

Modélisation du vieillissement de la drosophile en vue de l'identification des SNPs influents

Stage niveau M2 début 2021

Unité Maïage INRAE Maïage Jouy-en-Josas/CRI Université de Paris

Pour postuler, merci d'envoyer un CV, une lettre de motivation et vos derniers relevés de notes à estelle.kuhn@inrae.fr.

Contexte

Les études GWAS (Genome Wide Association Studies) sont un outil puissant pour explorer les liens entre variants génétiques et phénotypes. Le modèle drosophile possède une importante ressource GWAS appelée Drosophila Genetic Resource Panel (DGRP)[1]. Ces dernières années, nous avons utilisé les lignées du DGRP pour aborder la question du rôle des polymorphismes nucléotidiques uniques (SNPs) dans un nouveau cadre théorique du vieillissement, le modèle de vieillissement en deux phases [2], basé sur le phénotype Smurf [3]. Grâce à cette nouvelle approche de l'étude du vieillissement, nous avons pu détecter une forte association entre SNPs et la longévité des individus.

Le présent projet vise à développer un modèle prédictif de la longévité des individus de lignées DGRP non-testées en se basant uniquement sur leur génotype.

Les données disponibles sont des données de longévité (survie) d'environ 27000 individus femelles obtenues à partir de 117 des 205 lignées DGRP ainsi que celles d'environ 6000 supplémentaires réparties sur 40 lignées DGRP afin de valider le modèle. Pour chacune de ces lignées, nous disposons d'environ 2.5 millions de SNPs avec l'information de l'allèle parental et celui se trouvant dans ladite lignée. La durée de vie de chaque individu est connue à +/-48h.

Objectifs du stage

- * proposer un modèle statistique et un estimateur par critère pénalisé pour identifier les SNPs influents,
- * implémenter un algorithme stochastique d'estimation et le valider sur données simulées,
- * proposer un prédicteur pour l'espérance de vie à partir des SNPs influents et le valider sur données simulées,
- * ajuster ce modèle aux données réelles,
- * valider les gènes identifiés par cette approche pour leur rôle dans la détermination de la longévité à l'aide du modèle de la drosophile et d'une dérégulation conditionnelle de leur expression grâce au système de commutation génétique et aux outils génétiques disponibles.

Aspects mathématiques L'approche envisagée en priorité reposera sur les modèles de survie à effets aléatoires (modèles de fragilité) incluant un effet "lignée" et des SNPs comme variables explicatives de grande dimension. L'estimation des paramètres se fera à partir de critères pénalisés type LASSO [4] pour sélectionner les SNPs influents. Des algorithmes stochastiques seront mis en oeuvre pour calculer l'estimateur maximisant le critère choisi [5].

Profil recherché

Formation niveau BAC+5 (Master 2 ou école d'ingénieurs), connaissance en statistiques théoriques et appliquées, ayant un fort intérêt pour les applications en sciences du vivant ; maîtrise d'un langage de programmation indispensable (Python et/ou R de préférence) ; rigueur scientifique, curiosité intellectuelle, facilité de communication.

Modalités pratiques

Le stage se déroulera d'une part sur le centre INRAE de Jouy-en-Josas et d'autre part sur le site du Centre de Recherche Interdisciplinaire Paris de l'Université de Paris. La durée du stage sera de cinq ou six mois, entre février et septembre 2021. La gratification mensuelle est d'environ 550 euro (taux légal). L'encadrement sera réalisé par Estelle Kuhn (INRAE, MaIAGE) et Michael Rera (Université de Paris, CRI). Le stage pourra éventuellement déboucher sur un sujet de thèse impliquant modélisation mathématique, statistique et validations expérimentales.

Références bibliographiques

- [1] Mackay, T. F. et al. The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482, 173-8 (2012).
- [2] Tricoire, H., Rera, M. A New, Discontinuous 2 Phases of Aging Model: Lessons from *Drosophila melanogaster*. *PLoS One* 10, e0141920 (2015).
- [3] Rera, M. et al. Modulation of longevity and tissue homeostasis by the *Drosophila* PGC-1 homolog. *Cell Metab* 14, 623-34 (2011).
- [4] Lemler, S., Oracle inequalities for the Lasso in the high-dimensional Aalen multiplicative intensity model. *Annales de l'IHP*, Vol. 52-2, pp 981–1008, (2016).
- [5] Kuhn E., El-Nouty C., On one convergent stochastic estimation algorithm for frailty models. *Statistics and Computing*, Vol. 23, Issue 3, pp 413-423, (2013).