

# [GS-qualité] Établissement d'un modèle de Sélection Génomique pour la Qualité Boulangère des blés

Bruno POUPARD<sup>1\*</sup>, Sophie BOUCHET<sup>2</sup>, François Guion<sup>3</sup>, Matthieu BOGARD<sup>4</sup>

1 - LIMAGRAIN EUROPE – CS50005 ST BEAUZIRE, 63360 GERZAT

2 - INRA - 147 rue de l'université, 75007 PARIS

3 - ANMF - 66 rue La Boétie, 75008 PARIS

4 - ARVALIS Institut du végétal - 3 rue Joseph et Marie Hackin, 75016 PARIS

\* *Coordinateur* : Bruno POUPARD, bruno.poupard@limagrain.com

## 1. Introduction et objectifs du projet

### 1.1 - Introduction

L'offre variétale de blé tendre actuelle répond globalement bien aux besoins des utilisateurs. C'est la conséquence d'une volonté de la filière traduite par les critères d'inscription des variétés au CTPS. Le niveau de tous les critères technologiques (Poids Spécifique, force boulangère et Note Totale de Panification) a augmenté depuis 15 ans (Méléard, 2010). Toutefois, le progrès génétique, notamment en matière de Note Totale de Panification, pourrait notablement être amélioré. En effet, la complexité de l'élaboration de cette note, le besoin important en grain qu'elle requiert, ainsi que son coût, impliquent que ce critère n'est évalué dans les programmes de sélection que dans les étapes finales, juste avant le positionnement pour l'inscription au catalogue.

Dans une publication fondatrice, Lande and Thompson (1990) ont introduit la théorie de la Sélection Assistée par Marqueurs (SAM). Un score moléculaire est calculé comme la somme des effets aux allèles des QTL (Quantitative Trait Loci) détectés. Son intérêt pour l'amélioration de caractères quantitatifs chez les plantes a été démontré expérimentalement (ex. Eathington *et al.* 2007, Blanc *et al.* 2008) et ces méthodes sont aujourd'hui largement utilisées par les grandes entreprises de sélection. Néanmoins, l'efficacité de cette SAM peut être limitée par la première étape de détection de QTL, dont la puissance est faible pour les QTL d'effets faibles avec les tailles usuelles des populations de sélection. En effet, pour des caractères complexes comme le rendement, dont on peut supposer que les QTLs à effets forts ont déjà été sélectionnés et fixés, le contrôle génétique est vraisemblablement assuré par un très grand nombre de gènes à effets très faibles (modèle « infinitésimal »). Comme ceux-ci sont difficiles à détecter, la variance expliquée par les QTL retenus dans le modèle est souvent faible. C'est pourquoi, les premiers, Whittaker *et al.* (2000) ont suggéré d'inclure dans le modèle de prédiction tous les marqueurs, sans étape préalable de détection de QTL. Comme le nombre de marqueurs est le plus souvent très supérieur au nombre d'observations, les méthodes classiques de régression multiple ne peuvent pas être utilisées. En conséquence, Whittaker *et al.* (2000) ont suggéré l'utilisation de régressions pénalisées (« Ridge Regression ») pour résoudre ce problème de sur-paramétrage. Cette approche, aujourd'hui communément nommée Sélection Génomique (GS) a été largement développée en génétique animale (Meuwissen *et al.* 2001), pour la prédiction des valeurs d'élevage (Breeding value). Cette estimation requiert des informations sur les génotypes aux marqueurs et les phénotypes dans une population de calibration pour calculer les estimations génomiques des valeurs génétiques (Genomic Estimation of Breeding Value, ou GEBV). Ces GEBV peuvent ensuite être estimées, à l'aide des seuls

marqueurs, dans n'importe quelle population cible, où la sélection peut donc être conduite sur les GEBV plutôt que sur les phénotypes.

Le principal intérêt de la Sélection Génomique est de pouvoir remplacer la sélection phénotypique quand l'évaluation du caractère est longue et/ou coûteuse. C'est bien le cas pour le test de panification, qui est cher (>100€), nécessite plusieurs kilos de grain, et n'est donc effectué que sur quelques lignées de générations avancées. L'efficacité relative de la GS par rapport à la sélection phénotypique dépend de la précision des prédictions (GEBV).

À notre connaissance, aucune étude n'a encore porté sur la valeur boulangère telle qu'estimée par le test BIPEA. Pour d'autres caractères liés à la qualité comme la teneur en protéine ou le rendement en farine, Heffner *et al.* (2011b) rapportent des précisions variant de 0.45 à 0.76. À ce jour, la seule publication sur la prédiction du test BIPEA à l'aide des gluténines (Oury *et al.* 2010) rapporte des précisions de prédiction (*r*, coefficient de corrélation) de l'ordre de 0.4, qui montent à 0.55 si on inclut les tests indirects (teneur en protéines, dureté, alvéographe) dans l'équation de prédiction.

Des études plus récentes concernent des paramètres physiques de la pâte qui ne sont pas utilisés dans le système d'inscription français (Michel *et al.*, 2017, 2018, Hayes *et al.*, 2017).

### 1.2 - Objectifs de ce projet

Notre hypothèse était que le développement d'un outil de Sélection Génomique permettrait de sélectionner pour la Note Totale de Panification plus en amont dans les programmes de sélection, et donc d'augmenter la réponse à la sélection pour ce caractère. Les objectifs de ce projet étaient les suivants :

- Construire une base de calibration de Sélection Génomique permettant de prédire la Note Totale de Panification, ainsi que ses composantes, et la teneur en protéines,
- Evaluer la précision de ces prédictions et déterminer par simulation les facteurs influençant cette précision,
- Évaluer la « portabilité » des prédictions d'un jeu de données à l'autre,
- Tester l'effet de l'inclusion dans les modèles GS des variables phénotypiques de l'alvéographe,
- Sur la base des résultats obtenus, proposer un protocole pour la construction d'une base de calibration de Sélection Génomique pour la qualité boulangère,
- Réaliser un test d'associations pour identifier les zones génomiques (QTL) ayant un effet sur la qualité boulangère et la teneur en protéine.

Afin de réaliser cette étude nous avons utilisé un panel de lignées de sélection du programme de Limagrain et utilisé les données d'essais post-inscription d'ARVALIS - Institut du végétal.

## 2. Matériel et méthode

### 2.1 - Matériel végétal du Panel Recherche

Le matériel végétal de ce panel était constitué de 2318 lignées de sélection, issues du programme de recherche de Limagrain dont l'objectif est la création de lignées commerciale pour le marché français. Les parents de ces croisements étaient soit des variétés commerciales françaises, soit des lignées de sélection performantes pour le marché français. Au total, 74% (1725) de ces lignées étaient des Haploïdes Doubles (HD), le reste (593) étant constitué des lignées obtenues par Single Seed Descend (SSD) au stade F3:5 (la dernière plante unique est une F3, mais les plantes cultivées sur les micro-parcelles d'essai sont des F5). L'année précédant leur mise en essai, ces lignées ont été sélectionnées sur des critères agronomiques (épiage, hauteur, résistances aux stress...) de manière à être adaptées au marché français, mais aucune sélection n'a été effectuée sur leur qualité ou leur rendement.

### 2.2 - Dispositif expérimental du Panel Recherche

Le dispositif expérimental initial prévoyait 12 environnements (4 lieux d'essais x 3 années), avec dans chaque environnement 192 lignées de sélection ainsi que 4 variétés témoins (observation sans répétition). Les conditions climatiques ayant eu un impact négatif sur la qualité boulangère pour certains environnements (germination sur pieds observée lors de la campagne 2013-2014, conditions climatiques très atypiques au Nord de la Loire lors de la campagne 2015-2016), nous avons dû adapter notre dispositif. Le dispositif final est décrit dans le Tableau 1. Chaque lignée de ce panel a donc été observée une seule fois, tandis que les témoins ont été répétés d'un environnement à l'autre et au sein de chaque environnement.

Lieu	Commune (Département)	2014	2015	2016	Total
CVCV	Chartainvilliers (Eure-et-Loir)		184 APACHE (2), AREZZO (2), CELLULE (2), RUBISKO (2)		184 (8)
CVTA	Aigrefeuille-d'Aunis (Charente-Maritime)	388 APACHE (2), AREZZO (2), SOLEHIO (2), SY MOISSON (2)	157 APACHE (2), AREZZO (2), CELLULE (2), RUBISKO (2)	396 AREZZO (2), CELLULE (2), RUBISKO (2), SOLEHIO (2), SY MOISSON (1)	941 (25)
CYRV	Revel (Haute-Garonne)	380 APACHE (2), AREZZO (2), SOLEHIO (2), SY MOISSON (2)		387 AREZZO (2), CELLULE (2), RUBISKO (2), SOLEHIO (1), SY MOISSON (1)	767 (16)
VRPE	Péronne (Somme)		209 APACHE (3), AREZZO (2), CELLULE (2), RUBISKO (2)		209 (9)
VRVR	Verneuil l'Étang (Seine-et-Marne)		217 APACHE (1), AREZZO (2), CELLULE (2), RUBISKO (2)		217 (7)
<b>Total</b>		<b>768 (16)</b>	<b>767 (32)</b>	<b>783 (17)</b>	<b>2318 (65)</b>

**Tableau 1** : Dispositif expérimental du Panel Recherche. Dans chaque cellule est indiqué le nombre de lignées observées (en gras), ainsi que les témoins présents et leur nombre de répétitions (entre parenthèses).

Dans chaque environnement, les lignées et les témoins ont été cultivés sur des micro-parcelles de 6m<sup>2</sup> selon un itinéraire technique non-limitant (fertilisation azotée, traitements fongiques...).

### 2.3 - Test de panification du Panel Recherche et transformation des données

Les tests de panification ont été réalisés par le laboratoire Qualtech sur des échantillons issus de la récolte des micro-

parcelles d'essai, selon le test BIPEA pain courant (NF V 03-716, Décembre 2008). L'Annexe 1 présente le détail des variables de ce test. Le teneur en protéine des farines a été mesurée par spectrométrie infrarouge.

Les notes composantes du test de panification BIPEA suivent un barème non linéaire : les scores -1, -4, -7, indiquent un déficit (-1 étant le plus extrême), tandis que les scores 7, 4 et 1 indiquent un excès (1 étant le plus fort). Le score de 10 traduit le niveau optimal de chaque variable. Pour des raisons arithmétiques, il a été nécessaire de linéariser selon la méthode présentée dans le Tableau 2.

SCORE BIPEA	SIGNIFICATION	RECODAGE ECHELLE 1-10
-1	Fort déficit	1
-4	Déficit	2.5
-7	Déficit léger	4
10	Optimal	5.5
7	Excès léger	7
4	Excès	8.5
1	Fort excès	10

**Tableau 2** : Recodage des scores du test BIPEA. Exemple : le score BIPEA de -7, qui correspond à un déficit léger par rapport à l'optimum, est recodé en 4. Ainsi, sur la nouvelle échelle, plus la valeur augmente et plus on passe d'un déficit à un excès.

### 2.4 - Matériel végétal du Panel Variétal et dispositif expérimental

Un panel de 192 variétés commerciales (« Panel Variétal ») a aussi été utilisé dans ce projet. Il s'agit de variétés étudiées dans le cadre de l'activité d'évaluation post-inscription d'ARVALIS - Institut du végétal. Ce panel comprend 3 des 6 témoins du Panel Recherche (APACHE, ARREZO et SOLEHIO). Les noms de variétés de ce panel sont présentés en Annexe 2.

Le dispositif expérimental de ce panel est fortement déséquilibré et les conditions environnementales et de culture sont extrêmement diverses. Les données de ce panel ont été acquises entre 2003 et 2013 avec des variétés communes entre les années. En moyenne, 57 variétés ont été testées par an. Le jeu de données comporte 158 lieux de récolte différents avec en moyenne 13 variétés par lieu. Les modalités de culture testées dans ces essais étaient principalement le niveau de fertilisation azotée et l'irrigation (essai pluvial versus irrigué). Le Tableau 3 donne plus de détails sur ce dispositif. Au final, les variétés ont été observées en moyenne 20 fois (médiane à 15), et sur 1 à 10 environnements (3.6 en moyenne).

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
2003	33										
2004	30	65									
2005	15	49	63								
2006	6	35	43	81							
2007	5	19	22	51	79						
2008	5	18	16	35	59	83					
2009	8	19	19	33	54	68	106				
2010	6	17	15	22	27	38	68	71			
2011	7	16	16	24	25	18	45	44	51		
2012	4	10	11	19	22	17	23	21	25	27	
2013	1	8	10	17	20	16	21	20	19	19	23

**Tableau 3** : Occurrence des variétés testées en fonction des années pour la Panel Variétal. Les chiffres représentent le nombre de variétés communes entre les deux années (ligne, colonne). Les chiffres sur la diagonale représentent les variétés présentes dans l'année correspondante.

Les caractères mesurés sur ce panel sont les variables d'alvéographe (W, G, P, L, le), les notes de panification et leurs

composantes, le taux de protéines, la dureté, le taux d'hydratation et le volume du pain. Les notes de panification sont attribuées selon la norme NF V03-716 en vigueur lors de la réalisation du test. Toutes les mesures ont été effectuées au laboratoire d'ARVALIS Institut du végétal de Boigneville (91).

Les scores du test BIPEA ont été linéarisés selon la méthode présentée dans le paragraphe 2.3

## 2.5 - Ajustement des données phénotypiques

Pour les 2 panels, les données phénotypiques ont été ajustées de manière à décomposer les effets génétiques des effets environnementaux. Pour ce faire, les différents facteurs environnementaux (facteurs « année » et « lieu » pour le Panel Recherche, facteurs « année », « conduite » et « lieu » pour le Panel Variétal) ont été combinés pour former un facteur « Environnement ».

Puis les moyennes ajustées ont été calculées via un modèle mixte en considérant les facteurs « Génotype » et « Environnement » comme aléatoires. Le modèle pour la variété  $i$  et l'essai  $k$  s'écrit ainsi :

$$Y_{ik} = \mu + G_i + E_k + \varepsilon_{ik}$$

Avec  $Y_{ik}$  la valeur phénotypique du Génotype  $i$  dans l'Environnement  $K$ ,  $\mu$  la moyenne générale du trait étudié,  $G$  l'effet Génotype,  $E$  le facteur Environnement et  $\varepsilon$  la résiduelle.

L'héritabilité ( $h^2$ ) a été calculée de la façon suivante :

$$h^2 = \frac{VG}{VG + VR}$$

avec  $VG$  la variance génétique et  $VR$  la variance résiduelle.

## 2.6 - Génotypage et nettoyage des données

L'ensemble du matériel (2318 lignées de sélection du Panel Recherche, 192 variétés du Panel Variétal, 3 témoins non compris dans le Panel Variétal) a été analysé au laboratoire de Génotypage de Limagrain (Chappes, 63), au moyen d'une puce de haute densité développée par Limagrain et comprenant 14592 marqueurs SNP. 97.8% de ces marqueurs ont une position de cartographie génétique (carte consensus interne à Limagrain, la répartition des SNP sur les chromosomes est présentée dans le Tableau 4).

Chromosome	A	B	D	Inconnu
1	876	948	251	
2	742	1034	285	
3	891	1150	337	
4	659	595	307	
5	945	1038	286	
6	577	859	221	
7	1004	990	276	
Inconnu				321
<b>Total</b>	<b>5694</b>	<b>6614</b>	<b>1963</b>	<b>321</b>

Tableau 4 : Répartition des 14592 marqueurs SNP utilisés sur le génome

Les données de génotypage ont été nettoyées de la façon suivante : les marqueurs monomorphes, ayant un taux de données manquantes supérieur à 30%, un taux d'hétérozygotie supérieur à 10% ou une fréquence de l'allèle minoritaire inférieure à 5% ont été écartés. Les individus ayant plus de 10% de données manquantes ont aussi été écartés.

## 2.7 - Modèle de Sélection Génomique, cross validation et portabilité des prédictions

Le modèle de prédiction génomique utilisé dans notre étude est un GBLUP (Meuwissen *et al.*, 2001) qui s'écrit de la façon suivante :

$$Y_j = \mu + G_j + \varepsilon_j$$

Avec  $Y_j$  la moyenne ajustée du génotype  $i$ ,  $\mu$  la moyenne générale,  $G$  l'effet Génétique avec  $G \sim N(0, K\sigma^2_g)$ ,  $K$  la matrice d'apparentement et  $\varepsilon$  la résiduelle.

Les calculs ont été effectués à l'aide de la fonction kin.blup du package 'rrBLUP' (Endelman, 2011). La matrice d'apparentement a été calculée avec la fonction A.mat du même package. Cette fonction calcule une matrice d'apparentement qui estime une IBS (Identity By State, Endelman and Jannink, 2012).

Les capacités prédictives des modèles GS ont été estimées par validation croisée de type k-fold avec  $k=10$ . Cette procédure a consisté à séparer le jeu de données en  $k$  échantillons. Le modèle a été calibré sur  $k-1$  échantillons puis les valeurs prédites pour l'échantillon  $k$  ont été calculées et comparées aux valeurs observées. Pour chaque échantillon  $k$ , le coefficient de corrélation entre valeurs prédites et observées a été calculé. La capacité prédictive de chaque modèle a été définie comme la moyenne des coefficients de corrélations des  $k$  échantillons.

La « portabilité » des équations de prédiction génomique obtenues sur ces 2 panels a été évaluée en réalisant des prédictions croisées. Autrement dit, les moyennes ajustées du Panel Recherche ont été prédites à l'aide du modèle calibré sur les données du Panel Variétal et vice versa.

## 2.8 - Influence des caractéristiques de la base de calibration et utilisation de covariables

L'influence du nombre de marqueurs pris en compte dans la calibration, ainsi que l'influence du nombre de génotypes, ont été testées sur le Panel Recherche. Pour cela, un sous échantillon de taille croissante (Nombre de Marqueurs = 50, 100, 1000, 5000, 8000, 1000 et Nombre de Génotypes = 50, 250, 500, 1000, 1500, 2000) a été tiré aléatoirement et une cross-validation a été effectuée selon la procédure définie au paragraphe 2.7. Cette procédure a été répétée 30 fois de manière à calculer la précision moyenne ainsi que son écart-type.

L'influence du nombre d'observations par génotype, quant à elle, a été testée sur le Panel Variétal selon la même procédure en prenant Nombre d'Observations = 1, 2, 3, 4, 5. Pour chaque valeur d'Observation la cross-validation a été itérée 10 fois.

L'intérêt d'inclure dans les modèles de prédiction génomique des covariables phénotypiques liées à la qualité a également été testé. Les covariables utilisées sont celles mesurées par l'alvéographe. Pour ce faire, les variables ont été ajoutées au modèle GBLUP décrit dans le paragraphe 2.7, et une procédure de cross-validation similaire a été utilisée. Les résultats obtenus ont été comparés aux résultats du GBLUP ainsi qu'aux résultats d'un modèle de régression linéaire (variable BIPEA =  $f(\text{variable alvéographe})$ ). L'effet de ces covariables n'a pu être testé que sur le Panel Variétal car elles n'ont pas été mesurées sur le Panel Recherche.

## 2.9 - Tests d'associations

Un test d'associations a été réalisé pour l'ensemble des variables sur chacun des 2 panels. La méthode utilisée correspondait à la méthode MLM proposée par Segura *et al.* (2012). Il s'agit d'une méthode itérative consistant à tester pour chaque SNP le modèle suivant :

$$Y = \mu + \alpha X + \beta SNP + \gamma G + \varepsilon$$

Avec  $Y$  la valeur phénotypique (moyenne ajustée),  $\mu$  la moyenne générale,  $\alpha X$  l'effet des SNP inclus en effets fixes,  $\beta SNP$  l'effet du SNP testé,  $G$  l'effet du fond génétique avec  $G \sim N(0, K\sigma^2_g)$ ,  $K$  la matrice d'apparentement et  $\varepsilon$  la résiduelle.

A chaque itération, le marqueur le plus significatif est inclus en effet fixe dans le modèle et la procédure se termine lorsque plus aucun marqueur ne présente une p-value inférieure ou égale à un seuil prédéfini (0.0001 dans le cas présent).

### 3. Résultats - Discussion

#### 3.1 - Données moléculaires et structuration

Après nettoyage des données de génotypage, le Panel Variétal comprend 188 variétés et 10482 marqueurs, et le Panel Recherche 2300 lignées et 10759 marqueurs.

La Figure 1 présente les résultats de l'Analyse en Composantes Principales réalisée sur les résultats de génotypage.

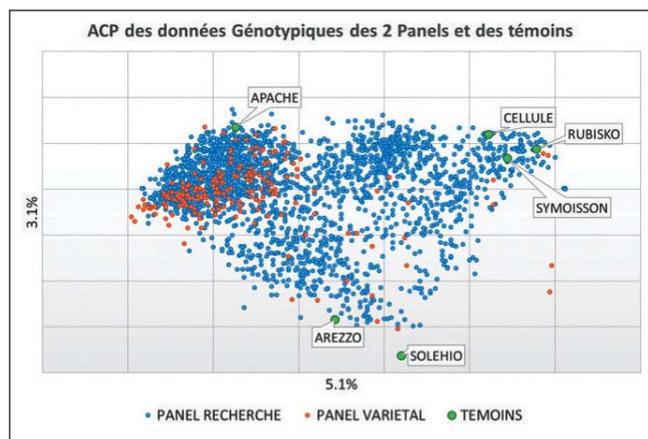


Figure 1 : ACP sur les données Génétiques

Il apparaît que les 2 panels sont peu structurés (Variance expliquée par les 2 premiers axes : 8.2%). Par ailleurs, le Panel Recherche semble bien recouvrir le Panel Variétal. L'inverse n'est cependant pas totalement vrai puisque l'on constate que la majorité des variétés de ce panel se situent dans la partie gauche du graphique

(proche du témoin APACHE). Les témoins quant à eux semblent se positionner dans ces nuages de points, mais en périphérie.

#### 3.2 - Données Phénotypiques

La Figure 2 montre que les 2 panels ont des performances moyennes très similaires pour la teneur en protéines et les Notes de Pâte, de Pain, de Mie et de Panification Totale. Cette dernière tourne autour de 250 sur 300 pour les 2 panels, avec de façon surprenante une variabilité moins importante sur le Panel Recherche, pourtant non sélectionné sur ce critère.

La Note de Pain, avec des valeurs moyennes comprises entre 65 et 68 pour les 2 panels, est la composante la plus faible et la plus variable de la Note Totale de Panification.

La Note de Pâte, dont la moyenne tourne de 85 pour les 2 panels, est la deuxième composante impactant le plus la Note Totale de Panification. Il est à noter que les variables entrant dans le calcul de la Note de Pâte présentent des corrélations fortes, notamment les variables liées à l'extensibilité, fortement corrélées entre elles et négativement corrélées aux variables d'élasticité (résultats non montrés).

La Note de Mie présente des valeurs très fortes (>95) et peu variables, de même que ses composantes (résultats non montrés). Ces variables ont donc souvent été écartées des analyses suivantes, à part la couleur de la mie (COUM). Un certain nombre d'autres variables (Collant au façonnage et à la mise au four, déchirement à l'apprêt...), elles aussi, ne présentent pas ou peu de variabilité et ont été écartées des analyses suivantes.

L'importance des effets environnementaux n'a pas été étudiée en détail, mais il est rappelé que nous avons écarté 4 environnements de cette étude à cause de conditions climatiques impactant négativement et de façon non désirée les résultats de panification. Pour le Panel Recherche, le Tableau 5 permet de voir que la Note Totale de Panification était au-dessus de 250 pour tous les environnements sauf deux où elle chute fortement (Aigrefeuille-d'Aunis 2015 = 238, Revel 2016 = 224) du fait de faibles Notes de Pain.

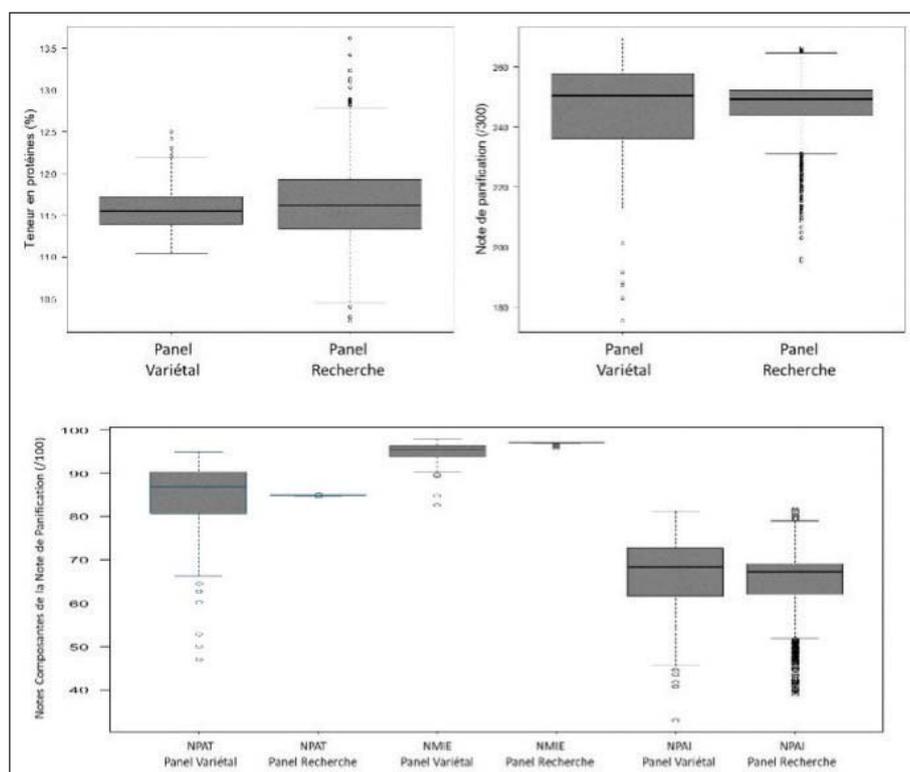


Figure 2 : Comparaison des distributions de la teneur en protéines, de la note de pain (NPAI), de mie (NMIE), de pâte (NPAT) et de la note totale de panification entre les panels Variétal et Recherche.

ANNEE	LIEU	PROT	VOL	NPAT	NPAI	NMIE	NPANI
2014	CVTA	10.80 (0.77)	1561 (193)	80.2 (9.6)	69.7 (13.0)	99.0 (3.7)	249.3 (19.6)
2014	CYRV	11.67 (0.85)	1617 (202)	81.5 (10.1)	72.1 (12.5)	99.1 (3.5)	253.1 (19.1)
2015	CVCV	11.53 (0.65)	1519 (143)	86.7 (12.0)	68.3 (17.9)	95.0 (4.4)	250.3 (28.6)
2015	CVTA	12.46 (0.74)	1501 (143)	84.4 (15.4)	58.0 (21.1)	95.6 (5.0)	238.5 (35.7)
2015	VRPE	10.99 (0.71)	1453 (160)	86.7 (11.4)	67.3 (16.4)	96.7 (4.1)	251.0 (26.9)
2015	VRVR	11.32 (0.66)	1511 (120)	87.8 (10.1)	70.6 (13.7)	96.0 (4.3)	254.7 (20.4)
2016	CVTA	11.72 (0.65)	1435 (131)	89.7 (10.0)	67.9 (16.5)	97.9 (3.4)	255.4 (23.7)
2016	CYRV	12.57 (0.71)	1412 (162)	81.1 (17.4)	46.8 (26.2)	96.6 (3.7)	224.4 (40.9)
<b>GLOBAL</b>		<b>11.63 (0.94)</b>	<b>1502 (179)</b>	<b>84.3 (12.7)</b>	<b>64.9 (19.7)</b>	<b>97.4 (4.1)</b>	<b>246.8 (29.5)</b>

**Tableau 5 :** Moyenne et écartype (entre parenthèses) pour chaque Environnement du Panel Recherche. Les variables présentées sont la Teneur en Protéine (%), PROT), le Volume (cm3, VOL), la Note de Pâte (NPAT), la Note de Pain (NPAI), la Note de Mie (NMIE) et la Note Totale de Panification (NPANI).

Quant aux interactions Génotype x Environnement, elles n'ont pas été formellement étudiées. Cependant, leur importance par rapport à la Variance Génétique peut être approchée via le calcul d'héritabilité effectué sur le Panel Variétal (voir Tableau 6). Si les Notes de Panification, la Teneur de Protéines et le Volume ont des héritabilités similaires (autour de 0.4), les variables affectant les Notes de Mie, de Pâte et de Pain ont des héritabilités plus faibles (autour de 0.3), bien plus faibles que les héritabilités des variables de l'alvéographe (0.7).

### 3.3 - Précision des modèles de GS

Les résultats des validations croisées sur les 2 panels sont présentés dans le Tableau 6. Pour le Panel Variétal, la précision des prédictions pour les critères de panification varie de -0.01 pour ELAM (élasticité de la mie) à 0.56 pour COUM (couleur de la mie), avec une précision moyenne de  $r=0.21$ , ce qui est globalement faible. La précision est en revanche meilleure pour les variables de l'alvéographe ( $r$  moyen = 0.41). Comme attendu, la précision de prédiction au carré ( $r^2$ ) d'un caractère est toujours inférieure à son héritabilité.

Pour le Panel Recherche, la précision des prédictions pour les critères de panification varie de 0.13 pour SOUM (souplesse de la mie) à 0.56 pour EXTF (extensibilité au façonnage), avec une moyenne à  $r=0.31$ , ce qui est meilleur que pour le Panel Variétal mais reste faible.

La teneur en protéines et le Volume sont prédits avec des précisions similaires sur les 2 panels ( $r \sim 0.4$ ), tandis que la prédiction de la Note Totale de Panification est nettement meilleure avec le Panel Recherche (0.35) qu'avec le Panel Variétal (0.14). Il est à noter que les variables liées à l'extensibilité et à l'élasticité au pétrissage et au façonnage (EXTP, ELAP, EXTF, ELAF) ainsi qu'à la couleur de la mie (COUM) semblent plutôt bien prédites par rapport aux autres variables du test de panification.

### 3.4 - Portabilité des modèles de Sélection Génomique

Les résultats montrent que les capacités prédictives des modèles calibrés sur le Panel Variétal sont faibles ( $-0.11 < r < 0.13$ ). Elles sont, en revanche, plus élevées pour les modèles calibrés sur le Panel Recherche ( $-0.28 < r < 0.43$ , voir le Tableau 7). Les prédictions obtenues avec les modèles calibrés sur le Panel Recherche apparaissent donc plus robustes bien que le niveau de précision soit relativement modéré. Ces différences sont à mettre en relation avec la taille des jeux de données de calibration (2300 lignées pour le Panel Recherche contre 190 variétés inscrites pour le Panel Variétal), facteur connu comme ayant un impact significatif sur les capacités prédictives des modèles et leur robustesse. Par ailleurs, bien que le niveau de précision soit relativement modéré, ces modèles peuvent néanmoins trouver une application pratique en sélection étant donné que les caractères analysés résultent de tests relativement coûteux et qui ne peuvent être mis en œuvre que tardivement dans les schémas de sélection.

Catégorie	Variables	Héritabilité Panel Variétal	Validation croisée Panel Variétal	Validation croisée Panel Recherche
Variables test BIPEA	Note de panification	0.41	0.14	0.35
	Note de mie	0.30	0.09	
	COUM	0.52	0.56	0.50
	SOUM	0.24	0.00	0.13
	ELAM	0.11	-0.01	
	EPAL	0.37	0.12	
	Note de pâte	0.46	0.16	0.32
	LISP	0.37	0.26	0.32
	COLP	0.45	0.26	0.33
	EXTP	0.31	0.32	0.48
	ELAP	0.23	0.16	0.45
	RELP	0.40	0.14	0.24
	REPO	0.34	0.15	
	EXTF	0.61	0.43	0.56
	DECF	0.22	0.24	0.15
	ELAF	0.43	0.35	0.51
	ACTP	0.34	0.17	0.23
	DECA	0.24	0.11	
	COMF	0.22	0.18	
	TENU	0.35	0.17	0.23
	COLF			0.22
	Note de pain	0.33	0.22	0.35
	SECT	0.45	0.34	0.30
	COUP	0.39	0.23	0.23
	DVCL	0.33	0.11	0.31
	RECL	0.15	0.14	0.16
	DECL	0.24	0.38	0.21
	Variables alvéographe	W	0.68	0.37
G		0.63	0.39	
P		0.79	0.42	
PL		0.67	0.35	
L		0.60	0.40	
le		0.79	0.53	
Autres variables	Durété	0.85	0.25	
	Hydratation	0.51	0.32	
	Volume	0.41	0.41	0.37
	Protéine (%)	0.37	0.41	0.43

**Tableau 6 :** Héritabilité des variables mesurées sur le Panel Variétal et capacités prédictives ( $r$ ) estimées par validation croisée des modèles de sélection génomique

Variable	Calibration Panel Variétal	Calibration Panel Recherche
Teneur en protéines	0.13	0.38
Note totale de panification	0.07	0.26
Note de pain	0.08	0.33
Note de pâte	0.11	0.28
Volume	0.16	0.43
LISP	-0.11	-0.28
COLP	0.09	0.14
EXTP	0.05	0.18
ELAP	0.08	0.17
EXTF	0.09	0.27
DECF	0.1	0.25
DVCL	0.07	0.27
DECL	0.08	0.28
COUM	0.04	0.08

**Tableau 7 :** Portabilité des calibrations GS. La capacité prédictive des équations de prédiction génomique calibrées sur le Panel Variétal a été évaluée en calculant la corrélation ( $r$ ) entre valeurs prédites et observées sur le Panel Recherche, et vice versa

### 3.5 - Influence des caractéristiques de la base de calibration sur la précision des modèles

Les Figures 3 et 4 illustrent l'effet du nombre de génotypes et de marqueurs sur la précision ( $r$ ) des prédictions dans le Panel Recherche. On observe sur la Figure 3 que plus on augmente le nombre de génotypes dans la base de calibration, plus les prédictions sont précises. Il semble que même avec 2000 génotypes dans la base on n'ait pas encore atteint de plateau dans les prédictions.

A contrario, la Figure 4 montre que l'on atteint un plateau dans les prédictions avec 2000 marqueurs pour la Note de Pain et la Note de Panification, et 5000 marqueurs pour la teneur en Protéine. Ces chiffres sont donc bien inférieurs au nombre de marqueurs disponibles dans cette étude (10000) ou disponibles

sur des puces couramment utilisées en blé (9K, 35K, 420K...). Seules 3 variables ont été présentées pour des questions de lisibilité mais ces tendances se retrouvent pour tous les critères étudiés. Par ailleurs, ces observations sont en accord avec ce qui est reporté dans la littérature et avec des résultats obtenus sur un 3<sup>e</sup> Panel (Panel INRA-AO, résultats non inclus dans cet article).

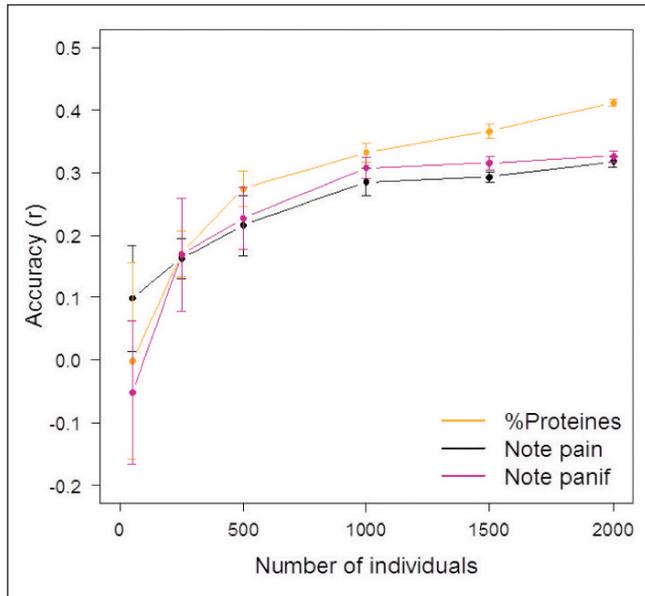


Figure 3 : Capacité de prédiction ( $r$ ) en fonction du Nombre d'Individus pris en compte dans la Calibration. Les variables présentées sont la teneur en protéine (%Protéines), la Note de Pain (Note pain) et la Note Totale de Panification (Note panif).

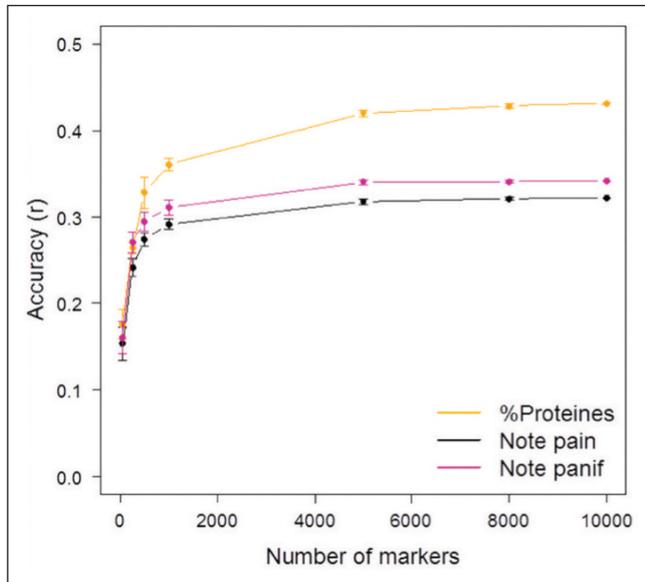


Figure 4 : Capacité de prédiction ( $r$ ) en fonction du Nombre de marqueurs pris en compte dans la Calibration. Les variables présentées sont la teneur en protéine (%Protéines), la Note de Pain (Note pain) et la Note Totale de Panification (Note panif).

La Figure 5, quant à elle, montre un important gain de précision lorsque les génotypes sont évalués dans 2 ou 3 environnements plutôt qu'un seul. Les environnements supplémentaires apportent un gain beaucoup moins important. Peu de génotypes ont été mesurés dans plus de trois environnements dans ce panel, d'où un léger décrochement de la courbe. Ces résultats sont conformes à la littérature et corroborés par les résultats obtenus sur le Panel INRA-AO (résultats non présentés).

Il semble donc qu'au global le Panel Recherche permette une calibration plus précise et plus robuste que le Panel Variétal, et ce alors que le set de marqueurs est le même pour les 2 panels, et qu'il

y a beaucoup plus de données phénotypiques pour le Panel Variétal que pour le Panel Recherche (près de 3800 observations contre 2300). Comme les données Phénotypiques sont vraisemblablement plus précises pour le Panel Variétal (du fait du nombre d'observations par génotype) que le pour le Panel Recherche, notre hypothèse est que la précision supérieure du Panel Recherche vient du fait qu'il est constitué de 10 fois plus de génotypes que le Panel Variétal (2300 contre près de 200).

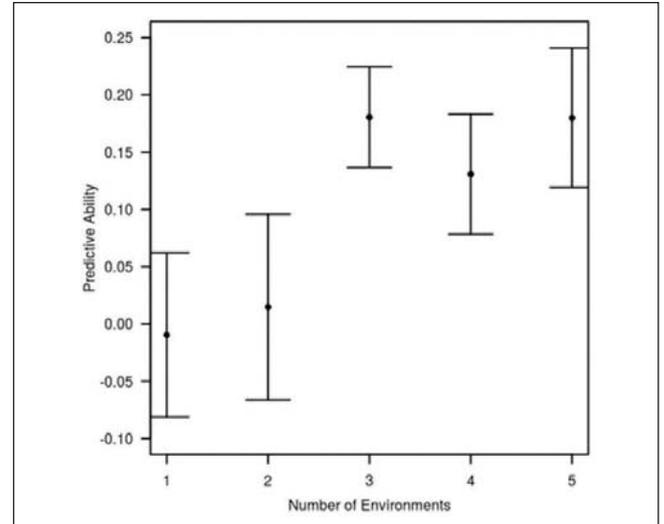


Figure 5 : Précision de prédiction ( $r$ ) en fonction du Nombre d'Observations (Environnements) par génotype dans la Calibration du Panel Variétal, toutes variables confondues.

### 3.6 - Apport des covariables d'alvéographe dans les modèles GS

Sur le Panel Variétal, les résultats obtenus montrent que les capacités prédictives du modèle GBLUP sont relativement limitées pour les caractères principaux (Teneur de Protéines, Note de Pain, Note de Pâte, Note de Mie et Note Totale de Panification,  $0.10 < r < 0.40$ ). Cependant, il a été possible de tester sur ce panel l'effet de l'ajout dans le modèle GBLUP des covariables mesurées par l'alvéographe. Cet ajout a permis d'augmenter significativement les capacités prédictives des modèles ( $0.40 < r < 0.70$ , Figure 6).

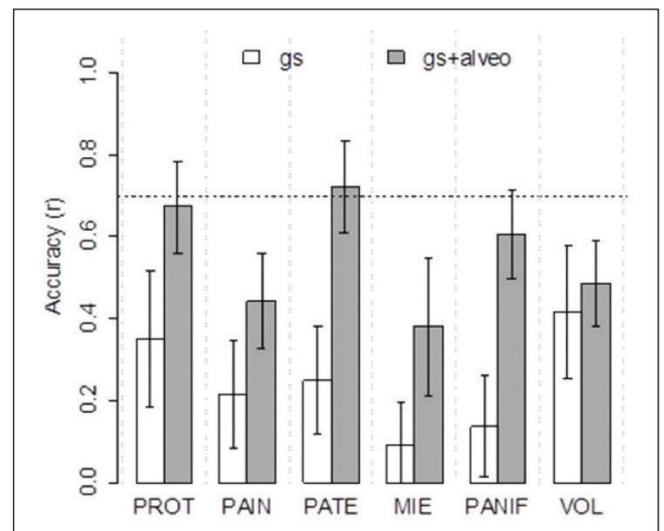


Figure 6 : Apport des covariables d'alvéographe à la capacité prédictive des modèles GS. Les capacités prédictives ont été estimées par validations croisées sur le Panel Variétal pour la teneur en protéines, la note de pain, la note de pâte, la note de mie et la note totale de panification. Les 2 modèles comparés sont (1) un modèle GBLUP ("gs") et (2) un modèle GBLUP auquel est ajouté les variables issues du test alvéographe ("gs+alveo"). La ligne tirée correspond à un coefficient de corrélation  $r=0.70$ .

Un tel accroissement des capacités prédictives grâce à l'utilisation de mesures indirectes de la qualité boulangère peut s'expliquer par le fait que ces mesures capturent des effets génétiques mais aussi des effets environnementaux. Le modèle GBLUP, qui utilise uniquement les données de génotypage, ne peut pas prendre en compte ces effets environnementaux. Ce résultat paraît intéressant du point de vue du sélectionneur puisque la quantité de grains nécessaire à la réalisation de ce type de test est relativement réduite et pourrait permettre d'appliquer ce type de modèle de manière relativement précoce dans les schémas de sélection.

### 3.7 - Tests d'associations

Au total, 49 associations significatives ont été identifiées sur le Panel Recherche et 57 sur le Panel Variétal (voir le détail en Annexe 3). La plupart des marqueurs identifiés étaient situés sur les chromosomes 1D (respectivement 14 et 10 marqueurs pour les Panels Variétal et Recherche), 1B (respectivement 8 et 5 marqueurs pour les Panels Variétal et Recherche) et 6B (respectivement cinq et quatre marqueurs pour les Panels Variétal et Recherche). L'analyse des colocalisations pour les deux panels a permis de mettre en évidence :

- une zone située entre 31 et 33cM sur le chromosome 1B associée au collant de la pâte au pétrissage et à l'extensibilité au façonnage.
- une zone située à 66cM sur le chromosome 1D associée au collant de la pâte au pétrissage, au déchirement et à l'extensibilité au façonnage.
- une zone située à 34cM sur le chromosome 6B associée au volume de la pâte.

Les zones situées sur les chromosomes 1B et 1D correspondent à la localisation des gènes de gluténines de haut poids moléculaire. Concernant la zone sur le 6B, un candidat possible est le gène de gliadine Gli-B2 bien que cela n'ait pu être démontré dans cette étude. De manière générale, la plupart des zones identifiées dans un panel diffèrent de celles identifiées dans l'autre panel, à l'exception des trois cas mentionnés ci-dessus.

## 4. Conclusion : protocole de calibration GS pour la qualité boulangère

A l'issue de ce projet, nous sommes en mesure de proposer le protocole suivant pour créer et utiliser une calibration pour la qualité boulangère des blés :

- Taille de la base de calibration : ce point est crucial pour s'assurer d'une bonne calibration. Nos résultats montrent que plus de 2000 génotypes n'est pas encore suffisant pour prédire précisément les caractères principaux,
- Génotypage : quelques milliers de marqueurs SNP (2000 à 5000), bien répartis sur le génome, semblent suffisant.
- Phénotypage : ce point est aussi crucial pour s'assurer d'une bonne calibration. Il est conseillé de s'intéresser à tous les critères du test BIPEA (et pas seulement aux notes globales), car tous peuvent présenter un intérêt et être plus ou moins bien prédits. Compte tenu des risques d'accidents climatiques et de l'importance des interactions GxE, un réseau d'essai multi-local et multi-annuel doit être envisagé, en prenant en compte un fort risque de perte d'essais. Contrairement au choix fait dans cette étude, nous recommandons d'observer au moins 2 à 3 fois chaque génotype (dans des environnements différents). Enfin, il faut s'assurer d'un bon niveau de variabilité pour les caractères, et donc ne pas utiliser uniquement des individus « bons » en termes de qualité.
- Nous recommandons aussi d'inclure les résultats d'alvéographe, ou tout autre test indirect de la qualité boulangère, en tant que covariables quand cela est possible, car cela améliore significativement la précision des modèles de GS.

Enfin, quant à l'utilisation de ces prédictions dans les programmes de sélection, nous pensons que la précision observée dans cette étude ne permet qu'un tri grossier. Si ce tri intervient dans les stades précoces de la sélection, il est conseillé d'appliquer une pression de sélection modérée. Dans les stades plus avancés, les mesures indirectes de la qualité boulangère, tel l'alvéographe, peuvent être combinées aux données de génotypage pour obtenir une prédiction beaucoup plus précise de la qualité boulangère.

Tous nos remerciements aux partenaires, stagiaires et collègues ayant contribué à ce projet : Gilles Charmet, Magali Thierry, Pierre Colin, Antonin Galien, Nicolas Perardel, Benoît Méléard, Céline Duque, Sébastien Deshayes, Valérie Herteman et David Gouache.

**Annexe 1** : Liste des composantes de la note de panification selon l'étape dans la fabrication du pain et selon la catégorie (définie par la note principale correspondante), et détail du calcul des notes.

Catégorie (Note Principale)	Type	Observations	Abréviations
Note de pâte	PETRISSAGE	Rapidité lissage	LISP
		Collant	COLP
		Consistance	CONP
		Extensibilité	EXTP
		Elasticité	ELAP
	POINTAGE	Détente relâchement	RELP
			REPO
	FACONNAGE	Extensibilité	EXTF
		Déchirement	DECF
		Elasticité	ELAF
	APPRET	Collant	COLF
		Activité fermentaire	ACTP
MISE AU FOUR	Déchirement	DECA	
		COMF	
Note de pain	ASPECT DU PAIN	Collant	TENU
		Section	SECT
		Couleur	COUP
		Épaisseur	EPAP
	COUP DE LAME	Croustillant	CROP
		Développement	DVCL
		Régularité	RECL
		Déchirement	DECL
Note de mie	ASPECT DE LA MIE	Couleur	COUM
		Souplesse	SOUM
	TEXTURE	Elasticité	ELAM
		Collant	COLM
	ALVEOLAGE	Régularité	REAL
		Épaisseur	EPAL
		Saveur et arôme	FLAV

Formules de calcul des notes principales à partir de composantes.  
Le coefficient est attribué en fonction de la note du critère pondérant (en gras dans la formule).

**Note de pâte =**  
 $(0.5 \times LISP + COLP \times 0.5 + EXTP \times 0.5 + ELAP \times 0.5 + RELP + 0.5) \times coef(*) + REPO + (EXTF \times 0.5 + DECF \times 0.5 + ELAF \times 0.5 + COLF) \times coef(*) + (ACTP \times 0.5 + DECA \times 0.5) + (COMF + TENU \times 2) \times coef(*)$

**Note de pain =**  
 $(SECT + COUP \times 2 + EPAP \times 0.5 + CROP \times 0.5 + DVCL + RECL + DECL) \times coef(*) + VOL**$

**Note de mie =**  
 $(SOUM \times 2 + ELAM + COLM) + (REAL + EPAL \times 2 + FLAV \times 3)$

**Note Totale (0-300) =**  
 $Note\ de\ pâte(0-100) + Note\ de\ pain(0-100) + Note\ de\ mie(0-100)$

Note du critère pondérant (en gras)	10	7	4	1
coef(*) correspondant	1	0.75	0.5	0.25

\*\* Un score basé sur le Volume du Pain (VOL) est utilisé dans ce calcul.

**Annexe 2** : Liste des 192 variétés du Panel Variétal.

ABAQUE, ACCOLADE, ACCOR, ACCROC, ACIENDA, ACOUSTIC, ADAGIO, ADEQUAT, ADHOC, AEROBIC, AGAPE, AGRESTIS, AGUILA, AKRATOS, ALCAZAR, ALDRIC, ALIGATOR, ALIXAN, ALIZE, ALLEZY, ALTAMIRA, ALTIGO, ALVAREZ, AMADOR, AMBELLO, AMBITION, AMUNDSEN, ANDALOU, ANDINO, ANTONIUS, APACHE, APRILIO, ARACK, ARAMIS, ARDELOR, AREZZO, ARISTOTE, ARKEOS, ARLEQUIN, AROBASE, ASDECOEUR, ASTUCE, ATHLON, ATTITUDE, ATTLASS, AUBUSSON, AUDI, AUTAN, AUTENTIC, AVANTAGE, AXIMA, AZIMUT, AZZERTI, AZZURO, BAGATELLE007, BAROK, BATTANT, BERMUDE, BIANCOR, BOISSEAU, BOKARO, BOREGAR, BOSPHOR, BRENTANO, BUENNO, CAMPERO, CAPHORN, CARIBOU, CARRE, CCBINEDIT, CCBINGENIO, CCBPREFERENCE, CELESTIN, CENTENAIRE, CEZANNE, CHAGALL, CHARGER, CHEVALIER, CHEVRON, CIENTO, COMODOR, COMPIL, CONTREFOR, CORDIALE, COURTOT, CROISADE, DIALOG, DINOSOR, EMERALD, EPHOROS, EPIDOC, EUCLIDE, EXELCIOR, EXOTIC, EXPERT, FARINELLI, FARMEUR, FIORETTO, FLAMENKO, FLAUBERT, FLUOR, FOLKLOR, FORBAN, FORBLANC, FOURMI, GALACTIC, GALIBIER, GALOPAIN, GARANTUS, GARCIA, GAUGAIN, GLADIATOR, GONCOURT, GRAINDOR, GRETHEL, GULLIVER, HATTRICK, HAUSSMANN, HEKTO, HONNOR, HOURRA, HYBERY, HYMACK, HYSORE, HYSTAR, HYSUN, HYXO, ILLICO, INCISIF, INNOV, INOUI, INSPIRATION, INSTINCT, INTACT, INTERET, IRIDIUM, ISENGRAIN, ISTABRAQ, JBASANO, JBDIEGO, KALANGO, KALYSTAR, KARILLON, KETCHUM, KLEBER, KORELI, LANCELOT, LIMES, LORD, MANAGER, MARCELIN, MATRIX, MAXWELL, MAXYL, MAYFAIR, MELKIOR, MENDEL, MENESTREL, MERCATO, MINOTOR, MIROIR, MUSIK, NIRVANA, NOGAL, NUAGE, NUCLEO, OAKLEY, OCTET, OLIVART, ORVANTIS, OXEBO, PAKITO, PALADAIN, PALEDOR, PAROLI, PEPIDOR, PERFECTOR, PERICLES, PHARE, PICADOR, PLAINEDOR, PLAYER, POTENZIAL, PR22R20, PR22R58, PREMIO, PREVERT, QUEBON, RACINE, RAGLAN, RAISON, RASPAIL, RAZZANO, RICHPAIN, RIMBAUD, RIZE, ROCHFORD, RODRIGO, ROSARIO, ROYSSAC, RUSTIC, SAINTEX, SAMURAI, SANKARA, SCENARIO, SCOR, SEBASTO, SELEKT, SEYRAC, SIRTAKI, SOBBEL, SOGOOD, SOISSONS, SOKAL, SOLEHIO, SOLLARIO, SOLUTION, SOPHYTRA, SORRIAL, SPONSOR, SUBLIM, SUBTIL, SUMO, SWKUNGSJET, SWEET, SWINGGY, SYALTEO, SYMATTIS, TAMARO, TIAGO, TIMBER, TIMING, TITLIS, TOISONDOR, TRAPEZ, TULIP, USKI, VALODOR, VELOURS, VERLAINE, VOLONTAIRE, WALLABY, WELLAND

Annexe 3 : Description des associations significatives détectées dans les 2 Panels.

Variable	Marqueur	Chrom.	Position (cM)	-log(p)	Allèle référence	Allèle alternatif	Effet Allélique	R2	Panel	Colocalisation
ACTP	WC0222309	5B	115.5	5.03	AA	TT	0.25	0.01	Recherche	
COLF	WC0216089	1B	92.7	4.86	CC	TT	0.14	0.01	Recherche	
COLF	WC0224844	1D	66.2	8.52	AA	GG	0.17	0.02	Recherche	
COLF	WC0221092	4A	98.1	4.78	AA	CC	-0.15	0.01	Recherche	
COLF	WC0222489	6A	72	4.15	AA	GG	0.09	0.01	Recherche	
COLP	WC0218313	1A	4.4	4.82	CC	GG	0.20	0.02	Recherche	
COLP	WC0218779	1B	26.1	4.36	CC	TT	-0.86	0.12	Variétal	
COLP	WC0218765	1B	30.8	6.07	GG	TT	0.96	0.1	Variétal	*
COLP	WC0218674	1B	33.3	7.79	AA	GG	0.35	0.02	Recherche	*
COLP	WC0224463	1B	93.9	4.32	AA	CC	-0.16	0.01	Recherche	
COLP	WC0224844	1D	66.2	11.07	AA	GG	0.23	0.03	Recherche	*
COLP	WC0224869	1D	66.2	6.89	CC	TT	0.66	0.11	Variétal	*
COLP	WC0227693	5B	96.2	5.18	AA	GG	1.11	0.11	Variétal	
COMF	WC0224869	1D	66.2	7.65	CC	TT	0.32	0.14	Variétal	
COMF	WC0228175	6A	24.8	4.58	AA	GG	-0.35	0.1	Variétal	
COMF	WC0228148	6A	101.8	4.30	CC	TT	-0.19	0.05	Variétal	
CONP	WC0225687	3A	140.8	4.60	AA	TT	0.57	0.09	Variétal	
CONP	WC0216555	3B	60.9	4.32	AA	GG	0.61	0.09	Variétal	
COUM	WC0219202	2A	21.9	4.90	GG	TT	0.23	0.01	Recherche	
COUM	WC0228994	7A	45.6	4.03	AA	TT	-0.28	0.01	Recherche	
COUM	WC0217442	7A	128	4.20	AA	CC	-0.20	0.01	Recherche	
COUM	WC0217329	7A	158.7	5.84	AA	GG	0.19	0.01	Recherche	
COUM	WC0223874	7B	79.4	4.12	CC	TT	0.12	0.01	Recherche	
COUM	WC0217003	7B	83.7	14.13	AA	GG	-0.28	0.03	Recherche	
COUP	WC0214377	3B	143.1	4.56	CC	GG	-2.52	0.13	Variétal	
COUP	WC0216424	3D	37.9	5.71	AA	GG	3.11	0.12	Variétal	
COUP	WC0221568	5A	117.7	4.56	AA	GG	2.69	0.09	Variétal	
DECF	WC0218314	1A	7.8	4.67	CC	TT	0.10	0.01	Recherche	
DECF	WC0224846	1D	66	10.12	GG	TT	-0.71	0.25	Variétal	*
DECF	WC0224844	1D	66.2	12.07	AA	GG	-0.20	0.03	Recherche	*
DECF	WC0219669	2B	51.1	4.25	AA	GG	-0.41	0.06	Variétal	
DECF	WC0222445	5D	60.4	4.27	AA	GG	0.48	0.08	Variétal	
DECF	WC0228383	6B	76.2	4.28	GG	TT	0.20	0.01	Recherche	
DECL	WC0218778	1B	26.3	5.40	AA	TT	1.07	0.12	Variétal	
DECL	WC0218809	1D	65.7	9.91	CC	GG	0.26	0.02	Recherche	
DECL	WC0225492	2D	140.3	4.71	AA	GG	-0.31	0.01	Recherche	
DECL	WC0221413	4B	40.6	4.85	GG	TT	0.72	0.09	Variétal	
DECL	WC0213174	6A	9	4.41	CC	TT	-1.07	0.08	Variétal	
DECL	WC0215903	6B	33.9	5.62	AA	GG	-0.93	0.12	Variétal	
DECL	WC0228288	6B	52.2	4.09	GG	TT	1.09	0.06	Variétal	
DUR	WC0214377	3B	143.1	5.07	CC	GG	-14.08	0.12	Variétal	
DUR	WC0216424	3D	37.9	10.50	AA	GG	25.79	0.22	Variétal	
DVCL	WC0222162	5B	120.4	4.28	CC	GG	2.49	0.09	Variétal	
DVCL	WC0215903	6B	33.9	4.06	AA	GG	1.89	0.11	Variétal	
DVCL	WC0228764	7A	92.1	4.28	AA	GG	-0.17	0.01	Recherche	
ELAF	WC0218731	1B	30.8	4.72	CC	TT	-3.15	0.06	Variétal	
ELAF	WC0224869	1D	66.2	4.72	CC	TT	2.25	0.1	Variétal	
ELAP	WC0218731	1B	30.8	4.09	CC	TT	-1.97	0.09	Variétal	*
ELAP	WC0218674	1B	33.3	9.64	AA	GG	0.61	0.02	Recherche	*
ELAP	WC0224869	1D	66.2	7.83	CC	TT	0.39	0.02	Recherche	
ELAP	WC0228096	6A	104	4.15	AA	CC	0.79	0.04	Variétal	
EPAL	WC0224869	1D	66.2	4.21	CC	TT	0.34	0.05	Variétal	
EPAL	WC0219824	2D	24.6	7.16	CC	TT	0.67	0.16	Variétal	

Variable	Marqueur	Chrom.	Position (cM)	-log(p)	Allèle référence	Allèle alternatif	Effet Allélique	R2	Panel	Colocalisation
EPAP	WC0224560	1B	86.4	4.24	AA	CC	-1.19	0.09	Variétal	
EPAP	WC0220154	3A	19.7	6.04	AA	CC	1.32	0.16	Variétal	
EXTF	WC0224841	1D	55.5	4.27	CC	TT	-0.29	0	Recherche	
EXTF	WC0224874	1D	65.2	5.72	GG	TT	1.69	0.12	Variétal	*
EXTF	WC0224869	1D	66.2	8.91	CC	TT	-0.50	0.02	Recherche	*
EXTF	WC0225282	2B	95.6	5.09	AA	GG	-0.73	0.01	Recherche	
EXTF	WC0227751	5B	36.6	4.98	AA	GG	2.45	0.13	Variétal	
G	WC0224846	1D	66	5.28	GG	TT	-1.45	0.13	Variétal	
G	WC0227324	5A	80.2	4.06	CC	TT	1.33	0.11	Variétal	
le	WC0224869	1D	66.2	13.13	CC	TT	4.81	0.21	Variétal	
LISP	WC0224869	1D	66.2	13.80	CC	TT	-0.31	0.05	Recherche	
MIE	WC0222070	5B	43.2	4.87	CC	TT	-2.17	0.1	Variétal	
NPAI	WC0216750	2B	14.9	4.84	CC	TT	-3.32	0.01	Recherche	
NPAI	WC0215903	6B	33.9	4.78	AA	GG	1.99	0.01	Recherche	
NPAI	WC0228733	7A	92.2	5.11	GG	TT	2.56	0.01	Recherche	
NPAI	WC0229499	7D	8.9	5.40	AA	GG	3.11	0.01	Recherche	
NPANI	WC0218674	1B	33.3	5.27	AA	GG	3.74	0.01	Recherche	
NPANI	WC0224844	1D	66.2	5.06	AA	GG	2.14	0.01	Recherche	
NPANI	WC0216750	2B	14.9	5.20	CC	TT	-5.08	0.02	Recherche	
NPANI	WC0229502	7D	7.9	4.62	GG	TT	-2.60	0.01	Recherche	
P.L	WC0224846	1D	66	4.70	GG	TT	0.30	0.1	Variétal	
PANIF	WC0224869	1D	66.2	4.30	CC	TT	8.36	0.05	Variétal	
PANIF	WC0219824	2D	24.6	5.84	CC	TT	14.08	0.12	Variétal	
PANIF	WC0221946	5A	22.5	4.43	AA	GG	-10.73	0.1	Variétal	
PATE	WC0218806	1B	26.9	5.12	AA	GG	-8.73	0.09	Variétal	
PATE	WC0224869	1D	66.2	7.26	CC	TT	6.28	0.12	Variétal	
PROT	WC0214404	2A	7	4.58	CC	GG	-0.13	0.01	Recherche	
PROT	WC0226542	4A	94.8	4.36	CC	TT	0.18	0.01	Recherche	
PROT	WC0215996	4A	141.5	5.43	GG	TT	-0.21	0.02	Recherche	
PROT	WC0226868	4B	54.5	12.32	AA	GG	0.04	0	Recherche	
PROT	WC0213051	4D	25.5	12.44	AA	GG	0.14	0.03	Recherche	
PROT	WC0227190	4D	30.2	4.72	AA	GG	-0.23	0.17	Variétal	
PROT	WC0227400	5A	175.8	4.43	CC	GG	0.13	0.02	Recherche	
PROT	WC0228421	UNM	49	4.44	CC	TT	0.03	0	Recherche	
REAL	WC0222969	6B	57.3	4.06	CC	GG	-0.91	0.12	Variétal	
RELP	WC0220871	4A	141.5	5.23	CC	TT	-0.92	0.13	Variétal	
RELP	WC0227399	5A	73.3	4.51	CC	TT	-0.24	0.01	Variétal	
REPO	WC0224869	1D	66.2	8.92	CC	TT	0.60	0.17	Variétal	
REPO	WC0225626	2D	24.6	5.93	CC	TT	-0.56	0.11	Variétal	
SECT	WC0218731	1B	30.8	4.66	CC	TT	-2.80	0.07	Variétal	
SECT	WC0218809	1D	65.7	4.89	CC	GG	1.92	0.12	Variétal	
SECT	WC0226602	4A	47.5	4.37	AA	CC	1.44	0.06	Variétal	
TENU	WC0224846	1D	66	7.49	GG	TT	2.27	0.17	Variétal	
TENU	WC0219824	2D	24.6	4.34	CC	TT	2.13	0.11	Variétal	
VOL	WC0213718	3B	64.6	4.74	AA	GG	31.43	0.02	Recherche	
VOL	WC0215903	6B	33.9	7.71	AA	GG	43.45	0.03	Recherche	*
VOL	WC0215903	6B	33.9	5.98	AA	GG	96.96	0.23	Variétal	*
VOL	WC0228764	7A	92.1	4.93	AA	GG	-38.79	0.02	Recherche	
VOL	WC0229499	7D	8.9	4.66	AA	GG	57.87	0.02	Recherche	
VOLG	WC0216017	1D	63.6	4.10	GG	TT	11.56	0.01	Recherche	
VOLG	WC0217247	2B	34.3	4.63	CC	TT	15.36	0.02	Recherche	
VOLG	WC0215903	6B	33.9	9.09	AA	GG	22.43	0.04	Recherche	
W	WC0225377	2B	42.6	4.12	AA	GG	-52.95	0.11	Variétal	

## Références bibliographiques

- Blanc G., Charcosset A., Veyrieras J. B., Gallais A., and Moreau L.** (2008). Marker-assisted selection efficiency in multiple connected populations: a simulation study based on the results of a QTL detection experiment in maize. *Euphytica* 161, 71–84.
- Eathington, S. R., T. M. Crosbie, M. D. Edwards, R. S. Reiter, and J. K. Bull.** (2007). Molecular Markers in a Commercial Breeding Program. *Crop Sci.* 47(Suppl3):S-154-S-163.
- Endelman, J.B.** (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal* 4, 250.
- Endelman J. B. and J.-L. Jannink** (2012) Shrinkage estimation of the realized relationship matrix. *G3 Genes Genomes Genet.* 2: 1405–1413
- Hayes B., J. Panozzo, C. Walker, A. Choy, S. Kant, et al.** (2017) Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet.* 130: 2505–2519.
- Heffner E.L., Sorrells M.E., Jannink J-L** (2009) Genomic Selection for Crop Improvement. *Crop Science* 49:1-12
- Heffner E.L., Lorenz A.J., Jannink J-L. Sorrells M.E.** (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci* 50: 1681-1690.
- Heffner E.L., Jannink J-L. Sorrells M.E.** (2011a) Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. *The Plant Genome* 4:65–75.
- Heffner E.L., Jannink J-L., Iwata H., Souza E., Sorrells M.** (2011b) Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci* 51:2597–2606.
- Lande R, Thompson R** (1990) Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genetics* 124 : 743-756
- Méléard B.** (2010). Quelles évolutions qualitatives dans l'offre française de blé tendre ? Bilan de 15 années d'amélioration. *Industries des Céréales*, 170: 14-19.
- Meuwissen THE, Hayes B, Goddard ME** (2001) Prediction of total genetic value using genome-wide dense marker maps; *Genetics* 157:1819-1829
- Michel S., M. Gallee, F. Löschenberger, H. Buerstmayr, and C. Kummer** (2017) Improving the baking quality of bread wheat using rapid tests and genomics: The prediction of dough rheological parameters by gluten peak indices and genomic selection models. *J. Cereal Sci.* 77: 24–34.
- Michel S., C. Kummer, M. Gallee, J. Hellinger, C. Ametz, et al.** (2018) Improving the baking quality of bread wheat by genomic selection in early generations. *Theor. Appl. Genet.* 131: 477–493.
- Oury FX, Chiron H, Faye A, gardet O, Giraud A, Heumez E, Rolland B, Rousset M, Trottet M, Charmet G** (2010) The prediction of bread wheat quality: joint use of the phenotypic information brought by technological tests and the genetic information brought by HMW and LMW glutenin subunits. *Euphytica* 171:87-109.
- Segura V., Vilhjálmsson B.J., Platt A., Korte A., Seren Ü., Long Q., Nordborg M.** (2012) An efficient multi-locus mixed model approach for genome-wide association studies in structured populations. *Nat Genet* 44, 825–830.
- Whittaker, J.C., R., Thompson, M.C. Denham.** (2000) Marker-assisted selection using ridge regression. *Genet. Res.* 75:249-252.