

Atelier de prédiction et sélection génomique

J. David & T. Flutre

17/02/2016

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 de Montpellier SupAgro. Il accueille en 2016 des étudiants de la spécialisation PIST d'AgroParisTech.

Contents

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

1 Date, lieu et contacts

- Du lundi 15 au vendredi 19 février 2016.
- Montpellier SupAgro, [bâtiment 18](#) (annexe de Mandon).
- [Jacques David](#) (Montpellier SupAgro, [UMR AGAP](#), équipe [GE2POP](#)) et [Timothée Flutre](#) (INRA, [UMR AGAP](#), équipe [DAAV](#)), avec l'aide de Julie Fiévet et Philippe Brabant d'[AgroParisTech](#).

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

- 1.1. Simulation de données phénotypiques selon différentes architectures génétiques
- 1.2. Présentation des principales méthodes d'inférence des effets génétiques
- 1.3. Exploration de la qualité des prédictions en fonction de paramètres tels que la structure du matériel végétal, la densité de génotypage, etc.

2. Le *journal club*

- 2.1. Distribution d'articles en début d'atelier
 - 2.2. Lecture personnelle & préparation de présentations
 - 2.3. Présentation & discussion collective
3. La simulation d'un cycle de sélection
- 3.1. Constitution d'équipes de sélection
 - 3.2. Elaboration d'une stratégie de sélection mêlant essais et génotypage
 - 3.3. Réalisation du schéma de sélection sur une succession de 8 périodes simulant la temporalité d'un véritable programme de sélection
 - 3.4. Mise en commun, examen des résultats et comparaison des stratégies

3 Pré-requis

Les étudiants doivent avoir révisé assez sérieusement les notions suivantes.

1. Les bases de la génétique quantitative
 - 1.1. Notion d'héritabilité
 - 1.2. Décomposition de la variance
 - 1.3. Effets additifs d'un allèle, modèle bi-allélique de Fisher
 - 1.4. Prédiction de l'effet additif par la régression du nombre d'allèles favorables
 - 1.5. 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 Ecriture d'une fonction
 - 4.3 Boucles & stockage de valeurs dans un objet
 - 4.4 Ecriture de rapports au format Rmd

Afin de vérifier que ces pré-requis ont été révisés, un petit QCM sera distribué en début d'atelier. Il servira pour une partie de la note de l'atelier.

4 Équipement et configuration

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

1. système d'exploitation Windows, Mac ou Linux;
2. logiciel R (version ≥ 3) et au moins les paquets suivants:
 - [knitr](#), [rmarkdown](#): pour les rapports dynamiques;
 - [MASS](#): pour les simulations multivariées;
 - [QTLRel](#), [qqman](#): pour la génétique d'association;
 - [rrBLUP](#), [BGLR](#): pour la prédiction génomique.

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/timflutre/atelier-prediction-genomique>. 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

6.1.1 9h - 12h

Introduction à la modélisation statistique via la simulation en prenant la régression linéaire simple comme exemple (document “premiers-pas.pdf”).

6.1.2 14h - 17h

Extension au modèle linéaire mixte, puis exploration par simulation de la prédiction génomique (document “prediction-genomique.pdf”).

6.2 Mardi

6.2.1 9h - 12h

Investigation par groupe des aspects influençant la précision de la prédiction.

6.2.2 14h - 17h

Suite du matin, puis présentation des résultats.

6.3 Mercredi

6.3.1 9h - 12h

Lecture libre, individuelle ou collective, des articles suivants. Il est encouragé que tous les étudiants aient lu au moins le résumé des articles qu'ils n'ont pas à étudier.

- Le commentaire d'auteurs clarifiant leur utilisation pionnière de marqueurs génétiques pour estimer l'héritabilité de la taille dans l'espèce humaine

Visser, P. M., Yang, J., Goddard, M. E., Dec. 2010. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010) 13 (6), 517-524. [URL](#)

- Un article sur la manière d'optimiser la répartition des ressources pour un progrès génétique maximal par unité de temps

Riedelsheimer, C., Melchinger, A. E., Nov. 2013. Optimizing the allocation of resources for genomic selection in one breeding cycle. Theoretical and Applied Genetics 126 (11), 2835-2848. [URL](#)

- Une comparaison de différents modèles statistiques utilisés en prédiction génomique

Heslot, N., Yang, H.-P., Sorrells, M. E., Jannink, J.-L., 2012. Genomic selection in plant breeding: a comparison of models. Crop Science 52 (1), 146+. [URL](#)

- Une revue détaillant les bonnes pratiques à suivre pour les comparaisons de méthodes

Daetwyler, H. D., Calus, M. P., Pong-Wong, R., de Los Campos, G., Hickey, J. M., Feb. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193 (2), 347-365. [URL](#)

- Une revue détaillant les stratégies de différents schémas de sélection

Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., Prasanna, B., Feb. 2012. Whole-genome strategies for marker-assisted plant breeding. Molecular Breeding 29 (4), 833-854. [URL](#)

6.3.2 14h - 17h

Discussion des articles après présentation de chacun des groupes.

6.4 Jeudi

6.4.1 9h - 12h

- présentation de l'exercice de sélection (document "programme-selection.pdf")
- constitution des groupes
- réflexions stratégiques intra-groupe

6.4.2 14h - 17h

- lancement de l'exercice
- toutes les heures les équipes viennent soumettre leur propositions aux animateurs qui envoient en retour les nouveaux croisements, génotypes et/ou phénotypes demandés

6.5 Vendredi

6.5.1 9h - 12h

Le processus de sélection continue avec un pas horaire.

6.5.2 13h30 - 16h

- les animateurs délivrent les résultats: meilleure variété & meilleur matériel
- discussions croisées sur les différentes stratégies suivies

6.5.3 16h - 17h

Discussion sur les acquis réalisés durant l'atelier et sur les améliorations à apporter.