Analyse statistique de l'expression des gènes exprimés par les micro-organismes au cours des procédés de fabrication et d'affinage de fromage

Contexte

Le fromage est une matrice biologiquement et métaboliquement dynamique dans lequel l'activité et le niveau de population des micro-organismes sont fortement influencés par les conditions environnementales et les pratiques industrielles.

Le stage s'inscrit dans le cadre du projet TANGO, porté par Hélène Falentin (UMR Science & Technologie du lait & et de l'œuf, INRA Rennes), dont l'objectif principal est d'appréhender, lors d'une fabrication de pâte pressée non cuite modèle, l'impact de différents itinéraires technologiques sur le métabolisme d'une communauté bactérienne simplifiée (3 souches). Les enjeux à terme sont l'amélioration des propriétés organoleptiques des fromages grâce à une meilleure maîtrise des choix d'itinéraires technologiques.

La métatranscriptomique permet l'étude de l'expression des gènes de l'ensemble des micro-organismes composant un écosystème (ici fromager) et donc l'étude du profil fonctionnel d'une communauté de micro-organismes tout au long de la fabrication et de l'affinage (données en cinétique).

Les données de métatranscriptomique sont représentatives de l'expression de chaque gène dans les conditions biologiques d'intérêt. Elles sont issues d'un séquençage de l'ARN et correspondent pour chaque échantillon biologique à un nombre de séquences alignées sur une région génomique d'intérêt. Ces données sont par nature discrète, positive et peuvent prendre des valeurs dans un intervalle très grand. On s'intéresse alors à détecter les régions génomiques différentiellement exprimées entre des conditions biologiques d'intérêt (deux itinéraires technologiques par exemple). Le profil cinétique d'expression des gènes impliqués dans la synthèse d'une dizaine de composés d'arôme sera ensuite corrélé à la production temporelle de composés d'arôme.

Objectif du stage

Le stagiaire devra dans un premier temps se familiariser avec les données de métatranscriptomique issues du séquençage de l'ARN (RNA-Seq) et s'appropriier les méthodes d'analyse statistique (basée sur le modèle linéaire généralisé) ainsi que les outils pour faire ces analyses.

Il veillera notamment à sélectionner les méthodes les plus appropriées à l'analyse de données métatranscriptomiques et adaptera le cas échéant les méthodes existantes à la particularité des données du projet.

Dans un second temps, l'étudiant devra analyser les données et pourra intégrer les données de transcriptomique aux données physico-chimiques ou métaboliques disponibles.

L'implémentation des méthodes et l'analyse des données se fera avec le langage R dans l'environnement RStudio.

Une présentation des résultats et des échanges réguliers avec l'équipe du porteur du projet (UMR STLO, Rennes) est prévue. Deux déplacements sur Rennes sont à prévoir pendant la durée du stage.

Compétences recherchées

Master 1 ou Master 2 en statistiques ou équivalent en école d'ingénieur

Bonnes connaissances théoriques en statistiques, notamment du modèle linéaire et ses extensions.

Maîtrise du langage R.

Aucune connaissance dans le domaine d'application n'est demandé mais un intérêt pour les applications statistiques en biologie est nécessaire.

Anglais (lecture articles scientifiques)

Unité d'accueil

Le stage se déroulera au sein de l'UMR Mathématiques Informatique Appliquées de Paris (Paris 5^{ème}) <https://www6.inra.fr/mia-paris>.

L'UMR **Mia-Paris**, AgroParisTech, Inra, regroupe des statisticiens et des informaticiens spécialisés dans la modélisation et l'apprentissage statistique et informatique pour la biologie, l'écologie, l'environnement, l'agronomie et l'agro-alimentaire. Leurs compétences portent sur les méthodes d'inférences statistiques (modèles complexes, modèles à variables latentes, inférence bayésienne, apprentissage, sélection de modèle...), et algorithmiques (généralisation, transfert de domaine, représentation des connaissances).

L'unité développe des méthodes statistiques et informatiques originales génériques ou motivées par des problèmes précis en science du vivant. Ses activités s'appuient sur une bonne culture dans les disciplines destinatrices : écologie, environnement, agro-alimentaire, biologie moléculaire et biologie des systèmes.

Trois équipes de recherche structurent l'unité Mathématiques et Informatique Appliquées : Statistique et Génome, Modélisation et Risque en Statistique Environnementale et Learning and Integration of Knowledge. Le stagiaire sera intégré dans l'équipe Statistique et Génome et travaillera sur les données produites dans le cadre du Projet TANGO. La durée du stage (entre 5 et 6 mois) et la date de commencement (mars/avril) peuvent être adaptées en fonction des contraintes du candidat.

Contact

Pour postuler envoyer une lettre de motivation, ainsi qu'un CV et Julie Aubert, UMR 518 AgroParisTech / INRA MIA, julie.aubert@agroparistech.fr