

Estimation de la capacité de prédiction génomique de la valeur d'un croisement (PrediCroit)

Contrat : 18 mois (à ajuster suivant le niveau de salaire), peut commencer entre octobre 2021 et avril 2022, se terminer entre mars 2023 et septembre 2023

Equipe d'accueil : GDEC, DIGEN (Sophie Bouchet)
Soutien méthodologique : GENEPHYSE, MG2 (Jean-Michel Elsen)

Contact : sophie.bouchet@inrae.fr

Compétences : PhD, génétique quantitative, programmation (R et bases bash), rédaction d'articles scientifiques en anglais

MOTS CLEFS

valeur de croisement / critère d'utilité / prédictions génomiques / blé tendre / rendement / protéine / date d'épiaison

OBJECTIFS:

Le choix du plan de croisement pour un sélectionneur est crucial. C'est de cette décision que dépendent la moyenne et la variance génétique observée dans la descendance, le gain génétique, la probabilité d'obtenir des individus supérieurs aux variétés inscrites, et donc la probabilité d'inscrire une variété ayant un succès commercial.

L'objectif de ce travail est d'estimer notre capacité de prédiction génomique de la valeur d'un croisement dans du matériel élite et dans du matériel plus diversifié de blé tendre, grâce à des simulations et à des données expérimentales issues de programmes de sélection INRAE-AO et Florimond-Desprez.

Depuis environ 10 000 ans, les hommes ont sélectionné plus ou moins consciemment les meilleurs épis pour contribuer à la génération suivante. Le caractère d'intérêt principal a longtemps été le rendement. Pour assurer un haut niveau dans la descendance et un gain génétique rapide, les sélectionneurs professionnels croisaient les meilleurs individus qu'ils avaient à disposition sur les caractères à valeur commerciale. Le principal risque est que ces meilleurs individus possèdent des combinaisons alléliques similaires, que le matériel élite soit de plus en plus apparenté de génération en génération et qu'il perde à moyen terme de nombreux allèles favorables, limitant le potentiel génétique sur le long terme.

Le choix des croisements a été basé longtemps sur des observations phénotypiques. Grâce aux prédictions génomiques, nous espérons pouvoir prédire la valeur des croisements, c'est-à-dire la distribution attendue dans la descendance, et surtout la valeur du meilleur individu potentiel. Il s'agit donc d'estimer la moyenne et la variance de la descendance. La moyenne de la descendance correspond à la moyenne des valeurs génétiques des parents. Par contre la variance est plus difficile à estimer. Elle dépend de la phase des allèles aux QTLs chez les parents, de la complémentarité allélique chez les parents, et du taux de recombinaison entre QTLs. Grâce au vecteur de recombinaison estimé sur l'ensemble du génome, la valeur d'un croisement peut

s'estimer soit par simulation de descendance soit par des formules analytiques. Alors que nous ne pouvons pas tester au champ tous les croisements possibles, nous pouvons estimer in-silico et rapidement les valeurs de nombreux croisements.

Les différents estimateurs de la valeur d'un croisement que nous proposons d'utiliser sont :

- La moyenne des Genomic Estimated Breeding Value (GEBV) des parents
- La valeur du meilleur descendant théorique (Optimal Haploid Value) (suppose une taille infinie de la population pour que le meilleur descendant obtenu ressemble au meilleur descendant théorique possible)
- La valeur des meilleurs descendants pour une taille de descendance fixée (simulation de descendants)
- Le critère d'utilité d'un croisement : la valeur génétique attendue de la fraction des descendants qui est sélectionnée (les 10% meilleurs par exemple) (formule analytique)
- La probabilité de produire un descendant meilleur que le meilleur parent (formule analytique)

Nous proposons de tester la capacité prédictive de ces différents estimateurs de la valeur d'un croisement grâce à

- (1) Des données expérimentales : comparer la valeur des estimateurs à la distribution des performances des descendants de 95 croisements des programmes de sélection INRAE-Agri-Obtentions et Florimond-Desprez observés au champ. Nous évaluerons la distribution de la date d'épiaison, la hauteur, le rendement et la teneur en protéine du grain. Le problème est que nous allons comparer ici deux estimateurs de la variance (phénotypiques et génétiques) sans valeur référence pour chaque croisement.
- (2) Des données simulées : des descendance seront simulées dans un panel historique de lignées inscrites (données CTPS), dans un panel de lignées européennes modernes non élites, dont les lignées ont été génotypées et phénotypées. Nous simulerons des caractères avec différentes architectures et héritabilités. Dans ce cas, nous pourrions comparer les estimateurs phénotypiques et génétiques de la valeur des croisements ainsi que les vraies valeurs génétiques des meilleurs descendants.

Nous regarderons également l'évolution de la diversité génétique dans la population élite historique et le matériel de sélection INRAE-AO, l'évolution des contributions parentales en particulier dans cette dernière population grâce à la connaissance des pedigrees.

La précision de la prédiction génomique des valeurs de croisements dépend de la précision de l'estimation des effets des marqueurs, donc de la qualité des génotypes et phénotypes de la population d'entraînement, du modèle de prédiction (qui peut dépendre de l'architecture génétique du caractère), de la composition allélique des populations d'entraînement et de la population à prédire, de la congruence entre DL et densité de marqueurs, et de la qualité de l'estimation du taux de recombinaison le long du génome. Nous testerons les effets de plusieurs stratégies de construction de la population d'entraînement, de plusieurs vecteurs de recombinaison (estimé sur population bi-parentale ou panel de diversité) et de plusieurs modèles de prédiction.

L'objectif sera ensuite de valider les capacités d'amélioration du gain génétique pour des situations plus complexes (sélection multi-caractères, croisements multi-parentaux, optimisation d'un plan de croisements à l'échelle d'un programme de sélection) en utilisant un algorithme génétique.

Perspectives de valorisation

Nous publierons les résultats dans une revue à comité de lecture qui comparera l'efficacité de différents critères d'utilité pour prédire la valeur d'un croisement, sur données simulées et réelles, panels élites et diversifiés.

Les algorithmes de prédiction de la valeur des croisements développés seront intégrés au pipeline de simulation de schémas de sélection développé dans le cadre des projets Selgen Predhaptor et FSOV Predicropt.

Publications des partenaires du projet sur le sujet

Prédictions Génomiques

Ben-Sadoun S, Rincenc R, Auzanneau J, Oury F-X, Rolland B, Heumez E, Ravel C, Charmet G, Bouchet S. (2020) Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality. *Theor. Appl. Genet.*

Fugerey-Scarbel A, Ben-Sadoun S, Bouchet S, Lemarié S. (2021) Analyzing the economic effectiveness of genomic selection relative to conventional breeding approaches. In: Ahmadi N, Bartholomé J (eds) *Genomic prediction of complex traits*. Springer Nature

Consortium R, Fugerey-Scarbel A, Bastien C, Dupont-Nivet M, and Lemarié S. (2021) Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience. *Front. Genet.*

Civan P, Rincenc R, Danguy-Des-Deserts A, Elsen J-M, Bouchet S. (2021) Population Genomics Along With Quantitative Genetics Provides a More Efficient Valorization of Crop Plant Genetic Diversity in Breeding and Pre-breeding Programs. In Rajora OP (ed) *Population Genomics: Crop Plants*. Springer Nature Switzerland AG.

Ben-Sadoun S, Fugerey-Scarbel A, Charmet G, Lemarié S, Bouchet S. Integration of genomic selection into winter-type bread wheat breeding schemes : a simulation pipeline including economic constraints. *Crop Breeding, Genetics and Genomics*. Under press.

Danguy des Déserts A, Servin B, Elsen J-M and Bouchet S. A pipeline to estimate cross value and design mating plan from parental genotypes and phenotypes. Under prep.

Raoul J, Palhière I, Astruc J-M, Swan A, Elsen J-M. (2018) Optimal mating strategies to manage a heterozygous advantage major gene in sheep. *animal* 12, 454–463.

Raoul J, Palhière I, Astruc J-M, Elsen J-M. (2016) Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs. *Journal of animal science* 94, 3663–3683.

Shumbusho F, Raoul J, Astruc J-M, Palhière I, Lemarié S, Fugerey-Scarbel A, Elsen, J-M. (2016) Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. *animal* 10, 1033–1041.

Recombinaison

Bouchet S, Olatoye M-O, Marla S-R, Perumal R, Tesso T, et al. (2017) Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics* 206 : 573–585.

Danguy des Déserts A, Bouchet S, Sourdille P, and Servin B (2021) Evolution of recombination landscapes in diverging populations of bread wheat. *Genome Biol Evol* evab152.

Publications de référence sur l'estimation de la variance génétique

Bijma P, Wientjes Y-C, and Calus M-P. (2020) Breeding top genotypes and accelerating response to recurrent selection by selecting parents with greater Gametic variance. *Genetics* 214: 91–107.

Lehermeier C, de Los Campos G, Wimmer V, and Schön C-C. (2017) Genomic variance estimates: With or without disequilibrium covariances? *Journal of Animal Breeding and Genetics* 134: 232–241.

Lehermeier C, Teyssède S, and Schön C-C. (2017) Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. *Genetics* 207: 1651–1661.

Zhong S, and Jannink J-L. (2007) Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics* 177: 567–576.