

[B-Dul] Blé Durable de Qualité : Sélection de blés productifs de qualité sous contrainte de nutrition azotée

Ellen GOUDEMANT-DUGUE^{1*}, Jérôme AUZANNEAU², Pierre LEMEUNIER^{1,2,3}, François-Xavier OURY³, Jacques BORDES³, Thierry DEMARQUET¹, Davis ALVAREZ³, Sibille PERROCHON³, Mireille DARDEVET³, Catherine RAVEL³

1 - FLORIMOND DESPREZ VEUVE & FILS, BP41 59242 Cappelle-en-Pévèle

2 - AGRI OBTENTIONS, 78280 GUYANCOURT

3 - INRA, UMR 1095 INRA UCA GDEC, 5 chemin de Beaulieu 63100 Clermont-Ferrand

* *Coordinateur* : Ellen GOUDEMANT-DUGUE, ellen.goudemand@florimond-desprez.fr

1. Introduction

Le blé tendre (*Triticum aestivum* L.) est une ressource essentielle pour l'alimentation humaine mondiale. La qualité d'usage de la farine de blé est définie par sa capacité à être transformée en une pâte aux propriétés rhéologiques appropriées, grâce à ses protéines de réserve, les gliadines et les gluténines qui forment le gluten. Les protéines de réserve, riches en acides aminés soufrés, à l'origine du gluten, constituent 80% des protéines totales présentes dans le grain (Shewry, 2009). Les gluténines représentent le plus couramment 35 à 50% des protéines du grain. Elles sont divisées en gluténines de hauts et faibles poids moléculaires. Les gliadines sont des monomères divisés en ω 5-, ω 1,2-, α/β - et γ -gliadines (Wieser, 2007). Elles représentent entre 30 et 40% des protéines du grain. Les protéines de réserve sont à l'origine du polymère de gluten. Les gluténines sont liées par des liaisons covalentes (des ponts disulfures). Les gliadines sont fixées sur ces polymères par des liaisons faibles (liaisons hydrogènes...). Les procédés de transformation secondaire du blé vont conduire à une restructuration de ce réseau : lors du mélange de la farine avec de l'eau, des liaisons disparaissent, de nouvelles se forment, ce qui donnera de nouvelles propriétés viscoélastiques (Schiedt *et al.*, 2013 ; Ortolan et Steel, 2017). Les gluténines, et particulièrement les sous-unités de haut poids moléculaire, confèrent l'élasticité à la pâte alors que les gliadines apportent la viscosité (MacRitchie, 1999 ; Shewry *et al.*, 2002). La qualité de la pâte résulte de la balance entre ces deux propriétés, due aux proportions de gliadines et gluténines (d'où l'importance du ratio gliadines/gluténines (gli/glu) pour la qualité technologique de la pâte). De plus, depuis 1979, le rôle clef des gluténines de haut poids moléculaire dans la force de la pâte a été démontré (Payne *et al.*, 1979). Une augmentation des gliadines par rapport aux gluténines va réduire la force de la pâte, son élasticité, sa stabilité au cours du temps ainsi que le temps nécessaire pour qu'elle acquière ses propriétés optimales. Le plus souvent, des pâtes fortes (faible rapport gli/glu) sont souhaitées pour la boulangerie et des pâtes plus visqueuses (fort gli/glu) pour la fabrication de pâtisseries.

Si le rapport gliadines sur gluténines est important pour la qualité technologique de la farine, sa prise en compte dans les programmes de sélection n'est pas acquise. En effet, ce trait est difficile et coûteux à mesurer car son phénotypage nécessite plusieurs étapes pour extraire et doser ces protéines (Plessis *et al.*, 2013). L'introduction de ce caractère en sélection repose donc sur des méthodes alternatives basées sur l'utilisation de marqueurs comme la sélection assistée par marqueurs (SAM) ou la sélection génomique (SG) qui permettrait de le prédire.

Quelques études ont été réalisées pour établir le déterminisme génétique (régions chromosomiques impliquées dans le caractère ou QTL et allèles favorables qui peuvent ensuite être sélectionnés). C'est le cas d'une étude de génétique d'association à l'échelle du

génomique (GWAS, Genome Wide Association Study) réalisée par Plessis *et al.* (2013). Ces auteurs ont montré que le QTL qui contrôlait la plus grande variation du rapport gli/glu dans le grain expliquait seulement moins de 10% de la variabilité totale du caractère. Dans ce contexte, la SG serait une alternative plus adaptée pour prédire ce rapport, car capable de prendre en compte un nombre important de QTL à effet faible impliqués dans un caractère (Meuwissen, 2001). Les modèles de SG sont établis à partir d'une population d'entraînement (TP), génotypée et phénotypée pour le caractère d'intérêt. Ce modèle est ensuite appliqué pour prédire ce caractère au sein d'une population à sélectionner, uniquement génotypée. Ajouter certains QTL majeurs en effet fixe au modèle de SG peut améliorer sa précision (Bernardo, 2014 ; Spindel *et al.*, 2016 ; Michel *et al.*, 2018).

Dans ce projet, d'une part, des marqueurs moléculaires, facilement utilisables d'un laboratoire à un autre, ont été développés pour permettre aux sélectionneurs d'intégrer ces caractères dans leurs programmes. Ces marqueurs résultent de QTL identifiés par GWAS. D'autre part, la capacité de prédiction du rapport gli/glu a été étudiée en utilisant la SG. Différentes possibilités d'amélioration ont été étudiées comme l'optimisation de la population d'entraînement ou l'ajout de QTL majeur dans le modèle. C'est cette seconde partie du travail que nous avons choisie de présenter.

2. Matériel et méthode

► Matériel végétal et expérimentation

Trois populations de blé tendre d'hiver ont été utilisées. Ces populations ont été phénotypées dans différents environnements (lieux et années). La première population, CC196, est composée de 196 génotypes issus de la core collection de 372 accessions établie par Balfourier *et al.* (2007). Ces 196 génotypes sont les plus courts et les plus productifs de la collection. La moitié d'entre eux ont été inscrits après 1960 (Bordes *et al.* 2013). Le panel CC196 a été phénotypé à Clermont-Ferrand en 2005-2006 et au Moulon en 2006-2007. La seconde population est constituée de 186 génotypes, issus de la population BWP3 du projet Breedwheat. La population BWP3 est une core collection comprenant les variétés les plus modernes sélectionnées au sein des 4600 accessions du projet Breedwheat représentant la diversité mondiale. La population BWP3 a été phénotypé en 2016-2017 à Clermont-Ferrand. Enfin, la troisième population BDUL est composée de 94 génotypes élites en cours de sélection fournis par les sociétés Agri-Obtentions et Florimond Desprez. Ces génotypes étaient en 2015 soit dans le cycle d'inscription, soit un ou deux ans avant l'inscription. Le panel BDUL a été phénotypé en 2015-2016 dans trois lieux (Clermont, Cappelle-en-Pévèle et Orsonville) dans le cadre du projet FSOV. Le panel BWP3 partage

8 génotypes avec le panel CC196 et 1 génotype avec BDUL. Les panels CC196 et BDUL ne partagent aucun génotype.

Dans tous les essais, les blés ont été cultivés dans des conditions représentatives de ce qui est fait en France.

Site	Year			
	2006	2007	2016	2017
CF	CC196		BDul	BWP3
LM		CC196		
Or			BDul	
CeP			BDul	

Tableau 1 : Répartition des collections dans les six environnements.

► Quantification des gliadines et des gluténines

Une extraction séquentielle des gliadines et des gluténines a été réalisée à partir de 5mg de farine de chaque génotype selon la méthode décrite dans Plessis *et al.* (2013) pour le panel CC196. Pour les panels BWP3 et BDUL, le protocole d'extraction a été ajusté en réduisant le pH et la solution tampon pour améliorer l'extraction des gliadines. Une extraction par génotype et par environnement a été effectuée. La teneur en azote dans chaque fraction (gliadines puis gluténines) a été dosé avec un analyseur élémentaire de type DUMAS (Thermo Electron). Une multiplication par 5.7 de ces teneurs a permis de les « transformer » en % gliadines ou % de gluténines. Ces dernières ont servi à calculer le rapport gli/glu.

► Données de génotypage

Chaque lignée a été génotypée avec la puce Affymetrix Axiom 420k (Rimbert *et al.* 2018). Les marqueurs OTV, les marqueurs avec une MAF (minor allele frequency) inférieure à 0.1, les marqueurs avec un taux de données manquantes supérieur à 0.2 et les marqueurs avec un taux d'hétérozygotie supérieur à 0.15 ont été supprimés. Les génotypes avec un taux de données manquantes supérieur à 0.1 et un taux d'homozygotie supérieur à 0.05 ont été supprimés. Après l'application de ces filtres, le jeu de données était composé de 166 génotypes dans CC196, 167 génotypes dans BWP3, 88 génotypes dans BDUL et 101 402 marqueurs. Les allèles de chaque marqueur ont été codés 1 ou -1 pour chaque allèle présent à l'état homozygote, et 0 pour les hétérozygotes. Les données manquantes ont été imputées par la moyenne de chaque marqueur.

► Evaluation de la qualité du modèle prédictif

Les modèles de SG testés prédisent trois effets : génotype, environnement et génotype x environnement. Le but de cette étude est d'évaluer si les modèles proposés permettent une bonne prédiction de la valeur génétique du rapport gli/glu, quel que soit l'environnement. De ce fait, les valeurs prédites doivent être comparées avec les vraies valeurs génétiques des individus. Les modèles de SG testés ont été évalués selon deux critères : la précision et la RMSE (root mean square error). La précision est définie comme la corrélation entre les valeurs prédites par le modèle et les valeurs mesurées du caractère. La RMSE est la distance quadratique entre les valeurs mesurées et prédites. Ce critère reflète la capacité du modèle à correctement prédire la valeur de l'ensemble des individus, pas uniquement un classement comme c'est le cas de la précision.

Précision et les RMSE ont été estimés par validation croisée. CC196 et BW3 sont systématiquement utilisés dans la population d'entraînement pour calibrer les modèles. On y ajoute 2/3 de la population BDUL, qui contient les génotypes les plus récents. La population de validation sera constituée du

dernier tiers de la population BDUL. Le processus de validation croisée a été répété 60 fois afin de calculer une moyenne et une variance de la précision et de la RMSE.

► Les modèles de sélection génomique testés

Le premier modèle de sélection génomique utilisé pour prédire le ratio gli/glu est le modèle G-BLUP. Les QTL influençant le ratio gli/glu ont été détecté par Zhang *et al.* (2011) et Plessis *et al.* (2013). Nous avons donc exploré une possible amélioration du modèle de prédiction en ajoutant les effets de QTL au modèle G-BLUP. Ce modèle a été intitulé G-BLUP + de novo GWAS comme dans la publication de Spindel *et al.* (2016) et la procédure utilisée est la suivante :

- Modèle GWAS : test des 101 402 SNP avec le modèle GBLUP établi avec la population d'entraînement.
- Sélection des SNP les plus significatifs (FDR =0.5%)
- Regroupement des SNP les plus significatifs en fonction de leur position sur les chromosomes. Les SNP distants de moins de 6000 kb sont considérés comme appartenant au même QTL.
- Sélection d'un SNP représentatif par QTL, celui avec la plus petite Pvalue.
- Ajout de ces SNP en effet fixe au modèle G-BLUP
- Calcul de la précision et de la RMSE de ce modèle G-BLUP + de novo GWAS.

► Optimisation de la population d'entraînement

Plusieurs études ont démontré qu'un fort apparentement entre la population d'entraînement et la population à sélectionner permet d'obtenir de meilleures précisions de prédiction (Albrecht *et al.*, 2011 ; Ly *et al.*, 2013). Si une entreprise de sélection souhaite prédire un caractère par sélection génomique, elle doit construire une population d'entraînement avec des variétés issues de son propre programme de sélection. Si une entreprise possède une grande diversité de variétés génotypées et phénotypées pour le caractère d'intérêt alors, sélectionner seulement une partie d'entre elles peut permettre d'augmenter la capacité de prédiction du modèle. Pour tester cette hypothèse, nous avons comparé 5 méthodes d'optimisation de la population d'entraînement à une méthode de sélection aléatoire. Ces méthodes ont été testées en utilisant un modèle G-BLUP avec des populations d'entraînement de différentes tailles, allant de 20 génotypes au nombre maximum de génotypes disponibles, soit 382.

- La première méthode d'optimisation (OPT_MEAN) consiste en la sélection des génotypes de la population d'entraînement les plus apparentés en moyenne aux génotypes de la population à sélectionner. Certains génotypes seront fortement apparentés et d'autres moins.
- La seconde méthode (OPT_MAX) va sélectionner les génotypes avec l'apparentement maximal.
- La méthode OPT_MIN va sélectionner des individus avec les plus hauts apparentements minimaux.
- La méthode d'optimisation CD_MEAN va se baser sur le coefficient de détermination généralisé (CD) décrit par Laloë (1993), Rincenc *et al.* (2012) et Rutkoski *et al.* (2015). Un CDmean proche de 1 signifie que les contrastes prédits entre la population d'entraînement et la population à sélectionner sont fortement précis et donc les précisions de prédiction seront bonnes. Afin d'optimiser la population d'entraînement avec le critère CDmean, nous avons utilisé le script proposé par Rincenc *et al.* (2012) avec 3000 répétitions.
- La dernière méthode, OPT_IND, optimise la population d'entraînement pour chaque génotype de la population à sélectionner en choisissant les individus les plus apparentés.

Les calculs ont été effectués sous R (R Core Team 2017). Le package Sommer a été utilisé pour la SG et la recherche de QTL pour la partie GWAS (Covarrubias-Pazaran, 2016).

3. Résultats

► Apparentement entre les génotypes

L'apparentement entre les génotypes des différents panels est représenté sur l'analyse en composantes principales ci-dessous (Figure 1). L'axe 1 et l'axe 2 représentent 9.5% et 3.7% de la variance totale, respectivement. On peut clairement voir que l'apparentement entre les génotypes élités de BDUL est fort, avec une moyenne de 0.30. Les génotypes d'Agri-Obtentions sont plus apparentés aux core collections (BWP3 et CC196) que les génotypes Florimond Desprez. Les deux core collections sont relativement peu apparentées au panel BDUL, avec une moyenne de -0.09. Cependant, 9 et 3 génotypes des populations CC196 et BWP3 respectivement sont fortement apparentés à 1 ou plusieurs génotypes du panel BDUL (coefficient d'apparentement supérieur à 0.6).

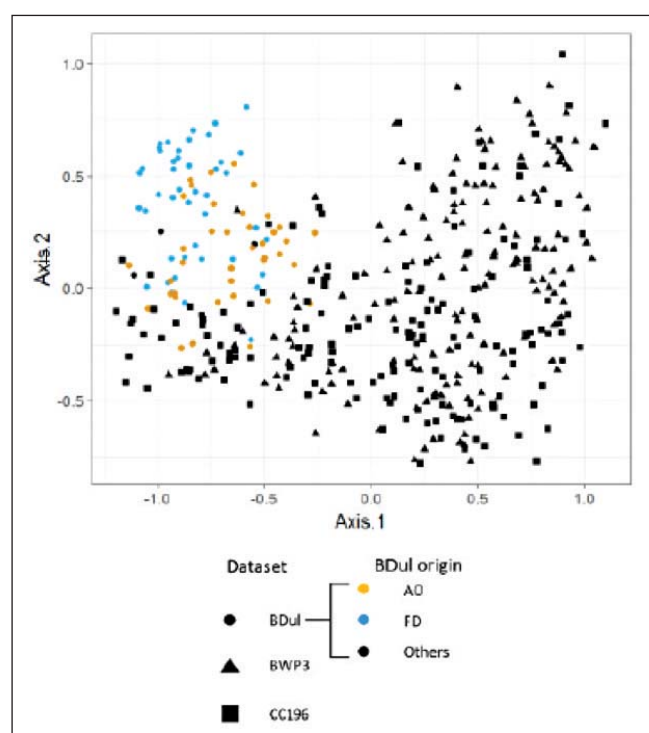


Figure 1 : Analyse en composantes principales des génotypes des 3 collections.

► Modèles de sélection génomique

Tout d'abord, les modèles G-BLUP et G-BLUP + de novo GWAS ont été comparés. De légères différences ont observées entre les deux modèles.

G-BLUP	Accuracy=0.543	(standard-deviance (sd)=0.144), RMSE=0.0606 (sd=0.0043)
G-BLUP + de novo GWAS	Accuracy=0.57	(sd=0.142), RMSE=0.0600 (sd=0.0042)

Le modèle G-BLUP + de novo GWAS intègre entre 0 et 4 QTL détectés par GWAS. A part deux répétitions sur les 60, les QTL détectés été systématiquement les suivants :

- QTL sur le chromosome 1A, effet moyen = -0.0275, position = 509 014 168 pb
- QTL sur le chromosome 1B, effet moyen = -0.0201, position entre 5 143 303 et 5 699 395 pb
- QTL sur le chromosome 7D, effet moyen = 0.0145, position = 571 785 471 pb

► Méthodes d'optimisation

Cinq méthodes d'optimisation de la population d'entraînement ont été testées et comparées à une méthode de sélection aléatoire. Le but était d'augmenter la capacité de prédiction en sélectionnant une population d'entraînement d'une taille limitée. Les précisions et les RMSE montrent que les méthodes d'optimisation de la population d'entraînement améliorent les capacités de prédiction, comparées à la méthode de sélection aléatoire (Figure 2). Toutes les méthodes de sélection montrent une amélioration de la capacité de prédiction lorsque la taille de la population d'entraînement augmente jusqu'à atteindre un seuil, minimisant la RMSE et maximisant la précision, qui se situe entre 80 et 140 génotypes. Une fois le maximum atteint, augmenter la taille de la population d'entraînement aboutit globalement à une diminution des capacités de prédiction jusqu'à atteindre une précision de 0.543 et une RMSE de 0.0606 lorsque le maximum d'individus est utilisé (382) (Figure 2).

La méthode de sélection aléatoire des individus présente une augmentation logarithmique de la précision (avec un minimum de 0.271 pour 20 génotypes dans la population d'entraînement) et une décroissance exponentielle de la RMSE (avec un maximum de 0.0818) avec la taille de la population. Les différences de capacité de prédiction entre les méthodes d'optimisation sont faibles. A 20 génotypes, la précision varie de 0.368 (OPT_IND) à 0.482 (CD_MEAN) mais la RMSE ne suit pas le même ordre en variant de 0.0723 (OPT_MIN) à 0.0618 (OPT_IND). La précision maximale (0.624) est obtenue avec la méthode OPT_IND avec 80 génotypes mais les quatre autres méthodes montrent des résultats proches. La méthode CD-MEAN présente la particularité de conserver une capacité de prédiction constante de 60 à 280 génotypes. On observe également que la méthode OPT_MAX présente également de faibles variations de la capacité de prédiction.

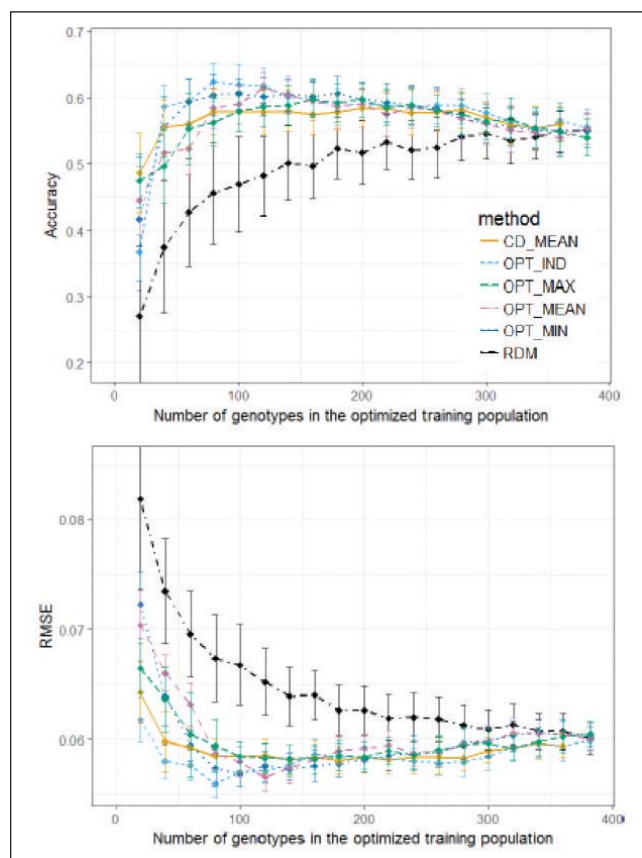


Figure 2 : Evolution de la précision (en haut) et du RMSE (en bas) selon la méthode de sélection de la population d'entraînement et sa taille. Méthode random (RDM) : les génotypes sont choisis au hasard.

4. Discussion

Le ratio gli/glu dans la farine de blé est un composant majeur de la qualité boulangère (Marchetti *et al.*, 2012 ; Barak *et al.*, 2014). Mais son phénotypage est cher et fastidieux. C'est pourquoi sa prédiction par sélection génomique pourrait être utilisée pour sélectionner plus précocement et plus précisément des blés de qualité.

► Le modèle G-BLUP est bien adapté pour prédire le ratio gliadine/gluténine

Le modèle G-BLUP qui a été appliqué prend en compte trois effets importants en sélection : le génotype, l'environnement et l'interaction génotype x environnement (Jarquin *et al.*, 2014). La covariance entre les effets des génotypes a été établie avec l'appareillement entre les individus, calculé avec les marqueurs. Cependant, aucune covariance n'a été considérée entre les environnements, qui aurait pu être calculée grâce à des covariables environnementales. Ces covariables influencent très certainement le ratio gli/glu et elles ajouteraient de l'information au modèle G-BLUP. Connaître ces informations et les inclure dans le modèle permettraient d'en augmenter la précision et donc la capacité de prédiction de l'effet génétique des individus.

La capacité de prédiction du modèle G-BLUP (précision de 0.543 et RMSE de 0.0606) était bonne comparée aux prédictions des caractères de qualité précédemment publiés. Avec un modèle RR-BLUP, Michel *et al.* (2018) ont obtenu des précisions entre 0.30 et 0.54 en fonction du caractère de qualité boulangère analysé. Cependant, la comparaison reste difficile puisque les structures des populations sont différentes.

Ajouter des QTL en effet fixe au modèle de sélection génomique peut améliorer la capacité de prédiction d'un caractère. Cela a déjà été démontré pour plusieurs caractères, comme les résistances aux maladies et la tolérance à la germination sur pied (Arruda *et al.*, 2016 ; Moore *et al.* 2017). Cependant, Michel *et al.* (2018) et Liu *et al.* (2016) ont observé qu'ajouter des marqueurs en effet fixe au modèle de sélection génomique améliore très peu, voire pas du tout, la capacité de prédiction des caractères de qualité. Ces résultats sont confirmés dans notre étude avec notre modèle G-BLUP + de novo GWAS qui permet d'améliorer que très faiblement la précision du modèle G-BLUP, alors que la RMSE reste constante. Cela confirme les résultats obtenus par Bernardo (2014) qui a observé que les marqueurs doivent expliquer une part importante de la variance phénotypique (supérieure à 10%) pour être intégrés au modèle de sélection génomique et améliorer les prédictions. Le ratio gli/glu est un caractère fortement polygénique et donc les QTL trouvés ont des effets très faibles ce qui ne permet pas d'améliorer le modèle de sélection génomique.

Les modèles Bayésiens sont également une alternative au modèle G-BLUP testé. Cependant, ces modèles semblent plus efficaces que le modèle G-BLUP uniquement si des QTLs majeurs influencent le caractère (Clark *et al.* 2011 ; Heffner *et al.* 2011).

► L'optimisation de la population d'entraînement permet d'aboutir à de meilleures capacités de prédiction

La grande diversité observée entre les génotypes des trois jeux de données implique que certains individus de la core collection sont génétiquement distants des génotypes modernes que nous souhaitons prédire. Inclure ces génotypes distants dans la population d'entraînement de notre modèle de prédiction peut diminuer la capacité à prédire la valeur des génotypes modernes. Pour contourner ce problème, nous avons testé cinq méthodes d'optimisation de la population d'entraînement. Toutes ces méthodes étaient basées sur l'appareillement entre les génotypes de la population à sélectionner et les individus de la population d'entraînement. Elles aboutissent à de meilleurs résultats que la sélection d'individus au hasard. La meilleure méthode d'optimisation sera celle qui sélectionnera une population d'entraînement avec la même diversité allélique que la population à sélectionner. On observe une diminution de la capacité de prédiction une fois que le maximum est atteint avec les méthodes d'optimisation. Cela s'explique par le fait que les derniers individus ajoutés sont trop distants génétiquement de la population à sélectionner, et ils apportent des allèles qui sont absents de la population à sélectionner. Isidro *et al.* (2015) suggéraient que capturer la plus grande partie de la variance phénotypique dans la population d'entraînement optimisée peut permettre d'augmenter la capacité de prédiction. Nos résultats vont à l'encontre de cette suggestion puisqu'ajouter des génotypes à la population d'entraînement augmente la variance phénotypique capturée mais diminue la capacité de prédiction. Ce qui certainement dû à notre objectif de prédire une population avec une faible diversité en comparaison avec la diversité de la population d'entraînement.

Les méthodes d'optimisation testées donnent des résultats similaires. Aucune n'apparaît vraiment meilleure que les autres. On ne peut donc que conseiller d'utiliser la méthode la plus rapide en termes de calcul sachant que la sélection des individus et la demande computationnelle étaient différentes entre les méthodes (la mise en œuvre de OPT_IND et CD-MEAN demande plus de 10 fois plus de temps que les autres méthodes). Rincet *et al.* (2017) ont testé les méthodes d'optimisation OPT_MEAN et CD_MEAN alors que Isidro *et al.* (2015) et Rutkoski *et al.* (2015) ont testé uniquement CD_MEAN. Leurs résultats montrent que l'efficacité de la méthode d'optimisation est fortement dépendante du caractère et de la structure de la population. Notre objectif était de prédire du matériel élite avec un panel de diversité, ce qui est un cas particulier. Nos conclusions doivent donc être mises en perspective.

OPT_IND aboutit à une relativement meilleure capacité de prédiction. Chaque génotype est prédit par son plus proche apparenté présent dans la population d'entraînement. La capacité de prédiction aurait été probablement bien meilleure que les autres méthodes si les génotypes de la population à sélectionner avaient été plus divers. Il n'en reste que cette méthode demande du temps de calcul.

Pour conclure, malgré le fait que nous ayons cherché à prédire du matériel élite avec des collections de diversité, le modèle G-BLUP développé montre des résultats satisfaisants ce qui montre l'intérêt de la sélection génomique pour la prise en compte des caractères de qualité dans les programmes de sélection.

Références bibliographiques

- Albrecht T, Wimmer V, Auinger HJ et al.** (2011) Genome-based prediction of test - cross values in maize. *Theor Appl Genet* 123:339–350.
- Arruda MP, Lipka AE, Brown PJ, et al.** (2016) Comparing Genomic Selection and Marker-Assisted Selection for Fusarium Head Blight Resistance in Wheat (*Triticum Aestivum* L.). *Molecular Breeding* 36:84.
- Balfourier F, Rousset V, Strelchenko P et al.** (2007) A Worldwide Bread Wheat Core Collection Arrayed in a 384-Well Plate. *Theor Appl Genet* 114:1265–75.
- Barak S, Mudgil D, Khatkar BS** (2014) Influence of Gliadin and Glutenin Fractions on Rheological, Pasting, and Textural Properties of Dough. *International Journal of Food Properties* 17:1428–1438.
- Bernardo R** (2014) Genomewide Selection When Major Genes Are Known. *Crop Science* 54:68–75.
- Bordes J, Ravel C, Jaubertie JP et al.** (2013) Genomic Regions Associated with the Nitrogen Limitation Response Revealed in a Global Wheat Core Collection. *Theoretical and Applied Genetics* 126:805–22.
- Clark SA, Hickey JM, van der Werf JH** (2011) Different models of genetic variation and their effect on genomic evaluation. *Genet Sel Evol* 43:18.
- Covarrubias-Pazarán G** (2016) Genome-Assisted prediction of quantitative traits using the *r* package sommer. *PLoS One* 11:1–15.
- Heffner EL, Jannink JL, Iwata H et al.** (2011) Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Science* 51:2597–2606. <https://doi.org/10.2135/cropsci2011.05.0253>
- Isidro J, Jannink JL, Akdemir D et al.** (2015) Training Set Optimization under Population Structure in Genomic Selection. *Theor Appl Genet* 128:145–158. <https://doi.org/10.1007/s00122-014-2418-4>
- Jarquín D, Crossa J, Lacaze X, et al.** (2013) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127:595–607. <https://doi.org/10.1007/s00122-013-2243-1>
- Laloë D** (1993) Precision and information in linear models of genetic evaluation. *Genet Sel Evol* 25:557–576.
- Liu G, Zhao Y, Gowda M et al.** (2016) Predicting Hybrid Performances for Quality Traits through Genomic-Assisted Approaches in Central European Wheat. *PLoS ONE* 11:e0158635. <https://doi.org/10.1371/journal.pone.0158635>
- Ly D, Hamblin M, Rabbi I et al.** (2013) Relatedness and Genotype × Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava. *Crop Sci* 53:1312–1325.
- MacRitchie F** (1999) Wheat proteins: characterization and role in flour functionality. *Cereal Foods World*, 44:188–193
- Marchetti L, Miguel C, Leda C, Cristina F** (2012) Effect of Glutens of Different Quality on Dough Characteristics and Breadmaking Performance. *Food Science and Technology* 46:224–31. <https://doi.org/10.1016/j.lwt.2011.10.002>
- Meuwissen THE, Hayes BJ, Goddard ME** (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Michel S, Ametz C, Gungor H et al.** (2016) Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theor Appl Genet*. 129:1179–89.
- Michel S, Kummer S, Galle M et al.** (2018) Improving the baking quality of bread wheat by genomic selection in early generations. *Theor Appl Genet*. 131:477–93.
- Moore J K, Manmathan HK, Anderson VA et al.** (2017) Improving Genomic Prediction for Pre-Harvest Sprouting Tolerance in Wheat by Weighting Large-Effect Quantitative Trait Loci. *Crop Sci* 57:1315–1324.
- Ortolan F, Steel CJ** (2017) Protein Characteristics that Affect the Quality of Vital Wheat Gluten to be Used in Baking: A Review. *Comprehensive reviews in food science and food safety* 16:369–381.
- Oury FX, Godin C** (2007) Yield and Grain Protein Concentration in Bread Wheat: How to Use the Negative Relationship between the Two Characters to Identify Favourable Genotypes? *Euphytica* 157:45–57.
- Payne PI, Corfield KG, Blackman JA** (1979) Identification of HMW-subunit of glutenin whose presence correlates with bread-making quality in wheat of related pedigree. *Theor appl genet*. 55:153–159.
- Rimbert H, Darrier B, Navarro J et al.** (2018) High Throughput SNP Discovery and Genotyping in Hexaploid Wheat. *PLOS ONE* 13:e0186329.
- Rincent R, Laloë D, Nicolas S, et al.** (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of Maize Inbreds (*Zea mays* L.). *Genetics* 192:715–728.
- Rutkoski J, Singh RP, Huerta-Espino J et al.** (2015) Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. *Plant Genome* 8:1.
- Plessis A, Ravel C, Bordes J et al.** (2013) Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. *J Exp Bot* 64:3627–3644.
- Schiedt B, Baumann A, Conde-Petit B, Vilgis TA** (2013) Viscoelastic Properties during Dough Development. *J Texture Stud* 44:317–332.
- Shewry PR** (2009) Wheat *Journal of Experimental Botany* 60: 1537–1553.
- Spindel JE, Begum H, Akdemir D et al.** (2016) Genome-Wide Prediction Models That Incorporate de Novo GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116:395–408.
- R Development Core Team** (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Wieser H** (2007) Chemistry of gluten proteins. *Food Microbiology* 24:115–119.
- Zhang Y, Tang J, Zhang Y et al.** (2011) QTL Mapping for Quantities of Protein Fractions in Bread Wheat (*Triticum Aestivum* L.). *Theor Appl Genet* 122:971–987