

# Atelier de prédiction et sélection génomique

J. David et équipe pédagogique : Vincent Segura, Friedrich Longin, Margaux Jullien, Charlotte Brault,

18/02/2021

## Abstract

Ce document présente les objectifs, les pré-requis et le programme de l'atelier de prédiction et sélection génomique. Cet atelier est destiné aux étudiants en spécialisation APIMET et SEPMET de Montpellier SupAgro et PIST d'AgroParisTech. Un peu d'histoire : En 2021, ce sera la 6ème édition de cet atelier construit au départ par Timothée Flûtre, Jacques David, Friedrich Longin et Philippe Brabant, fortement amélioré par Julien Diot pour le jeu sérieux. Cette année, l'équipe comporte Vincent Segura, Charlotte Brault, Margaux Jullien, Michel Colombo et Jacques David.

## Contents

<b>1</b>	<b>Date, lieu et contacts</b>	<b>1</b>
<b>2</b>	<b>Objectifs de l'atelier</b>	<b>2</b>
<b>3</b>	<b>Pré-requis</b>	<b>2</b>
<b>4</b>	<b>Equipement et configuration</b>	<b>3</b>
<b>5</b>	<b>Ressources en ligne</b>	<b>3</b>
<b>6</b>	<b>Programme de la semaine</b>	<b>3</b>

## 1 Date, lieu et contacts

- Du lundi 15 au vendredi 19 février 2021.

En raison des conditions sanitaires, l'atelier se déroulera en distanciel à l'adresse suivante. [Salle ZOOM](#)

- Habituellement ca se passe là : Montpellier SupAgro, campus La Gaillarde, bâtiment 14 de l'annexe de Mandon ([plan](#)).
- Contact [Jacques David](#) (Montpellier SupAgro, UMR AGAP, équipe GE2POP),  
Equipe pédagogique 2021 : Vincent Segura (Montpellier SupAgro, UMR AGAP, équipe DAAV Julie Fiévet (INRA, UMR GQE) Friedrich Longin de l'université d'Hohenheim Charlotte Brault de l'IFV (en thèse à l'INRA, UMR AGAP, équipe DAAV) Margaux Jullien, (INRA, UMR GQE) Michel Colombo (<mailto:michel.colombo@inrae.fr>) (Montpellier SupAgro, UMR AGAP, équipe GE2POP)

Crédits Timothée Flutre (INRA, UMR GQE) Philippe Brabant, Jacques David Julien Diot

## 2 Objectifs de l'atelier

Il vise à familiariser les étudiants avec la prédiction génomique de caractères quantitatifs et son utilisation dans les programmes de sélection. Il se déroule en trois étapes.

1. Les bases de la prédiction génomique, par Vincent Ségura (INRA) et Charlotte Brault (IFV, INRA)
  - Simulation de données phénotypiques selon différentes architectures génétiques
  - Présentation des principales méthodes d'inférence des effets génétiques
  - Exploration de la qualité des prédictions en fonction de paramètres tels que la structure du matériel végétal, la densité de génotypage, etc.
2. "Jouez au sélectionneur !"
  - Constitution d'équipes de sélection
  - Elaboration d'une stratégie de sélection mêlant essais au champ, génotypage et prédiction
  - Réalisation du schéma de sélection sur une succession de plusieurs années simulant la temporalité et le budget d'un véritable programme de sélection
  - Mise en commun, examen des résultats et comparaison des stratégies
3. "Optimization of breeding schemes with genomic selection in R", par Friedrich Longin (Univ. Hohenheim)
  - Présentation (en anglais) de la théorie
  - Mise en pratique via des simulations de schémas de sélection

## 3 Pré-requis

Les étudiants doivent avoir révisé assez sérieusement les notions suivantes.

1. Les bases de la génétique quantitative
  - 1.1. Décomposition de la variance et notion d'héritabilité
  - 1.2. Effets additifs d'un allèle, modèle bi-allélique de Fisher
  - 1.3. Prédiction de l'effet additif par la régression du nombre d'allèles favorables
  - 1.4. Covariance entre apparentés et construction d'une matrice de simple apparentement
2. Les outils statistiques suivants
  - 2.1 Modèle linéaire mixte (à effets fixes et aléatoires)
  - 2.2 Estimation des composantes de la variance par le modèle mixte
  - 2.3 Utilisation des matrices d'apparentement dans le modèle mixte
3. Avoir revu les TPs
  - 3.1 Estimation des composantes de la variance
  - 3.2 Génétique d'association
  - 3.3 Prédiction génomique
4. Logiciel R
  - 4.1 Bases de la programmation en R
  - 4.2 Boucles et stockage de valeurs dans un objet
  - 4.3 Ecriture d'une fonction
  - 4.4 Ecriture de rapports au format Rmd

## 4 Équipement et configuration

Les étudiants doivent venir chacun avec leur ordinateur portable avec la configuration suivante:

1. système d'exploitation Windows, Mac ou GNU/Linux;
2. logiciels [R](#) (version  $\geq 3$ ) et [RStudio](#) (version  $\geq 1$ ), et au moins les paquets suivants:
  - [selection](#): pour la prédiction du gain génétique;
  - [knitr](#), [rmarkdown](#): pour les rapports dynamiques;
  - [MASS](#): pour les simulations multivariées;
  - [lme4](#): pour l'analyse de données phénotypiques via des modèles mixtes
  - [rrBLUP](#), [BGLR](#): pour la prédiction génomique.
  - [cvTools](#): pour la validation croisée
  - [gtools](#): pour trier de manière alphanumérique
  - [MM4LMM](#): algorithme MM pour ajuster une régression par ML ou ReML

## 5 Ressources en ligne

Ce document, ainsi que les autres nécessaires à la réalisation de l'atelier, se trouve à l'adresse suivante: <https://github.com/jacquesdauid/SELGEN2021>. Le copyright appartient à Montpellier SupAgro et à l'Institut National de la Recherche Agronomique. Le contenu du répertoire est sous license [Creative Commons Attribution-ShareAlike 4.0 International](#). Veuillez en prendre connaissance et vous y conformer (contactez les auteurs en cas de doute).

## 6 Programme de la semaine

### 6.1 Lundi

- 9h - 12h: introduction à la modélisation statistique via la simulation en prenant la régression linéaire simple comme exemple, par V. Segura (document “premiers-pas.pdf”)
- 14h - 17h: extension au modèle linéaire mixte, puis exploration par simulation de la prédiction génomique, par V. Segura (document “prediction-génomique.pdf”)

### 6.2 Mardi

- 9h - 12h: investigation par simulation, en groupes, des aspects influençant la précision de prédiction
- 14h - 15h30: présentation des résultats et discussion
- 15h30 - 17h: présentation du jeu sérieux sur la sélection, constitution des équipes et réflexions stratégiques

### 6.3 Mercredi

- 9h - 12h: lancement du jeu (1 heure = 1 année)
- 14h - 17h: suite du jeu (1 heure = 1 année)

## 6.4 Jeudi

- 9h - 13h00: suite du jeu (40 minutes = 1 année)
- 13h: Envoi d'un mail avec les géotypes candidates à l'inscription
- 14h30 - 16h30: discussion sur les différentes stratégies suivies

## 6.5 Vendredi

- 9h - 12h: cours "Optimization of breeding schemes with GS in R", par F. Longin (document "Selgain\_Montpellier\_2018.pdf")
- 14h - 16h30: travaux pratiques via des simulations de schémas de sélection (document "Selgain\_Montpellier\_2018\_example.R").
- 16h30 - 17h: discussion sur les acquis réalisés durant l'atelier et les améliorations à apporter