

ExIGE : Evaluation multi-environnementale de blé tendre « exotique » - prédictions génomiques GxE et génétique d'association

Justin BLANCON¹, Jonathan KITT¹, Pauline LASSERRE-ZUBER¹, David ALVAREZ¹, Sibille PERROCHON¹, Jean-Noël THAUVIN², Céline DUQUE³, Sylvie DUTRIEZ⁴, Pascal GIRAudeau⁵, Ellen GOUEMAND-DUGUE⁶, Emmanuel HEUMEZ⁷, Christophe MICHELET², Patrice SENELLART⁸, Alexis OGER⁹, François BALFOURIER¹, Catherine RAVEL¹, Sophie BOUCHET^{1*}

- 1 - INRAE Université Clermont-Auvergne, UMR 1095, GDEC, 5 chemin de Beaulieu, 63100 Clermont-Ferrand, FRANCE
- 2 - RAGT, Rue Emile Singla, BP 3331 12033 Rodez Cedex 9, FRANCE
- 3 - Limagrain Europe, Centre de recherche de Chappes, 63720 Chappes, FRANCE
- 4 - Lidea Seeds, avenue Gaston Phoebus, 64230 Lescar, FRANCE
- 5 - Secobra Recherches, Centre de Bois-Henry, 78580 Maule, FRANCE

- 6 - FLORIMOND DESPREZ VEUVE & FILS, 59242 Cappelle-en-Pévèle, FRANCE
 - 7 - INRAE UE GCIE, 2 chaussée Brunehaut, Estrées Mons, BP50136,80203 Peronne Cedex, FRANCE
 - 8 - KWS MOMONT SAS, 7 Rue de Martinval, 59246 Mons-en-Pévèle, FRANCE
 - 9 - Syngenta France SA, 1228 Chemin de l'Hobit, 31790 Saint-Sauveur, FRANCE
- *Coordinatrice : Sophie BOUCHET, sophie.bouchet@inrae.fr

Introduction

Face à l'instabilité et la stagnation des rendements de blé tendre, il est nécessaire de réintroduire une diversité utile en sélection pour améliorer la tolérance aux stress biotiques et abiotiques mais aussi la qualité boulangère. En raison de l'impact majeur des interactions Génotype x Environnement, ces réintroductions doivent s'opérer dans les environnements adéquats pour maximiser le gain génétique. Dans le projet ExIGE, nous avons caractérisé un panel de diversité mondiale enrichi en accessions asiatiques (Figure 4), au champ (rendement, maladies, variables liées à la qualité boulangère) et au laboratoire (SNP Breedwheat, RNAseq DeBAT). L'objectif principal était de calibrer un modèle de prédiction génomique GxE permettant de prédire de nouveaux individus dans de nouveaux environnements pour le rendement. Le deuxième objectif est d'identifier les déterminants génétiques des caractères d'intérêt majeur du blé tendre (rendement, maladies, qualité boulangère).

Réseau d'essais

Au cours de ce projet nous avons complété un jeu de données phénotypiques issu de 12 essais menés lors du projet PIA BreedWheat en conditions fort vs faible azote ou pluvial vs irrigué, avec 12 nouveaux essais dans des conditions similaires (Figure 1). Le réseau constitué a permis d'observer une large gamme de rendement, avec des essais assez peu corrélés, mais sans structure d'interaction marquée.

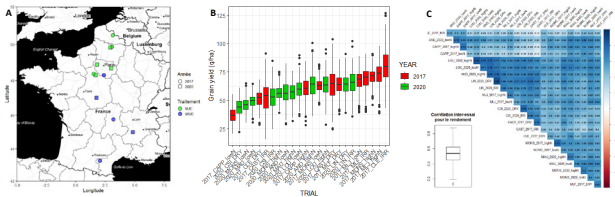


Figure 1 - (A) Répartition du réseau d'essais en France. La forme des points indique l'année et la couleur la contrainte abiotique appliquée ; (B) Gamme de variation pour le rendement. La couleur indique l'année ; (C) Matrice de corrélation entre chaque essai pour le rendement. Le boxplot montre la gamme de corrélation (r)

Caractérisation environnementale

Les données météorologiques, de description du sol et d'itinéraire technique ont été utilisées pour simuler 192 des indices environnementaux à l'aide du modèle éco-physiologique CHN (Arvalis). Ces indices ont été estimés par phase de développement à l'aide des données mesurées pour l'épiaison. L'évolution du confort hydrique et azoté a été estimée tout au long du cycle (Figure 2) et montre que la plupart des essais a subi un stress hydrique nul, faible ou tardif, à l'exception de CLE_2017_DRY, MAU_2020, MONS_2017 et MONS_2020. Le stress azoté a également été très limité dans le réseau et principalement présent sur les essais MONS_2017_lowN et MONS_2020_lowN.

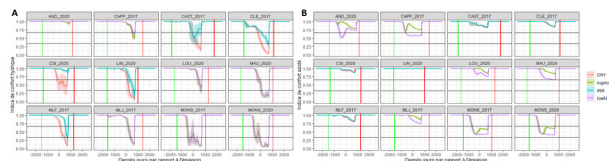


Figure 2 - Evolution de l'indice de confort hydrique (A) et azoté (B) au cours du cycle de la plante. Une valeur de 1 indique le confort et une valeur inférieure indique un stress : faible au-dessus de 0.66, modéré entre 0.66 et 0.33 et fort en-dessous

Prédictions génomiques GxE

Une approche de prédiction génomique basée sur un modèle AMMI (Additive Main effect and Multiplicative Interaction, (Gallab 1968)) a été évaluée à l'aide de 8 modèles différents (Table 1). Ces modèles se différencient par la modélisation de la matrice d'interactivité génétique. Ils utilisent une matrice Kinship (modèles 1, 2, 3, 4 et 6), une matrice issue de l'approche AMMI (modèles 5, 6, 7 et 8), une matrice construite sur les indices environnementaux (modèles 3, 4 et 5). La predictive ability de ces modèles estimée par cross-validation dans différents scénarios de prédiction (Table 2) indique que l'approche AMMI semble prometteuse avec un gain théorique de prédictive ability de +0,2 en moyenne dans les 4 scénarios par rapport au GBLUP. Mais en pratique le gain reste minime (gain maximum de 0,08) car les matrices d'interactivité génétique et environnementale sont difficiles à prédire.

| Modèles | scénarios de Cross-Validation | RMSE | r | RMSE | r |
|---------|-------------------------------|-------|------|-------|------|
| 1 | 1 | 0.072 | 0.82 | 0.086 | 0.48 |
| 2 | 1 | 0.076 | 0.82 | 0.086 | 0.48 |
| 3 | 1 | 0.073 | 0.82 | 0.086 | 0.48 |
| 4 | 1 | 0.070 | 0.84 | 0.080 | 0.56 |
| 5 | 1 | 0.071 | 0.83 | 0.081 | 0.49 |
| 6 | 1 | 0.072 | 0.82 | 0.086 | 0.48 |
| 7 | 1 | 0.071 | 0.83 | 0.081 | 0.49 |
| 8 | 1 | 0.070 | 0.84 | 0.080 | 0.56 |

Table 1 (gauche) - Liste des 8 modèles comparés et modélisation de la variance d'interaction GxE correspondante

Table 2 (droite) - Qualité de prédiction pour les 8 modèles testés. Les scénarios sont aGoE (old Genotype & old Environment), nGoE (new Genotype & old Environment), aGoE (old Genotype & new Environment) et nGoE (new Genotype & new Environment)

GWAS - Caractères agronomiques

Une analyse d'association a été réalisée avec le modèle K par l'approche LOCO (Rincent et al., 2014). Au total, 272 QTL ont été détectés, répartis sur l'ensemble des chromosomes (Figure 3). La qualité de l'analyse est confirmée par l'observation d'importantes colocalisations au niveau de gènes majeurs, comme PPD-D1 pour la précocité. Pour le rendement, 9 QTL ont été identifiés. Une colocalisation de QTL de composantes du rendement, indépendants de la précocité et de la hauteur semble intéressante sur le chromosome 7A et explique en moyenne 6,7% de la variance.

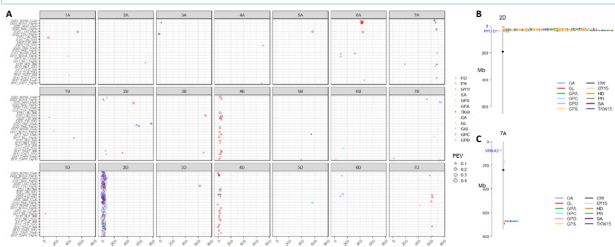


Figure 3 - (A) Répartition des QTL détectés sur l'ensemble du génome pour chaque caractère dans chaque essai ; (B) Colocalisation de QTL de précocité avec PPD-D1 ; (C) Colocalisation de QTL de composantes du rendement sur le 7A. Les caractères étudiés sont la précocité (HD), la hauteur (PH), le rendement à 15% d'humidité (CY15), le nombre d'épis par m² (SA), le nombre de grains par épi (GPS), le nombre de grains par m² (GPA), le poids de mille grains (TKW15), la surface, la longueur et la largeur du grain (GA, GL, GW), la teneur en protéine (GPC) et le Grain Protein Deviation (GPD).

Phénotype protéines de réserve / qualité boulangère

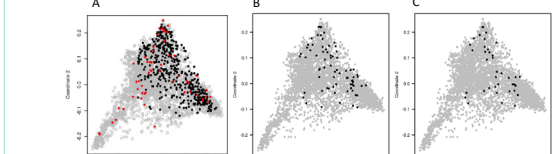


Figure 4 - (A) : panel BWP3 étendu (485), (B) panel gluténines (73), (C) panel alvéolographie (45)

Parmi les 4500 accessions Breedwheat, 450 correspondent à BWP3 (noires sur Figure 4A), 20 asiatiques et 25 présentant des prédictions génomiques intéressantes ont été ajoutées (rouges sur Fig. 4A) pour former BWP3 étendu. Parmi elles, 73 accessions représentatives ont été dosées et séquencées pour les protéines de réserve (noires sur Figure 4B). Parmi elles, 45 ont été évaluées pour les variables d'alvéolographie en conditions optimales et stressées (azote: N/NO, irrigation: DRY/IRR) (noires sur Figure 4C).

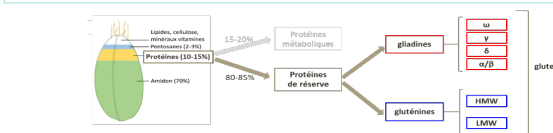


Figure 5 - Composition du grain de blé tendre, focus sur les protéines de réserve

Séquençage des gluténines

L'annotation des principales protéines de réserve (Figure 5), 18 gènes de gluténines et 44 gènes de gliadines, a été réalisée chez Chinese-Spring V2.1. Des fragments de 4500 à 8500 pb ont été amplifiés. Seules les amorces de gluténines ont fonctionné et permis le séquençage de 6 gènes de gluténines de haut poids moléculaire (HMW en fin des chromosomes du groupe 1) et 12 de faible poids moléculaire (LMW en début des chromosomes du groupe 1). Parmi les 2102 SNP, 1309 ont une MAF>0.05 et 53 sont en équilibre de liaison. On remarque que l'allèle « absent » est un allèle fréquent à considérer dans les associations avec les phénotypes (Table 2).

On observe du DL intra-gène. Au total, 25 marqueurs permettent de résumer la diversité des HMW et 29 pour les LMW. Chaque gène possède en moyenne 4 haplotypes (Figure 6).



Figure 6 - Exemple de 4 haplotypes chez 73 accessions pour un gène LMW

GWAS - Protéines de réserve

Table 2 - Associations protéines de réserve

le: élasticité, L: extensibilité, protTot: protéine totale, glu: gluténine, gli: gliadine, N: azote total

| Trait | #Chr | Chr | Position (CS V2.1, bp) | Pvalue | part variance | QTL |
|---------|------|-----|------------------------|----------|---------------|---------------------------|
| le | 1 | 1A | 509 017 352 | 9.40E-07 | 0.49 | Glu_HMW-A |
| L | 5 | 2B | 745 715 941 | 2.19E-07 | 0.56 | |
| L | 15 | 5D | 548 216 820 | 4.84E-07 | 0.52 | CW5-5D |
| protTot | 2 | 1B | 94 458 186 | 6.01E-09 | 0.37 | 1R5/1B(gli/LMW) |
| protTot | 2 | 1B | 105 538 550 | 6.01E-09 | 0.37 | 1R5/1B(gli/LMW) |
| gliW5 | 3 | 1D | 578 541 | 3.33E-09 | 0.39 | gli/LMW |
| gliTot | 7 | 3A | 728 650 452 | 6.02E-07 | 0.29 | TGW5 |
| Nfarine | 8 | 3B | 373 613 528 | 4.50E-07 | 0.32 | TalyC-B1 |
| Nfarine | 8 | 3B | 774 409 784 | 4.32E-07 | 0.32 | Red-3B |
| Nfarine | 10 | 4A | 263 086 190 | 4.72E-07 | 0.32 | introgression dicoccoides |
| gliW5 | 11 | 4B | 1 135 117 | 6.22E-09 | 0.38 | Low-B1 |
| gliW1.2 | 13 | 5A | 476 434 666 | 6.77E-07 | 0.30 | nac27 |
| gliW5 | 14 | 5B | 641 046 239 | 8.76E-07 | 0.30 | SHP |
| gliW5 | 14 | 5B | 703 647 253 | 7.25E-09 | 0.38 | |
| gliW5 | 17 | 6B | 10 754 025 | 7.96E-09 | 0.39 | |
| gliW5 | 17 | 6B | 514 027 602 | 4.38E-07 | 0.30 | SAD |

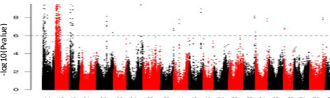


Figure 7 - Manhattan plot protéines de réserve

Toutes les variables liées aux protéines de réserves ont été projetées sur le même manhattan plot pour avoir une vision globale des QTL détectés

Au total, 16 QTL principal ont été détectés (Table 2). Ils sont en général associés à plusieurs caractères dans plusieurs environnements et conditions. Ils expliquent entre 30 et 50% de la variance du phénotype.

De nombreuses associations sont trouvées dans les régions des gliadines et LMW au début des chromosomes 1A, B et D (1,2 et 3) (Figure 7). L'introgression 1R5/1B du seigle sur le chromosome 1B est présente dans le panel, ce qui explique l'étendue du QTL sur le bras court.

GWAS - maladies

Au total, 54 QTL principaux ont été détectés sur la rouille jaune, la septoriose et la fusariose. Ils sont retrouvés dans plusieurs environnements et conditions. Ils expliquent entre 4 et 11% de la variance.

Les QTL les plus intéressants se situent sur les chromosomes 2B (5) et 6A (16) pour la rouille jaune (Figure 8). Certaines accessions présenteraient une introgression au niveau du chromosome 2B.

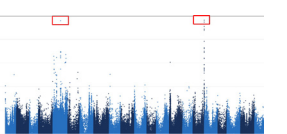


Figure 8 - Manhattan rouille jaune

L'association a été faite sur les individus d'un modèle avec hauteur et précocité en variable sans explication.

Perspectives

Un jeu de données de qualité de génotypes et de phénotypes a été produit sur un large panel de diversité (485). Des analyses poussées ont été faites sur les données rendement. Les analyses sont encore en cours pour les autres caractères. Un travail d'imputation et de méta-analyses avec d'autres jeux de données sera nécessaire pour valider les QTL, en particulier pour les caractères où nous avons évalué seulement 45 (alvéolographie) ou 75 (protéines de réserve) individus. L'objectif est de détecter des génotypes intéressants et de constituer un set de marqueurs diagnostiques pour optimiser le pre-breeding.

