

Proposition de stage de Master 2:

Développement des outils d'intelligence artificielle pour la prédiction de la germination des semences

Nourrir une population croissante dans les prochaines décennies nécessite de prendre en compte dès maintenant les conséquences du réchauffement climatique qui constitue une menace sérieuse pour la sécurité alimentaire mondiale. La productivité des plantes est à la base de l'alimentation mondiale. Elle dépend directement de la performance des semences pendant le semis et des températures plus élevées et une disponibilité réduite en l'eau vont avoir un impact négatif sur la phénologie de la germination et donc sur le rendement des cultures. Il existe donc un besoin urgent de mieux comprendre le lien entre l'environnement et les performances des semences dans un environnement en mutation.

Dans ce cadre de nombreux progrès liés au développement des approches -omiques ont été effectués et ils ont mieux permis de mieux comprendre les mécanismes moléculaires contrôlant la germination. De nombreux algorithmes ont été développés pour analyser les données d'expression chez les plantes mais il n'existe pas d'algorithme d'apprentissage automatique facilement disponible pour prédire la germination et détecter les gènes liés à sa réalisation dans différents contextes. L'utilisation de l'intelligence artificielle et du machine learning sont nouvelles en biologie végétale. L'IA a surtout été utilisée pour modéliser des analyses de phénotypage ou pour prédire la réponse des plantes à des stress. Dans le cadre de la germination l'IA a été utilisée pour construire des réseaux de gènes ou pour prédire la germination par analyse d'images. Cependant ces travaux sont très intéressants car ils suggèrent que les outils d'IA peuvent prédire et optimiser les processus de germination.

L'objectif de ce projet, de nature exploratoire mais à très fort potentiel d'application, est de s'appuyer sur les jeux de données de transcriptome disponibles dans le contexte de la germination pour créer le premier modèle prédictif de germination et pour permettre la détection de biomarqueurs de ce processus biologique.

Dans le cadre de ce stage, et sur la base des données disponibles à l'IBPS (plusieurs jeux de données de transcriptomique obtenus chez les semences d'*Arabidopsis thaliana*) ou dans le domaine public, plusieurs approches seront évaluées : modèles statiques, apprentissage non supervisé (clustering, réduction de dimension), apprentissage supervisé (k-NN, régression logistique, modèles linéaires, machines à vecteurs de support, forêts aléatoires, boosting etc.). Ce travail interdisciplinaire sera co-supervisé par un biologiste de l'IBPS et un chercheur IA pour le vivant du LCQB.

Contact :

Christophe Bailly - christophe.bailly@sorbonne-universite.fr

UMR7622 - Biologie des semences

Sorbonne Université

Bat C, 2ème étage, boîte 24

4, place Jussieu - 75252 Paris cedex 05

tél : + 33 1 44 27 59 29

<http://www.ibps.upmc.fr/en/research/developmental-biology-laboratory/seed-biology>

Références

Ayuso M, Landín M, Gallego PP, Barreal ME (2019) Artificial Intelligence Tools to Better Understand Seed Dormancy and Germination. In "Seed Dormancy and Germination" Edited by Jose Carlos Jimenez-Lopez. doi: 10.5772/intechopen.90374

- Ayuso M, Ramil-Rego P, Landin M, Gallego PP, Esther Barreal M** (2017) Computer-Assisted Recovery of Threatened Plants: Keys for Breaking Seed Dormancy of *Eryngium viviparum*. *Front Plant Sci* **8**: 2092 <https://doi.org/10.3389/fpls.2017.02092>
- Bassel GW, Glaab E, Marquez J, Holdsworth MJ, Bacardit J** (2011) Functional Network Construction in Arabidopsis Using Rule-Based Machine Learning on Large-Scale Data Sets. *The Plant Cell* **23**: 3101–3116 <https://doi.org/10.1105/tpc.111.088153>
- Genze N, Bharti R, Grieb M, Schultheiss SJ, Grimm DG** (2020) Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant Methods* **16**: 1–11 <https://doi.org/10.1186/s13007-020-00699-x>
- Gupta C, Ramegowda V, Basu S, Pereira A** (2021) Using Network-Based Machine Learning to Predict Transcription Factors Involved in Drought Resistance. *Frontiers in Genetics* **12**: 943 <https://doi.org/10.3389/fgene.2021.652189>
- Kang D, Ahn H, Lee S, Lee C-J, Hur J, Jung W, Kim S** (2019) StressGenePred: a twin prediction model architecture for classifying the stress types of samples and discovering stress-related genes in arabidopsis. *BMC Genomics* **20**: 1–13 <https://doi.org/10.1186/s12864-019-6283-z>
- Meng X, Liang Z, Dai X, Zhang Y, Mahboub S, Ngu DW, Roston RL, Schnable JC** (2021) Predicting transcriptional responses to cold stress across plant species. *PNAS*. doi: [10.1073/pnas.2026330118](https://doi.org/10.1073/pnas.2026330118)
- Nehoshtan Y, Carmon E, Yaniv O, Ayal S, Rotem O** (2021) Robust seed germination prediction using deep learning and RGB image data. *Sci Rep* **11**: 22030 <https://doi.org/10.1038/s41598-021-01712-6>
- Singh AK, Ganapathysubramanian B, Sarkar S, Singh A** (2018) Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in Plant Science* **23**: 883–898 <https://doi.org/10.1016/j.tplants.2018.07.004>