

Sujet proposé pour un stage niveau master 1 de 4 mois en bioinformatique

Caractérisation des différences de profils thermiques (en termes d'intensité, de durée et fréquence des événements thermiques) entre zones classiques et zones marginales de culture du sorgho.



Encadrante principale : Lauriane Rouan UMR AGAP Institut, Equipe PhenoMEn, Montpellier

Contact : lauriane.rouan@cirad.fr (Biostatisticienne)

Co-encadrants : Grégory Beurier (Chercheur en intelligence artificielle), Catherine Giauffret (Biologiste-Agronome).

Contexte général dans lequel le stage va se dérouler :

Les vagues de froid modéré (« chilling ») sont une source importante de stress pour les cultures d'été. Paradoxalement, dans un contexte de réchauffement climatique global, des espèces comme le maïs ou le sorgho tendent à être exposées à des épisodes de froids plus intenses en raison de semis très précoces ou de remontée des cultures vers le Nord pour échapper aux contraintes hydriques. Ce stage s'inscrit dans le cadre du projet SMACC « Exploring genetic resources of Sorghum and Maize for Adaptation to Complex Chilling scenarios » soutenu par le département de Biologie et d'Amélioration des Plantes d'INRAE, visant à i) définir des scénarios thermiques cibles pour ces nouvelles modalités de culture en France et ii) explorer les potentialités de ressources génétiques mondiales pour la sélection de ces deux espèces dans ces nouveaux scénarios.

Objectifs du stage :

A partir de l'expertise d'agronomes concernant les zones et périodes de culture classiques des deux espèces, les zones marginales (vers le Nord) et les périodes correspondant à des semis très précoces seront définies en amont du stage. Le stage demandera dans un premier temps de recueillir auprès de sites appropriés (base de données météo France, stations météorologiques d'unités expérimentales publiques...) les données climatiques associées à des lieux représentatifs des zones classiques et des zones marginales, ou, si elles ne sont pas disponibles, de télécharger les données fournies par la NASA aux points de coordonnées GPS correspondants.

De la même façon, les données climatiques correspondant aux lieux de collecte des différentes accessions maïs et sorgho du Centre de ressources biologiques (CRB) GAMÉT seront collectées.

Il s'agira ensuite, à partir de cette source de données, de :

- Mettre en évidence les événements thermiques extrêmes spécifiques des semis très précoces ou des zones marginales (scénarios cibles).
- Identifier des clusters d'environnements/périodes se basant sur les occurrences d'évènements thermiques extrêmes.
- Observer les similarités entre scénarios cibles et scénarios des sites de collecte des ressources génétiques.

Méthodologie :

Après une phase de collecte et de mise en forme des données climatiques, le stagiaire devra mettre en œuvre des techniques d'apprentissage machine (Machine learning) pour 1- l'imputation et l'interpolation des données météorologiques manquantes ou incomplètes (CNN, transformers, ...), 2- la classification et le clustering des séries climatiques en fonction des stress. Les scénarios de température des différentes zones pourront alors être comparés entre eux en utilisant, par exemple, des méthodes de quantification de similarité et de comparaison de distributions.

Outils utilisés : Logiciel R et Python