effet de pressions de sélection environnementale, ces traits sont partagés par beaucoup d'espèces et moins variables dans le temps.

La métagénomique « Whole Genome Shotgun », qui repose sur le développement récent des techniques de séquençage haut-débit, permet d'obtenir des inventaires de gènes provenant de communautés microbiennes impossibles à cultiver. Cette approche permet de quantifier des Traits Fonctionnels Agrégés (TFA) à l'échelle d'une communauté par l'abondance de familles de gènes marqueurs de fonctions (Raes 2011).

Cependant, même si on se limite à une fonction métabolique particulière à l'échelle d'un écosystème microbien, il arrive souvent que le nombre de traits fonctionnels agrégés impliqués dans sa description soit important (de l'ordre d'une à quelques centaines) et difficile à interpréter. L'objectif de ce travail a donc été de développer une approche mathématique permettant de déterminer un nombre limité de Combinaisons de Traits Fonctionnels Agrégés (CTFA) pour caractériser une fonction écosystémique, ainsi que l'intensité de ces CTFA dans l'écosystème.

Résultats :

Le choix de descripteurs pertinents pour des fonctions écosystémiques construits à partir d'abondance de gènes marqueurs est une question délicate. Le regroupement dans des voies arbitrairement prédéfinies comme celles de KEGG aide à simplifier la description du système en combinant des traits, cependant cette démarche n'exploite pas toute l'information contenue dans les données. A l'opposé, des techniques d'analyse de données non supervisées comme la décomposition en valeurs singulières sont très souvent utilisées pour extraire des descriptions de dimension réduite de ces données complexes. Ces approches n'exploitent pas la richesse des connaissances disponibles concernant la génomique et la physiologie et le métabolisme des microorganismes, à l'échelle individuelle ou collective. Nous avons donc proposé une approche innovante guidée par des modèles, où nous mêlons la biologie des systèmes et l'analyse statistique des données.

Notre approche s'appuie sur la factorisation en matrices positives (« Non-negative Matrix Factorization », ou NMF). Il s'agit d'une technique d'apprentissage automatique classique dans le domaine de l'analyse des données et des images, destinée à résoudre des problèmes de séparation de sources : les abondances de gènes marqueurs sont considérées comme un mélange de signaux provenant de plusieurs sources. Les sources sont modélisées comme des sous communautés dans l'écosystème, chacune étant caractérisée fonctionnellement par une CTFA, que nous cherchons à reconstruire. Nous avons développé une méthode originale de construction systématique de contraintes sur la structure attendue des CTFA à partir de connaissance a priori sur le métabolisme et la structure de génomes prévalent dans l'écosystème considéré, lorsque ce type d'information est disponible.

Nous avons appliqué notre approche à l'étude de la dégradation des fibres par le microbiote intestinal humain. Nous avons utilisé les abondances de gènes obtenues dans 1408 échantillons provenant des projets HMP (Methé et al. 2012), MetaHIT (Li et al. 2014) et MicroObes (Le Chatelier et al. 2013). Nous avons montré qu'il suffit de 4 CTFA pour caractériser ce processus majeur dans le microbiote. Grâce à notre approche nous obtenons des CTFA biologiquement cohérents et interprétables, qui éclairent d'un jour nouveau la structuration de la fonction de dégradation des fibres. Une première analyse nous a permis

d'observer des différences de composition en terme de CTFA entre des échantillons provenant de malades de Crohn et leurs contrôles.

Perspectives :

Nous allons maintenant valider nos observations concernant les malades de Crohn sur des données supplémentaires, par le biais d'une collaboration avec l'université de Warwick
Notre méthode est générique et peut être utilisée pour d'autres écosystèmes, comme par exemple les microbiotes digestifs animaux, et d'autres processus métaboliques.

Valorisation :

Le travail a donné lieu au développement d'un ensemble de codes en Python et C++ qui seront mis à disposition de la communauté.

Il a donné lieu à la publication de l'article de journal

S. Raguideau, S.Plancade, N. Pons, M. Leclerc and B. Laroche. *Inferring Aggregated Functional Traits from Metagenomic Data Using Constrained Non-negative Matrix Factorization: Application to Fiber Degradation in the Human Gut Microbiota*. To appear in Plos Computational Biology.

Références bibliographiques :

Raes J, Letunic I, Yamada T, Jensen LJ, Bork P. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Molecular Systems Biology*. 2011;7(1).

Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, et al. A framework for human microbiome research. *Nature*. 2012;486(7402)

Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. *Nature biotechnology*. 2014;32(8).

Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500(7464).

CLASSIFICATION

Axes du document d'orientation

- Intégration des performances économiques, sociales et environnementales de l'agriculture
- Développement de systèmes alimentaires sains et durables
- Atténuation de l'effet de serre et adaptation de l'agriculture et de la forêt au changement climatique
- Valorisation de la biomasse pour la chimie et l'énergie
- Sécurité alimentaire mondiale et changements globaux
- Approches prédictives pour la biologie
- Agro-écologie

Tripode

- Alimentation
- Agriculture
- Environnement

Domaine d'activités

- Animaux
- Végétaux
- Micro-organismes
- Procédés agro-industriels
- Mathématique - modélisation - informatique

Méta-programmes

- SMACH
- M2E-MEM
- GISA
- SELGEN
- DID'IT
- ACCAF
- EcoServ
- Glofoods