

# GPDUR : Génétique et Phénotypage du blé dur pour l'amélioration conjointe du rendement et de la teneur en protéines

**Delphine HOURCADE\*<sup>1</sup>, Boris DEMENOU<sup>1</sup>, Damien EYBALIN<sup>1</sup>, Matthieu BOGARD<sup>1</sup>, Thierry LEFEVRE<sup>2</sup>, Marion CHAPUS<sup>2</sup>, Carine ANDRE<sup>3</sup>, Andrea MASSI<sup>3</sup>, Jacques DAVID<sup>4</sup>, Pierre ROUMET<sup>4</sup>, Céline DUQUE<sup>5</sup>**

1 - ARVALIS - 6 Chemin de la Cote Vieille - 31450 Baziège

2 - RAGT 2n - 1740 Route du Canal - 82700 Montbartier

3 - SYNGENTA - 12 Chemin de l'Hobit, 31790 Saint-Sauveur / Società Produttori Sementi S.p.A. Via Macero, 1 - 40050 Argelato (Bologna) - Italy

4 - INRAe - Bâtiment Arcad, 10 rue Arthur Young, 34090 Montpellier

5 - LIMAGRAIN Europe - Ferme de l'Etang - BP3 - 77390 Verneuil l'Etang

\***Coordinateur : Delphine HOURCADE**, [d.hourcade@arvalis.fr](mailto:d.hourcade@arvalis.fr)

## 1 Introduction

La France est le 2<sup>ème</sup> producteur de l'Union Européenne (UE) de blé dur et c'est la 4<sup>ème</sup> céréale cultivée en France. Ainsi, 2 millions de tonnes sont produites sur le territoire dont 600 000 t/an sont utilisées par les industries de transformations (6 semouleries, 8 usines de fabrication de pâtes alimentaires, 4 de couscous et 2 de grains de blé dur précuits). L'autre partie de la récolte est exportée à parts égales vers les pays de l'UE (Italie et Benelux) et les pays du sud de la Méditerranée (Maghreb principalement).

Depuis quelques années, les surfaces françaises emblavées en blé dur diminuent et les exigences qualitatives des utilisateurs ont été à l'origine d'une structuration de la filière au sein de laquelle toutes les parties prenantes ont visé une production de haut niveau qualitatif, recherchant notamment de hauts niveaux de teneur en protéines.

Malheureusement, la base génétique du matériel végétal cultivé en France, étroite, semble arriver à ses limites. En effet, les résultats des dernières années d'inscription de variétés montrent que les gains de rendement se font systématiquement au détriment du taux de protéines. Ce second critère étant majeur pour l'accès au marché, le progrès génétique en matière de rendement est de fait très réduit. Or, l'absence de progrès en rendement sur le blé dur risque de mettre à mal la compétitivité de la culture vis-à-vis des autres cultures qui s'offrent au choix des producteurs, mettant en danger les volumes d'approvisionnement nécessaires au bon fonctionnement des industries d'aval. Evaluer une plus large diversité dans les conditions de culture française, afin d'identifier des sources génétiques pouvant amener à rompre cette corrélation négative entre rendement en grains et teneur en protéines semble donc essentiel. La relation négative entre rendement et protéine n'est pas une spécificité du blé dur. C'est également un sujet de préoccupation en blé tendre. Or, les travaux conduits par l'INRAe, les semenciers et Arvalis ont permis de développer des méthodes et des hypothèses permettant d'identifier les variétés non plus individuellement pour chacun des deux critères, mais conjointement pour les deux. Brièvement, il s'agit de caractériser la relation négative sur l'ensemble des données disponibles, puis de calculer, pour chaque variété, son écart à la tendance globale. Ce critère, dit « Grain Protein Deviation » (GPD) peut faire l'objet

d'études génétiques et d'amélioration variétale (Oury *et al.*, 2007, Bogard *et al.*, 2011). Sur blé tendre, la focalisation sur ce critère a également permis de montrer que l'absorption d'azote post-floraison est un facteur clé de la détermination du GPD (Bogard *et al.*, 2010), et que celle-ci est elle-même en partie liée à la dynamique de sénescence (Bogard *et al.*, 2011 ; Gaju *et al.*, 2013). Cette dynamique de la sénescence peut désormais être évaluée de manière fine sur des dispositifs de grandes tailles grâce aux outils de phénotypage haut débit (Comar *et al.*, 2012 ; Vilnius *et al.*, 2014). L'étude du critère GPD en blé dur n'a pas fait l'objet d'autant de travaux qu'en blé tendre. Bien que celui-ci soit relativement facile à calculer, par exemple sur les données d'inscription et post-inscription, il semble utile de mieux caractériser la diversité disponible aux sélectionneurs sur ce critère. La génétique d'association est une approche désormais utilisée en routine pour identifier des zones génomiques sous-tendant les variations de traits quantitatifs comme le rendement ou le GPD (Maccaferri *et al.*, 2011 ; Cormier *et al.*, 2014). Les panels d'association sont en général issus d'échantillonnage au sein de collections préétablies et présentent souvent une structuration génétique due à l'âge ou à l'origine du matériel (Buckler et Thornsberry 2002). Chez le blé dur, la structuration génétique est marquée, notamment entre les blés durs élites, les cultivars anciens et autres formes de *Turgidum* (Thuillet *et al.*, 2005, Maccaferri 2005). Plusieurs sources de diversité sont aujourd'hui disponibles pour constituer un panel de diversité génétique élite, adaptée ou rapidement adaptable aux conditions françaises. Depuis 2010, l'université de Bologne a constitué puis intensément étudié un panel de variétés adaptées aux conditions méditerranéennes, tant sur des critères de réponse au stress hydrique (Maccaferri *et al.*, 2011) que sur diverses résistances (rouilles - Maccaferri *et al.*, 2010 ; mosaïque des céréales SBCMV - Ratti *et al.*, 2006). L'intérêt majeur de ce panel réside à la fois dans sa représentation de la diversité élite (Royo *et al.*, 2010) et dans la fenêtre de floraison restreinte qui a été choisie (Maccaferri *et al.*, 2006, 2011) qui permet de limiter les effets de confusion lors des études sur le rendement et les stress abiotiques.

Le panel EPO (Evolutionary Pre-breeding Population) (David *et al.*, 2014) a été spécialement conçu dans le but d'intégrer un maximum de diversité élargie et originale par rapport au compartiment élite du blé dur (Thuillet *et al.*, 2005, Haudry *et al.*, 2007). Il est issu de



14 cycles de brassage avec un taux d'allogamie estimé de 20 % d'inter-croisements d'une population composite fondée par l'introgession d'accessions entre les sous espèces de blé dur *T. turgidum sp.* ayant divergé avant l'apparition de la forme moderne de blé dur (*T. turgidum durum*) correspondant aux différentes étapes évolutives de ce lignage. Ce composite de l'espèce permet d'avoir une base génétique élargie avec un déséquilibre de liaison peu étendu (David *et al.*, 2014, David *et al.*, 2015). Environ 200 lignées correspondant à du matériel génétique fixé, ayant des caractéristiques (taille, précocité, sensibilité à la verse, et facilité de récolte) compatibles avec les modalités expérimentales ont été caractérisées dans le projet et génotypées avec la puce Axiom Breedwheat 420K.

Ce projet a donc pour objectif d'évaluer un panel combinant lignées élites et lignées ancêtres améliorées, par la mise en place d'un réseau d'essais rendement/protéines pour caractériser le GPD des lignées. Par ailleurs des modèles de prédiction génomique et des études de génétique d'association ont été réalisés afin de faciliter la sélection de lignées d'intérêt et l'introgession d'allèles absents du compartiment élite.

## 2 Matériel et méthode

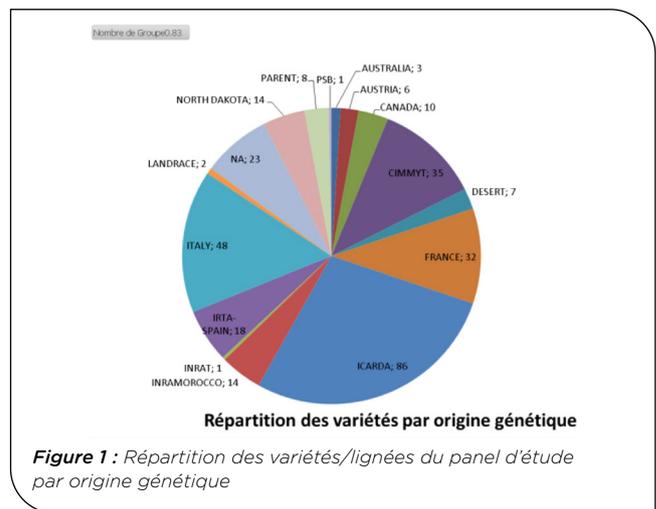
Un panel de 300 génotypes a été constitué à partir de 3 sources :

- le Panel EPO (UMR AGAP) (ci-après dénommé EPO) : 178 lignées issues de 14 cycles d'inter-croisements réalisés grâce à la présence d'un gène de stérilité mâle au sein d'une population présentant une base génétique large. Ces lignées ont été fixées et la stérilité mâle éliminée.
- Panel élite français (ci-après dénommé French) : 55 variétés, cela correspond à l'ensemble des lignées ayant été inscrites en France dans les 50 dernières années.
- Panel « Tuberosa » (ci-après dénommé Unibo) : ensemble de 216 lignées élites assemblées par le groupe de Roberto Tuberosa à l'Université de Bologne, ayant une fenêtre de date de floraison assez restreinte et adaptée aux conditions méditerranéennes.

La sélection des 300 génotypes est basée sur : la proximité génétique par analyse de similarité (IBS), la précocité (variation de 10 jours d'écart maximum), la hauteur (gamme 60 à 100 cm), le potentiel de rendement et ses composantes, les connaissances à dire d'expert pour le GPD.

Pour des raisons de taille de dispositif, le panel final évalué dans les essais rendement/protéines comprenait 238 variétés dont 77 variétés du panel Unibo, 51 variétés françaises et 110 lignées EPO. La répartition des origines génétiques des variétés sélectionnées du panel Tuberosa et variétés élites est représentée sur la figure 1.

Le panel a été évalué au champ entre 2016 et 2019. La campagne 2016-2017 était dédiée à la multiplication des semences des 300 génotypes sélectionnés pour assurer les essais rendement/protéines prévus ensuite et a été l'opportunité d'évaluer la tolérance vis-à-vis de la rouille jaune, la septoriose et la fusariose. Les années 2 et 3 ont été consacrées à la mise en place d'essais rendement/protéines sur 238 génotypes.



Ainsi 8 essais ont été conduits (Tableau 1). Les sites RAGT (Montbartier, 81), QUALPREST (Baignolet et Sancheville, 28) et SYNGENTA (Argelato, Nord Italie) ont été conduits en pluvial et en fertilisation azotée optimale, alors que dans les sites INRAE (Mauguio, 34) et ARVALIS (Gréoux-les-Bains, 04) qui sont des sites naturellement séchant, une modalité irriguée et une modalité pluviale ont été conduites. Deux répétitions ont été implantées dans chaque essai ainsi que 4 à 6 témoins de précocité.

	2016 -2017	2017 -2018	2017 -2018
<b>RAGT</b>	Multiplication : Rouille jaune et Septoriose	Rdt / Prot	Rdt / Prot
<b>Limagrain / QUALPREST</b>	Rouille jaune	Rdt / Prot	Rdt / Prot
<b>INRAE</b>			Rdt / Prot
<b>SYNGENTA</b>	Fusariose	Rdt / Prot	Rdt / Prot
<b>ARVALIS</b>			Rdt / Prot

**Tableau 1.** Récapitulatif des essais conduits sur le panel

Les notations réalisées sur chacun des sites sont: Rendement (GY), PMG (TKW), Poids spécifique (PS), Teneur en protéines (GPC), Indice de sédimentation SDS, Gluten Index et humide (WG), Mitadin (Yellow-Berry), Moucheture (Black-point), Hauteur, Epiaison (Z55). Spécifiquement, un suivi NDVI a été conduit sur les essais INRAE (Greenseeker), ARVALIS (Phénomobile) et SYNGENTA (Greenseeker). La teneur en protéines a été mesurée par l'INRAE par infrarouge pour tous les essais sauf pour celui d'ARVALIS 2019 pour lequel les échantillons ont été analysés par le laboratoire AUREA. Les taux de mitadin et moucheture ont été mesurés par SYNGENTA en Italie, le SDS et les mesures sur le Gluten par RAGT à Montbartier.

### ► Caractérisation des environnements

Afin d'identifier le type de stress subit dans chacun des sites, le modèle CHN (ARVALIS) a été utilisé. Les données météorologiques, les données sol (texture, réserve utile), de stades phénologiques, d'itinéraire cultural des essais ont été utilisées comme paramètres d'entrées du modèle afin de calculer des indices de stress hydriques et azotés tout au long du cycle de la culture.

### ► Méthode d'évaluation des essais, du réseau et calcul du GPD

Pour calculer une moyenne par variété dans le réseau d'essai, différentes étapes se sont succédé : 1 - Calcul des moyennes ajustées par essai ; 2 - Calcul du GPD par essai ; 3 - Calcul de moyennes ajustées sur le réseau.

#### 1 - Calcul des moyennes ajustées par essai

Les sites d'ARVALIS, RAGT, QUALPREST sont des dispositifs en bloc complet, bloc incomplet pour l'INRAe (un bloc avec 40 % des variétés). Par ailleurs, nous disposons des informations sur les caractéristiques du terrain notamment la réserve utile de chacune des microparcelles pour le site de Gréoux-les-bains. La réserve utile (RU) peut avoir un impact sur certains caractères tel que le rendement ou le taux de protéines. La sélection du meilleur modèle (avec ou sans RU) est basée sur le critère d'information d'Akaike ou AIC.

Le modèle testé est :

$$Y(ij) = g(i) + rep(j) + \varepsilon(ij)$$

avec  $i$ , l'indice des variétés et  $j$ , l'indice des répétitions.  $g(i)$  effet fixe du génome et  $rep(j)$  effet aléatoire des répétitions, avec ou sans la RU en effet fixe.

Le plan expérimental mis en place par Syngenta est un faux lattice carré ce qui implique la création d'un modèle sur mesure pour cet essai.

Le modèle utilisé est :

$$Y(ijkl) = g(i) + rep(j) + gp(k) + rep \times gp(jk) + rep \times gp \times incbloc(jkl) + \varepsilon(ijkl)$$

avec  $g$  l'effet fixe du génotype,  $rep$  l'effet fixe de la répétition,  $gp$  l'effet fixe du sous-bloc,  $rep \times gp$  l'interaction entre la répétition et le sous-bloc et enfin un effet aléatoire  $rep \times gp \times inc bloc$  avec « inc bloc » qui représente l'effet de chaque colonne dans la répétition.

Tous les caractères pour lesquels les postulats de normalité des erreurs et d'homoscédasticité ne sont pas vérifiés pour réaliser l'analyse de variance, ne sont pas pris en compte dans l'analyse. Des moyennes ajustées sont ainsi calculées sur chacun des traits par essais.

#### 2 - Calcul du GPD par essai

Pour calculer le GPD, la méthode de Oury *et al.*, 2007 a été utilisée. Cette méthode itérative permet de limiter l'effet de levier de certaines observations sur le calcul des paramètres de la droite de régression. Elle consiste à ajuster une droite de régression puis à mettre de côté les observations dont la valeur absolue des résidus standardisés est supérieure à seuil (ici, 1.96). Ces étapes sont répétées jusqu'à ce qu'aucun point ne présente un résidu standardisé supérieur au seuil. Les paramètres de la droite de régression sont alors déterminés et les résidus (GPD) sont calculés pour l'ensemble des observations.

#### 3 - Calcul des moyennes ajustées sur le réseau

Un modèle mixte a été utilisé avec en effet fixe le génotype et en effet aléatoire un effet essai, selon la formule :

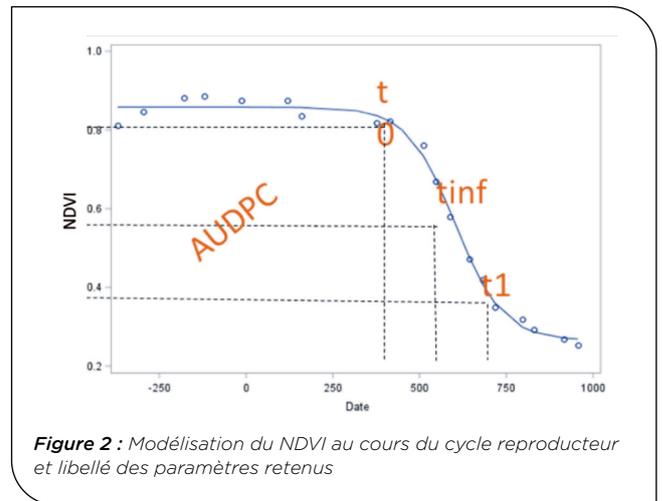
$$Y(ij) = g(i) + essai(j) + e$$

Une valeur par variété a ainsi été obtenue.

### ► Modélisation des données NDVI

Les données NDVI acquises par l'INRAe sont utilisées pour ajuster une courbe logistique afin d'en extraire les

paramètres indiqués sur la figure 2. Le paramètre  $t0$  est la valeur de NDVI au début de la sénescence,  $t1$  la valeur de NDVI à maturité et  $tinf$ , le point d'inflexion entre ces 2 bornes.



Les données NDVI acquises par Arvalis sont modélisées avec une courbe double logistique afin de prendre en compte la montaison.

### ► Génotypage du panel d'étude

Le panel a été génotypé à l'aide de la puce TaBW420K développée dans le cadre du projet français Breedwheat (Rimbert *et al.*, 2018). Une petite partie du panel a été génotypée avec la puce TaBW35K issue de la 420K puis imputée avec le logiciel BEAGLE. Une matrice d'apparentement a été calculée et utilisée dans le modèle de génétique d'association (Yu *et al.*, 2006). Les marqueurs utilisés en génétique d'association ont été sélectionnés de manière à ne garder que les plus informatifs, ce qui a conduit à éliminer : les marqueurs avec une MAF (« minor allele frequency ») inférieure à 5 %, les marqueurs avec un taux d'hétérozygotie supérieur à 20 % (Turner *et al.*, 2017) (Edae *et al.*, 2014). Ainsi, un set de 68 093 SNPs a été sélectionné.

### ► Analyse d'association génétique (GWAS)

La procédure de GWAS utilisée est similaire à celle proposée par (Segura *et al.*, 2012). L'association génétique entre le GPD et les autres traits mesurés et les marqueurs SNP de la matrice finale est réalisée. Le modèle statistique Q+K a été utilisé :

$$Y = \mu + X\beta + S\tau + Zg + \varepsilon$$

Avec  $Y$  le vecteur des valeurs phénotypiques,  $\mu$  la moyenne générale,  $g$  un vecteur d'effets aléatoires de variance  $K\sigma^2$  (avec  $K$  la matrice d'apparentement des variétés),  $\tau$  l'effet du SNP testé en effet fixe,  $\beta$  l'effet du contraste entre les différentes populations constituant le panel et  $\varepsilon$  la variance résiduelle. Dans ce modèle  $X$ ,  $Z$  et  $S$  sont des matrices de contrastes pour les composantes respectives du modèle.

Le seuil de significativité utilisé usuellement est calculé sur la méthode de correction de Bonferroni soit  $-\log_{(10)}(0.05 / 68\ 093) = 6$  qui est très stringente notamment sur les traits complexes étudiés dans ce projet, un seuil de 4 a donc été choisi comme cela a été fait par Rapp *et al.*, 2018. Les marqueurs SNP

significatifs pour chacun des traits ont été identifiés, de même que les QTL après l'analyse de déséquilibre de liaison (LD) entre marqueurs significatifs pour le même trait. Les effets et les proportions de variances expliquées des meilleurs QTL ont été estimés.

### ► Prédiction génomique (GS) des traits étudiés

Pour inclure et valoriser l'effet de tous les QTL y compris les QTL à effet mineurs (expliquant une proportion faible de variance phénotypique), une approche de prédiction génomique a été utilisée à l'aide du package R BGLR (Bayesian Generalized Linear Regression) (Perez and De Los Campos, 2014). Différents modèles ont été comparés et le meilleur modèle a été identifié par validation croisée. Une validation croisée en 10 ensembles sans remplacement a été utilisée, où neuf des dix ensembles ont été utilisés comme jeu d'entraînement pour construire le modèle et estimer les effets de marqueurs pour prédire les GEBV (genomic estimated breeding values) du dixième ensemble. La précision de chaque modèle a été estimée comme la moyenne des valeurs de capacité de prédiction.

## 3 Résultats

### ► Réseau d'essai

Les essais pépinières de la campagne 2016-2017 conduits sur 300 génotypes ont permis, au-delà de multiplier les semences, de collecter des notations maladie.

Les pépinières Rouille jaune (Verneuil/Limagrain), Septoriose (Montbartier/RAGT) et Fusariose (Saint Sauveur/Syngenta) ont subi une forte pression maladie et les symptômes se sont bien développés.

A l'inverse, la pression rouille brune et jaune était très faible à Montbartier / RAGT.

La variabilité de notations sur le panel permet d'identifier des lignées potentiellement tolérantes, mais une évaluation supplémentaire permettrait de confirmer cette caractérisation.

Concernant les essais Rendement / Protéines, les données d'itinéraire cultural collectées sur l'ensemble des essais ont été rassemblées dans des bases de données après validation des valeurs individuelles (Tableau 2).

### ► Caractérisation des environnements

Les indices de stress en Azote et en Eau ont été calculés pour chacun des sites. Ainsi 3 scénarii de stress ont été identifiés :

- Pas de stress : Montbartier 2018, Argelato 2019, Greoux IRR 2019 ;
- Stress Hydrique à Floraison : Baignolet 2018, Argelato 2018, Gréoux SEC 2019, Montbartier 2019 ;
- Stress Hydrique dès Montaison : Mauguio IRR et SEC 2019 .

Un clustering des essais sur la base de la typologie de stress subi a été effectué mais n'a pas permis de mettre en évidence un effet sur la variation du rendement observé dans les essais. L'analyse est donc poursuivie à l'échelle du réseau complet.

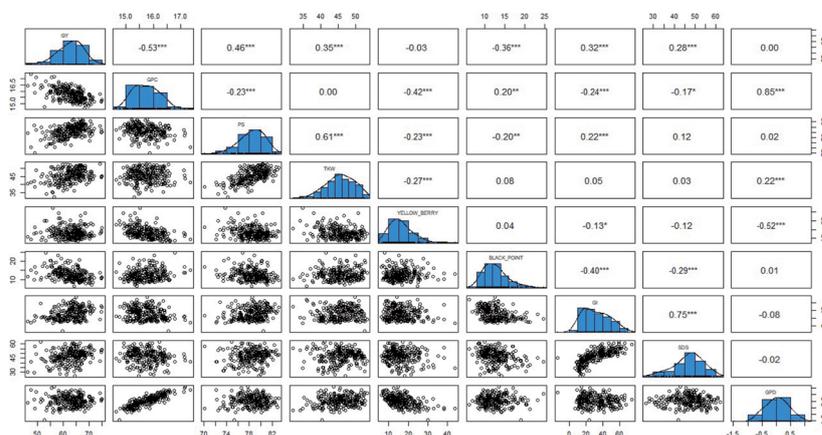
### ► Analyse du réseau d'essai

Selon les modèles d'analyse de variance décrits dans les matériels et méthodes, une moyenne ajustée a été calculée pour tous les traits mesurés pour l'ensemble du réseau d'essai. La corrélation négative rendement-protéines est significative avec un coefficient de 0,53 (Figure 3).

Récolte	Partenaire	Lieux	Semis	Parcelle récoltée (m <sup>2</sup> )	Fongicides	Pesticides	Fertilisation Azotée (U)	Irrigation (mm)
2018	Syngenta	Argelato	03/11/2017	6	oui	oui	260	0
2018	Qualprest	Baignolet	26/10/2017	13.6	oui	oui	246	35
2018	Ragt	Montbartier	26/10/2017	7.5	oui	oui	150	0
2019	Arvalis	greoux_irr	19/11/2018	11.88	oui	oui	250	210
2019	Arvalis	greoux_sec	19/11/2018	11.88	oui	oui	250	0
2019	Inrae	Mauguio_optimal	11/12/2018	5.6	oui	oui	240	60
2019	Inrae	Mauguio_limitant	11/12/2018	5.6	oui	oui	60	30
2019	Syngenta	Argelato	17/11/2018	6	oui	oui	130	0
2019	Qualprest	Sancheville	19/11/2018	13.6	oui	oui	246	35
2019	Ragt	Montbartier	29/10/2018	7.5	oui	oui	150.7	0

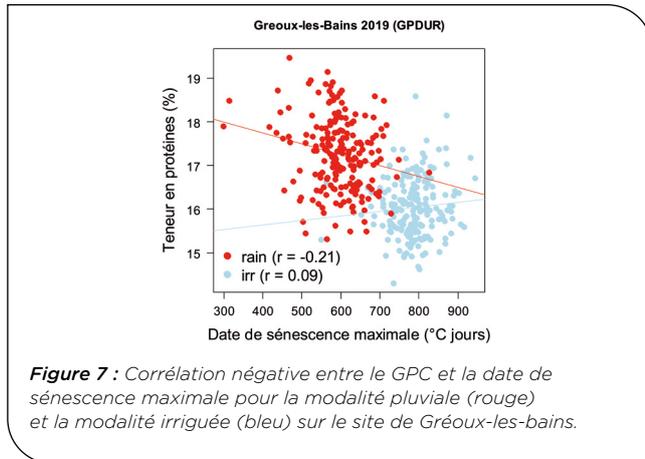
Tableau 2. Itinéraires culturaux des essais Rendement/Protéines du réseau

Figure 3 : Corrélation phénotypique entre le rendement (GY), la teneur en protéines (GPC), le GPD et les autres traits étudiés





Une relation négative significative a été identifiée entre la date de sénescence maximale et la teneur en protéines en condition de stress hydrique (Figure 7). Cela pourrait traduire que des variétés ayant une sénescence précoce ont tendance à favoriser l'accumulation de protéines.



**Figure 7 :** Corrélation négative entre le GPC et la date de sénescence maximale pour la modalité pluviale (rouge) et la modalité irriguée (bleu) sur le site de Gréoux-les-bains.

### ► Génétique d'association

Au total, 12 QTL ont été identifiés pour quatre traits parmi les onze étudiés. Pour le trait "Hauteur de la plante", 1 QTL (CFN2441455) identifié sur le chromosome 4B avec un LOD de 12.2 expliquant une variance phénotypique de 19 % a été identifié. Neuf QTL ont été identifiés pour le SDS sur les chromosomes 1A, 1B, 2B, 4A, 5B et 6B avec un LOD allant de 7.45 à 14.46 pour une variance phénotypique expliquée de près de 15 %. Un QTL a été identifié respectivement pour le PMG (CFN1520694 ; chr. 2A) et le Mitadin (CFN0559919 ; chr. 1B) avec un LOD de 10 et 6.1 pour une variance phénotypique expliquée respectivement de 21 % et 15 %. Pour les sept autres traits y compris le GY, GPD, GPC et Z<sub>55</sub>, aucun QTL n'a été identifié (Figure 8).

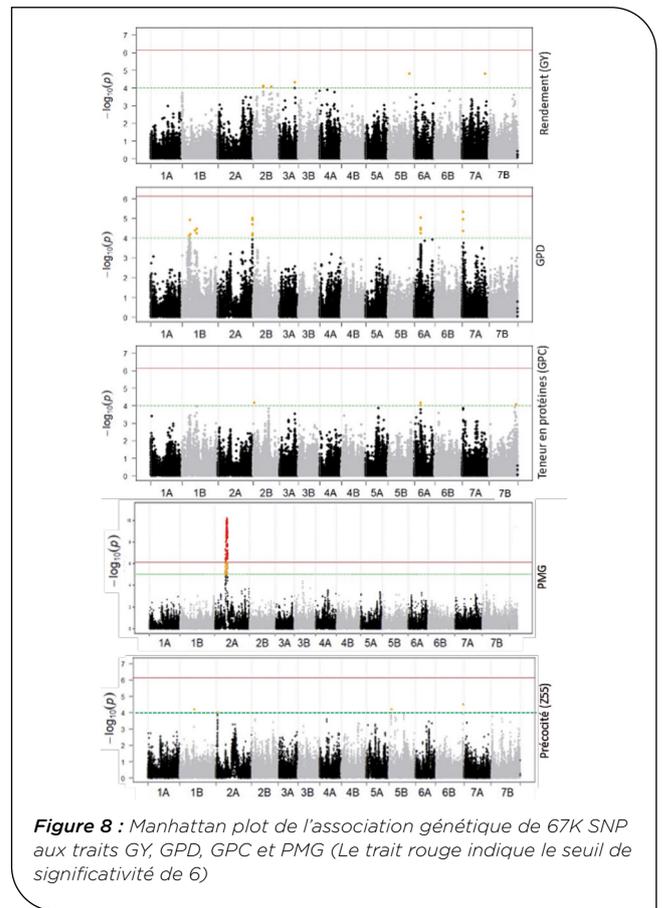
Au seuil de 4, de 3 à 26 QTL sont obtenus pour les traits étudiés. Quatre QTL sont obtenus pour le GPD avec des LOD de 4,91 à 5.33 sur les chromosomes 1B, 2A, 6A et 7A. Trois QTL pour le GPC et 5 pour le GY sont également obtenus. Un QTL associé à un seuil de 4.3 pour Z<sub>55</sub> sur le chromosome 2A (CFN0378505) colocalise avec le QTL majeur connu *Ppd-A1*.

L'effet allélique du meilleur QTL (LOD >4) estimé pour les traits étudiés, comme pour le GY, GPD, GPC, PMG et Z<sub>55</sub> (Figure 9) permet d'identifier pour ces QTL les allèles favorables au trait (à l'exception du GPC). En effet pour le PMG par exemple, il y a une différence significative entre les états alléliques A et G, avec l'allèle A en faveur des meilleurs poids de mille grains ( $46.4 \pm 3.45$  contre  $40.3 \pm 3.31$  pour le génotype GG).

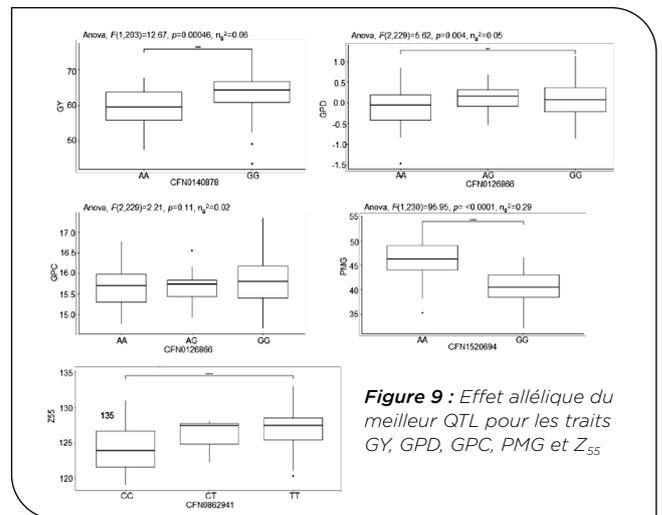
### ► Prédiction génomique

Les QTL identifiés au seuil de 4 ont été utilisés en effet fixe pour les 2 modèles l'incluant (QTL et GBLUP-QTL). La prédiction génomique des traits étudiés révèle des capacités prédictives moyennes comprises entre 0.61 et 0.78 avec le modèle GBLUP-QTL (Tableau 3).

L'intégration des marqueurs génétiques dans le modèle de prédiction génomique améliore ainsi les capacités prédictives pour tous les traits (Tableau 3).



**Figure 8 :** Manhattan plot de l'association génétique de 67K SNP aux traits GY, GPD, GPC et PMG (Le trait rouge indique le seuil de significativité de 6)



**Figure 9 :** Effet allélique du meilleur QTL pour les traits GY, GPD, GPC, PMG et Z<sub>55</sub>

Trait	$r_{(QTL)}$	$r_{(GBLUP)}$	$r_{(QTL+GBLUP)}$
Moucheture	0.44	0.43	0.61
Teneur en protéines (GPC)	0.21	0.52	0.63
GPD	0.34	0.45	0.63
Rendement (GY)	0.32	0.56	0.61
Hauteur	0.48	0.23	0.48
Poids spécifique (PS)	0.40	0.71	0.77
SDS	0.50	0.49	0.65
Poids de mille grains (PMG)	0.51	0.68	0.75
Gluten humide	0.60	0.46	0.64
Mitadin	0.55	0.37	0.63
Précocité (Z55)	0.62	0.72	0.78

**Tableau 3.** Capacité prédictive ( $r$ ) du GPD et des dix autres traits étudiés. Trois modèles QTL, GBLUP et QTL-GBLUP ont été testés.

## 4 Discussion

Cette étude avait pour objectif d'identifier de la variabilité génétique pour le caractère GPD pour l'amélioration conjointe du rendement et de la teneur en protéines du blé dur dans les programmes de sélection. Un panel constitué de génotypes de trois origines différentes (EPO, French et Unibo) a été évalué dans des environnements optimaux et sous stress hydrique sur 3 années et 5 sites (France et Italie), soit un total de 8 essais.

La mise en place du réseau d'essai a permis de générer des données suffisantes pour caractériser le GPD des variétés. Cette étude met en évidence que la sélection française est déjà efficace sur ce critère et qu'il reste compliqué d'identifier des géniteurs marquants pour ce caractère. Néanmoins les lignées EPO, parce qu'elles disposent d'un fond génétique ancestral et sûrement plus diversifié que celui des variétés inscrites, peuvent présenter un réel potentiel vis-à-vis de l'équilibre entre métabolismes azoté et carboné. Cela représente une source de variabilité génétique intéressante à intégrer dans les programmes de sélection.

Une étude complémentaire est nécessaire pour décortiquer les processus physiologiques qui sous-tendent la capacité de certains génotypes à mieux valoriser l'azote disponible sans pénaliser le métabolisme carboné. Le jeu de données généré dans le cadre de ce projet ne permet pas notamment de valider l'hypothèse de l'influence de l'azote absorbé post-floraison sur la GPD démontrée sur blé tendre.

Le choix de constituer un panel d'association diversifié de 300 accessions à partir de différentes origines génétiques (notamment EPO versus Unibo et French) est donc intéressant, mais peut par ailleurs générer quelques contraintes de structuration au sein du panel. Il a été nécessaire de prendre en compte cet effet de structuration et donc d'adapter le modèle d'analyse d'association génétique. Avec notre jeu de données, aucun QTL n'a été identifié pour le GPD pour le seuil de significativité de 6. Quatre QTL au seuil de LOD de 4 ont par ailleurs été trouvés et faiblement associés au trait GPD ; avec un LOD compris entre 4.9 et 5.33 sur les chromosomes 1B, 2A, 6A et 7A. Ce type de résultats a été déjà mentionné dans de précédentes études : Nigro *et al.* (2019) n'ont pu identifier que 4 QTL avec des associations faibles (entre 3 et 5.9) sur un panel de 240 accessions de blé dur évalué dans 7 environnements ; Rapp *et al.* (2018) quant à eux sur deux panels de 189 et 159 génotypes ont pu identifier respectivement 6 QTL sur cinq chromosomes (5A, 5B, 6A, 6B et 7A) et 5 QTL sur 4 chromosomes (2A, 3B, 4A et 7A). Ces résultats sont un peu attendus étant donné que les traits dérivés (ou secondaires) de traits primaires comme le GPD montrent généralement un pouvoir de détection de QTL plus faible. D'autres études ont signalé ce type de résultats sur d'autres traits dérivés (Rapp *et al.* 2018) ou encore sur d'autres cultures comme le maïs (Wang *et al.* 2012). Cependant pour les 4 QTL concernant la GPD, une proportion de variance phénotypique expliquée pouvant aller jusqu'à 11 % chacun ( $R^2$  compris entre 8.5 et 10.6) est observée, avec un effet de substitution d'allèles positif sur le GPD (allant de 0.15 à 0.29).

Au seuil de LOD de 4, 3 QTL ont été identifiés pour le GPC parmi lesquels le QTL CFN0126866 (6A) détecté également pour le GPD. Ce QTL explique une variance phénotypique de 7.4 % pour le GPC et 10.6 % pour le GPD avec un effet allélique positif sur les deux traits, ce qui implique qu'il pourrait augmenter le GPC sans effet négatif sur le rendement. Pour ce même QTL il y a une différence significative entre les états alléliques A et G (allèle G en faveur du GPD +) ; il existe une différence non significative entre la forme hétérozygote AG (allèle avec un effet meilleur que l'allèle G) et A et G. Les variétés French présentent une fréquence allélique proche des EPO pour l'allèle G (44 % contre 50 %) mais une fréquence allélique pour AG nettement meilleure que les EPO (38 % contre 1 %). Des observations similaires sont également faites par exemple pour le PMG (QTL CFN1445884) et d'autres traits. Ce qui confirme que les variétés françaises sont déjà assez efficaces sur le GPD, mais également sur d'autres traits analysés. Toutefois, la taille des sous-panel d'origine pourrait induire un biais d'interprétation étant donné que le sous-panel EPO est surreprésenté dans le panel d'étude.

Enfin, nous avons évalué un autre moyen de valoriser ces QTL à effets faibles. La prédiction génomique du GPD donne une valeur de capacité prédictive moyenne de 0.63 ; (Figure 10).

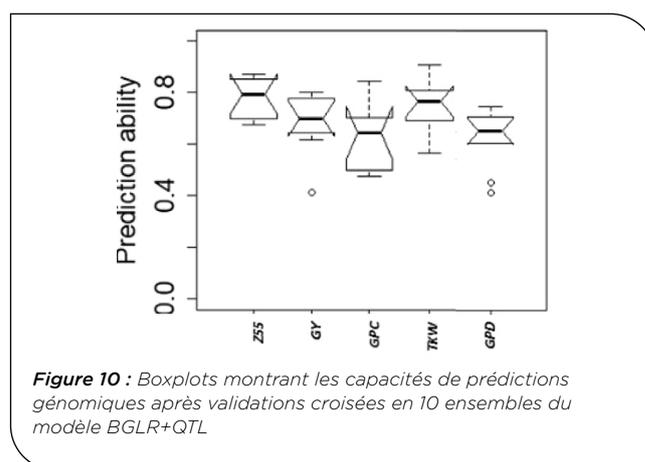


Figure 10 : Boxplots montrant les capacités de prédictions génomiques après validations croisées en 10 ensembles du modèle BGLR+QTL

Un niveau suffisamment intéressant pour son utilisation par les sélectionneurs pour une évaluation génétique précoce des jeunes plants dans les programmes de sélection. Pour les autres traits, les capacités prédictives plus élevées (> 0.75) ont été obtenues par exemple pour le PMG, la précocité et le poids spécifique, montrant une plus forte capacité prédictive.

Ces valeurs de capacités prédictives obtenues restent toutefois à affiner par l'utilisation de jeux de données complémentaires pour la calibration ou la validation des modèles.

## Annexe

Marqueurs significatifs ( $\text{LOD} \geq 4$ ) obtenus par l'association marqueur-trait pour le GPD et dix autres traits (modèle MLMM + K + Q) sur une collection de blé dur.

Chr. Chromosome ; LOD :  $-\log_{10}(\text{p-value de l'association génétique})$  ;  $R^2$  (%) : Proportion de la variance phénotypique expliquée par le QTL ;  $\alpha$ -effet : effet de la substitution d'allèle du QTL sur le trait.

Trait	Markers	Chr.	Position physique (bp)	Alleles	LOD	R2	$\alpha$ -effect	
Moucheture	CFN0600943	2A	5 692 141	G/T	4.38	0.07	-1.79	
	CFN0736436	4A	24 825 935	C/T	4.03	0.06	-0.78	
	CFN0768832	5A	706 492 258	C/T	4.03	0.09	-0.95	
	CFN0384675	5B	598 154 425	A/G	4.06	0.07	-1.74	
	CFN3365544	7A	68 853 295	T/C	4.25	0.08	-2.02	
	CFN0886687	7A	689 946 443	T/C	4.45	0.07	-1.55	
GPC	CFN0146921	2B	17 891 360	G/A	4.15	0.07	0.13	
	CFN0126866	6A	35 352 128	G/A	4.16	0.07	0.17	
	CFN0925222	7B	736 139 240	C/T	4.05	0.05	-0.20	
GPD	CFN0166791	1B	231 665 909	T/A	4.91	0.11	0.29	
	CFN1414835	2A	775 368 565	A/G	5.00	0.08	0.14	
	CFN0126866	6A	35 352 128	G/A	5.03	0.11	0.16	
	CFN0398866	7A	13 550 455	T/C	5.33	0.10	0.28	
GY	CFN0630403	2B	161 418 539	G/T	4.12	0.05	2.77	
	BOBWHITEC 18540351	2B	682 848 604	A/G	4.07	0.07	1.38	
	CFN0140878	3A	700 901 902	A/G	4.32	0.08	2.30	
	CFN2082330	3B	760 594 177	C/G	4.25	0.05	2.54	
	RAC875REPC 91630132	5B	622 555 946	A/G	4.81	0.07	1.61	
	CFN0885410	7A	699 297 937	A/C	4.81	0.07	2.64	
	Hauteur	CFN0639167	2B	175 458 307	C/T	4.45	0.08	-4.70
CFN2281473		4A	720 135 097	C/G	4.92	0.07	-3.49	
CFN2242431		4A	723 732 515	G/A	4.49	0.07	-2.07	
CFN0219360		4B	28 739 411	C/G	4.17	0.07	-3.12	
CFN2441455		4B	31 123 857	C/T	12.22	0.19	-5.67	
CFN2427933		4B	36 158 061	C/T	4.23	0.07	-2.44	
CFN2510590		5A	567 889 764	A/G	4.41	0.07	-2.90	
CFN0220091		5A	588 751 574	G/A	4.14	0.07	-3.41	
CFN0825318		6A	608 117 162	G/C	4.11	0.06	-3.15	
CFN0859003		6B	662 028 930	T/A	4.08	0.07	-4.88	
CFN0102075		7B	33 797 593	T/C	4.42	0.07	-3.47	
CFN3533081		7B	120 394 333	G/A	4.50	0.06	1.88	
CFN0917559		7B	639 278 318	T/C	4.35	0.06	-2.37	
PS		CFN0539874	1B	675 765 706	C/A	4.31	0.06	-0.67
		CFN1192478	1B	678 235 993	G/A	5.52	0.06	-0.62
	CFN0542607	1B	678 319 040	C/G	4.88	0.05	-0.82	
	CFN0534696	1B	687 071 879	T/C	5.70	0.07	0.58	
	CFN0602122	2A	82 130 364	C/T	4.00	0.05	0.57	
	CFN0636494	2B	133 019 317	T/C	4.28	0.05	0.53	
	CFN0609085	2B	742 172 008	A/G	4.81	0.05	-0.67	
	CFN0614490	2B	751 314 123	G/C	4.47	0.04	-0.73	
SDS	CFN0530094	1A	2 476 089	C/T	4.45	0.08	2.16	
	CONTIG24763566							
	BSO0085346	1A	7 184 769	A/T	7.45	0.14	4.06	
	CFN0500149	1A	534 544 595	T/C	4.00	0.08	2.68	
	CFN0354778	1B	730 263	C/T	4.55	0.08	3.26	
	CFN0550291	1B	5 054 319	T/C	11.72	0.15	5.12	
	CFN0338945	1B	6 791 669	T/C	14.46	0.28	4.78	
	RAC875C636241002	1B	11 859 931	A/G	4.90	0.05	3.43	
	IAAV2838	1B	41 642 534	A/G	5.21	0.11	2.52	
	KUKKRIC19181074	1B	548 968 557	A/G	4.46	0.07	-2.43	
	CFN0292064	1B	553 545 277	C/A	4.78	0.07	-2.46	
	CFN1137302	1B	554 581 714	T/C	8.16	0.14	2.96	
	CFN1155111	1B	557 418 957	A/G	4.22	0.06	2.20	
	CONTIG475478							
	1BSO0037246	1B	668 200 656	A/C	4.70	0.06	4.01	
	CFN0600533	2A	4 949 841	G/A	4.29	0.07	2.25	
	CFN0426975	2B	2 532 314	G/A	4.44	0.05	2.26	
	CFN1671235	2B	48 291 276	T/C	10.70	0.21	4.73	
	CFN1630005	2B	647 614 398	T/G	5.59	0.08	4.43	
	CFN2327938	4A	687 201 240	G/A	8.52	0.19	4.19	
	CFN0803022	5B	641 027 197	A/T	10.68	0.21	4.73	
	CFN2676300	5B	700 000 172	A/G	4.33	0.05	2.71	
	CFN0845828	6A	13 735 202	C/G	4.14	0.05	3.82	
	CFN0825310	6A	608 118 538	T/A	4.07	0.05	3.94	
	CFN0869301	6B	19 056 495	T/G	4.50	0.05	-3.54	
	CFN0294642	6B	22 107 347	T/C	5.39	0.09	-3.43	
	CFN0874001	6B	22 231 683	T/G	6.69	0.08	-5.19	
	CFN0294258	6B	28 059 431	A/G	7.43	0.11	5.93	
	PMG	CFN1520694	2A	149 394 582	G/A	9.99	0.21	2.64
		CFN0746355	4B	660 727 474	G/C	4.02	0.04	1.33
CFN0332040		6A	554 567 996	G/A	4.07	0.06	-1.68	
Gluten humide	CFN0515986	1A	592 312 894	T/A	4.04	0.05	-0.88	
	CFN0364249	1B	555 237 630	A/C	5.13	0.09	-0.86	
	CFN0533509	1B	615 126 866	T/G	4.88	0.09	0.84	
	CFN0608940	2B	610 021 362	G/A	4.40	0.09	-0.87	
	CFN0710183	3B	28 540 220	T/G	5.27	0.07	1.15	
	CFN0735095	4A	710 412 088	A/T	4.40	0.08	1.11	
Mitadin	BOBWHITEC 2314684	7A	448 098 049	A/G	4.84	0.09	1.32	
	CFN0559919	1B	212 662 522	T/G	6.04	0.15	-4.50	
	CFN0585285	2A	699 512 193	G/T	4.67	0.08	-4.50	
	CFN1676428	2B	73 068 593	T/C	5.69	0.10	-3.55	
	CFN0697901	3B	470 853 819	A/T	4.79	0.10	-3.57	
	CFN0191527	7A	15 678 656	G/A	4.46	0.09	-4.02	
	CFN0884952	7A	679 897 363	A/G	4.74	0.09	3.46	
CFN0922193	7B	743 476 186	G/C	4.06	0.08	-1.95		
Z55	CFN0545134	1B	390 384 518	A/G	4.34	0.05	0.90	
	CFN0378505	2A	36 691 989	A/G	4.28	0.05	-0.79	
	CFN0816949	5B	23 367 796	C/G	4.57	0.06	-0.84	
	CFN0862941	6B	712 270 064	C/T	4.56	0.09	0.85	

## Références bibliographiques

- Bogard, M., Allard, V., ..., Gate P. Le Gouis, J.** (2010). Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *Journal of experimental botany*.
- Bogard, M.,..., Ravel, C., Le Gouis, J.** (2011). Anthesis date mainly explained correlations between post-anthesis leaf senescence, grain yield, and grain protein concentration in a winter wheat population segregating for flowering time QTLs, 1-16.
- Buckler, E.S.T. and J.M. Thornsberry,** Plant molecular diversity and applications to genomics. *Curr Opin Plant Biol*, 2002. 5(2): p. 107-11.
- Comar, A., Burger, P., de Solan, B., Baret, F., Daumard, F., & Hanocq, J. F.** (2012). A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results. *Functional Plant Biology*.
- Cormier, F., Le Gouis, J., Dubreuil, P., Lafarge, S., & Praud, S.** (2014). A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat. *Theoretical and Applied Genetics*, 127(12), 2679-2693.
- Cormier, F.** (2015). Nitrogen use efficiency in wheat in bread wheat : breeding & gene discovery (Doctoral dissertation, Université Blaise Pascal-Clermont-Ferrand II).
- David, J., Holtz Y., Ranwez V., ...,Tavaud-Pirra M.** (2014). Genotyping by sequencing transcriptomes in an evolutionary pre-breeding durum wheat population. *Molecular Breeding*, 34(4) :1531-1548.
- Gaju, O., Allard, V., Martre, P., Le Gouis, J., Moreau, D., Bogard, M., ... Foulkes, M. J.** (2013). Nitrogen partitioning and remobilization in relation to leaf senescence, grain yield and grain nitrogen concentration in wheat cultivars. *Field Crops Research*.
- Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S, David J** (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Molecular biology and evolution* 24 (7):1506-1517.
- Kang, H. M. et al.** Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–54 (2010).
- Maccaferri, M., Sanguineti, M. C., ..., Tuberosa, R.** (2011). Association mapping in durum wheat grown across a broad range of water regimes. *Journal of experimental botany*, 62(2), 409–38.
- Maccaferri, M., Sanguineti, M. C., Mantovani, P., Czembor, J. H., Ezrati, S., & Tuberosa, R.** (2010). Association mapping of leaf rust response in durum wheat, 189–228.
- Maccaferri, M., Sanguineti, M. C., Natoli, V., Ortega, J. L. A., Salem, M. Ben, Bort, J., Tuberosa, R.** (2006). A panel of elite accessions of durum wheat (*Triticum durum* Desf.) suitable for association mapping studies. *Plant Genetic Resources: characterization and utilization*, 4(1), 79–85.
- Maccaferri M, Sanguineti M.C., Noli E., R Tuberosa** 2005. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding*
- Maccaferri, M., Ricci, A., Salvi, S., ..., Ammar, K.** (2015). A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant biotechnology journal*, 13(5), 648–663.
- Mackay, I. and W. Powell,** Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci*, 2007. 12(2): p. 57-63.
- Mangin, B., et al.,** Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* (Edinb), 2012. 108(3): p. 285-91.
- Nigro, D., Gadaleta, A., Mangini, G. et al.** (2019). Candidate genes and genome-wide association study of grain protein content and protein deviation in durum wheat. *Planta* 249, 1157–1175.
- Oury, F. X., & 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(1-2), 45-57.
- Perez P., de los Campos G.** (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198(2) :483-495.
- Rapp, M., Lein, V., Lacoudre, F., Lafferty, J., Müller, E., Vida, G., Leiser, W.L.** (2018) Simultaneous improvement of grain yield and protein content in durum wheat by different phenotypic indices and genomic selection. *Theor Appl Genet* 131(6):1315-1329.
- Ratti, C. et al.** 2006. Reaction of 111 cultivars of *Triticum durum* Desf. from some of the world's main genetic pools to soil-borne cereal mosaic virus. *Journal of Plant Diseases and Protection*, 113(4), 145-149.
- Royo, C., Maccaferri, M., Álvaro, F., Moragues, M., Corinna, M., Tuberosa, R., ... Villegas, D.** (2010). Understanding the relationships between genetic and phenotypic structures of a collection of elite durum wheat accessions, 119, 91-105.
- Segura, V. et al.** An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–30 (2012).
- Taulemesse, F.** (2015). Analyse écophysologique et génétique de l'absorption d'azote post-floraison chez le blé tendre en relation avec la concentration en protéines des grains (Doctoral dissertation, Université Blaise Pascal-Clermont-Ferrand II).
- Thuillet, A.C., et al.,** Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. *Genetics*, 2005. 169(3): p. 1589-99.
- Vilmus, I., Ecarnot, M., Verzelen, N., & Roumet, P.** (2014). Monitoring Nitrogen Leaf Resorption Kinetics by Near-Infrared Spectroscopy during Grain Filling in Durum Wheat in Different Nitrogen Availability Conditions. *Crop Science*, 54(1), 284-296.
- Wang, S., Wong, D., ..., Mastrangelo, A. M.** (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant biotechnology journal*, 12(6), 787-796.
- Yu, J., et al.,** A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, 2006. 38(2): p. 203-8.

# GPDUR : Génétique et Phénotypage du blé dur pour l'amélioration conjointe du rendement et de la teneur en protéines

Delphine HOURCADE<sup>1</sup>, Boris DEMENU<sup>1</sup>, Damien EYBALIN<sup>1</sup>, Matthieu BOGARD<sup>1</sup>, Thierry LEFEVRE<sup>2</sup>, Marion CHAPUS<sup>2</sup>, Carine ANDRE<sup>3</sup>, Andrea MASSI<sup>3</sup>, Jacques DAVID<sup>4</sup>, Pierre ROUMET<sup>4</sup>, Céline DUQUE<sup>5</sup>

- 1 - ARVALIS - 6 Chemin de la Cote Vieille - 31450 Bazège
  - 2 - RAGT 2n - 1740 Route du Canal - 82700 Montbartier
  - 3 - SYNGENTA - 12 Chemin de l'Hobit, 31790 Saint-Sauveur / Società Produttori Sementi S.p.A. Via Macero, 1 - 40050 Argelato (Bologna) - Italy
  - 4 - INRAe - Bâtiment Arcad, 10 rue Arthur Young, 34090 Montpellier
  - 5 - LIMAGRAIN Europe - Ferme de l'Etang - BP3 - 77390 Verneuil l'Etang
- \*Coordinateur : Delphine HOURCADE, d.hourcade@arvalis.fr

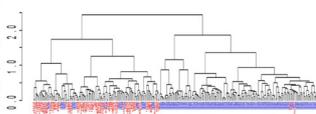
## Introduction :

Les exigences qualitatives des utilisateurs ont été à l'origine d'une structuration de la filière au sein de laquelle toutes les parties prenantes ont visé une production de haut niveau qualitatif, recherchant notamment de hauts niveaux de teneur en protéines. Malheureusement, la base génétique du matériel végétal cultivé en France, étroite, semble arriver à ses limites. Le taux de protéines est un critère majeur pour l'accès au marché, le progrès génétique en matière de rendement est de fait très réduit. Evaluer une plus large diversité dans les conditions de culture française, afin d'identifier des sources génétiques pouvant amener à rompre cette corrélation négative entre rendement et protéine semble donc essentiel. L'analyse de la GPD (Grain protein deviation) permet d'évaluer l'écart à la tendance globale de chaque variété à partir de la régression Rendement/Protéine. Il semble utile de mieux caractériser la diversité disponible aux sélectionneurs sur ce critère. Plusieurs sources de diversité étaient disponibles pour constituer un panel de diversité génétique élite, adaptée ou rapidement adaptable aux conditions françaises: le panel de diversité élite mondiale constitué par l'Université de Bologne et la population EPO (Evolutionary Pre-breeding Population) qui a été spécialement conçu dans le but d'intégrer un maximum de diversité élargie et originale par rapport au compartiment élite du blé dur. Il est issu de 14 cycles de brassage avec un taux d'allogamie estimé de 20% d'inter-croisements d'une population composite fondée par l'introgression d'accessions entre les sous espèces de blé dur *T. turgidum* sp. ayant divergé avant l'apparition de la forme moderne de blé dur (*T. turgidum durum*) correspondant aux différentes étapes évolutives. Le panel sera complété par un choix de variétés élites cultivées ou plus anciennes. L'ensemble de ces génotypes est phénotypé dans plusieurs sites, sur 3 ans puis génotypé pour être analysé avec des approches de génétique d'association et de prédiction génomique.

## Définition du panel

Différents critères ont permis de sélectionner les génotypes constitutifs du panel d'étude :

- Analyse de diversité basée sur les données de génotypage, puis sélection des génotypes représentant des groupes génétiques différents (Graphique 1).
- Précocité (variation de 10 jours maximum)
- Hauteur : gamme 60 à 100 cm
- Potentiel rendement
- Connaissance a priori GPD + ou GPD -



Graphique 1 : Classification hiérarchique de la population EPO (en bleu) et des variétés françaises (en rouge)

## Phénotypage du panel

Tableau 1 : Réseau d'essai du panel

	RAGT	Limagrain puis QUALPREST	INRAe	SYNGENTA	ARVALIS
2016-2017	Multiplication: Rouille jaune et Septoriose	Rouille Jaune		Fusariose	
2017-2018	Rendement /Protéines	Rendement /Protéines		Rendement /Protéines	
2018-2019	Rendement /Protéines	Rendement /Protéines	Rendement /Protéines	Rendement /Protéines	Rendement /Protéines

La campagne 2016-2017 a été consacrée à la multiplication des semences (300 variétés) en pépinière et à été l'opportunité d'évaluer la tolérance aux maladies. Les campagnes 2017-2018 et 2018-2019 ont été consacrées à la mise en place de 8 essais Rendement/Protéines sur 235 variétés.

- Les sites RAGT (Montbartier, 81), QUALPREST (Baignolet et Sancheville, 28) et SYNGENTA (Argelato, Nord Italie) sont adaptés en pluvial et fertilisation Azotée optimale
- Les sites INRAe (Mauguio, 34) et ARVALIS (Gréoux-les-Bains, 04) naturellement séchant, une modalité irriguée et une modalité pluviale sont conduites.
- 2 répétitions ont été implantées dans chaque essai ainsi que 4 à 6 témoins de précocité.
- Mesures : Rendement (GY), PMG (TKW), Poids spécifique, Teneur en protéines (GPC), Indice de sédimentation SDS, Gluten humide (WG), Mitadin (Yellow-Berry), Moucheture (Black-point), Hauteur, Epiaison (Z55). Spécifiquement, un suivi NDVI a été conduit sur les essais INRAe et Syngenta (Greenseeker) et ARVALIS (Phénomobile).

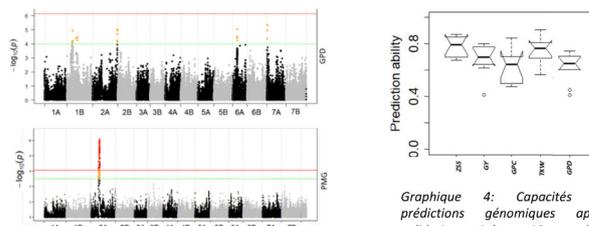


## Génétique d'association et prédiction génomique

Le panel a été génotypé à l'aide de la puce TaBW420K développée dans le cadre du projet français Breedwheat. Le set de 124923 marqueurs de départ est filtré selon plusieurs critères (MAF (« minor allele frequency ») < 5%, taux d'hétérozygotie > 20%). La matrice finale contient 68093 SNPs. La procédure de GWAS utilisée est similaire à celle proposée par (Segura et al., 2012). Le modèle statistique Q+K a été utilisé prenant en compte la matrice de Kinship et l'appartenance à l'une des 3 origines génétiques.

Le seuil de significativité utilisé usuellement est calculé sur la méthode de correction de Bonferroni soit  $-\log_{10}(0.05 / 68093) = 6$  mais il est très stringent notamment sur les traits complexes étudiés dans ce projet. Par conséquent un seuil de 4 est communément choisi pour GY, GPD et GPC.

Au total, 12 QTL sont identifiés pour quatre traits parmi les onze étudiés. Pour le trait "Hauteur de la plante", 1 QTL (CFN2441455) détecté sur le chromosome 4B avec un LOD de 12.2 expliquant une variance phénotypique de 19%. Neuf QTL sont identifiés pour le SDS sur les chromosomes 1A, 1B, 2B, 4A, 5B et 6B avec un LOD allant de 7.45 à 14.46 pour une variance phénotypique expliquée de près de 15%. Un QTL est identifié respectivement pour le PMG (CFN1520694, chr. 2A; Graphique 3) et le Mitadin (CFN0559919 ; chr. 1B) avec un LOD de 10 et 6.1 pour une variance phénotypique expliquée respectivement de 21% et 15%.



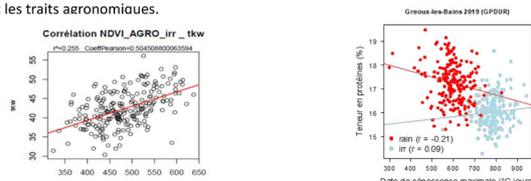
Graphique 3 : Manhattan plots issus des analyses de génétique d'association sur les traits GPD et TKW

Pour un seuil de LOD de 4, de 3 à 26 QTL sont observés. Quatre QTL sont obtenus pour le GPD avec des LOD de 4.91 à 5.33 sur les chromosomes 1B, 2A, 6A et 7A (Graphique 3). Trois QTL pour le GPC et 5 pour le GY sont également obtenus. Le QTL CFN0126866 (6A) est détecté pour le GPC et le GPD; avec une variance phénotypique expliquée de 7.4% et 10% respectivement, et un effet allélique positif sur les deux traits, impliquant qu'il pourrait augmenter le GPC sans effet négatif sur le rendement. Un QTL associé à un seuil de 4.3 pour Z55 sur le chromosome 2A colocalise avec un QTL majeur de précocité connu Ppd-A1.

Un autre moyen de valoriser ces QTL à effet faible a été évalué. La prédiction génomique du GPD donne une valeur de capacité prédictive moyenne de 0.63 (Graphique 4) et montre que l'intégration des marqueurs génétiques améliore ces capacités. Cette valeur de prédiction du GPD est suffisamment intéressante pour son utilisation par les sélectionneurs dans les programmes de sélection.

## Phénotypage haut débit

Concernant le site INRAe, une analyse d'ajustement de courbe logistique a été réalisée par l'INRAe sur les données NDVI pour en extraire différents paramètres. Des corrélations entre les valeurs de ces paramètres et les variables agronomiques ont été calculées. Elles sont nulles avec le PMG quel que soit le paramètre de la courbe étudié mais certaines sont intéressantes avec le GPD par exemple, c'est le cas pour 2 points, l'un correspondant au maximum de surface verte (t0) et l'autre au débit d'inflexion (tinf) pour la modalité irriguée (Graphique 5). En d'autres termes, une variété qui maintient un appareil végétatif vert le plus longtemps possible favorise le remplissage de ses grains. Pour la modalité stressée, les corrélations sont quasi inexistantes pour tous les paramètres et les traits agronomiques.



Graphique 5 : Relation linéaire entre les paramètres de courbes calculées sur les données NDVI et les traits agronomiques mesurés à Mauguio (INRAe) et Gréoux-les-bains (Arvalis)

Concernant le site de Gréoux-les-bains, un ajustement de courbe double-logistique est réalisé. Différents paramètres sont également extraits de ces courbes et sont mis en relation avec les variables agronomiques mesurées sur l'essai. Une relation négative significative existe entre la date de sénescence maximale et la teneur en protéines en condition de stress hydrique, traduisant qu'une sénescence précoce des variétés favorise une accumulation de protéines.

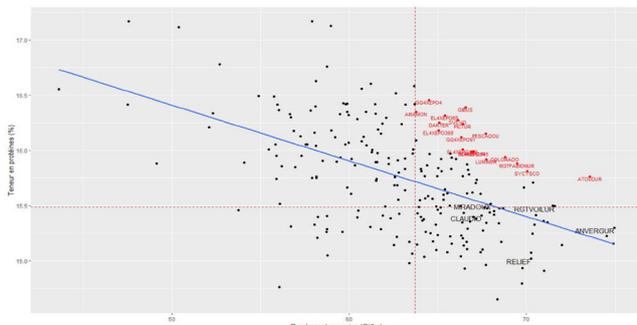
## Evaluation du critère GPD

Pour calculer le caractère GPD (grain protein deviation), une méthode itérative est utilisée afin de neutraliser l'effet des valeurs aberrantes sur la pente de la relation rendement-taux en protéines. A la fin, on obtient la régression linéaire qui explique le plus d'observations, on peut calculer les résidus de tous les individus.

Afin d'avoir des données moyennées à l'échelle du réseau d'essai, un modèle mixte a été utilisé avec en effet fixe le génotype et en effet aléatoire chaque essai, selon la formule:

$$Y_{(ij)} = g(i) + \text{essai}(j) \text{ Une valeur par variété et pour tout le réseau est ainsi obtenue.}$$

Pour identifier les génotypes qui s'écartent significativement de la régression globale, un seuil est fixé à 15% du meilleur rendement, à la teneur en protéine du meilleur témoin et  $GPD > 0.5$ .



Graphique 2 : Régression Rendement/Teneur en protéine à l'échelle du réseau d'essai

Sur le graphique 2, les génotypes positionnés au-delà de ces seuils et de la droite de régression sont considérés comme GPD+, car elles présentent un taux de protéines significativement supérieur pour un niveau rendement donné. Ces variétés ont un taux de protéines supérieur de l'ordre de 0.5 à 0.75 points de protéines par rapport à la variété référence du moment, RGT Voluir. Parmi ces génotypes, 6 variétés appartiennent au groupe EPO, 10 sont des variétés élites inscrites en France et 2 appartiennent au panel Unibo. Il n'existe pas vraiment de corrélation entre la nature des populations étudiées et leur écart à la relation « rendement-protéines ». On peut préciser toutefois que les lignées EPO présentent une variabilité plus importante des valeurs de GPD mais aussi les valeurs les plus élevées