

VetAgro Sup

Mémoire de fin d'études d'ingénieur

CONFIDENTIEL

Modélisation de la qualité boulangère chez le blé tendre (*Triticum aestivum* L.)

Magali Thierry

Option Génomique Ecophysiologie et Production végétale

Année 2017



Mémoire de fin d'étude

Modélisation de la qualité boulangère chez le blé tendre (*Triticum aestivum* L.)

THIERRY Magali

Soutenu à Clermont-Ferrand, 2017, 30 pages

STRUCTURE D'ACCUEIL ET INSTITUTIONS ASSOCIEES:

- ♦ ARVALIS Institut du végétal, Station inter-instituts, 6 Chemin de la côte vieille, 31 450 Baziège.

ENCADRANTS :

- ♦ Maître de stage : BOGARD, Matthieu (ARVALIS Institut du végétal)
- ♦ Tuteur pédagogique : MOUZEYAR, Said

OPTION : Génomique Ecophysiologie et Production végétale

RESUMÉ

Le blé tendre est une des céréales les plus cultivées dans le monde et un des principaux débouchés de cette production est la fabrication de pain. La qualité boulangère ou aptitude à la panification est un critère d'inscription des variétés au catalogue officiel Français. Cependant, la prise en compte de ce critère dans les programmes de sélection est difficile car elle ne peut être réalisée que tardivement. L'utilisation d'un modèle de prédiction génomique pourrait contribuer à résoudre ce problème et à augmenter ainsi l'efficacité de la sélection. Les objectifs de l'étude présentée dans ce rapport consistent à mettre en évidence l'architecture génétique de variables liées à la qualité boulangère, à estimer les capacités prédictives de modèles de sélection génomique et à quantifier l'apport de variables issues de tests indirects dans ces modèles de prédiction génomique. Le matériel végétal utilisé était constitué d'un panel de 188 variétés inscrites et d'un panel de 1529 lignées en cours de sélection. Ces variétés/lignées ont été génotypées pour environ 15 000 marqueurs SNP. Des zones génomiques associées à la qualité boulangère ont été détectées sur les chromosomes 1B, 1D, 2B, 2D, 3D, 6B, 7A, 7B. Les différents modèles de prédiction génomique testés ont une capacité prédictive comprise entre 0 et 0.45 suivant les panels et les variables considérées. Pour certains caractères, l'utilisation d'un sous-ensemble de marqueurs associés (entre 1000 et 5000) pour calibrer le modèle de prédiction génomique permet d'atteindre la qualité de prédiction obtenue avec la totalité des marqueurs. Enfin, l'inclusion des variables d'alvéographe en plus des marqueurs SNP dans un modèle de prédiction génomique permet d'améliorer significativement les capacités prédictives (+0.20 en moyenne). Ces résultats contribuent à mieux comprendre l'architecture génétique de la qualité boulangère du blé tendre et à faciliter la prise en compte de ce caractère par les sélectionneurs.

Mots clés :

Qualité boulangère, sélection génomique, génétique d'association, *Triticum aestivum*

Mémoire de fin d'étude

Modelisation of baking quality for bread wheat (*Triticum aestivum* L.)

THIERRY Magali
Clermont-ferrand, 2017, 30 pages

HOSTING STRUCTURE:

- ♦ ARVALIS Institut du végétal

SUPERVISOR :

- ♦ Maître de stage : BOGARD, Matthieu (ARVALIS Institut du végétal)
- ♦ Tuteur pédagogique : MOUZEYAR, Said

OPTION : Génomique Ecophysiologie et Production végétale

ABSTRACT

Bread wheat is one of the most cultivate crop in the world and one of its main markets opportunities is bread making. Baking quality, or bread making ability, is a registration criterion of varieties in the French catalogue. However, the inclusion of this criterion in selection programs is complicated because it could be integrated only at the end of the programs. A genetic prediction tool could be useful to solve this problem and improve the selection efficiency for this trait. The study presented in this report aims to demonstrate the genetic architecture of parameters linked to baking quality, estimate predictive abilities of genomic prediction models and evaluate the inclusion of parameters from indirect tests in genomic prediction models. Two panels of 188 french cultivars and 1529 lines in selection were tested. Varieties were genotyped for approximatively 15 000 SNP markers. Genomic are were detected by Genomic Wide Association Study on chromosomes 1B, 1D, 2B, 2D, 3D, 6B, 7A, 7B. The genomic prediction model used in this study had a predictive ability ranging from 0 to 0.53 depending on the panels and the parameters considered. For some traits, the use of a subset of associated markers (from 1000 to 5000) in the genomic prediction model allowed to reach the same predictive ability as if all markers were included. The inclusion of alveograph parameters in the genomic prediction model significantly improved predictive abilities (+0.20 on average). These results contribute to better understand the genetic architecture of baking quality in wheat and could facilitate the use of this criterion by breeders.

Key words :

Bread making Quality, Genomic Selection, Genome Wide Associations Studies, *Triticum aestivum*

« L'étudiant conserve la qualité d'auteur ou d'inventeur au regard des dispositions du code de la propriété intellectuelle pour le contenu de son mémoire et assume l'intégralité de sa responsabilité civile, administrative et/ou pénale en cas de plagiat ou de toute autre faute administrative, civile ou pénale. Il ne saurait, en cas, seul ou avec des tiers, appeler en garantie VetAgro Sup. »

Remerciements

Je tiens à remercier particulièrement mon maître de stage, Matthieu Bogard, pour son aide précieuse et le temps qu'il m'a accordé. Je n'aurais pu mener à bien ni ce stage ni ce rapport sans ses conseils.

Je voudrais également remercier les encadrants du projet FSOV dans lequel s'inscrit mon stage notamment Bruno Poupard, Benoît Méléard et Sophie Bouchet qui m'ont permis d'obtenir les données nécessaires à la réalisation de ce stage, qui ont supervisés mes résultats tout au long du stage et m'ont conseillée.

Je remercie également Renaud Rincent d'avoir fourni un script pour le calcul du nombre de marqueurs indépendants.

Je souhaite remercier Sophie Vallade d'avoir permis la réalisation de ce stage dans cet établissement. Je remercie aussi tout le personnel de la station agronomique inter-institut de Baziège qui m'a accueillie.

Je remercie mon tuteur de stage Said Mouzeyar d'avoir encadré le déroulement de ce stage.

Enfin, je remercie Nathan, Maité et Clémence qui m'ont apporté rire et soutien autour de nombreux cafés.

Table des matières

Introduction	1
I. Contexte	2
1) Origine du blé : aspects botaniques et génomiques	2
2) Importance économique de la culture du blé	2
3) Aptitude à la panification du blé tendre	3
4) Détermination de la qualité boulangère	4
a. Définition et importance de la qualité boulangère	4
b. Evaluation de la qualité boulangère	4
5) Facteurs influençant la qualité boulangère	6
a. Conduite de culture	6
b. Processus de panification	7
c. Effet variétal	7
6) Déterminisme génétique de la qualité boulangère	7
7) Modélisation génétique de la qualité boulangère	8
a. Modélisation basée sur des QTL et sélection assistée par marqueurs	8
b. Sélection génomique	9
II. Problématique de recherche	11
III. Objectifs du stage	12
IV. Matériels et méthodes	13
1) Matériel végétal	13
a. Données ARVALIS Institut du végétal	13
a. Données Limagrain	13
2) Préparation des données phénotypiques	14
3) Données génotypiques	14
4) Analyse descriptive des données	15
5) Recherche des déterminants génétique de la qualité boulangère	15
a. La méthode mixed model (MM) forward	15
6) Prédiction génomique de la qualité boulangère	17
a. Modèles utilisés	17
b. Etude de la capacité prédictive du modèle de prédiction génomique	18
V. Résultats	20
1) Analyse descriptive des données	20
2) Génétique d'association	21
3) Prédiction génomique	23
4) Etude de l'inclusion des marqueurs associés pour améliorer la prédiction génomique	

5)	Contribution des variables issues du test d'alvéographe un modèle de prédiction génomique de la qualité boulangère	24
VI.	Discussion	26
1)	Estimation de l'héritabilité	26
2)	Distribution des marqueurs sur le génome	26
3)	Architecture génétique : quelques QTL majeurs, de nombreux QTL mineurs	26
4)	Perspectives d'utilisation en SAM	27
5)	Intérêts de la prédiction génomique pour la qualité boulangère.....	27
	Conclusion.....	30

Table des illustrations

Figure 1 : Structure de l'épi, de l'épillet et de la fleur du blé tendre.

Figure 2 : Phylogénie du blé tendre.

Figure 3 : Structure schématique du grain de blé avec ses différentes enveloppes.

Figure 4 : Classification des protéines de l'albumen selon la classification d'(Osborne, 1907).

Figure 5 : Schéma de formation du réseau gluténique.

Figure 6 : Alvéographe de Chopin de la marque CHOPIN Technologies™.

Figure 7 : Alvéogramme.

Figure 8 : Profil de HPLC (High Performance Liquid Chromatography) des différentes protéines de l'albumen du grain de blé.

Figure 9 : Exemple de journée lors d'un test de panification au laboratoire Arvalis de Boigneville.

Figure 10 : Formule de calcul des notes principales à partir de composantes.

Figure 11 : Loci impliqués dans la synthèse des gliadines (Gli) et des gluténines (Glu).

Figure 12 : Schéma des tests effectués lors de l'évaluation des variétés candidates à l'inscription au catalogue officiel Français par le Comité Technique Permanent de la Sélection (CTPS).

Figure 13 : Nombre d'essais par variétés.

Figure 14 : Présentation du recodage effectué pour la notation des composantes du test BIPEA.

Figure 15 : Correspondance entre la valeur en panification d'un critère et sa note pour le test BIPEA

Figure 16 : Analyse qualité des marqueurs SNP du panel Arvalis.

Figure 17 : Schéma représentant les étapes de validation croisée afin d'évaluer la prédiction génomique.

Figure 18 : Schéma représentant la méthode de validation croisée afin d'étudier la capacité prédictive du modèle en fonction du nombre de marqueurs associés au caractère inclus dans le modèle de prédiction.

Figure 19: Représentation de la carte génétique où sont positionnés les marqueurs SNP utilisés dans cette étude.

Figure 20: Représentation graphique des QTL détectés sur les chromosomes 1B, 2B et 2D.

Figure 21: Représentation graphique des QTL détectés sur le chromosome 1D.

Figure 22 : Représentation graphique des QTL détectés sur les chromosomes 3D, 6B, 7A et 7B.

Figure 23: Capacités prédictives des notes principales, de la note de panification, de la teneur en protéines, de la dureté, du taux d'hydratation et du volume du pain et des composantes de la note de pâte prédits selon trois modèles différents.

Figure 24 : Capacités prédictives des notes composantes de la note de pain (en haut) et mie (en bas) prédits selon trois modèles différents : modèle basé sur les variables d'alvéographe seules (en rose), modèle basé sur les marqueurs génétique (prédiction génomique ; en vert) et modèle combinant les variables d'alvéographe et les marqueurs génétiques (en orange). Les noms des caractères sont répertoriés dans le tableau 4.

Table des tableaux

Tableau 1 : Production de blé tendre (en million de tonnes) pour les années 2010 et 2014 en France, en Union Européenne (UE) et dans le monde.

Tableau 2 : Correspondance entre les paramètres de l'alvéographe et la valeur boulangère en panification française.

Tableau 3 : Echelle de notation des composantes des notes principales.

Tableau 4 : Liste des composantes de la note de panification (et leurs abréviations).

Tableau 5 : Nombre de lignées testées par lieux et par année pour le panel Limagrain avec le total du nombre de variétés par année.

Tableau 6: Variétés témoins présentes par lieux et par année.

Tableau 8 : Héritabilités des variables mesurées sur le panel Arvalis (ARV) ou le panel Limagrain (LIM), capacités prédictives estimées par validation croisée ou par validation sur un panel indépendant des modèles de sélection génomique calibrés sur chacun des panels pour les différentes variables étudiées.

Tableau 8 : Nombre de marqueurs significatifs par caractères.

Table des annexes

Annexe 1 : Grille de notation du test BIPEA selon la norme NF V03-716.

Annexe 2 : Corrélations entre les variables d'alvéographe et les notes de pâte, pain, mie et totale. Cercle de corrélation des variables d'alvéographe avec les composantes de la note de pâte.

Représentation graphique des QTL détectés sur le chromosome 1B.

Annexe 3: Evolution du déséquilibre de liaison sur le panel Arvalis en fonction de la distance entre marqueurs.

Annexe 4: Occurrence des variétés testées en fonction des années.

Annexe 5:

A : Distribution du volume du pain pour les deux classes alléliques au marqueur WC0215903 localisé sur le chromosome 6B pour le panel Limagrain (à gauche) et le panel Arvalis (à droite).

B : Distribution de la note de pâte pour les deux classes alléliques au marqueur WC0218806 localisé sur le chromosome 1B (à gauche) et au marqueur WC0224844 localisé sur le chromosome 1D (à droite) pour le panel Arvalis.

C : Distribution de la note de panification pour les deux classes alléliques au marqueur WC0219824 localisé sur le chromosome 2D (à gauche) et au marqueur WC0216254 localisé sur le chromosome 3D (à droite) pour le panel Arvalis.

Liste des abréviations

CTPS : Comité Technique Permanent de la Sélection

BIPEA : Bureau Interprofessionnel d'Etudes Analytiques

FSOV : Fond de Soutien à l'Obtention Végétal

NIRS : Near Infrared Spectroscopy

QTL : Quantitative Trait Loci

GWAS : Genomic Wide Association Studies

rrBLUP: ridge regression Best Linear Unbiased Prediction

SNP: Single Nucleotide Polymorphism

ACTP: Activité fermentaire au pétrissage

COLP: Collant au pétrissage

CONP : Consistance au pétrissage

EXTP : Extensibilité au pétrissage

ELAP : Elasticité au pétrissage

RELP : Relâchement au pétrissage

REPO : Détente relâchement au pointage

EXTF : Extensibilité au façonnage

DECF : Déchirement au façonnage

ELAF : Elasticité au façonnage

COLF : Collant au façonnage

DECA : Déchirement à l'apprêt

COMF : Collant mise au four

TENU : Tenue mise au four

SECT : Section

COUP : Couleur du pain

EPAP : Epaisseur du pain

CROP : Croustillant

DVCL : Développement au coup de lame

RECL : Régularité au coup de lame

DECL : Déchirement au coup de lame

COUM : Couleur de la mie

SOUM : Souplesse de la mie

ELAM : Elasticité de la mie

COLM : Collant de la mie

Introduction

Le blé tendre est la céréale la plus cultivée en France (Agreste, 2017). Plus de la moitié de la production nationale de blé tendre est utilisée après transformation meunière pour l'alimentation humaine. Dès lors, la part importante de ce débouché fait de la qualité boulangère un des critères importants lors de l'évaluation des variétés en cours d'inscription par le Comité Technique Permanent de la Sélection (CTPS). Cela incite donc les sélectionneurs à tenir compte de l'aptitude à la panification des blés lors de la création variétale. Cependant, l'application de critères de sélection pour ce caractère pose des problèmes. En effet, l'évaluation de la qualité boulangère demande d'obtenir une quantité de grains importante (plus d'un 1kg) qui n'est disponible qu'à la fin d'un programme de sélection (génération F6 ou F7). De plus, d'un point de vue génétique, la qualité boulangère est un caractère polygénique c'est-à-dire expliqué par un grand nombre de facteurs génétiques, généralement à effets faibles. Il est cependant nécessaire de trouver une solution afin d'intégrer les critères de qualité dans les premières étapes des programmes de sélection dans le but d'augmenter la réponse à la sélection. Une alternative connue pour évaluer des critères complexes, longs et coûteux à phénotyper est la sélection génomique. Cette méthode, qui a d'abord été développée chez les bovins laitiers, est apparue plus récemment en création variétale pour les espèces de grandes cultures. Le projet FSOV « *Etablissement d'un modèle de Sélection Génomique de la Qualité Boulangère des blés* », dans lequel s'inscrit ce stage, résulte d'un partenariat entre Limagrain, l'UMR GDEC de l'INRA de Clermont-Ferrand, ARVALIS Institut du végétal et l'Association Nationale de la Meunerie Française (ANMF). Le principal objectif de ce projet est de fournir à la communauté scientifique et à celle des sélectionneurs français en particulier un protocole permettant d'établir un modèle de prédiction génomique pour la qualité boulangère du blé tendre. Les objectifs de mon stage sont :

1. identifier les QTL associés aux composantes de la qualité boulangère,
2. établir un modèle de prédiction génomique sur différents panels de variétés ou de lignées
3. tester l'apport des variables issues de test alvéographe dans le modèle de prédiction génomique.

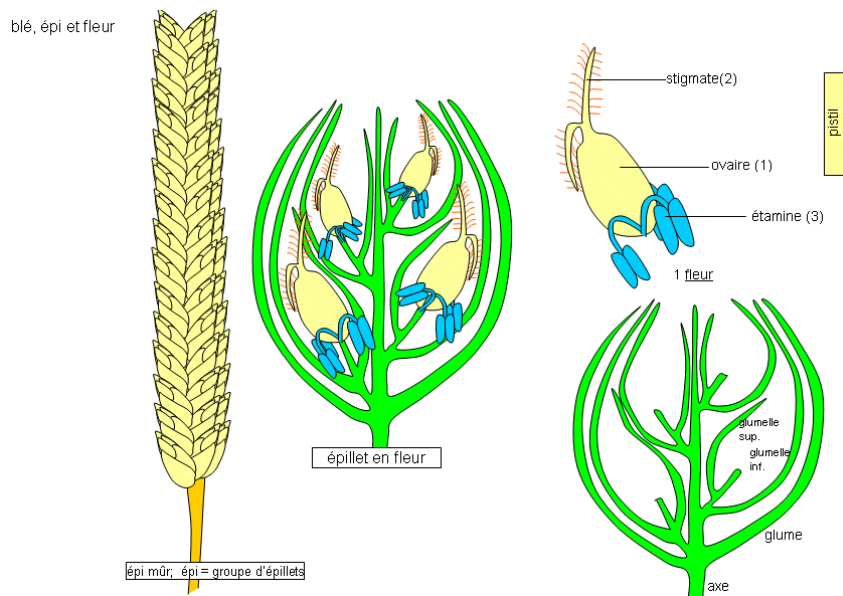


Figure 1 : Structure de l'épi, de l'épillet et de la fleur du blé tendre. Source : Académie de Dijon, 2008.

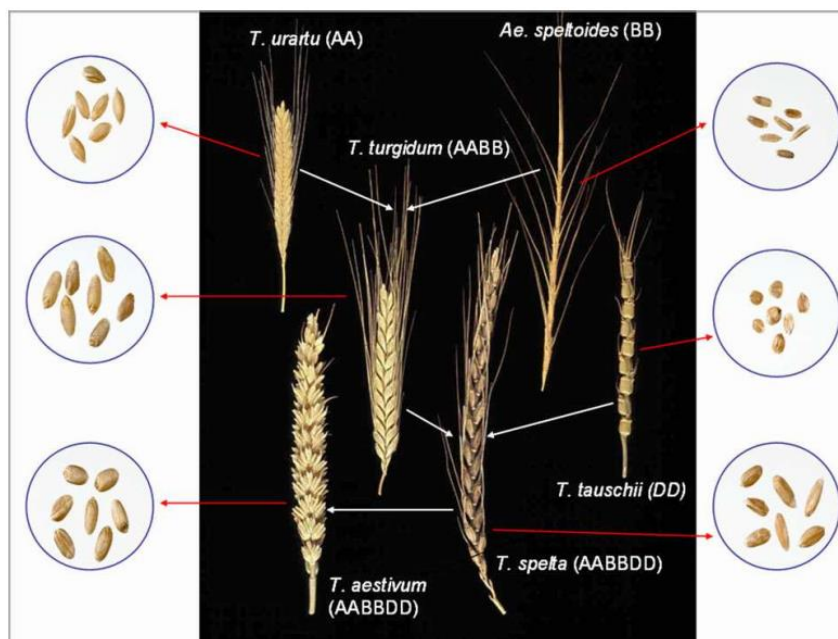


Figure 2 : Phylogénie du blé tendre. L'espèce porteuse du génome A est le *Triticum urartu*, celle porteuse du génome B est *Aegilops speltoides* et celle porteuse du génome D est *Triticum tauschii*. Source : (Shewry, 2009)

Tableau 1 : Production de blé tendre (en million de tonnes) pour les années 2010 et 2014 en France, en Union Européenne (UE) et dans le monde. Sources : FAO, Eurostat, Agreste, 2016

	2010	2014
Monde	678.5	732.9
UE	129.5	149.1
France	35.2	37.5

I. Contexte

1) Origine du blé : aspects botaniques et génomiques

Le blé tendre (*Triticum aestivum*) est une plante herbacée monocotylédone appartenant à la famille des Poacées du genre *Triticum*. Il représente une des principales espèces de blé cultivées avec le blé dur. Les espèces de blé cultivées de nos jours sont apparues suite à des hybridations entre espèces diploïdes et à des transformations dues à des phénomènes de domestication (Feillet, 2000). Le blé tendre et le blé dur se différencient par la dureté de leur grain. En alimentation humaine, le blé tendre est utilisé pour la panification tandis que le blé dur est utilisé pour la fabrication de pâtes et de semoules (Bergez and Abecassis, 2009).

Le blé tendre est une graminée annuelle de hauteur moyenne (un mètre environ ; (Lersten, 1987). Chaque tige possède cinq à sept nœuds ainsi que trois ou quatre feuilles véritables et produit un épi terminal composé d'épillets aplatis. Les épillets sont attachés à l'axe principal de l'épi (Figure 1). Chaque épillet est un axe reproducteur contenant deux à cinq fleurs. La fleur possède trois étamines et un pistil. Le régime de reproduction dominant chez le blé est l'autogamie. En effet, le pollen est généralement relâché avant même que la fleur ne s'ouvre ; on parle de cleistogamie. Des différences variétales avec des degrés d'autogamie divers existent cependant.

Le grain de blé est un fruit sec indéhiscent composé de différents tissus : les tissus enveloppes, le tissu de réserve (l'albumen) et le tissu embryonnaire. L'albumen représente quantitativement la plus grande part du grain de blé (80 à 85 % du grain). Il est composé d'une partie amylacée et de la couche à aleurone. La partie amylacée est composée de 70-80 % d'amidon et entre 10 à 18 % de protéines. Les protéines de réserve (gluténines et gliadines) représentent 80% des protéines de l'albumen (Feillet, 2000). La couche à aleurone est quant à elle très riche en protéines et en minéraux. .

Le génome du blé contient 16 milliards de paires de base. La grande taille de ce génome s'explique d'abord par son caractère polyploïde. En effet, le blé tendre est apparu suite à des croisements interspécifiques entre trois espèces ancestrales diploïdes menant à des événements de polyploïdisation (Figure 2). Le premier croisement interspécifique entre deux espèces diploïdes (*Triticum urartu* et *Aegilops speltoides*) a eu lieu il y a environ 500 000 ans et a conduit à l'apparition du blé dur tétraploïde (*Triticum turgidum*). Le deuxième événement de polyploïdisation a eu lieu il y a environ 9000-12000 ans entre le blé dur cultivé (tétraploïde) et *Triticum tauschii* et a donné le blé tendre panifiable actuel (*Triticum aestivum* L.). Le blé tendre comporte 6 jeux de chromosomes, on dit qu'il est allohexaploïde (Bonjean, 2001; Chantret et al., 2005). Chaque génome (nommé A, B et D) des trois espèces ancestrales diploïdes possèdent sept paires de chromosomes ce qui revient donc à un total de 21 paires de chromosomes ($2n = 6x = 42$, AABBDD). (Brenchley et al., 2012) Néanmoins, lors de la méiose, les chromosomes des différents génomes s'apparient entre eux et ségrégent indépendamment ce qui fait que le blé peut être considéré, d'un point de vue de l'hérédité des caractères, comme une espèce diploïde (les gamètes issues de la méiose contiennent toutes un exemplaire de chacun des chromosomes des différents génomes).

2) Importance économique de la culture du blé

D'un point de vue mondial, le blé tendre est la troisième céréale la plus cultivée dans le monde derrière le maïs (*Zea Mays*) et le riz (*Oryza sativa*) (voir Tableau 1). Elle représente 20 % des calories consommées dans l'alimentation humaine et une source importante de protéines (FAOSTAT, 2011). La France est le premier producteur et exportateur de blé tendre

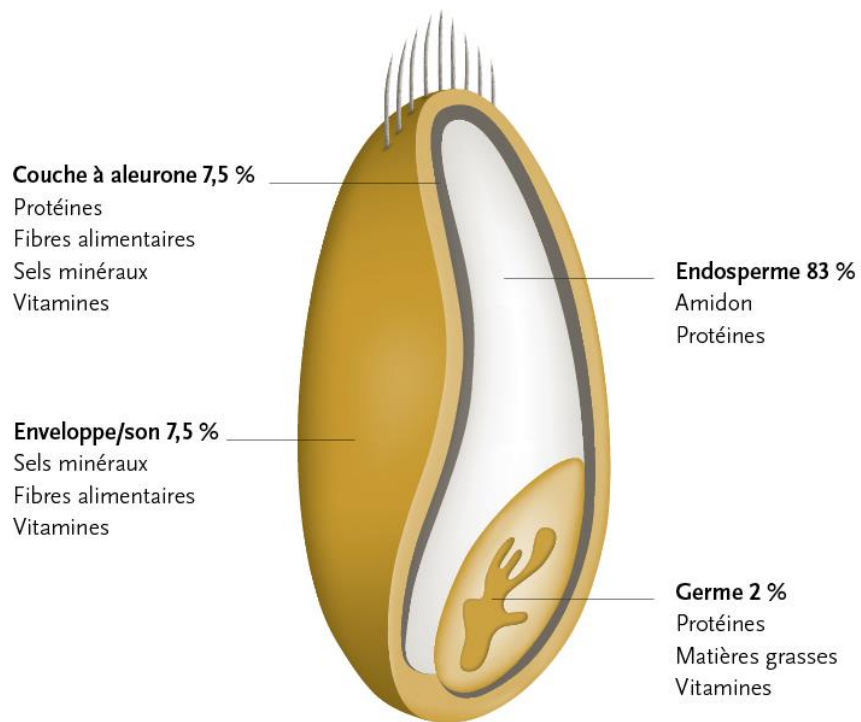


Figure 3 : Structure schématisique du grain de blé avec ses différentes enveloppes. Source : Planet-Vie, 2017.

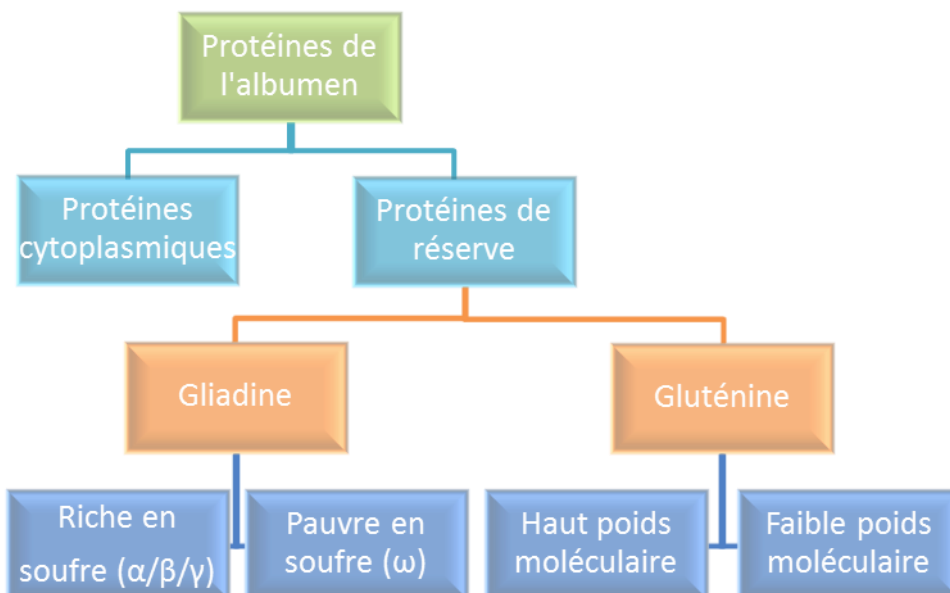


Figure 4 : Classification des protéines de l'albumen selon la classification d'Osborne (1907). D'après Casale (2015).

dans l'Union Européenne. En 2015, 150 Mt de blé ont été produites en Europe dont 37.5 Mt en France. Sur toute la production européenne, 23 % ont été exportées dont 1/3 en l'Europe et 2/3 vers des pays tiers. Malgré une situation exceptionnelle en 2016 avec une baisse de 24% par rapport à la moyenne 2011-2015, la culture de blé est une filière à enjeux pour la France (Agreste, 2016a).

Sur la part de production française exportée, qui représente environ 50% de la production totale, 80% sont utilisés en alimentation humaine. Pour la moitié restant en France, la première utilisation est l'alimentation animale, suivie par l'alimentation humaine et l'industrie. En fin de compte, environ 60 % de la production française est utilisée après une première transformation, la transformation meunière (Agreste, 2016b). En moyenne, chaque année, 4.12 Mt de farine sont produites en France. Le chiffre d'affaire du secteur de la meunerie en France s'élève à 1,77 milliard d'euro (ANMF, 2016). Il s'agit donc clairement d'un secteur qui présente un enjeu économique important.

3) Aptitude à la panification du blé tendre

L'aptitude à la panification dépend de la composition et de la teneur en protéines des farines issues de la mouture des grains de blé. Dans le grain, ces protéines sont localisées majoritairement dans l'albumen (Figure 3). Lors du mélange de la farine et de l'eau, l'interaction entre les protéines de réserve engendre la formation d'un réseau appelé gluten (Shewry et al., 2002). Le gluten est formé de 70% de protéines (les protéines de réserve), d'amidon et de lipides. La formation de ce réseau est due à la liaison des gliadines et des sous-unités de gluténines de haut poids moléculaire par des ponts disulfures (Figure 5 ; Branlard and Dardevet, 1985; Cauvain, 2003). Ce réseau détermine largement les propriétés rhéologiques (physiques) de la pâte comme la ténacité, la consistance, l'extensibilité et la force. C'est grâce au gluten que la pâte contient les gaz produits lors de la fermentation (CO₂) et engendre ainsi la levée de la pâte (Rousset, 1978).

Les protéines de réserve peuvent être classées en fonction de leur extractibilité selon la classification d'Osborne, (1907)(Figure 4). Cette classification permet de distinguer les gliadines qui représentent 40% des protéines de réserve et les gluténines qui constituent les 60% restants. Ces deux groupes ont par la suite été classés en fonction de leur poids moléculaire et de leur teneur en acides aminés soufrés.

- Les gluténines sont des molécules polymériques dont les sous-unités se décomposent en deux groupes : les molécules de haut poids moléculaire pauvres en soufre et celles de faible poids moléculaire riches en soufre. Les gluténines sont des molécules linéaires. Elles peuvent se lier par des ponts disulfures.
- Les gliadines sont des molécules monomériques classées en quatre groupes : ω -gliadines (pauvres en soufre), $\alpha/\beta/\gamma$ -gliadines (riches en soufre). Les protéines comprises dans ces classes sont nombreuses et diffèrent tant au point de vue de leur structure que de leur composition (Lindsay and Skerritt, 1999; Rasheed et al., 2012).

En ce qui concerne l'effet de ces différentes fractions protéiques sur les propriétés rhéologiques de la pâte, les gliadines apportent de la viscosité et de l'élasticité tandis que les gluténines sont responsables de la ténacité de la pâte (Méléard, 2014). Les gliadines apportent l'élasticité grâce aux nombreuses liaisons covalentes de type pont disulfure. Le caractère tenace est conféré par les liaisons hydrogènes de sous-unités de haut poids moléculaire de gluténines (Feillet, 2000).

La structure du réseau gluténique nécessite un équilibre complexe. Le blé doit avoir une quantité et une composition adaptée en protéine de réserve pour être panifiable (Branlard and Dardevet, 1985). Des études réalisées sur des farines riches en gluténines ou riches en

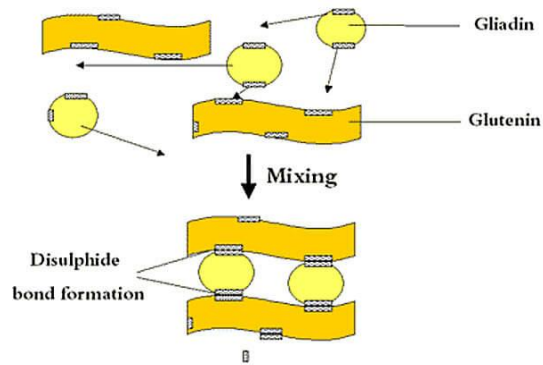


Figure 5 : Schéma de formation du réseau gluténique. Le mélange des gliadines et des gluténines entraîne la formation de ponts disulphures entre les protéines formant ainsi le gluten. Source : Perten Instruments, 2017.



Figure 6 : Alvéographe de Chopin de la marque CHOPIN Technologies™, 2017. L'appareil est équipé d'un pétrin qui permet la formation des pâtons (à gauche), d'un système de mise sous pression et d'un tambour enregistreur (à droite).

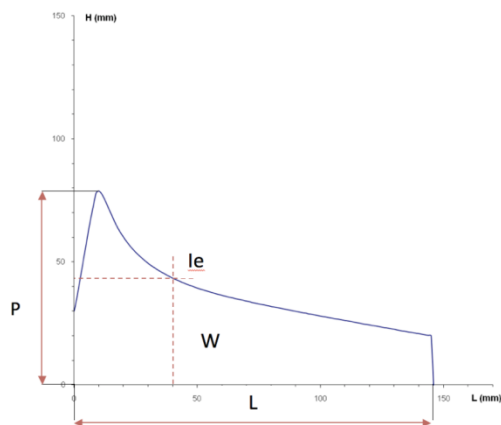


Figure 7 : Alvéogramme représentant le temps (en abscisse) et la pression (en ordonnée). La variable P représente la ténacité (en mm), L l'élasticité (en mm), le l'indice d'élasticité (sans unité) et W la force boulangère (en 10^{-4} J). Source : site internet Laboratoire CGAC

gliadines montrent que le pain qui en découle ne lève pas bien. Le ratio gliadine/gluténine est un élément important dans l'établissement de la qualité boulangère (Gupta et al., 1992; Weegels et al., 1996).

Les sous-unités de gluténines jouent un rôle particulièrement important dans l'établissement de la qualité boulangère (Branlard and Dardevet, 1985; Gupta et al., 1992). Il a été montré que la composition en gluténines est un indicateur de la qualité boulangère (Weegels et al., 1996) mais ne permet pas à elle seule de prédire la valeur en panification (Oury et al., 2010). Plusieurs allèles codant pour des sous-unités spécifiques de gluténines sont corrélés positivement à la qualité boulangère (Branlard et al., 2001). Bien que la part des sous-unités de gluténines de haut poids moléculaire ne représente que 10 % du total des protéines de réserve, ces protéines semblent être les plus corrélées à la qualité boulangère (Gupta et al., 1992; Wrigley, 1982; Wrigley et al., 2015). Cependant, il ne faut pas négliger non plus l'importance des sous-unités de faible poids moléculaire car leur perte engendre une baisse de qualité de la pâte (Zhen et al., 2014).

4) Détermination de la qualité boulangère

a. Définition et importance de la qualité boulangère

La notion de qualité change selon le devenir final du produit, ce qui la rend complexe à appréhender. De plus, la qualité est une notion subjective, les critères changeront si l'on se place du point de vue d'un industriel ou d'un consommateur mais également en fonction des débouchés. Nous nous intéressons ici aux critères de qualité boulangère représentant l'aptitude à la panification du blé. Ce type de qualité est mesurée par des critères technologiques (collant de la pâte par exemple), rhéologiques (visco-élasticité) et sensoriels (odeur de la mie par exemple ; Branlard and Autran, 1986a).

La qualité boulangère est un critère à fort impact sur le marché français. Il ne faut cependant pas confondre cette qualité technologique avec le taux de protéines. Les exigences du marché liées à l'export reposent sur le taux de protéines uniquement. Le taux minimum pour considérer un blé de bonne qualité est de 11.5% de protéines totales (Association générale des producteurs de blé et al., 2013). Le taux de protéines est un indicateur important de la force boulangère mais pas suffisant (Méléard, 2014). En effet, il n'existe pas de corrélation strictement positive entre la teneur en protéines et la qualité boulangère. Cette dernière dépend davantage du type de protéines présentes que de leur abondance.

b. Evaluation de la qualité boulangère

Méthodes indirectes

Il existe différents types de méthodes indirectes. La première consiste à mesurer la valeur rhéologique de la pâte (Feillet, 2000). Entre autres, l'alvéographe inventé par Chopin en 1927 (Figure 6) permet de mesurer l'extensibilité et la résistance à l'étirement de la pâte. Cette méthode est désormais uniformisée selon la norme AFNOR NF ISO 5530 (AFNOR, 2015).

L'alvéographe permet d'évaluer les propriétés d'élasticité et de ténacité de la pâte en mesurant la pression nécessaire pour déformer et faire éclater un échantillon de pâte. Pour réaliser ce test, des pâtons doivent être découpés de façon uniforme à l'aide d'un emporte-pièce. Un pâton est ensuite introduit dans le système de mise sous pression de l'alvéographe au sein de laquelle la température et l'hygrométrie sont contrôlées pour ne pas influencer sur la

Tableau 2 : Correspondance entre les paramètres de l'alvéographe et la valeur boulangère en panification française (Bouquelet, 2016)

Force W	Indice élasticité Ie	Ténacité P	Extensibilité L	Valeur boulangère
< 150	< 35	<40	< 20	Insuffisante
150 – 180	35 – 45	40 – 60	20 – 22	Moyenne
180 – 220	45 – 55	60 – 80	22 – 24	Bonne
> 220	> 55	> 80	> 24	Elevée

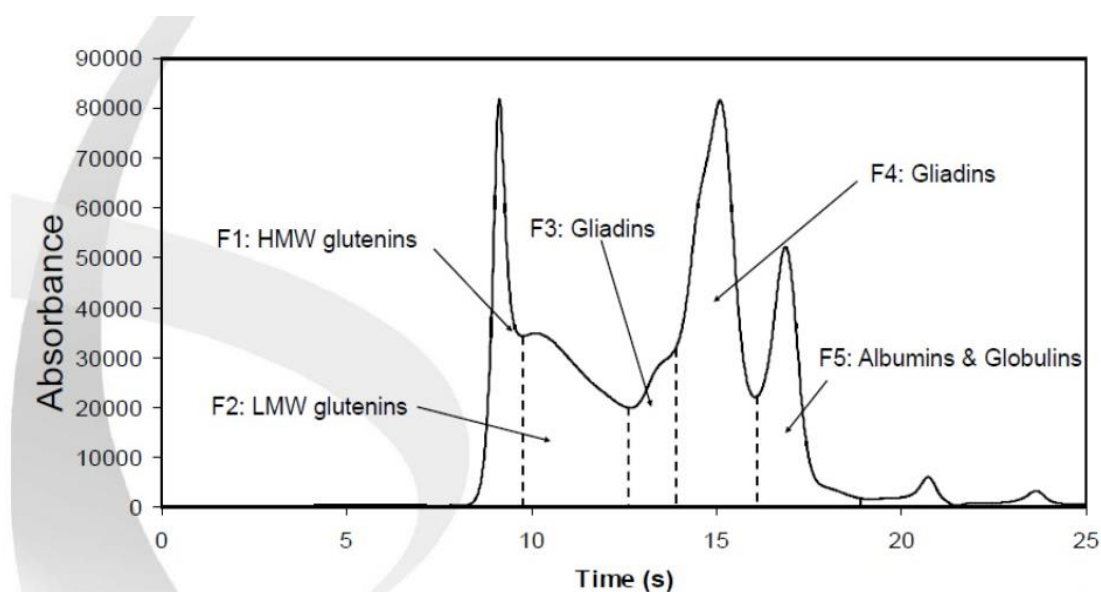


Figure 8 : Profil de HPLC (High Performance Liquid Chromatography ; mesure de l'absorbance en fonction du temps) des différentes protéines de l'albumen du grain de blé. Le premier pic correspond à la gamme d'absorbance des gluténines de haut poids moléculaire, le second aux gluténines de faible poids moléculaire, les deux suivants aux gliadines et le dernier aux albumines et globulines. Source : Wan et al., 2012

consistance de la pâte. De l'air est ensuite insufflée dans le pàton. La pression de l'air à l'intérieur du pàton est mesurée tout au long du test et transcrite par une courbe nommée alvéogramme (Figure 7). La première phase du test permet de mesurer la variable P (en mm) qui correspond à la tenacité. Lors de cette phase, l'air insufflé est retenu par la pâte et la pression augmente jusqu'à ce que le pàton se déforme (formation d'une « bulle » de pâte). Une fois que la pâte commence à se déformer, la pression à l'intérieur du pàton diminue. La bulle de pâte augmente progressivement de volume sous l'action de l'air jusqu'à ce qu'elle éclate et la pression chute alors brutalement. Cette phase permet la mesure de l'élasticité qui correspond au temps nécessaire pour faire éclater la bulle de pâte. L'élasticité est représentée par la variable L (en mm) sur l'alvéogramme. L'alvéogramme permet également d'obtenir la valeur de la force boulangère. Cette force, représentée par la variable W (en 10^{-4} J), est obtenue par le calcul de l'aire sous la courbe. Elle représente le cumul de pression qu'il a fallu appliquer sur le pàton afin de le faire éclater. Deux autres variables peuvent aussi être déduites de ce test : l'indice d'élasticité et de gonflement. L'indice d'élasticité I_e (sans unité) est obtenu par le calcul : P_{200}/P où P_{200} est la pression à 47 cm du début de la courbe, soit lorsque 200 cm^3 d'air ont été insufflés (seuil défini arbitrairement). Cet indice représente un point d'inflexion de la courbe et traduit l'élasticité de la pâte comme la variable L. L'indice de gonflement est liée à la variable L par la relation $G = 2,22\sqrt{L}$ (Cauvain, 2003; Dubois, 2016; Sahin and Sumnu, 2007).

Une autre méthode de mesure de la rhéologie de la pâte consiste à évaluer l'évolution des caractéristiques de la pâte au cours de sa formation. Les deux principaux appareils effectuant ce type de mesure sont le farinographe (Brabender™) et le mixographe. Le mixographe mélange la farine et l'eau en suivant un cycle élongation-rupture-relaxation (Cauvain, 2003). Il mesure et enregistre la résistance de la pâte lors du pétrissage. La courbe obtenue par cette méthode, appelée mixogramme, donne le même type d'information que l'alvéogramme, comme par exemple l'évolution de la tolérance de la pâte à l'extensibilité au cours de sa formation. Le farinographe est un pétrin enregistreur qui va permettre de mesurer la force nécessaire au pétrissage d'une pâte. La courbe obtenue, appelée farinogramme, permet en particulier d'apprécier la stabilité de la consistance de la pâte et l'intensité de son affaiblissement après une durée de pétrissage de 12min. (Feillet, 2000). La rhéologie de la pâte peut également être évaluée indirectement par l'indice de sédimentation, ou indice de Zényly, qui évalue la quantité de gluten. Il se mesure en centrifugeant la pâte et en évaluant la proportion de gluten retenue sur la grille de la centrifugeuse (Cauvain, 2003).

La teneur en protéines de réserve (gluténines, gliadines) est également un indicateur indirect de la qualité boulangère. Des études ont notamment démontré que la force boulangère augmente en même temps que la teneur en protéines de réserve (Stevenson, 1987; Uthayakumaran et al., 2000). Par ailleurs, la composition protéique que l'on peut mesurer par le ratio gliadine/gluténine, a un impact sur la force boulangère. Teneur et composition en protéines de réserve sont donc des indicateurs importants pour évaluer la qualité boulangère (Gupta et al., 1992). La technique de spectroscopie dans le proche infra-rouge (Near Infrared Reflectance Spectroscopy ; NIRS (Bokobza, 1998) est une méthode de détermination de la quantité de protéines. Cette méthode non destructive permet d'effectuer une bonne estimation de la quantité de protéines du grain (Feillet, 2000). Plusieurs méthodes existent pour évaluer la composition en protéines. La méthode de chromatographie en phase liquide à haute performance (high performance liquid chromatography ; HPLC ; Figure 8) permet de séparer les protéines en fonction de leur solubilité dans un solvant. Le seuil de solvabilité des protéines étant connu, cette technique permet de déduire la nature des protéines de réserves (Snyder et al., 2012).



Figure 9 : Exemple de fournée lors d'un test de panification au laboratoire Arvalis de Boigneville.

Tableau 3 : Echelle de notation du test de panification BIPEA allant de -7 à 10. La note 10 est attribuée lorsque le caractère est à l'optimum. Si le caractère est en excès, les notes +7, +4 et +1 peuvent être attribuées selon le degré d'excès. Si le caractère est en défaut, les notes -7, -4 et -1 peuvent être attribuées selon le degré de défaut (-1 étant attribué si le défaut est très marqué). Source : Roussel et al., 2010

INSUFFISANCE (-)				EXCES (+)		
- 1	-4	-7	10	+7	+4	+1
très insuffisant	insuffisant	un peu insuffisant		un peu excessif	excessif	très excessif
Intensité du défaut						
Très marqué	marqué	Peu marqué	normal	peu marqué	marqué	Très marqué

Méthodes directes

L'évaluation de manière directe consiste tout simplement à fabriquer du pain à partir de la farine et à évaluer différents critères lors de la fabrication du pain. Un exemple de fournée lors d'un test de panification au laboratoire Arvalis de Boigneville est en figure 9. Une trentaine de composantes (tableau 4) servent à élaborer 3 grandes notes (notes de pain, de pâte et de mie) allant de 0 à 100 dont la somme constitue la note de panification totale allant de 0 à 300. Les notes concernant la pâte permettent d'évaluer le développement, le rendement et la machinabilité (mesure de l'élasticité et de la résistance de la pâte pour le travail en machines) de la pâte (élasticité et tenue lors du pétrissage, collant au façonnage, activité fermentaire, tolérance de la mise au four...). Les notes relatives au pain concernent le développement du pain et son aspect extérieur (volume, couleur, croustillance...). Pour la mie, ce sont les qualités organoleptiques qui sont évaluées (alvéolage, odeur, souplesse...) (Roussel et al., 2010). Cette notation a été mise en place par la filière dans un souci d'harmonisation du contrôle de la qualité boulangère. Le Bureau Interprofessionnel d'Etudes Analytiques (BIPEA) a élaboré avec l'Association française de normalisation (AFNOR) la norme NF V03-716 définissant les règles d'attribution de la note de panification (Annexe 1, AFNOR, 2015). Ce test est communément appelé test BIPEA. Les conditions de préparation du pain représentent au mieux les pratiques des boulangers.

La notation choisie, assez complexe, est présentée dans le tableau 3. Les notes possibles sont 10, 7, 4 et 1. Les notes composantes obtiennent une valeur de 10 lorsque le critère est à son optimum. Plus la note est éloignée de 10, plus le critère est en défaut ou en excès. Un « - » est ajouté devant le chiffre lorsque le critère est en défaut, un « + » lorsqu'il est en excès. Ces notes sont pondérées différemment selon leur importance (Figure 10) et l'intensité du défaut. Ainsi, les notes principales (pain, pâte, mie) augmentent lorsque l'optimum de chaque composante est approché (Roussel et al., 2010).

Une harmonisation de la notation est nécessaire pour donner du sens à la norme. La majorité des notes ne sont pas des mesures mais bien des notes soumises à la subjectivité de l'expérimentateur. Les critères précis de notation sont définis dans la norme pour chaque composante des notes principales (Roussel et al., 2010). De plus, le BIPEA organise des regroupements ou des tests pour comparer les pratiques inter-instituts. Ainsi, les expérimentateurs harmonisent régulièrement leur façon de noter (BIPEA, 2017).

5) Facteurs influençant la qualité boulangère

a. Conduite de culture

Les protéines de réserve synthétisées et stockées par la plante dans les grains sont à l'origine de l'aptitude à la panification. Les protéines étant constituées d'azote, leur synthèse est directement impactée par la capacité de la plante à absorber l'azote et à le remobiliser vers les grains lors de leur remplissage (effet variétal). Les quantités et les dates d'apport en fertilisant jouent un rôle important sur la teneur en protéines des grains (effet des pratiques agronomiques). En particulier, un apport d'azote tardif est conseillé pour augmenter la teneur en protéines (Gate, 1995).

Les gluténines et la majorité des gliadines étant composés d'acides aminés soufrés, la fertilisation soufrée est également un levier agronomique pour augmenter la teneur en protéines de réserve. Au même titre que l'azote, les agriculteurs doivent tenir compte du moment où la plante absorbera et allouera le soufre aux protéines de réserve pour que la nutrition soit efficace (Zhao et al., 1999).

Tableau 4 : Liste des composantes de la note de panification (et leurs abréviations) selon l'étape dans la fabrication du pain et selon la note principale correspondante.

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

$$\begin{aligned}
 \text{Note de pâte} = & (0.5 \times LISP + \mathbf{COLP} \times 0.5 + EXTP \times 0.5 + ELAP \times 0.5 + \mathbf{RELP} + 0.5) \times \text{coef}(\ast) \\
 & + \text{REPO} + (\mathbf{EXTF} \times 0.5 + \text{DECF} \times 0.5 + ELAF \times 0.5 + \mathbf{COLF}) \times \text{coef}(\ast) \\
 & + (\text{ACTP} \times 0.5 + \text{DECA} \times 0.5) + (\mathbf{COMF} + \mathbf{TENU} \times 2) \times \text{coef}(\ast)
 \end{aligned}$$

$$\begin{aligned}
 \text{Note de pain} = & (\text{SECT} + \text{COUP} \times 2 + \text{EPAP} \times 0.5 + \text{CROP} \times 0.5 + \mathbf{DVCL} + \text{RECL} + \text{DECL}) \\
 & \times \text{coef}(\ast) + \text{VOL}
 \end{aligned}$$

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

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

Figure 10 : Formule 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).

b. Processus de panification

La qualité boulangère évolue suivant le type de farine utilisée pour faire le pain. Majoritairement, le pain en France est fabriqué à partir de farine blanche. Cette farine est obtenue par mouture du grain de blé sans son enveloppe (le son) ni la couche à aleurone. Or, ces deux tissus comportent des protéines et des enzymes qui ont tendance à diminuer la qualité boulangère. Les farines les moins épurées sont plus nutritives mais de moins bonne qualité technologique (Astier, 2016).

Les boulangers ont la possibilité d'améliorer la qualité de leur farine par ajout d'additifs si le réseau gluténique n'est pas assez bien formé. Le choix de l'additif à utiliser dépend du type de défaut de la pâte et de son intensité. Par exemple, afin d'augmenter la force boulangère, du gluten peut directement être ajouté à la pâte. L'ajout de farine de fève entraîne une augmentation du volume du pain. L'extensibilité de la pâte peut être augmentée par ajout de lécithine de soja ou d'oxydants. Une oxydation des groupements thiols (—SH) libres du gluten permet de former des ponts disulfures entre les chaînes de gluténines augmentant ainsi l'élasticité des pâtes en panification. Les oxydants utilisés en panification sont l'acide ascorbique ou les hydroperoxydes des farines de fève ou de soja (Partners and Food in Emergency and Development Aid, 1997).

c. Effet variétal

La qualité boulangère varie en fonction de l'environnement (conditions de culture) et de la variété utilisée. Les variétés diffèrent par leurs concentrations en protéines de réserve et par la composition de leurs sous-unités. Il est possible de différencier les variétés grâce à leur fraction protéique (Cauvain, 2003). Le choix variétal est un levier agronomique simple à mettre en place qu'il faut coordonner avec les pratiques culturales. Par exemple, la fertilisation soufrée améliore la qualité boulangère pour les variétés de qualité intermédiaire mais peut diminuer la qualité de variétés déjà de bonne qualité (Lévy and Brabant, 2014).

6) Déterminisme génétique de la qualité boulangère

La qualité boulangère est un caractère quantitatif déterminé par de nombreux gènes (ou QTL ; Quantitative Trait Loci) à effets faibles (caractère polygénique) et dépendant de facteurs environnementaux et d'interaction génotype × environnement. A ce titre, c'est un caractère complexe, difficile à évaluer et dont la décomposition en caractères plus « simples » peut permettre de mieux appréhender son déterminisme. Ainsi, la valeur d'usage de la farine étant dépendante de la composition et de la concentration en protéines de réserve (G. Branlard et al., 2001), l'étude du déterminisme génétique de ces caractères est un moyen d'étudier indirectement le déterminisme génétique de la qualité boulangère.

La synthèse des gluténines est contrôlée par 6 loci sur les chromosomes 1A, 1B et 1D (figure 11). Les protéines de haut poids moléculaire sont codées par un locus *Glu1* localisé sur le bras long de ces chromosomes tandis que les gluténines de faible poids moléculaire sont codées par un locus *Glu3* sur le bras court de ces chromosomes. Une partie des gliadines est contrôlée également par des loci (*Gli1*, *Gli3*, *Gli5*, *Gli6*) situés sur le bras court des chromosomes 1A, 1B et 1D (Branlard, 2012; Branlard and Autran, 1986b). Un autre loci, *Gli2*, impliqués dans la synthèse des gliadines se trouve sur les chromosomes 6A, 6B et 6D.

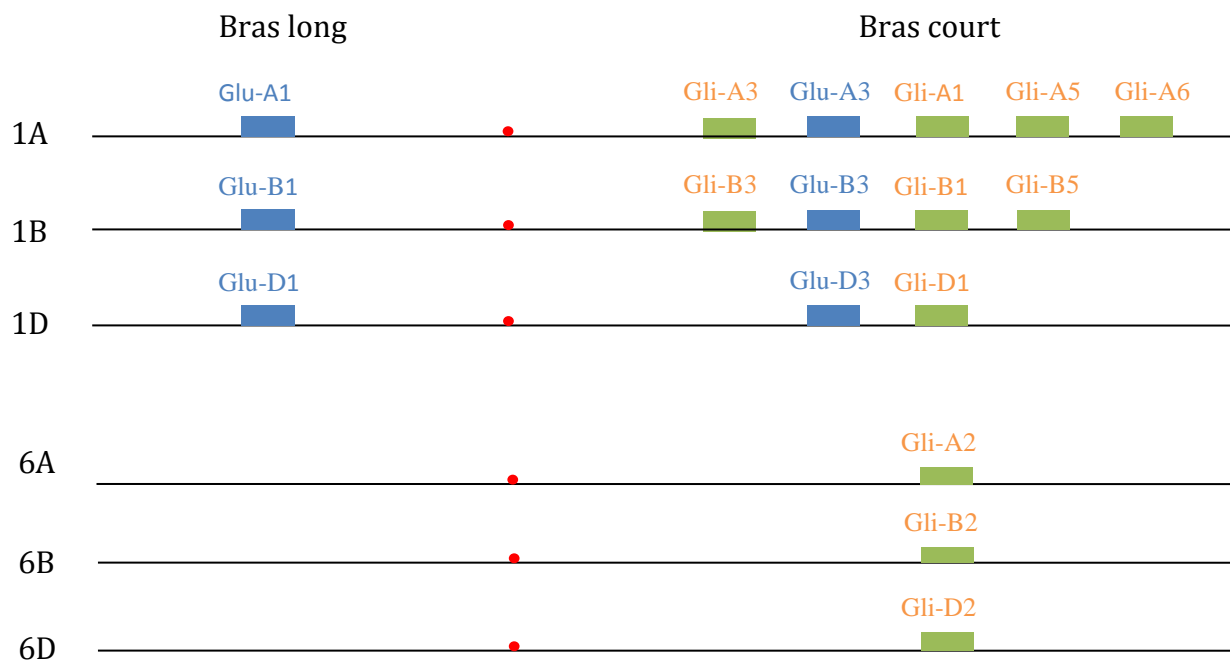


Figure 11 : Loci impliqués dans la synthèse des gliadines (Gli) et des gluténines (Glu). Les gènes de synthèse des gluténines sont localisés sur les chromosomes du groupe 1A, B et D. Les gènes de synthèse des gliadines sont localisés sur les chromosomes 1A, 1B, 1D et 6A, 6B et 6D. D'après Branlard & Autran (1986).

Les loci ont de multiples formes alléliques conduisant à des phénotypes différenciés au niveau rhéologique (Branlard and Le Blanc, 1985).

Le déterminisme génétique de la composition et de la concentration en protéines de réserve a été étudié par (Plessis et al., 2013). Les caractères phénotypiques analysés dans cette étude étaient liés à la composition et à la concentration en protéines représentant la part de gluténines et de gliadines dans la part totale des protéines de réserve mais aussi à la proportion de chaque sous-unité au sein de chacune de ces deux classes. 44 loci dispersés sur la quasi-totalité du génome ont été identifiés comme fortement associés à ces caractères. Il a été montré dans cette étude que les gènes régulant la composition des sous-unités d'une des deux classes de protéines (gluténines ou gliadines) sont différents et agissent indépendamment des gènes régulant la proportion des sous-unités de l'autre classe. Il apparaît donc possible de modifier indépendamment la composition en gluténines et en gliadines mais aussi la proportion des sous-unités de ces deux classes. Une série de transgène a été conçue avec un taux variable de sous-unités de gluténines à haut poids moléculaire (Feillet, 2000). La variation de ce taux engendre une variation de l'élasticité de la pâte. Ceci a permis d'identifier la composition en gluténines qui engendre une élasticité optimale.

D'autres facteurs génétiques déterminent la qualité boulangère. Ainsi, les gènes d'assimilation et d'absorption de l'azote (Yara, 2015; Zhao et al., 1999) déterminent indirectement la quantité et la qualité des protéines de réserve. Une analyse génétique réalisée sur les critères de qualité a mis en évidence une forte implication du chromosome 1B et des effets plus faibles sur les chromosomes 3A, 3B, 5B et 7D (Groos and Charmet, 2001).

7) Modélisation génétique de la qualité boulangère

a. Modélisation basée sur des QTL et sélection assistée par marqueurs

La connaissance des déterminants génétiques d'un caractère permet potentiellement de le modéliser à partir des QTL détectés. Les premiers travaux fondateurs d'utilisation de ce type de modélisation génétique ont été conduits par (Lande and Thompson, 1990) qui ont développé le concept de sélection assistée par marqueurs. Ce type de sélection utilise quelques QTL majeurs (fortement associés) afin d'accélérer la sélection de caractères complexes. Ces quelques QTL sont analysés dans les premières étapes du programme de sélection et permettent d'identifier les variétés qui possèdent les allèles favorables aux différents QTL ciblés. La sélection assistée par marqueurs a été utilisée dans des programmes nationaux en Amérique, Australie et Mexique sur le blé (Dubcovsky, 2004; Gupta et al., 2008).

L'étape préalable à la sélection assistée par marqueur est l'identification des QTL. Ceci peut être réalisé par génétique d'association (*Genome Wide Association Study, GWAS*). Cette méthode repose sur le fait qu'il existe un lien statistique entre les formes alléliques des QTL et la valeur phénotypique du caractère. La majorité des variétés ayant une bonne valeur phénotypique porteront la forme allélique favorable au QTL et inversement. Les QTL sont donc des zones génomiques associées statistiquement au caractère. Le lien statistique entre la forme allélique de tous les QTL et le caractère est étudié via des marqueurs génétiques répartis sur tout le génome.

L'utilisation des marqueurs pour rechercher des QTL exploite le principe de déséquilibre de liaison local. Le déséquilibre de liaison consiste en l'association non indépendante entre les allèles de deux zones génomiques ou *loci*. Cette statistique est calculée comme la différence

entre la fréquence des allèles des deux zones étudiées par rapport à celle issue d'une association aléatoire de ces allèles (égale au produit des fréquences des deux allèles de ces deux zones ; (Feingold, 1991). Les marqueurs en déséquilibre de liaison avec un QTL portent une information sur la forme allélique probable du QTL (Roldan, 2011). Une bonne couverture du génome est nécessaire pour que les marqueurs soient proches des QTL. La connaissance de la forme allélique d'un très grand nombre de marqueurs répartis sur tout le génome est donc nécessaire pour la détection des QTL. Grâce au développement récent de puces à ADN, l'analyse simultanée d'un grand nombre de marqueurs (plusieurs centaines de milliers) est désormais possible chez le blé (Chen et al., 2014).

Une fois que les QTL associés sont détectés, il est possible d'analyser leur forme allélique sur les lignées en sélection et d'éliminer les lignées possédant les formes alléliques défavorables.

b. Sélection génomique

L'efficacité de la sélection assistée par marqueurs peut être limitée par l'étape de détection des QTL. En effet, le niveau de résolution ne permet de détecter que les QTL à effet relativement fort. Or, le contrôle génétique des caractères quantitatifs est généralement dû à une multitude de QTL à effets faibles. Les QTL majeurs seront pris en compte rapidement et fixés (Moreau et al., 2011). C'est pourquoi, (Meuwissen et al., 2001) ont développé une méthode dite de sélection génomique qui permet de prédire la performance phénotypique d'un individu à partir de la totalité des marqueurs répartis sur le génome. La sélection se fait alors sur les valeurs phénotypiques estimées. Ce type de sélection permet de ne pas avoir à phénotyper la ou les populations en sélection et ainsi de limiter le coût du phénotypage.

La prédiction génomique nécessite une population de référence sur laquelle le modèle est calibré. Théoriquement, le modèle calibré peut ensuite être appliqué sur un autre panel. La calibration permet d'établir le lien entre l'information génétique et les valeurs phénotypiques selon un modèle statistique. Une fois le modèle calibré, les données de génotypage d'un individu à prédire permettront d'obtenir les valeurs phénotypiques estimées de cet individu. Les méthodes d'estimation par régression multiple ne sont pas adaptées à ce type de données. Le nombre de paramètres à estimer (effet de chaque marqueur) est largement supérieur au nombre d'observations. Cette perte de degré de liberté conduit à une impossibilité de résoudre le système (Moreau et al., 2011). Des méthodes adaptées ont été mises en place telles que la régression « ridge » et les méthodes bayésiennes. La méthode de régression « ridge », communément appelée rrBLUP (*ridge regression Best Linear Unbiased Predictor*), repose sur l'hypothèse que tous les QTL ont un effet identique et faible (modèle infinitésimal). Les marqueurs sont tous considérés comme expliquant une même part de variance. La méthode kinship-BLUP qui découle de la méthode précédente consiste à prédire les valeurs phénotypiques à partir de la ressemblance génétique entre les lignées. Cette méthode estime directement les génotypes de chaque individu à partir de la matrice d'apparentement (kinship) sans passer par l'estimation des marqueurs. Il existe un lien fort entre la méthode rrBLUP et la méthode kinship-BLUP (Moreau et al., 2011). Ces deux méthodes supposent que les QTL ont un effet infinitésimal. Or, pour certains caractères, la détection de QTL met en évidence des zones génomique de QTL à effet fort. D'autres méthodes existent (ex : méthodes bayésiennes) et permettent de mieux refléter les différences d'architecture génétique entre caractères

La qualité de prédiction est estimée par le calcul de la corrélation entre les valeurs phénotypiques prédites et les valeurs observées. Comme la GWAS, la sélection génomique repose sur le principe de déséquilibre de liaison. Un fort déséquilibre de liaison entre un marqueur et un QTL représente une information stable dans le temps et à travers les générations, ce qui assure son utilisation pour la sélection génomique (Liu et al., 2015).

L'efficacité de la sélection génomique est fortement liée à l'héritabilité. L'héritabilité représente la proportion de la variance génétique expliquant la variance phénotypique, c'est-à-dire l'importance des effets génétiques dans l'établissement d'un caractère phénotypique. Il permet donc d'estimer la réponse à la sélection.

La sélection génomique a fait ses preuves chez les bovins laitiers (Goddard and Hayes, 2007). Chez le blé tendre, la sélection génomique a été testée pour le rendement (Crossa et al., 2010), le poids des grains, la date de floraison (Charmet et al., 2014) en utilisant dans les deux études différentes méthodes de prédiction (rrBLUP ou méthodes bayésiennes). Les précisions pour ces différents caractères vont de 0.35 à 0.60. Une comparaison de 10 méthodes de prédiction a été effectuée et montre une faible différence de résultats entre les méthodes (Heslot et al., 2012). La sélection génomique sur des critères de qualité tels que la capacité de rétention en eau ou la teneur en protéines conduit à des précisions de 0.27 à 0.68 avec la méthode rrBLUP (Heffner et al., 2011). A ce jour, aucune étude n'a testée la capacité de prédiction génomique sur le test de panification.

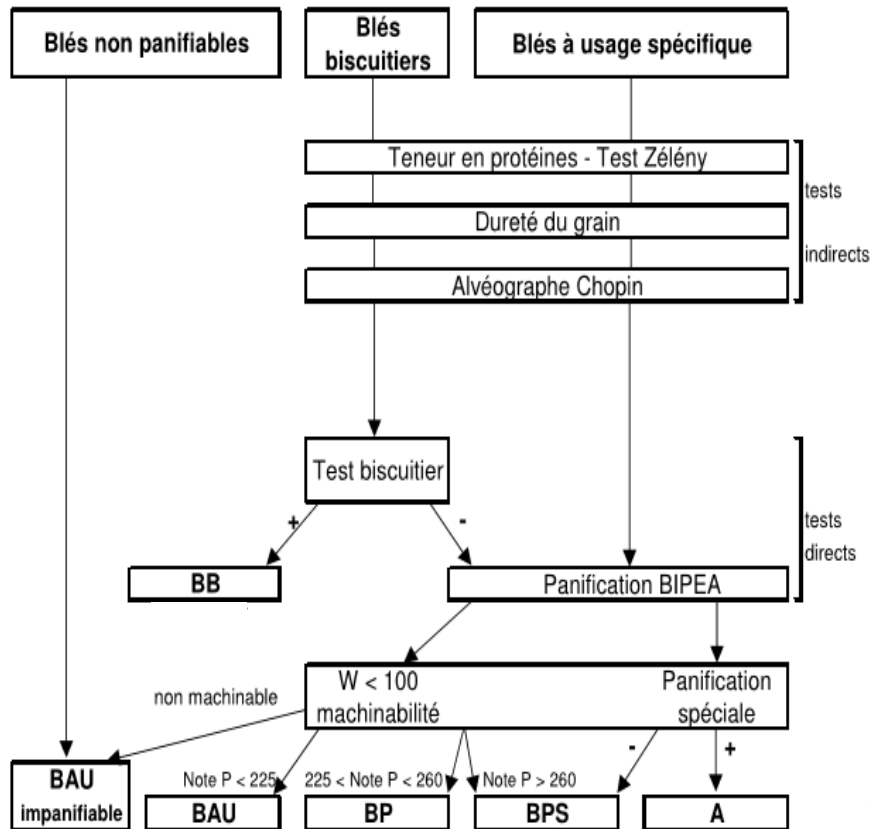


Figure 12 : Schéma des tests effectués lors de l'évaluation des variétés candidates à l'inscription au catalogue officiel Français par le Comité Technique Permanent de la Sélection (CTPS). Source : Lecomte, 2005

II. Problématique de recherche

Lors de l'inscription au catalogue officiel Français par le Comité Technique Permanent de la Sélection (CTPS), les variétés de blé sont classées suivant 4 catégories : blé améliorant ou de force (BAF), blé panifiable supérieur (BPS), blé panifiable (BP) et blé à autres usages (fourrager ou biscuitier ; BAU) (voir Figure 12). Les variétés candidates à l'inscription au catalogue font l'objet de deux années d'expérimentation. Pour être inscrites au catalogue officiel, les nouvelles variétés doivent être distinctes des variétés déjà inscrites, homogènes et stables, c'est-à-dire garder leurs caractéristiques phénotypiques de génération en génération. Elles doivent également posséder une valeur agronomique, technologique et environnementale suffisante par rapport aux variétés les plus utilisées du moment. Une série d'essais agronomiques est effectuée visant à évaluer leurs performances et à les comparer à des variétés témoins. Les critères d'évaluation comprennent le rendement, les facteurs de régularité du rendement (la tolérance aux maladies, précocité d'épiaison, résistance au froid...) et la qualité d'usage (Feillet, 2000). Afin d'évaluer la qualité d'usage, la mesure de la teneur en protéines, de la dureté du grain, de l'indice de Zélény et du test d'alvéographe sont effectués lors de la première année d'évaluation. Ces premiers tests permettent d'effectuer un classement provisoire. Suite à ce classement, en deuxième année d'évaluation, le test de panification est effectué sur les variétés potentiellement panifiables. Les autres variétés sont soumises au test biscuitier ou au test de machinabilité européen pour les BAU. Suivant les résultats de cette seconde série de test, certaines variétés peuvent être sur ou sous classées et soumises au test de leur nouvelle classe (Lecomte, 2005). La note totale de panification servira au CTPS à évaluer l'aptitude à la panification de la variété considérée. Les blés classés en BPS doivent obtenir une note de panification supérieure à 250 (FranceAgriMer and Arvalis, 2014).

Du point de vue de l'agriculteur, le choix variétal est un levier agronomique simple à prendre en compte et potentiellement efficace à moindre coût. Pour les sélectionneurs, ce critère n'est cependant pas simple à intégrer dans les programmes de sélection. Le test direct de panification requiert quelques kilos de farine qu'il n'est possible d'obtenir qu'à la fin d'un programme de sélection (en génération F6 ou F7). Ce test est par ailleurs coûteux en argent (environ 120€; (Paulic Minotiers, 2016) et en temps (Kleijer, 2002). Le test d'alvéographe de Chopin est moins cher (environ 35€) et requiert une plus petite quantité de farine (200g) mais ne reflète pas entièrement la qualité boulangère et ne peut donc pas remplacer le test de panification. La teneur en protéines, qui pourrait servir d'indicateur de la qualité, est fortement impactée par l'environnement ce qui rend la sélection sur ce critère difficile (Branlard and Autran, 1986b). Par ailleurs, il n'existe pas de lien direct en la teneur en protéines et la qualité boulangère car la composition en protéines est également un facteur clé.

L'utilisation de la génétique pour prédire les performances des variétés faciliterait l'intégration de ce critère dans les programmes de sélection. En effet, l'information génétique peut être obtenue dès les premières générations dans les programmes de sélection (un échantillon de feuille suffit à réaliser un test ADN) ce qui permettrait d'augmenter l'intensité de la sélection pour la qualité boulangère à moindre coût.

III. Objectifs du stage

Le premier objectif de mon stage était d'identifier des déterminants génétiques de la qualité boulangère. L'identification de QTL majeurs est une information importante pour les sélectionneurs souhaitant pratiquer la sélection assistée par marqueurs pour la qualité boulangère. Ces études de génétiques d'association permettent en outre d'avoir une idée de l'architecture génétique de la note de panification (nombre de QTL impliqués, effets des QTL, présence de QTL majeur ...). Cela peut permettre de définir le formalisme d'un modèle de prédiction qui serait le plus adapté à ce caractère. Enfin, grâce à l'identification des zones de QTL fortement associés aux caractères, il est possible de comparer les zones de contrôle génétique de la qualité avec d'autres zones renseignées dans la littérature comme par exemple celles des gènes de gluténines.

Le second objectif de ce stage était de calibrer un modèle de prédiction génomique puis de valider ses capacités prédictives par une méthode de validation croisée et par une validation sur un panel de variétés ou de lignées indépendantes. Ce travail a été réalisé pour la note de panification et l'ensemble de ces composantes. L'objectif était bien sûr de rechercher un modèle combinant à la fois robustesse et bonnes capacités prédictives.

Enfin, l'opportunité d'utiliser des covariables issues de tests indirects (alvéographe Chopin) dans le modèle de prédiction a également été étudiée. Il peut en effet être intéressant de valoriser les variables issues de ce test indirect (qui est beaucoup plus facile à mettre en œuvre et moins coûteux qu'un test de panification) afin de prédire l'aptitude à la panification des variétés / lignées de blé.

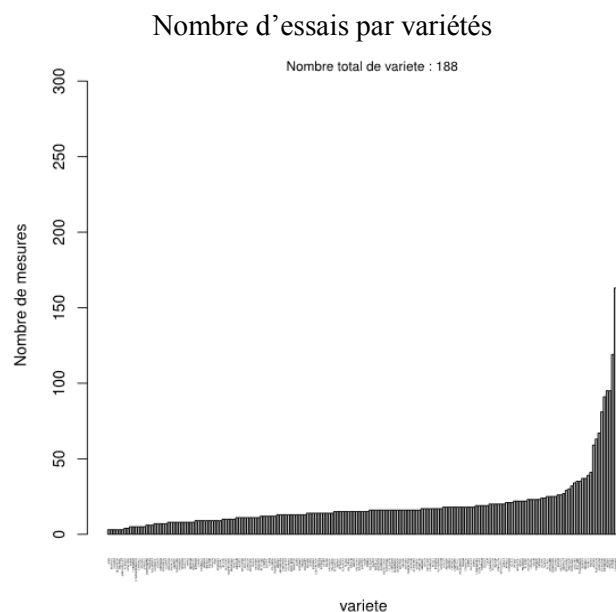


Figure 13 : Nombre d'essais par variétés.

Tableau 5 : Nombre de lignées testées par lieux (en ligne) et par année (en colonne) pour le panel Limagrain avec le total du nombre de variétés par année.

	2014	2015
CYRV	378	0
CVTA	384	157
CVCV	0	184
VRPE	0	209
VRVR	0	217
TOTAL	762	767

IV. Matériels et méthodes

1) Matériel végétal

a. Données ARVALIS Institut du végétal

Le jeu de données fournis par ARVALIS Institut du végétal est un regroupement de données acquises sur 11 ans dans le cadre de l'activité d'évaluation post-inscription des variétés. Le panel constitué à partir de ces données contient 188 variétés inscrites de blé tendre testées pour leur qualité boulangère. Ces données n'ayant pas été acquises dans le but de former un panel homogène, les conditions environnementales et de culture sont extrêmement diverses. Ce panel est nommé « panel Arvalis » dans ce rapport.

Les caractères mesurés sur ce panel sont les variables d'alvéographe, 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 étudiées ont été obtenues au laboratoire d'ARVALIS Institut du végétal de Boigneville.

Les variétés ont été testées en moyenne dans 20 expérimentations (médiane à 15, figure 13). Ces mesures ne sont pas des répétitions car toutes les modalités environnementales varient. Les données de ce panel ont été récoltées entre 2003 et 2013 avec des variétés communes entre les années (annexe 4). En moyenne, 57 variétés ont été testées par an (médiane de 64). Le jeu de données comporte 158 lieux de récolte différents avec en moyenne 13 variétés par lieux (médiane à 6). Les modalités de culture prises en compte dans l'analyse portent sur l'irrigation (pluvial ou irrigué) et la fertilisation azotée. En moyenne, 42 variétés ont été testées par modalité (médiane de 28).

a. Données Limagrain

Le jeu de données du programme de sélection de Limagrain est composé de 1529 lignées en sélection. Ces lignées ont été obtenues par haplo-diploïdisation, cette technique permet de fixer rapidement le matériel génétique (matériel homozygote). Les caractères étudiés pour ce panel sont les notes de panification obtenues selon la méthode BIPEA par le laboratoire QUALTECH. Le taux de protéines, le taux d'hydratation et le volume du pain ont également été mesurés. A la différence du panel Arvalis, ce jeu de données n'a pas été testé pour les variables d'alvéographe.

A l'origine du projet, les lignées devaient être testées dans 4 lieux identiques chaque année. En 2014, les contraintes expérimentales ont conduit à la perte de données sur 3 lieux. Cependant, des données de tests dans un autre lieu dans le sud de la France (Revel, Haute-Garonne) ont pu être récupérées pour compenser cette perte (Figure 5). Cette inclusion de nouvelles lignées a mené à l'incorporation d'un biais dans le jeu de données. En effet, les lignées testées sur ce lieu avaient été sélectionnées pour leur bon potentiel de performances vis-à-vis de la qualité boulangère. Les lignées intégrées ont donc globalement une bonne qualité boulangère.

Les tests ont été réalisés en présence de 6 témoins (Tableau6). Pour chaque essai (lieu × année), 4 témoins ont été testés avec 2 répétitions. Aucune variété témoin n'a été testée deux

Tableau 6: Variétés témoins (en ligne) présentes par lieux et par année (en colonne). Une croix représente la présence du témoin dans l'essai, une case vide représente l'absence.

		2014		2015			
Lieux Témoins		CVTA	CYRV	CVTA	CVCV	VRPE	VRVR
	Apache		X	X	X	X	X
Arezzo		X	X	X	X	X	X
Cellule				X	X	X	X
Rubisko				X	X	X	X
Solehio		X	X				
Symoïsson		X	X				

	Insuffisance			Optimal	Excès		
Score BIPEA	-1	-4	-7	10	7	4	1
Recodage	1	2.5	4	5.5	7	8.5	10

Figure 14 : Présentation du recodage effectué pour la notation des composantes du test BIPEA. Le recodage va de 1 à 10. Si le critère est à l'optimum, la note attribuée sera 5.5.

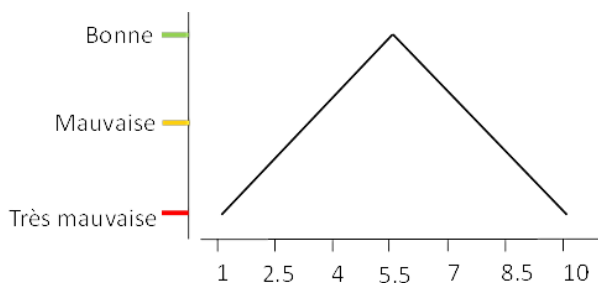


Figure 15 : Correspondance entre la valeur en panification d'un critère (en ordonnée) et sa note pour le test BIPEA recodé comme vu dans la figure précédente.

fois sur le même lieu. En revanche, deux variétés témoins ont été testées durant les deux années. Les variétés témoins sont des variétés inscrites dont les performances sont bien connues. Les tests sur les lignées en sélection n'ont pas été répétés, les caractères ont été mesurés une fois par lignée.

2) Préparation des données phénotypiques

Les notes composantes du test de panification sont notées sur un barème non linéaire comportant les scores -1, -4, -7, 10, 7, 4 et 1 (Tableau 3), avec 10 le score qui traduit le niveau optimal de chaque variable. Pour des raisons arithmétiques, il a été nécessaire de linéariser cette notation et les notes ont été recodées de 1 à 10 (Figure 14). Après transformation, l'optimum est à 5.5 et les notes traduisant un défaut (-1, -4 et -7) sont inférieures à 5.5 tandis que les notes traduisant un excès (+7, +4 et +1) sont supérieures à 5.5. Ainsi, l'échelle de notation reflète une variable selon une échelle croissante (par exemple, la note croît avec le niveau d'extensibilité de la pâte). Il faut cependant noter que ce nouveau codage n'est pas linéaire vis-à-vis de la valeur en panification. En effet, pour obtenir une bonne valeur en panification, la note qui traduit l'optimum est 5.5 (Figure 15). Cela a des conséquences dans l'interprétation des résultats de certaines analyses.

Des statistiques descriptives ont permis de réaliser une analyse qualitative des données afin de repérer des valeurs aberrantes. Ces analyses ont également permis d'évaluer la variance de chaque critère.

Afin de corriger les variations phénotypiques dues aux facteurs environnementaux, les moyennes par variétés pour les deux panels ont été calculées en prenant en compte un effet « environnement ». Les facteurs environnementaux ont été regroupés pour créer une nouvelle modalité « Essai ». Cette nouvelle modalité comprend les facteurs « année », « conduite » et « lieu » pour le panel Arvalis et les facteurs « année », et « lieu » pour le panel Limagrain. Les moyennes ajustées ont été calculées en considérant les facteurs génotype et essai comme aléatoires selon la méthode BLUP (*Best Linear Unbiased Predictor*). Le modèle pour la variété i et l'essai k s'écrit ainsi:

$$Y_{ik} = \mu + \underline{G}_i + \underline{E}_k + \underline{\varepsilon}_{ik}$$

Avec Y la valeur phénotypique de l'individu pour le trait étudié, μ la moyenne générale du trait étudié, G l'effet variété, E le facteur essai et ε la résiduelle. Les composantes soulignées étaient en effet aléatoire.

Après ajustement, une nouvelle étude de répartition des observations a mis en évidence une faible variabilité pour certains caractères. Les variables ayant une variabilité inférieure à 0.04 et les variables pour lesquelles au moins 90 % des variétés prennent la même note ont été écartées.

3) Données génotypiques

Les lignées ont été génotypées avec une puce à ADN développée par Limagrain comprenant 14 592 marqueurs SNP. Le taux de valeurs manquantes maximum accepté par marqueur a été fixé à 30% (Figure 16). Les SNP monomorphes, les marqueurs avec un seul type d'allèle homozygote ainsi que ceux possédant un taux d'hétérozygotie supérieur à 10% ont été supprimés. Les marqueurs avec une fréquence de l'allèle minoritaire inférieure à 5% ont été écartés. Au total, les analyses ont été effectuées sur 188 variétés avec 10 482

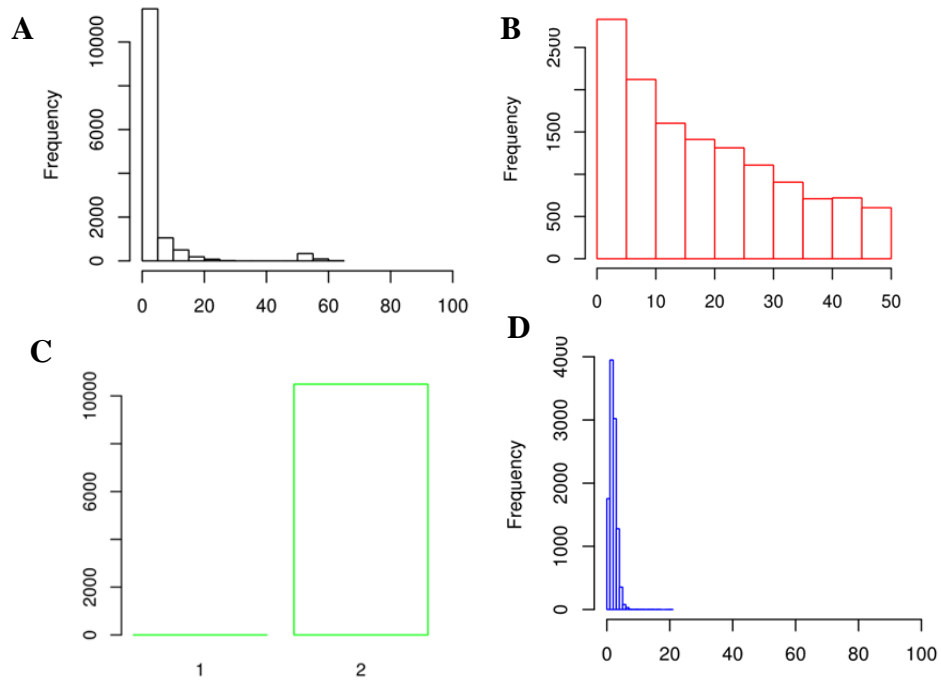


Figure 16 : Analyse qualité des données génotypiques du panel Arvalis.

A : Fréquence des valeurs manquantes des marqueurs SNP du panel Arvalis. La fréquence est en ordonnée et le taux de valeur manquante (en %) est en abscisse.

B : Fréquence de l'allèle minoritaire des marqueurs SNP du panel Arvalis. La fréquence est en ordonnée et le taux de présence de l'allèle minoritaire (en %) est en abscisse.

C : Fréquence des marqueurs SNP du panel Arvalis n'ayant qu'un seul ou deux types d'homozygote.

D : Fréquence du taux d'hétérozygotie des marqueurs SNP du panel Arvalis. La fréquence est en ordonnée et le taux d'hétérozygotie (en %) est en abscisse.

marqueurs pour le panel Arvalis et 1529 lignées avec 10 802 marqueurs pour le panel Limagrain.

Les allèles des marqueurs ont été codés en valeurs numériques pour les analyses. Le recodage est différent suivant les analyses. Pour les analyses de génétique d'association, les marqueurs homozygotes pour l'allèle majoritaire dans un panel ont été codés en 2, les marqueurs hétérozygotes en 1 et les marqueurs homozygotes pour l'allèle minoritaire en 0. Pour les analyses de prédiction génomique, les marqueurs homozygotes pour l'allèle majoritaire dans un panel ont été codés en 1, les marqueurs hétérozygotes en 0 et les marqueurs homozygotes pour l'allèle minoritaire en -1.

4) Analyse descriptive des données

La structuration phénotypique et génotypique de ces jeux de données a été évaluée par des Analyses en Composante Principale (ACP) en utilisant une fonction provenant du package FactoMineR (Lê et al., 2008; Patterson et al., 2006). La structuration du panel Arvalis a été également étudiée en fonction de l'organisme obtenteur des variétés.

L'héritabilité de chaque caractère pour les deux panels a été estimée selon la formule :

$$h^2 = \frac{VarG}{VarG+VarR}$$

avec VarG la variance génétique et VarR la variance résiduelle. Ces variances ont été obtenues à l'aide du modèle linéaire mixte utilisé pour le calcul des moyennes ajustées. La variance résiduelle inclut la part de variance environnementale qui n'a pas été prise en compte dans le modèle d'ajustement des moyennes. Cette grandeur représente donc le ratio entre la variance génétique et la variance due aux interactions génotype x environnement (non prise en compte de manière explicite dans le modèle) d'une part et aux erreurs de mesures d'autre part. En particulier, elle nous renseigne sur la stabilité du classement des variétés/lignées pour un caractère donné.

Le déséquilibre de liaison a été évalué par le coefficient de corrélation calculé entre deux marqueurs une fois les scores génotypiques transformé en valeurs numériques. Il est de plus en plus proche de 0 lorsque les allèles des deux loci tendent à être répartis aléatoirement. (Hill and Robertson, 1968)

5) Recherche des déterminants génétique de la qualité boulangère

a. La méthode mixed model (MM) forward

Les études d'association pour identifier des QTL associés aux caractères ont été réalisées selon le modèle mixte développé par Yu et al. (2006). Le modèle de régression mixte consiste à tester un par un l'effet de chaque SNP en effet fixe en incluant dans le modèle un effet dû au fond génétique. L'effet dû au fond génétique est apporté par la matrice d'apparentement (ou kinship). La matrice d'apparentement, calculée sur la base des données de génotypage, traduit la proximité des variétés entre-elles. Elle a été obtenue selon la méthode *identical-by-state* (IBS)(Tamura et al., 2013). Cette méthode consiste à calculer le pourcentage d'allèle en commun entre deux lignées. La kinship est obtenue en multipliant la matrice de génotypage par sa transposée. Pour l'individu i , le modèle s'écrit de la façon suivante :

$$Y_i = \mu + x_i\beta + \underline{G}_i + \underline{\varepsilon}_i$$

Avec Y le phénotype, μ la moyenne phénotypique, $x\beta$ le terme de l'effet du marqueur testé en effet fixe, G la matrice d'apparentement représentant le fond génétique en effet aléatoire et ε la résiduelle.

Cette méthode a été effectuée avec la fonction GWAS du package rrBLUP (Endelman, 2011) sous R. Ce modèle permet d'intégrer la matrice kinship seule ou avec la matrice de structure. L'inclusion de la matrice d'apparentement et de la matrice de structure permet de diminuer le taux de faux positif et faux négatif. (Hoffman, 2013; Yu et al., 2006) Les études de structure des panels n'ayant pas conduit à la mise en évidence d'une structuration forte, la matrice de structure n'a pas été intégrée dans le modèle de GWAS.

La multiplicité des tests effectués lors d'une GWAS implique d'utiliser un seuil de significativité corrigé, adapté aux tests multiples. La correction de Bonferroni, usuellement utilisée, revient à diviser le seuil fixé pour un test simple (5% usuellement) par le nombre de test à réaliser (ici, le nombre de marqueur). Cette correction est cependant très conservatrice (tend à réduire très fortement le nombre de marqueurs significatifs). Une alternative proposée par (Li and Ji, 2005) consiste à estimer le nombre de marqueurs indépendants et à utiliser ce nombre pour effectuer une correction du seuil de significativité. Ce travail a été effectué à l'aide d'un script développé par Renaud Rincant, chargé de recherche à l'UMR INRA GDEC. Le seuil obtenu est de $-\log_{10}(\text{p-value}) = 4.78$.

Cette méthode a été appliquée de manière itérative afin de bâtir un modèle de régression *forward*. A chaque itération, l'ensemble des SNPs sont testés avec le modèle décrit précédemment et le marqueur le plus significatif est inclus en effet fixe dans le modèle. L'algorithme stoppe lorsqu'aucun marqueur ne dépasse le seuil de significativité (ici, $-\log_{10}(\text{p-value})=4.78$). Ce type de régression permet de révéler des marqueurs qui auraient été masqué par la présence de QTL fortement associés.

Les résultats de cette étude ont ensuite été représentés graphiquement avec MapChart (Voorrips, 2002). Cette représentation a permis de visualiser directement les zones d'intérêt :

- les zones robustes (co-localisation de QTL détectés pour le même caractère dans les deux panels),
- les zones pleïotropes (co-localisation de QTL associés à différents caractères).

A cela s'ajoute les QTL à effet fort qui constituent également des zones d'intérêt. L'effet allélique des QTL a été estimé en calculant la différence entre la moyenne des individus homozygotes pour l'allèle majoritaire et la moyenne des individus homozygotes pour l'allèle minoritaire. Le recodage des notes de panification engendre une interprétation particulière des effets alléliques des QTL. En effet, la notation est linéaire vis-à-vis de l'intensité du caractère mais pas vis-à-vis de la valeur en panification (les valeurs variant de 1 à 10 mais avec un optimum à 5.5 pour la panification). Les QTL augmentent ou diminuent l'intensité du caractère (de l'extensibilité par exemple) mais on ne peut pas parler de QTL qui améliorent la performance au sens stricte. Dans ce cadre, les allèles seront « bons » ou « mauvais » selon la moyenne phénotypique de la population pour ce caractère.

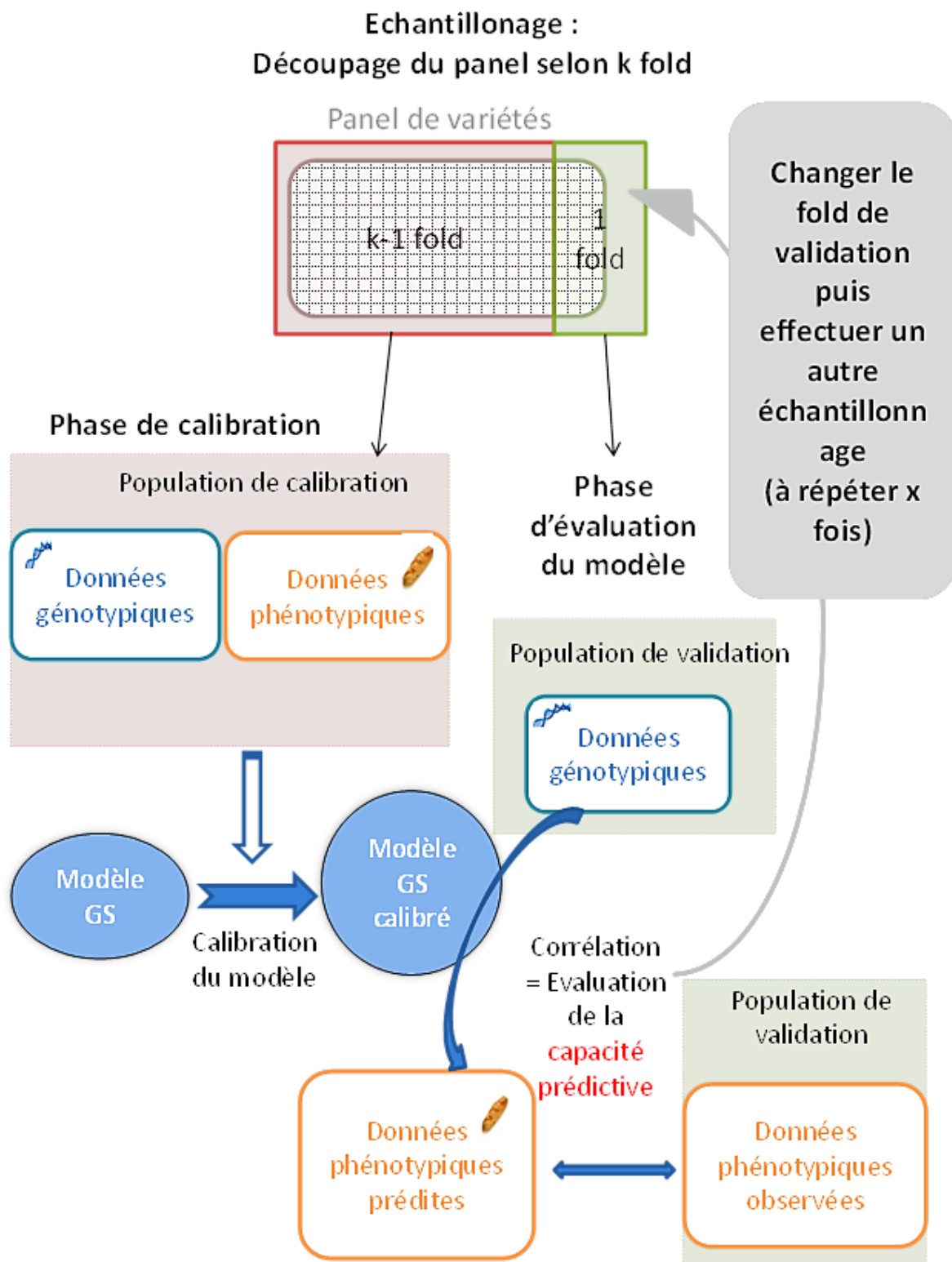


Figure 17 : Schéma représentant les étapes de validation croisée afin d'évaluer la prédiction génomique. Le panel de variétés est d'abord découpé en k sous-parties dont k-1 servent à calibrer le modèle et la sous-partie restante sert à valider le modèle en comparant les valeurs phénotypiques prédites par le modèle et les valeurs phénotypiques observées.

6) Prédiction génomique de la qualité boulangère

a. Modèles utilisés

Modèle de prédiction génomique

Le modèle de prédiction génomique utilisé était le modèle kinship-BLUP, également appelée GBLUP ((Meuwissen et al., 2001). Cette méthode d'estimation s'appuie sur l'hypothèse que les marqueurs ont tous un effet identique et faible. La méthode kinship-BLUP établit les prédictions à partir de la matrice d'apparentement. Les données de génotypage sont utilisées pour calculer une matrice d'apparentement. Cette matrice permet de prendre en compte la covariance entre les individus. (Gondro et al., 2013) ainsi que les effets non additifs tels que l'épistasie des QTL (Vieira et al., 2017). Le modèle de GBLUP s'écrit pour un individu i :

$$Y_i = \mu + G_i + \varepsilon_i$$

Avec Y le phénotype, μ la moyenne phénotypique, G la kinship en effet aléatoire et ε la résiduelle. Un terme de covariable peut être ajouté en effet fixe à ce modèle. (Endelman and Jannink, 2012). Dans notre étude, la prédiction génomique a été effectuée en 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 la matrice d'apparentement selon la méthode IBS.

Le modèle multi-locus mixed-model (MLMM)

Un des défauts majeurs de l'approche de sélection génomique présentée ci-dessus est de faire l'hypothèse implicite que l'ensemble des marqueurs génétiques ont un effet égale sur la qualité en panification puisqu'ils sont tous utilisés avec le même poids pour le calcul de la matrice d'apparentement. Or, il paraîtrait plus judicieux d'utiliser les marqueurs associés à des QTL de qualité boulangère pour estimer la proximité entre lignées et bâtir un modèle de prédiction génomique.

Il a donc été testé l'opportunité d'utiliser les marqueurs associés à des QTL pour bâtir un modèle GBLUP en utilisant la méthode multi-locus mixed-model (MLMM ; (Segura et al., 2012) pour rechercher des QTL par génétique d'association. Cette méthode avait par ailleurs été utilisée avec succès par des partenaires du projet. Cette méthode permet de sélectionner un marqueur significatif par zone de QTL associée au caractère.

Elle repose sur le même modèle que la méthode MM. C'est une méthode itérative mais l'itération n'est pas réalisée de la même manière. La régression effectuée est de type «*stepwise forward-backward*». La régression *forward* consiste à ajouter en cofacteur les marqueurs expliquant le plus de variance. A chaque étape, les marqueurs sont testés indépendamment. Celui qui possède la p-value la plus basse, donc celui qui est le plus associé, est inclus dans le modèle en cofacteur. Les marqueurs sont ensuite testés de nouveau avec le nouveau modèle incluant le marqueur en cofacteur. L'étape *forward* se termine lorsque la variance résiduelle est proche de zéro. Une régression *backward* est ensuite effectuée. Elle élimine un par un les marqueurs les moins associés. Pour chaque modèle, un critère statistique est calculé pour évaluer la vraisemblance du modèle. Le modèle MLMM propose deux critères : *multiple Bonferroni criterion* (mBonf) développé par (Segura et al., 2012) et l'*extended BIC* (Chen and Chen, 2008). Ces critères permettent de comparer tous les modèles testés lors de la régression stepwise avec les différents marqueurs en cofacteur et de déterminer le modèle le plus vraisemblable. Le critère mBonf est moins conservateur que l'*extended BIC* : tout en gardant un taux d'inclusion de faux positifs raisonnable, il discrimine moins de marqueurs dans le modèle (Segura et al., 2012). C'est ce critère qui a été utilisé dans

I - Echantillonnage : Découpage du panel selon k fold

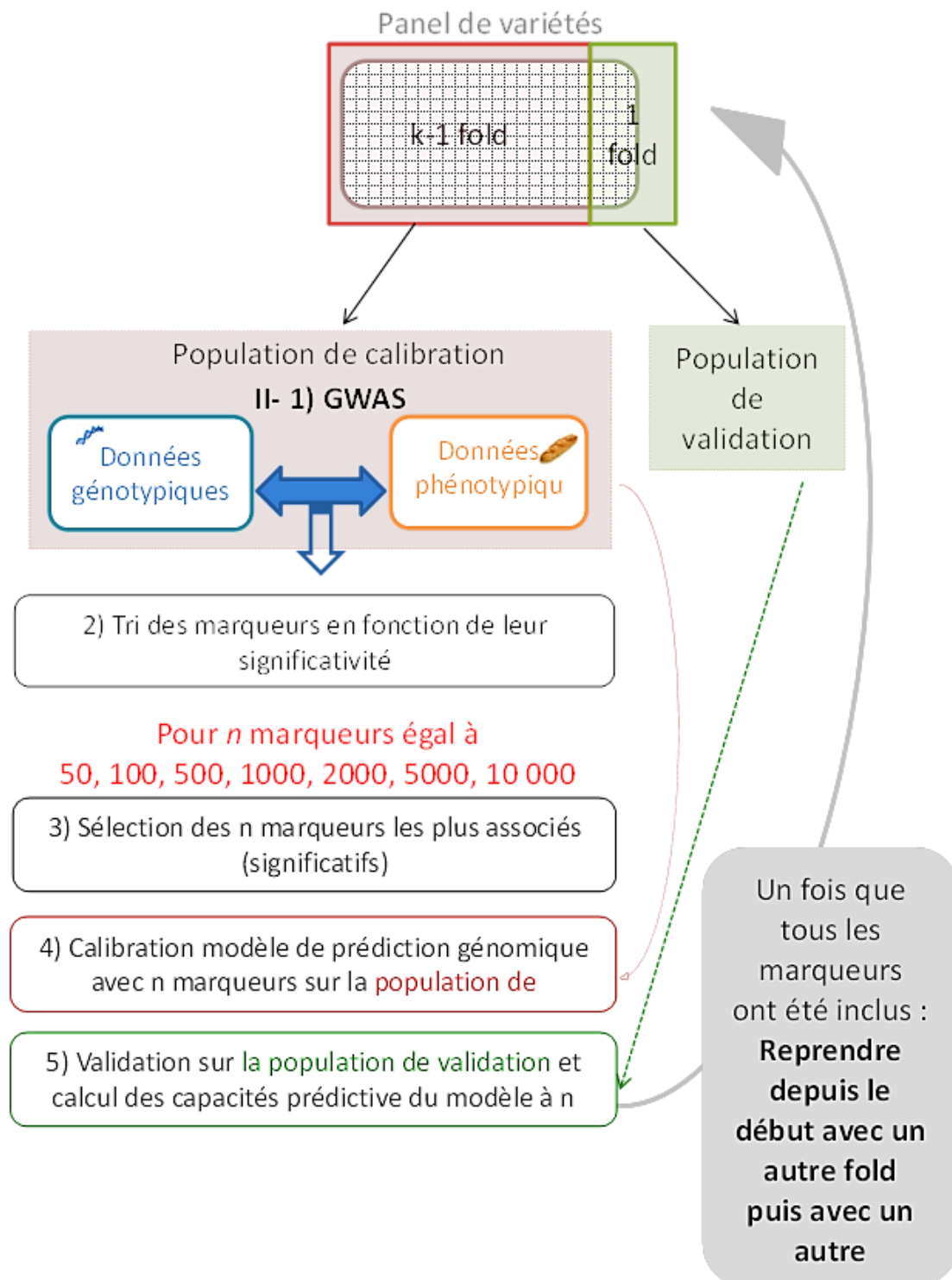


Figure 18 : Schéma représentant la méthode de validation croisée afin d'étudier la capacité prédictive du modèle en fonction du nombre de marqueurs associés au caractère inclus dans le modèle de prédiction. L'échantillonnage se fait dès le début et l'étude de génétique d'association est réalisée sur la population de calibration.

notre étude. Cette méthode a été effectuée grâce à la fonction mlmm disponible sur <https://github.com/Gregor-Mendel-Institute/mlmm> ((Segura, 2017).

b. Etude de la capacité prédictive du modèle de prédiction génomique

Evaluation de la capacité prédictive par validation croisée

La validation croisée est une technique de ré-échantillonnage aléatoire qui permet d'estimer la capacité prédictive d'un modèle sur un panel (Figure 18). Le panel est séparé en deux panels (un panel de calibration et un de validation). Afin de ne pas influencer les résultats par le choix des variétés à inclure dans le panel de calibration ou de validation, un échantillonnage aléatoire est nécessaire. La première étape de la validation croisée consiste en un échantillonnage pour répartir les variétés en k sous-populations (appelées « folds »). Parmi ces sous-populations, $k-1$ folds servent de population de calibration et le fold restant sert de population de validation. Le modèle est calibré sur la population de calibration et permet de prédire les performances des lignées de la population de validation. La corrélation entre les valeurs phénotypiques observées et les valeurs phénotypiques prédites sur le jeu de données de validation permet d'estimer la qualité de prédiction. La procédure est répétée jusqu'à ce que tous les folds aient servis une fois de population de validation. Dans notre étude, le nombre de folds a été fixé à 10 et les échantillonnages ont été répétés 30 fois. La validation croisée permet d'obtenir une estimation moyenne de la capacité prédictive du modèle pour le panel étudié.

Evaluation de la capacité prédictive par validation sur un jeu de données indépendant

Les différents modèles ont également été testés en évaluant leurs capacités prédictives sur un jeu de données indépendant. Cela consistait simplement à calibrer le modèle sur un de nos deux panels puis à prédire le second et inversement. Pour cela, nous avons rassemblé les données génotypiques et phénotypiques des deux panels. Les données génotypiques ont été filtrées (taux de valeurs manquantes maximum à 30%, taux d'hétérozygotie maximum à 10%, fréquence de l'allèle minoritaire minimum à 5%). Le modèle a été calibré sur un des deux panels puis validé sur l'autre et inversement. Cette analyse a été réalisée sur 32 caractères pour 10844 marqueurs.

Etude du nombre de marqueurs à inclure dans le modèle

L'étude de Pierre Colin, (2016) a montré que le modèle de prédiction génomique peut obtenir une meilleure capacité prédictive en fonction des marqueurs inclus dans le modèle. En effet, la prédiction serait meilleure en incluant dans le modèle seulement les marqueurs les plus associés aux caractères. Afin d'étudier cet aspect, la capacité prédictive en incluant dans le modèle une sous-population de marqueurs associés a été comparé avec la capacité prédictive incluant la totalité des marqueurs. Nous avons fait varier la taille de la sous-population de marqueurs ($n= 50, 100, 500, 1000, 2000, 5000$ marqueurs) pour étudier l'évolution de la capacité prédictive du modèle en fonction du nombre de marqueurs inclus (Figure 19). L'association des marqueurs avec les caractères a été testée par génétique d'association selon la méthode MLM. Cette étude a été conduite en validation croisée. Le panel de départ a été séparé en 10 folds. La génétique d'association a été réalisée sur la population de calibration (9 folds) . Les n marqueurs les plus associés ont été utilisés pour calibrer un modèle de prédiction génomique et celui-ci a été utilisé pour prédire les performances des individus du fold restant. La corrélation entre les valeurs prédits et les

Tableau 7 : Héritabilités des variables mesurées sur le panel Arvalis (ARV) ou le panel Limagrain (LIM), capacités prédictives estimées par validation croisée ou par validation sur un panel indépendant des modèles de sélection génomique calibrés sur chacun des panels pour les différentes variables étudiées.

Catégorie	Variables	Capacités prédictives					
		Héritabilités		Validation croisée		Validation panel indépendant	
		ARV	LIM	ARV	LIM	ARV	LIM
Variables test BIPEA	Note de panification	0.41	0.56	0.14	0.31	0.20	0.07
	Note de mie	0.30		0.09			
	COUM	0.52	0.46	0.56	0.49	0.51	0.31
	SOUN	0.24		0.00			
	ELAM	0.11		-0.01			
	EPAL	0.37		0.12			
	Note de pâte	0.46	0.64	0.16	0.35	0.25	0.11
	LISP	0.37	0.50	0.26	0.35	-0.25	-0.15
	COLP	0.45	0.33	0.26	0.31	0.18	0.12
	EXTP	0.31	0.39	0.32	0.43	0.10	0.08
	ELAP	0.23		0.16			
	RELP	0.40		0.14			
	REPO	0.34		0.15			
	EXTF	0.61	0.66	0.43	0.51	0.53	0.31
	DECF	0.22	0.52	0.24	0.12	0.24	0.09
	ELAF	0.43	0.69	0.35	0.44	0.40	0.24
	ACTP	0.34		0.17			
	DECA	0.24		0.11			
	COMF	0.22		0.18			
	TENU	0.35		0.17			
	COLF		0.71		0.21		
	Note de pain	0.33	0.53	0.22	0.26	0.27	0.08
	SECT	0.45	0.62	0.34	0.22	0.32	0.13
	COUP	0.39	0.70	0.23	0.22	0.13	0.07
	DVCL	0.33	0.45	0.11	0.22	0.13	0.02
	RECL	0.15	0.43	0.14	0.16	0.10	0.03
DECL	0.24	0.43	0.38	0.21	0.36	0.13	
Variables alvéographe	W	0.68		0.37			
	G	0.63		0.39			
	P	0.79		0.42			
	P.L	0.67		0.35			
	L	0.60		0.40			
	le	0.79		0.53			
Autres variables	Dureté	0.85		0.25			
	Hydratation	0.51	0.34	0.32	0.01	-0.11	-0.01
	VOL	0.41	0.69	0.41	0.33	0.44	0.20
	Prot	0.37	0.66	0.41	0.36	0.38	0.15

valeurs observées a été calculée. Ce processus a été répété 30 fois et les capacités prédictive du modèle à n marqueurs a été calculée comme la moyenne des coefficients de corrélation obtenus pour les 10 folds \times 30 répétitions.

Apport des variables d'alvéographe dans les modèles de prédiction génomique

Il est possible de faire l'hypothèse que l'inclusion de variables issues de tests indirects comme les variables issues du test d'alvéographe dans les modèles de prédiction génomique pourrait permettre d'améliorer les capacités prédictives de la qualité boulangère. En effet, ces tests apportent une information objective sur la qualité de la pâte en particulier. Par ailleurs, la prédiction par les marqueurs est susceptible de révéler des caractéristiques que ces tests ne permettent pas de mettre en évidence (couleur de la mie). Les mesures d'alvéographe et les données de génotypage étant disponible sur le panel Arvalis, cette hypothèse a donc pu être testée.

La première étape consistait à choisir les variables d'alvéographe à utiliser pour modéliser chaque caractère. Pour cela, une méthode itérative de type régression linéaire multiple en « stepwise » a été appliquée. Elle consistait à tester différents modèles à partir des 6 variables issues des tests d'alvéographe. Pour chaque modèle, le critère d'information d'Akaike (AIC ; (Akaike, 1974) a été calculé. Ce critère statistique permet de sélectionner le modèle le plus vraisemblable. Il représente un compromis entre la qualité de l'ajustement et la complexité du modèle, en pénalisant les modèles ayant un grand nombre de paramètres afin de limiter le phénomène de sur-ajustement. Pour chaque caractère étudié, le modèle avec le plus petit AIC (modèle le plus vraisemblable) a été sélectionné. Ensuite, un modèle de prédiction génomique a été calibré afin de prédire les résidus de ce modèle. Là encore, une validation croisée de type k-fold (10 folds, 30 répétitions) incluant l'étape de modélisation par les variables issues du test d'alvéographe et la sélection génomique a été réalisée afin de mesurer la capacité prédictive de cette approche.



Figure 19: Représentation de la carte génétique où sont positionnés les marqueurs SNP utilisés dans cette étude. Sur chaque chromosome, les marqueurs sont représentés pas un trait horizontal.

Tableau 8 : Nombre de marqueurs significatifs par caractères.

Nombre de marqueurs significatifs	5	4	3	2	1	0
Caractères	ELAF	DECL	EXTP	COLF	G	hydratation
		COLP	SECT	LISP	le	W
		EXTF	durete	DECF	L	P
		note.pate	VOL	COMF	P.L	ELAP
		COUM	note.panif	REPO	EPAL	ACTP
			prot	TENU	RELP	DECA
				DVCL	note.mie	DVCL
				note.pain	COUP	RECL
						SOUM
						ELAM

V. Résultats

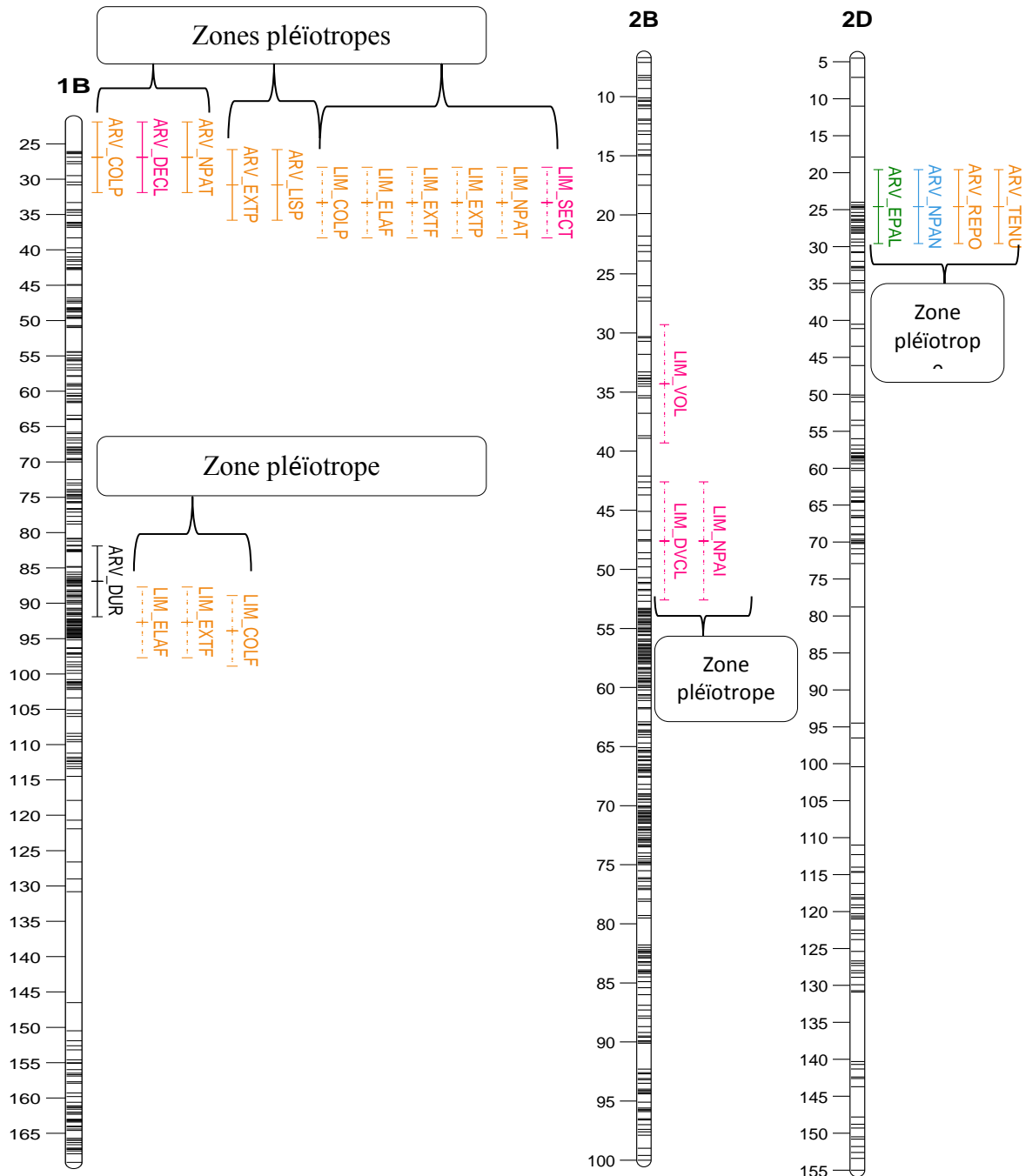
1) Analyse descriptive des données

L'analyse descriptive des données a conduit à la suppression de 7 variables pour le panel Arvalis et 17 pour le panel Limagrain du fait d'une variance faible voire nulle. Parmi les 7 variables supprimées pour le panel Arvalis, 2 étaient des composantes de la note de pâte, 2 des composantes de la note de pain et 3 des composantes de la note mie. Pour le panel Limagrain, la note de mie a été supprimée ainsi que 8 composantes de la note de pâte, 2 composantes de la note de pain et toutes les composantes de la note de mie, excepté la couleur de la mie (COUM). Les variables mesurées sur le panel Arvalis ont globalement une variance plus élevée que celles mesurées sur le panel Limagrain.

L'analyse par ACP a permis d'évaluer le lien entre les notes principales et leurs composantes pour chaque panel afin de déterminer le poids de chaque composante dans l'élaboration des notes principales (Annexe 2). Pour le panel Arvalis, la note de mie est très corrélée avec la souplesse de la mie. La couleur de la mie est en revanche très peu corrélée à la note principale. Sur les deux panels, la note de pain est très corrélée avec le volume du pain, la régularité du coup de lame (RECL) et le développement du coup de lame (DECL). Pour les deux panels, la note de pâte est corrélée avec l'élasticité au façonnage (ELAF) et inversement corrélée au collant au pétrissage (COLP). On observe une bonne corrélation entre les variables I_e , W et la note de pâte ($r > 0.6$) d'une part et la note de panification d'autre part ($r = 0.5$). La note de pâte étant la note la plus corrélée aux variables d'alvéographe, le lien entre ses composantes et les variables d'alvéographe a été étudié. Les variables G et L , représentant respectivement le gonflement et l'élasticité de la pâte, sont très corrélées avec l'extensibilité de la pâte au pétrissage (EXTP). Les variables I_e et W , représentant l'indice d'élasticité et la force boulangère, sont corrélées avec l'élasticité au façonnage (ELAF), l'élasticité au pétrissage (ELAP) et à la tenue lors de la mise au four (TENU). Les variables P et $P.L$, relatives à la ténacité, sont corrélées au déchirement au façonnage (DECF) qui est inversement corrélé à l'extensibilité au pétrissage.

L'héritabilité (h^2) a été calculée pour tous les caractères sur les deux panels (Tableau 7). Pour la majorité des caractères, l'héritabilité est plus grande sur le panel Limagrain (h^2 variant de 0.33 à 0.71) que sur le panel Arvalis (h^2 variant de 0.11 à 0.85). Sur le panel Limagrain, la note de pain, de pâte et la note totale de panification ont une héritabilité respectivement égale à 0.53, 0.64 et 0.56. Sur le panel Arvalis, la note de pain, de pâte, de mie et la note totale de panification ont une héritabilité respectivement égale à 0.33, 0.46, 0.30 et 0.41. Parmi les notes principales, la note de pâte est celle qui a la plus forte héritabilité pour les deux panels. Sur les deux panels, les 3 composantes de la note de pain qui ont les plus grandes héritabilités sont le volume du pain (VOL; $h^2 = 0.69$ pour le panel Limagrain, $h^2 = 0.39$ pour le panel Arvalis), la couleur du pain (COUP; $h^2 = 0.70$ pour le panel Limagrain, $h^2 = 0.39$ pour le panel Arvalis) et la section du pain (SECT; $h^2 = 0.62$ pour le panel Limagrain, $h^2 = 0.45$ pour le panel Arvalis). Sur le panel Arvalis, l'extensibilité de la pâte au façonnage (EXTF) et la couleur de la mie (COUM) ont une héritabilité forte (respectivement $h^2 = 0.61$ et $h^2 = 0.52$). Concernant le panel Limagrain, six composantes de la note de pâte ont également une héritabilité élevée : l'extensibilité de la pâte au façonnage (EXTF; $h^2 = 0.66$), élasticité au façonnage (ELAF; $h^2 = 0.43$), le collant de la pâte au façonnage (COLF; $h^2 = 0.71$).

Figure 20 : Représentation graphique des QTL détectés sur les chromosomes 1B, 2B et 2D. Les marqueurs associés sont représentés par un trait horizontal. Le trait vertical représente un pseudo intervalle de confiance de 10 centiMorgans. Les traits tiretés représentent les QTL détectés pour le panel Limagrain et les traits pleins les QTL pour le panel Arvalis. Le nom des QTL correspond à la concaténation du nom du panel (ARV, Arvalis ; LIM, Limagrain) et du nom du caractère (tableau 4).



2) Génétique d'association

Le nombre de marqueurs par chromosome et par génome a été étudié (Figure 20). Le chromosome qui a le plus de marqueurs est le chromosome 3B. La comparaison du nombre de marqueurs par génome montre un important déficit de marqueurs sur le génome D. La chute du déséquilibre de liaison (DL) a été estimée en calculant la corrélation entre l'ensemble des couples de marqueurs (annexe 3). L'étude de la chute du DL montre une grande similarité pour les deux panels. Pour les deux panels, 80% des couples de marqueurs avec un $R^2 > 0.2$ sont à moins de 6.2cM. Lorsqu'on étudie deux marqueurs, on peut s'attendre à avoir un DL faible entre eux s'ils sont à plus de 6.2cM. L'étude de la structuration génétique des individus n'a pas mis en évidence de structuration particulière dans les deux panels. En particulier, il n'a pas pu être observé de structuration liée à l'origine des variétés.

L'étude de génétique d'association a conduit à la détection de 33 QTL sur le panel Limagrain et de 35 QTL sur le panel Arvalis. Aucun QTL n'a été détecté pour 10 des caractères étudiés dont le taux d'hydratation, deux variables d'alvéographe, trois composantes de la note de pâte, deux composantes de la note de mie et deux composantes de la note de pain (Tableau 9). Le nombre de QTL détectés pour les autres caractères varie de 1 à 5 (Tableau 8). Le seul caractère pour lequel 5 QTL ont été détectés est l'élasticité de la pâte au façonnage (ELAF). En moyenne, deux QTL sont détectés par caractère.

Onze zones de co-localisations de QTL robustes (détectés sur les deux panels) et/ou pléiotropes (pour différents caractères) ont été détectées. Quatre d'entre elles se trouvent sur le chromosome 1B et les autres se trouvent sur les chromosomes 1D, 2B, 2D, 3D, 6B, 7A, 7B. Pour certaines zones, les marqueurs détectés ont une position proche mais ne sont pas identiques. Les effets alléliques pour les QTL d'une même zone ont été calculés à partir du même marqueur (annexe 5). Les effets alléliques présentés ont été classés en fonction de leur valeur absolue puisque le signe indique seulement quel allèle fait augmenter l'intensité du caractère.

Un QTL isolé a été détecté sur le chromosome 1A pour le déchirement au coup de lame (DECL, composante de la note de pain) avec un effet relativement important (0.54).

Les zones de co-localisation sur le chromosome 1B (Figure 20) comportent deux QTL pour la note de pâte et 10 QTL pour ses composantes. Ce chromosome semble également fortement lié aux propriétés de la pâte. L'effet allélique sur la note de pâte est de 8.95 sur le panel Arvalis et de 4.54 pour le panel Limagrain. Ce QTL de note de pâte est celui qui a le plus grand effet allélique détecté dans cette étude. Les effets les plus importants sont trouvés pour la rapidité de lissage (LISP) et le déchirement au coup de lame (DECL). Le DECL est une composante de la note de pain. Un QTL isolé a également été détecté sur le chromosome 1B pour la dureté. Ce QTL a un effet allélique relativement important (environ 14 alors que l'écart-type de la dureté est de 13).

La zone de co-localisation la plus importante se trouve sur le chromosome 1D (Figure 21) à la position 66.2. Cette zone comporte 23 QTL dont 6 robustes (COLP, DECF, ELAF, EXTF, note de pâte, SECT). Cette zone semble très liée aux propriétés de la pâte. En effet, la note de pâte et 10 de ses composantes ainsi que les variables d'alvéographe G, Ie et L sont associés à cette zone. Les effets alléliques de la note de pâte sont de 5.75 pour le panel Arvalis et de 2.89 pour le panel Limagrain. La composante de la note de pâte pour laquelle l'effet

Figure 21: Représentation graphique des QTL détectés sur le chromosome 1D. Les marqueurs associés sont représentés par un trait horizontal. Le trait vertical représente un pseudo intervalle de confiance de 10 centiMorgans. Les traits tiretés représentent les QTL détectés pour le panel Limagrain et les traits pleins les QTL pour le panel Arvalis. Le nom des QTL correspond à la concaténation du nom du panel (ARV, Arvalis ; LIM, Limagrain) et du nom du caractère (tableau 4).



allélique de cette zone est le plus fort est l'extensibilité au façonnage (EXTF) qui a un delta de 1.16 sur le panel Arvalis et de 0.72 sur le panel Limagrain. Tous les QTL détectés pour les variables d'alvéographe se trouvent sur cette zone.

La zone pléiotrope sur le chromosome 2B (Figure 20) comporte un QTL pour la note de pain et un QTL pour le développement du coup de lame (DVCL). L'effet allélique de la note de pain est de 3.26 (le plus important détecté dans cette étude pour cette note) et celui de DVCL est de 0.17.

La zone pléiotrope sur le chromosome 2D (Figure 20) comporte une QTL pour la note totale de panification, un QTL pour une composante de la note de mie (épaisseur de l'alvéolage de la mie, EPAL) et pour des composantes de la note de pâte (détente relâchement au pointage, REPO, et la tenue à la mise au four, TENU). Cette zone est liée à différents aspects de la note de panification. L'effet allélique sur la note de panification est de 13.0, ce qui signifie que le passage de l'état homozygote d'un allèle à l'autre à ce locus permet en moyenne d'augmenter la note de panification de 13 points. L'allèle défavorable pour ce QTL est l'allèle minoritaire qui a une fréquence de 22% dans le panel Arvalis.

Un QTL isolé a été détecté sur le chromosome 3B pour la dureté. Ce QTL a un effet allélique relativement important (de 14.5 alors que l'écart-type de la dureté est de 13).

La zone pléiotrope sur le chromosome 3D (Figure 22) comporte un QTL pour la note de panification et un QTL pour la dureté détectés sur le panel Arvalis. Les effets alléliques sont de 27.51 pour la dureté et de 23.78 pour la note de panification. Ce QTL est celui qui a le plus grand effet allélique pour la note de panification totale dans cette étude. Cependant, l'allèle minoritaire qui est également l'allèle défavorable pour ce caractère est présent dans 6 % des variétés dans ce panel, ce qui représente 5 ou 6 variétés.

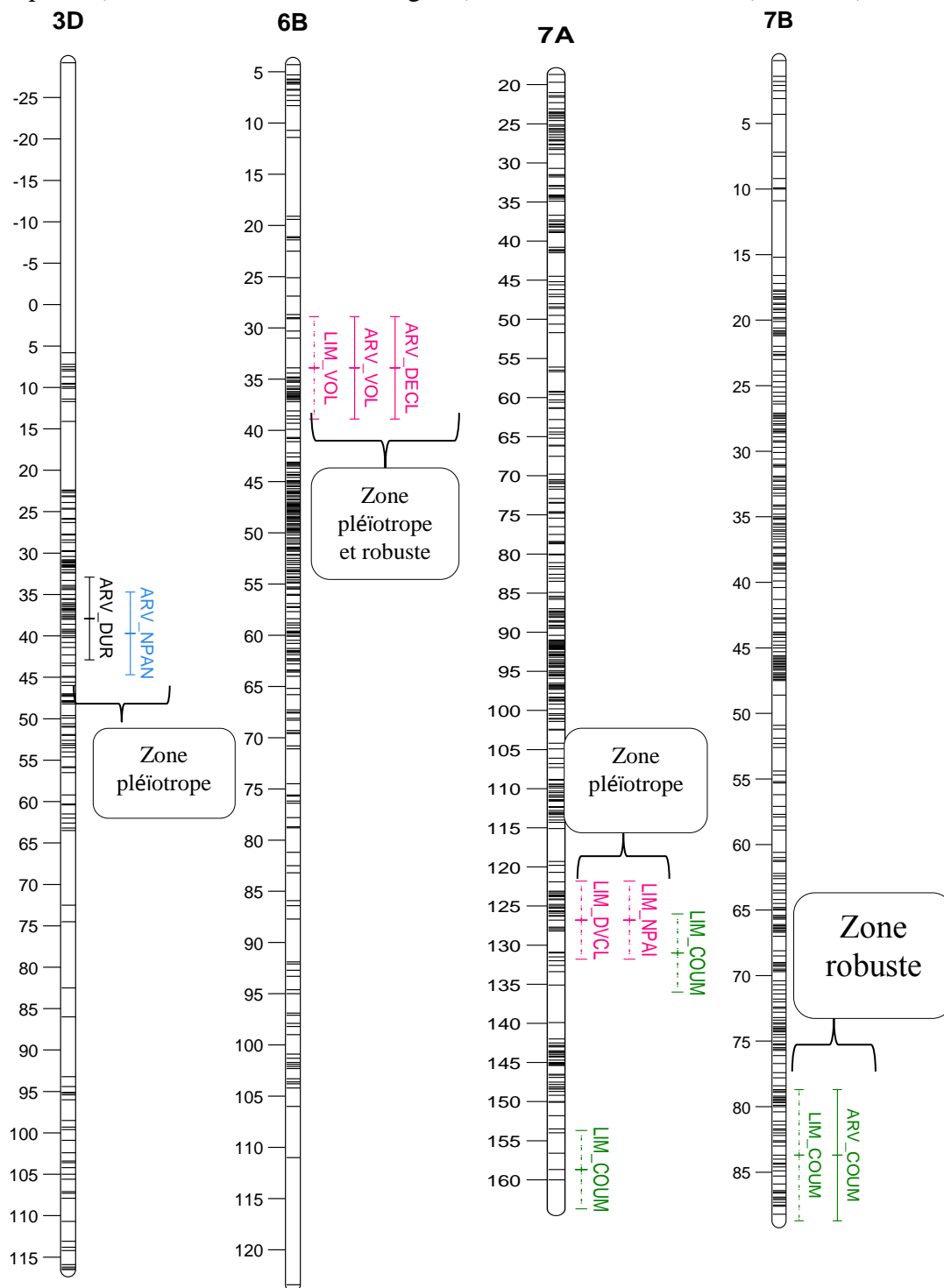
Le seul QTL trouvé pour la note de mie se trouve sur le chromosome 5B avec un effet allélique de 2.32.

La zone pléiotrope détectée sur le chromosome 6B (Figure 22) comporte un QTL pour le déchirement du coup de lame (DECL) et un QTL robuste pour le volume du pain. Les effets alléliques pour le volume sont de -96.73 cm^3 pour le panel Arvalis, de -55.57 cm^3 sur le panel Limagrain et de -0.47 pour DECL. L'écart-type du volume du pain allant de 80 pour le panel Arvalis à 119 pour le panel Limagrain, ces effets alléliques sont importants. L'allèle GG est favorable à l'augmentation du volume du pain. La proportion de cet allèle pour le volume du pain est de 23% pour le panel Limagrain et de 22 % pour le panel Arvalis. L'introduction de l'allèle GG pourrait permettre de corriger des défauts de volume chez certaines variétés.

La zone pléiotrope sur le chromosome 7A (Figure 22) comporte un QTL pour la note de pain et une de ses composantes (développement du coup de lame, DVCL). L'effet allélique de la note de pain est de 2.32 et l'effet pour DVCL est de 0.12.

Une zone robuste pour la couleur de la mie a été détectée sur le chromosome 7B (Figure 22) avec des effets alléliques de -0.87 pour le panel Arvalis et -0.47 pour le panel Limagrain.

Figure 22: Représentation graphique des QTL détectés sur les chromosomes 3D, 6B, 7A et 7B. Les marqueurs associés sont représentés par un trait horizontal. Le trait vertical représente un pseudo intervalle de confiance de 10 centiMorgans. Les traits tiretés représentent les QTL détectés pour le panel Limagrain et les traits pleins les QTL pour le panel Arvalis. Le nom des QTL correspond à la concaténation du nom du panel (ARV, Arvalis ; LIM, Limagrain) et du nom du caractère (tableau 4).



3) Prédiction génomique

Les capacités prédictives du modèle calibré sur les données du panel Arvalis en validation croisée sont très variables selon les caractères, allant de -0.09 pour l'élasticité de la mie (ELAM) à 0.55 pour la couleur de la mie (COUM) (Tableau7). La note de panification a une capacité prédictive de 0.13, celle de pain de 0.21, celle de pâte de 0.16 et celle de mie de 0.09. La composante de la note pâte qui atteint la meilleure capacité prédictive est l'extensibilité de la pâte au façonnage (EXTF) avec une capacité de 0.40. Les composantes de la note de pain ont une capacité prédictive moyenne à faible, allant de 0.11 (pour le développement au coup de lame) à 0.38 (pour le déchirement au coup de lame). Les composantes de la note de mie ont toute une faible capacité de prédiction sauf COUM (0.56). La teneur en protéines et le volume du pain (VOL) ont une capacité prédictive de 0.40. Les variables d'alvéographe ont globalement une bonne capacité prédictive, le minimum étant de 0.35 pour P.L.

Les capacités prédictives du modèle calibré sur les données du panel Limagrain estimée en validation croisée vont de 0.01 pour le taux d'hydratation à 0.51 pour l'extensibilité au façonnage (Tableau7). La note de pain a une capacité prédictive de 0.26, la note de pâte de 0.36 et la note panification de 0.31. Les capacités prédictives du modèle pour les composantes de la note de pâte vont de 0.12 pour le déchirement au façonnage à 0.51 pour l'extensibilité au façonnage (meilleure capacité obtenue pour ce panel). Les capacités prédictives du modèle pour l'élasticité au façonnage (ELAF) et l'extensibilité de la pâte au pétrissage (EXTP) sont de 0.44 et 0.43 et respectivement. Les capacités prédictives du modèle pour les composantes de la note de pain varient de 0.16 pour la régularité du coup de lame à 0.22 pour la section. Le modèle pour la composante de la note de mie évaluée sur ce panel (la couleur de la mie) a une bonne capacité prédictive (0.51). Le taux d'hydratation a la capacité prédictive la plus basse (0.01).

En comparant les résultats des validations croisées sur les deux panels (Tableau7), on s'aperçoit que l'extensibilité de la pâte au façonnage (EXTF) et la couleur de la mie (COUM) sont les deux variables les mieux prédites. Par ailleurs, on observe généralement de meilleures capacités prédictives du modèle calibré sur les données du panel Limagrain. Les résultats présentés sont les moyennes des capacités prédictives obtenues lors des différentes répétitions et des différents folds. Avant de réaliser la moyenne, l'étude de l'écart-type des capacités prédictives pour un même caractère montre que la variabilité des résultats est plus forte sur le panel Arvalis que sur le panel Limagrain. En moyenne, l'écart-type des résultats de capacité de prédiction sur le panel Arvalis est de 0.20 tandis que celui du panel Limagrain est de 0.08.

L'estimation des capacités prédictives des modèles à l'aide d'un jeu de données indépendant (c'est-à-dire la calibration du modèle sur le panel Limagrain suivi d'une validation sur le panel Arvalis et inversement) confirme la bonne capacité à prédire l'extensibilité au façonnage (EXTF), la couleur de la mie (COUM), le volume du pain (VOL) et l'élasticité au façonnage (ELAF) quel que soit le panel utilisé pour la calibration. De manière générale, ce sont les modèles calibrés sur le panel Limagrain qui semblent les plus robustes. Par ailleurs, la comparaison des capacités prédictives estimées par validation croisée et celles obtenues par validation sur un jeu de données indépendant montre une légère surestimation des capacités prédictives par la méthode de validation croisée quel que soit le panel utilisé pour la calibration bien que le classement des variables les mieux prédites reste sensiblement le même quel que soit le type de validation.

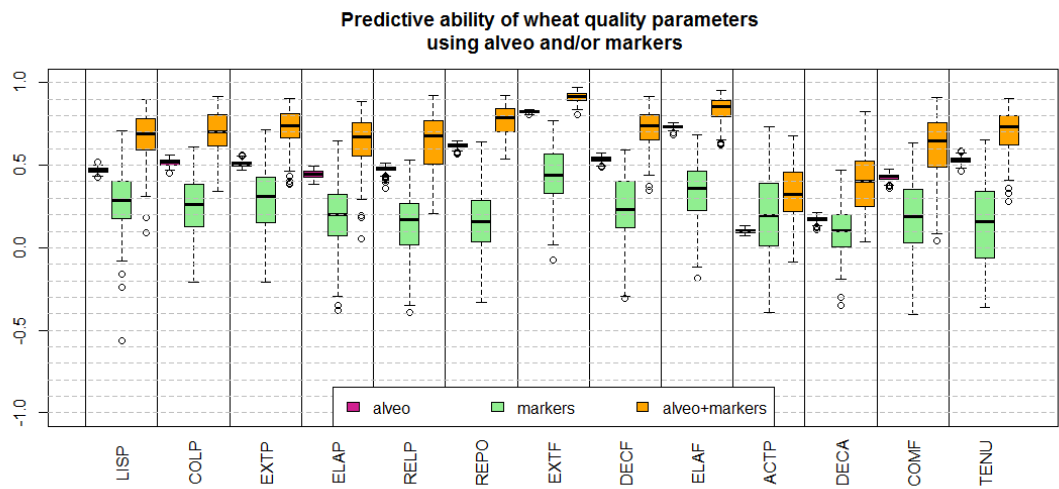
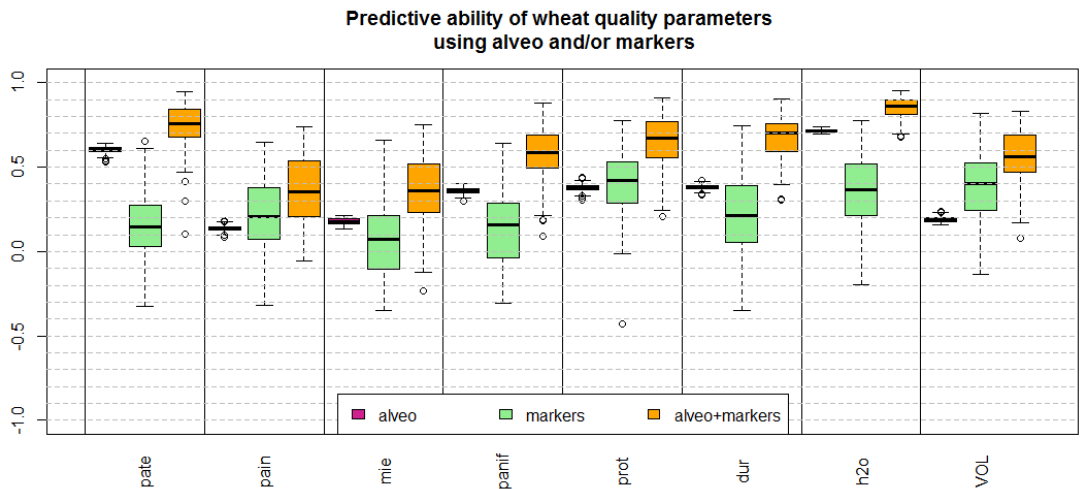


Figure 23: Capacités prédictives des notes principales (pate, pain, mie), de la note de panification (panif), de la teneur en protéines (prot), de la dureté (dur), du taux d'hydratation (h2o) et du volume du pain (VOL) (en haut) et des composantes de la note de pâte (en bas) prédits selon trois modèles différents : modèle basé sur les variables d'alvéographe seules (en rose), modèle basé sur les marqueurs génétique (prédiction génomique ; en vert) et modèle combinant les variables d'alvéographe et les marqueurs génétiques (en orange).

4) Etude de l'inclusion des marqueurs associés pour améliorer la prédiction génomique

Les résultats de cette étude sont présentés sous forme de graphiques représentant pour chaque caractère la capacité prédictive estimée après validation croisée en fonction du nombre de marqueurs inclus dans le modèle de prédiction génomique. Chaque point représente la moyenne des capacités prédictives et les barres d'erreurs correspondent aux écart-types obtenus après validation croisée.

Globalement, trois tendances ont été observées :

1. Pour des caractères comme la note de pâte sur le panel Limagrain, la capacité prédictive augmente de manière continue avec le nombre de marqueurs. Ce type de profil montre que la capacité de prédiction pour ces caractères est meilleure en incluant la totalité des marqueurs.
2. Pour des caractères comme la note de pain ou la couleur de la mie (COUM) pour le panel Limagrain, les capacités prédictives augmentent rapidement avec le nombre de marqueurs inclus et atteignent un plateau ou sont stables dès le début. Ces profils montrent que la capacité de prédiction est aussi bonne avec une sous-sélection de marqueurs.
3. Le dernier profil montre une capacité prédictive qui diminue fortement avec l'inclusion de marqueurs. Ces profils signifient que la capacité de prédiction est meilleure avec l'inclusion de 50 marqueurs plutôt que la totalité. Cette déduction paraît cependant peu probable.

Pour le panel Arvalis, une sous-sélection de marqueurs améliore la capacité pour les caractères suivants : P.L, Ie, la teneur en protéines, l'extensibilité au façonnage (EXTF) et l'élasticité au façonnage (ELAF). Cependant, cette amélioration est très faible (en moyenne de 0.03, maximum de 0.06 pour l'indice d'extensibilité). De nombreux caractères atteignent une capacité de prédiction aussi bonne avec une sous-sélection de marqueurs qu'avec la totalité. Globalement, la sous-sélection de marqueurs doit comporter entre 1 000 et 5 000 marqueurs. Parmi ces notes, on retrouve les notes de panification et de pain pour le panel Limagrain qui atteignent une aussi bonne prédiction avec 1 000 marqueurs qu'avec 10 000. L'utilisation de la totalité des marqueurs améliore la prédiction pour seulement 2 caractères : la note de pâte (+0.02 sur le panel Limagrain) et DECL (+0.08 pour le panel Arvalis). Une sous-sélection de marqueurs (entre 1 000 et 5 000 suivant le caractère) permet d'atteindre voire de dépasser la capacité prédictive atteinte avec tous les marqueurs.

5) Contribution des variables issues du test d'alvéographe un modèle de prédiction génomique de la qualité boulangère

La capacité prédictive du modèle calibré avec les variables d'alvéographe seules est de 0.34 pour la note de panification (Figure 23). Concernant les notes principales, c'est la note de pâte qui est la mieux prédite (0.60) suivie de la note de mie (0.36) et de la note de pain (0.35). Les composantes de la note de pâte sont globalement bien prédites par les variables d'alvéographe seules, particulièrement la variable de détente, le relâchement au pointage de la pâte (REPO), l'extensibilité de la pâte au façonnage (EXTF), l'élasticité de la pâte au façonnage (ELAF). En revanche, l'activité fermentaire et le déchirement à l'apprêt de la pâte sont mal prédites. Globalement, les composantes de la note de pain et de mie (Figure 24) ne sont pas très bien prédites par les variables d'alvéographe sauf la section du pain qui a une capacité de prédiction de 0.60.

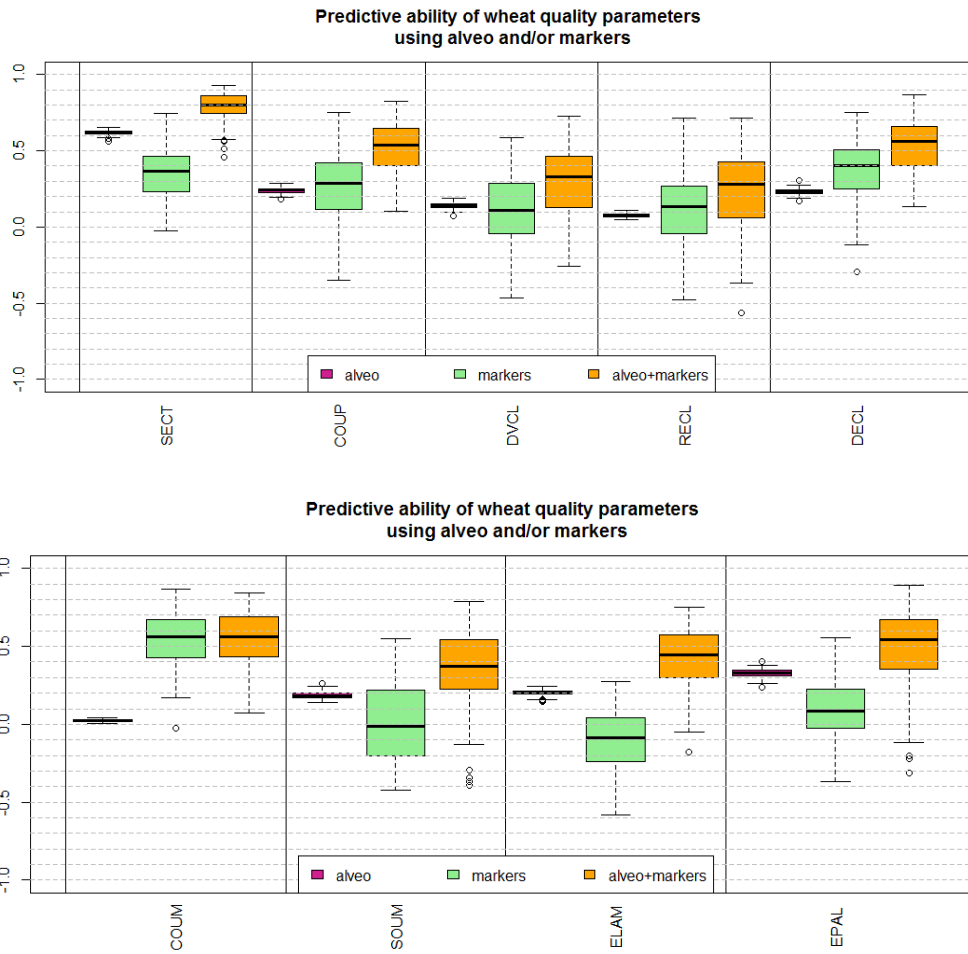


Figure 24: Capacités prédictives des notes composantes de la note de pain (en haut) et mie (en bas) prédits selon trois modèles différents : modèle basé sur les variables d'alvéographe seules (en rose), modèle basé sur les marqueurs génétique (prédiction génomique ; en vert) et modèle combinant les variables d'alvéographe et les marqueurs génétiques (en orange). Les noms des caractères sont répertoriés dans le tableau 4.

Pour toutes les variables, le modèle combinant les variables d'alvéographe et la prédiction génomique améliore la prédiction par rapport au modèle avec les marqueurs uniquement (+1.1 pour le relâchement au coup de lame à +6 pour la note de pâte +0.19 en moyenne). Avec ce nouveau modèle, la capacité prédictive de la note de pâte est de 0.75, celle de la note de panification est de 0.60. La capacité de prédiction du modèle pour la teneur en protéines, l'hydratation, la dureté et le volume du pain sont toutes supérieures à 0.50. La capacité prédictive de toutes les composantes de la note de pâte sauf ACTP et DECA ainsi que SECT, COUP et DECL (composantes de la note de pain) et COUM et EPAL (composantes de la note de mie) dépasse les 0.50. Ce modèle a donc de bonnes capacités prédictives pour la majorité des caractères étudiés, en particulier pour la note de pâte et ses composantes.

VI. Discussion

1) Estimation de l'héritabilité

La structure des données acquises sur le panel Limagrain ne laissait que deux témoins présent dans l'ensemble des essais pour effectuer un ajustement des effets environnement. Les effets environnement ne sont donc estimés que sur la base de ces deux témoins et seuls ces témoins contribuent à la variance résiduelle qui est dès lors plus faible que sur le panel Arvalis pour lequel d'avantage de variétés sont présentes dans les différents essais. L'estimation de l'héritabilité des variables mesurées sur le panel Limagrain a donc probablement été biaisée. La méthode d'estimation de l'héritabilité utilisée ici étant un ratio entre variance génétique et variance résiduelle, la première étant sensiblement identique dans les deux panels et la seconde étant sensiblement réduite dans le panel Limagrain, les héritabilités calculées ainsi étaient donc plus élevées sur le panel Limagrain.

2) Distribution des marqueurs sur le génome

Concernant les données génotypiques, le nombre de marqueurs présents sur le génome D est inférieur aux nombres de marqueurs présents sur les génomes A et B. Cette différence est due au faible taux de polymorphisme sur le génome D par rapport aux deux autres. En effet, le génome D étant le génome inclus le plus récemment, celui-ci a subi moins de mutations et présente donc moins de polymorphisme. Il semble que le polymorphisme de ce génome contribue néanmoins aux différences de qualité boulangère puisque des QTL ont pu être identifiés sur les chromosomes 1D, 2D et 3D. La production de blés synthétiques (croisement entre des ancêtres du blé tendre tel que des espèces de blé durs et des *Aegilops tauschii*) est une approche parfois proposée pour augmenter la variabilité génétique sur le génome D et pourrait également contribuer à augmenter la variabilité génétique pour la qualité boulangère (Guzmán et al., 2014; Li et al., 2014; Nedelkou et al., 2017).

3) Architecture génétique : quelques QTL majeurs, de nombreux QTL mineurs

L'étude de génétique d'association conduite sur les panels Limagrain et Arvalis a permis d'identifier plusieurs QTL d'intérêt pour la qualité boulangère et les variables associées. Le travail d'analyse des co-localisations et des effets alléliques des QTL a permis notamment d'identifier quelques QTL majeurs et de nombreux QTL mineurs (à effets faibles).

Les chromosomes 1A, 1B, 1D sont connus pour comporter des gènes de synthèse des gluténines (Branlard, 2012). Ces gènes sont des gènes candidats pour les QTL détectés sur les chromosomes 1B et 1D. La détection de zones particulièrement liées aux propriétés de la pâte montre une conformité des résultats par rapport à ce qui était déjà connu. Pour confirmer cette hypothèse, il pourrait être envisagé de génotyper les marqueurs pour les gènes de synthèse des gliadines et des gluténines des chromosomes du groupe 1 sur le panel Arvalis et le panel Limagrain. Une autre solution serait de comparer des cartes génétiques.

Cette étude a conduit à la détection d'un QTL de dureté sur le chromosome 1B. Une autre étude de détection de QTL conduite par Li et al., (2009) avait mené à la détection d'un QTL majeur sur le 1B pour la dureté qui pourrait s'avérer être le même QTL.

Un QTL pour la note de panification avec un effet allélique fort a été détecté sur le chromosome 3D. Le gène candidat correspondant à cette zone pourrait être un gène de synthèse des gluténines de faible poids moléculaire (Rosa Filho, 1997). Là encore, il pourrait être envisagé de génotyper des marqueurs pour ce gène sur le panel Arvalis et/ou le panel

Limagrain afin de confirmer cette hypothèse. La présence d'un QTL pour la note de panification dans une zone qui ne comporte aucun QTL pour les notes principales ni leurs composantes peut s'expliquer par l'additivité des effets de chacune des composantes. Leurs effets sur cette zone sont probablement faibles et donc non détectés mais le cumul des effets conduit à un effet important sur la note totale qui est donc détectée.

Une étude de génétique d'association sur un panel indépendant permettrait de valider la détection des QTL. En particulier, il serait intéressant de confirmer le QTL du chromosome 3D pour la note de panification et la dureté. En effet, ce QTL apparaît comme ayant l'effet le plus fort sur la note de panification mais le taux de polymorphisme au niveau de ce QTL est faible (seules 5 ou 6 variétés sont porteuses de l'allèle défavorable). Il y a donc un risque que ce QTL soit un faux positif dû à un biais d'échantillonnage. De même, la comparaison des résultats de notre étude avec les résultats d'autres études réalisées sur du matériel génétique indépendant par une approche de méta-analyse de QTL (Arcade et al. 2004) permettrait de confirmer et d'affiner les résultats.

4) Perspectives d'utilisation en SAM

La sélection assistée par marqueurs sur les QTL détectés pour les caractères de qualité boulangère permettrait théoriquement d'accélérer la sélection sur les critères de qualité boulangère. Les allèles de gluténines sont déjà pris en compte dans les programmes de sélection assistée par marqueurs (Méléard, communication personnelle). Les autres QTL pourraient donc être de nouveaux indicateurs d'intérêt pour les sélectionneurs. Néanmoins, ceux-ci devraient d'abord être validés au préalable.

Les QTL augmentant l'intensité d'un caractère ne sont pas forcément les QTL les plus favorables étant donné que les caractères doivent atteindre un optimum et non un maximum. Dans l'hypothèse où un sélectionneur souhaiterait faire de la sélection assistée par marqueurs avec ces QTL, il serait nécessaire de tenir compte de la valeur moyenne du matériel sur lequel s'opère la sélection afin de choisir l'allèle à fixer et d'approcher ainsi au mieux l'optimum de la variable ciblée.

De manière générale, il serait également intéressant d'analyser la fréquence allélique des QTL d'intérêt dans le matériel inscrit en fonction de la date d'inscription au catalogue officiel. Cela permettrait d'évaluer la pression de sélection exercée sur les allèles et d'identifier par exemple des allèles qui sont déjà « fixés » dans le matériel inscrit ou d'autres qui mériteraient d'être introduit car nouveau ou peu fréquents.

Le calcul des effets alléliques a été étudié pour chaque QTL indépendamment. Il serait intéressant d'évaluer le niveau de qualité que l'on pourrait atteindre avec une variété comportant tous les allèles favorables. Cette étude peut cependant devenir complexe étant donné que c'est un optimum qui est recherché pour les notes composantes. Les QTL des notes composantes pouvant être favorables ou non suivant la forme allélique des autres QTL.

5) Intérêts de la prédiction génomique pour la qualité boulangère

Les résultats de validation ont montré que les capacités prédictives du modèle calibré sur le panel Limagrain sont meilleures. Cette différence peut provenir d'une différence de taille des panels. Le panel Limagrain comprenant 1529 lignées a probablement une plus grande puissance statistique que le panel Arvalis qui ne comporte que 188 variétés. Cependant ce résultat est à nuancer à cause de l'estimation biaisée de l'héritabilité. En effet, les capacités prédictives estimés dans cette étude correspondent à la corrélation entre les valeurs phénotypiques ajustées par la méthode BLUPS (Estimated Breeding Values; EBV) et les valeurs phénotypiques prédites (Genomic Estimated Breeding Values ; GEBV). Or, c'est la

corrélation entre les GEBV et les vraies valeurs phénotypiques (TBV, Trait breeding Value) qui intéressent le sélectionneur. Les TBV sont inconnues et notre capacité à les estimer dépend de l'héritabilité du caractère étudié : $h^2 = r(EBV:TBV)$. Ainsi, la précision du modèle ou Accuracy correspond en réalité au produit entre de la corrélation GEBV:EBV et l'héritabilité :

$$r(GEBV:EBV) = r(GEBV:TBV) \times r(EBV:TBV)$$

Or, l'héritabilité pour les variables du panel Limagrain a sans doute été sur-estimée. En réalité, on peut faire l'hypothèse que les capacités prédictives du modèle calibré sur le panel Limagrain sont proches des capacités prédictives du modèle calibré sur le panel Arvalis.

Le modèle de prédiction génomique utilisé conduit à des capacités prédictives pour la note totale de panification de 0.07 à 0.31 suivant le type de validation et le panel utilisé. La meilleure capacité prédictive obtenue ($r^2=0.31$) est obtenue sur le panel Limagrain en validation croisée. Les capacités prédictives sur le panel Limagrain sont probablement sur-estimées comme vu précédemment. Il est donc probable que la capacité prédictive maximum réellement atteinte soit inférieure à 0.31. Or, une étude de prédiction de la note de panification par un modèle combinant des tests indirects (taux de protéines, indice de dureté et test alvéographe) avec l'information portée par les gènes de synthèse des gluténines permet d'atteindre un niveau de prédiction supérieur ($r \approx 0.30$ d'où $r^2 \approx 0.55$; Oury et al., 2010). L'utilisation d'un grand nombre de marqueurs n'apparaît donc pas être avantageuse pour ce caractère.

Les niveaux de capacités prédictives varient d'un caractère à l'autre. Pour des caractères difficiles et chers à phénotyper, la prédiction génomique représente une alternative intéressante même si le niveau de capacité prédictive est faible. L'utilisation de cette méthode engendre un nécessaire compromis entre précision des prédictions et pression de sélection: il peut être avantageux d'utiliser un modèle relativement peu précis s'il s'agit de prédire une variable difficile à phénotyper. Il s'agit alors d'établir un seuil pour sélectionner des individus de manière à éliminer au maximum les moins performants tout en évitant d'éliminer ceux dont les performances sont adéquates. Nécessairement, cela conduit à diminuer la pression de sélection lorsqu'un modèle peu précis est utilisé mais même dans cette situation, cela peut contribuer à améliorer l'efficacité de la sélection pour des caractères dont le coût du phénotypage est élevé. Le calcul de ce seuil pour la note totale de panification permettrait d'estimer si les niveaux de prédiction atteints dans cette étude sont suffisants pour justifier l'utilisation de la sélection génomique sur ce critère.

Pour poursuivre cette étude, il serait intéressant de tester le niveau de capacité prédictive qu'il est possible d'obtenir avec une autre méthode de prédiction génomique. En effet, la méthode utilisée (kinship-blup) repose sur l'hypothèse d'une structure génétique où tous les marqueurs ont un effet identique et faible. La détection de QTL majeurs pour certains caractères montre que cette hypothèse de base n'est pas forcément vérifiée. L'utilisation d'une autre méthode plus adaptée à l'architecture génétique permettrait peut-être d'améliorer les capacités de prédiction.

Enfin, l'inclusion des variables d'alvéographe dans le modèle de prédiction génomique mène à des capacités de prédiction très encourageantes. D'un point de vue économique, la réalisation du test d'alvéographe et du génotypage est avantageuse par rapport à la réalisation du test de panification. En effet, un test d'alvéographe coûte environ 35€ et le génotypage 15€ tandis que le test de panification vaut 120€ environ. De plus, le test d'alvéographe nécessite moins de grains que le test de panification ce qui permettrait d'intégrer variables relatives à la qualité boulangère plus précocement dans les programmes

de sélection. Une étude réalisée par (Michel et al., 2017) avait testé l'utilisation de tests indirects combinés avec la prédiction génomique. Les tests indirects utilisés sont relatifs à la quantité de gluten. Les capacités prédictives basées sur ces deux méthodes ensemble obtenues dans cette étude sont meilleures que les capacités prédictives des deux tests séparément (amélioration de 0.1 environ par rapport à la sélection génomique seule). Les résultats d'utilisation de tests indirects combinés avec la prédiction génomique sont prometteurs. Cette méthode permettrait d'intégrer les critères de qualité boulangère plus précocement dans les programmes de sélection tout en atteignant de bons niveaux de prédiction.

Conclusion

L'étude de génétique d'association a permis de mettre en évidence des zones génomiques associées à la qualité boulangère sur les chromosomes 1B, 1D, 2B, 2D, 3D, 6B, 7A, 7B. Les zones sur les chromosomes 1B et 1D sont probablement à associer avec les gènes de synthèse des gluténines qui ont un impact fort sur la qualité boulangère. La qualité boulangère semble être associée génétiquement à quelques QTL majeurs et de nombreux QTL mineurs. La prédiction génomique représente une opportunité pour les sélectionneurs de faciliter la prise en compte de critères longs et fastidieux à phénotyper. Les différents modèles de prédiction génomique testés permettent d'atteindre une capacité prédictive de 0.45 dans le meilleur des cas. Il est difficile d'estimer à partir de quel seuil de capacité prédictive la sélection génomique serait avantageuse mais ces résultats ne semblent cependant pas très prometteurs. Pour certains caractères, l'utilisation d'un sous-ensemble de marqueurs associés (entre 1000 et 5000) pour calibrer le modèle de prédiction génomique permet d'atteindre la qualité de prédiction obtenue avec la totalité des marqueurs. L'inclusion des variables d'alvéographe en plus des marqueurs SNP dans un modèle de prédiction génomique permet d'améliorer significativement les capacités prédictives (+0.20 en moyenne). Grâce à l'utilisation combinée des résultats de tests indirects et de la prédiction génomique, la qualité boulangère pourrait être intégrée de manière plus précoce dans les programmes de sélection (dès lors que la quantité de farine nécessaire à la réalisation de tests indirects est disponible). Le niveau de prédiction pourrait permettre de discriminer les lignées en sélection avec une faible qualité boulangère pour intensifier la sélection sur ce critère.

Bibliographie

- AFNOR, 2015. NF EN ISO 27971 - Céréales et produits céréaliers - Blé tendre (*Triticum aestivum* L.) - Détermination des propriétés alvéographiques d'une pâte à hydratation constante de farine industrielle ou d'essai et méthodologie pour la mouture d'essai [WWW Document]. URL <https://www.boutique.afnor.org/norme/nf-en-iso-27971/cereales-et-produits-cerealiers-ble-tendre-triticum-aestivum-l-determination-des-proprietes-alveographiques-d-une-pate-a-/article/818859/fa181958> (accessed 8.10.17).
- Agreste, 2017. Grandes cultures et fourrages.
- Agreste, 2016a. Agreste Bilan conjoncturel 2016.
- Agreste, 2016b. Produits agroalimentaires. GraphAgri 120–131.
- Akaike, H., 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19, 716–723. doi:10.1109/tac.1974.1100705
- ANMF, 2016. Découvrir la meunerie - Meunerie Française [WWW Document]. URL <http://www.meuneriefrancaise.com/Default.aspx?lid=1&rid=115&rvid=115> (accessed 8.9.17).
- Association générale des producteurs de blé, Association Nationale de la meunerie française, Coop de France-Métiers du Grain, 2013. Accord interprofessionnel, qualité du blé tendre.
- Astier, M., 2016. Quel pain voulons-nous? Le Seuil.
- Bergez, J.-É., Abecassis, J., 2009. Les filières céréalières: Organisation et nouveaux défis. Editions Quae.
- BIPEA, 2017. Page d'accueil | BIPEA [WWW Document]. URL <https://www.bipea.org/fr/content/page-daccueil> (accessed 8.10.17).
- Bokobza, L., 1998. Near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 6, 3–18.
- Bonjean, A., 2001. Histoire de la culture des céréales et en particulier de celle du blé tendre (*Triticum aestivum* L.). Dossier de l'environnement de l'INRA 21, 29–37.
- Branlard, 2012. La qualité du gluten: variations de sa composition et de ses propriétés. *Médecine et nutrition* 48, 21–25.
- Branlard, Dardevet, Saccomano, Lagoutte, Gourdon, 2001. Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica* 119, 59–67.
- Branlard, G., Autran, J.-C., 1986a. L'amélioration génétique de la qualité technologique du blé tendre.
- Branlard, G., Autran, J.-C., 1986b. L'amélioration génétique de la qualité technologique du blé tendre.
- Branlard, G., Dardevet, 1985. Diversity of grain protein and bread wheat quality II. Correlation between high molecular weight subunits of glutenin and flour quality characteristics. *Journal of Cereal Science* Pages 329-343.
- Branlard, G., Dardevet, M., Saccomano, R., Lagoutte, F., Gourdon, J., 2001. Genetic Diversity of Wheat Storage Proteins and Bread Wheat Quality pp 59–67.
- Branlard, Le Blanc, 1985. Les sous-unités gluténines de haut poids moléculaire des blés tendres et des blés durs cultivés en France 467–477.
- Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L.A., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M.-C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F.X., Edwards, K.J., Bevan, M.W., Hall, N., 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491, 705–710. doi:10.1038/nature11650
- Cauvain, S.P., 2003. Bread Making: Improving Quality. Elsevier.

- Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M.-F., Cattolico, L., Beckert, M., Aubourg, S., Weissenbach, J., Caboche, M., Bernard, M., Leroy, P., Chalhou, B., 2005. Molecular Basis of Evolutionary Events That Shaped the Hardness Locus in Diploid and Polyploid Wheat Species (*Triticum* and *Aegilops*). *Plant Cell* 17, 1033–1045. doi:10.1105/tpc.104.029181
- Charmet, G., Storlie, E., Oury, F.X., Laurent, V., Beghin, D., Chevarin, L., Lapierre, A., Perretant, M.R., Rolland, B., Heumez, E., Duchalais, L., Goudemand, E., Bordes, J., Robert, O., 2014. Genome-wide prediction of three important traits in bread wheat. *Molecular Breeding* 34, 1843–1852. doi:10.1007/s11032-014-0143-y
- Chen, J., Chen, Z., 2008. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* 95, 759–771. doi:10.1093/biomet/asn034
- Chen, L., Li, C., Sargolzaei, M., Schenkel, F., 2014. Impact of Genotype Imputation on the Performance of GBLUP and Bayesian Methods for Genomic Prediction. *PLOS ONE* 9, e101544. doi:10.1371/journal.pone.0101544
- Colin, P., 2016. Evaluation de la précision des modèles de prédiction de la qualité boulangère du blé tendre (*Triticum aestivum* L.). CFR Rennes.
- Crossa, J., Campos, G. d. I., Perez, P., Gianola, D., Burgueno, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., Braun, H.-J., 2010. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186, 713–724. doi:10.1534/genetics.110.118521
- Dubcovsky, J., 2004. Marker-assisted selection in public breeding programs: the wheat experience. *Crop Science* 44.
- Dubois, M., 2016. *AlveoConsistograph Handbook*. Elsevier.
- Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal* 4, 250. doi:10.3835/plantgenome2011.08.0024
- Endelman, J.B., Jannink, J.-L., 2012. Shrinkage Estimation of the Realized Relationship Matrix. *G3: Genes|Genomes|Genetics* 2, 1405–1413. doi:10.1534/g3.112.004259
- FAOSTAT, 2011. FAOSTAT [WWW Document]. URL <http://www.fao.org/faostat/en/#data> (accessed 8.9.17).
- Feillet, P., 2000. Le grain de blé: Composition et utilisation. Quae.
- Feingold, J., 1991. Le déséquilibre de liaison. *médecine/sciences* 7, 161–168.
- FranceAgriMer, Arvalis, 2014. Qualité des blés français.
- Goddard, M.E., Hayes, B.J., 2007. Genomic selection. *Journal of Animal breeding and Genetics* 124, 323–330.
- Gondro, C., van der Werf, J., Hayes, B. (Eds.), 2013. *Genome-Wide Association Studies and Genomic Prediction, Methods in Molecular Biology*. Humana Press, Totowa, NJ. doi:10.1007/978-1-62703-447-0
- Groos, C., Charmet, G., 2001. Analyse génétique de critères de qualité en panification française chez le blé tendre (*Triticum aestivum* L.) = Genetic analysis of French bread-making quality in bread wheat (*Triticum aestivum* L.).
- Gupta, Batey, MacRitchie, 1992. Relationships between protein composition and functional properties of wheat flours. *Cereal Chemistry* 69, 125–131.
- Gupta, P.K., Mir, R.R., Mohan, A., Kumar, J., 2008. Wheat Genomics: Present Status and Future Prospects. *Int J Plant Genomics* 2008. doi:10.1155/2008/896451
- Guzmán, C., Medina-Larqu e, A.S., Velu, G., Gonz alez-Santoyo, H., Singh, R.P., Huerta-Espino, J., Ortiz-Monasterio, I., Pe na, R.J., 2014. Use of wheat genetic resources to develop biofortified wheat with enhanced grain zinc and iron concentrations and

- desirable processing quality. *Journal of Cereal Science* 60, 617–622. doi:10.1016/j.jcs.2014.07.006
- Heffner, E.L., Jannink, J.-L., Iwata, H., Souza, E., Sorrells, M.E., 2011. Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Science* 51, 2597. doi:10.2135/cropsci2011.05.0253
- Heslot, N., Yang, H.-P., Sorrells, M.E., Jannink, J.-L., 2012. Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science* 52, 146. doi:10.2135/cropsci2011.06.0297
- Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. *TAG Theoretical and Applied Genetics* 38, 226–231.
- Hoffman, G.E., 2013. Correcting for Population Structure and Kinship Using the Linear Mixed Model: Theory and Extensions. *PLoS One* 8. doi:10.1371/journal.pone.0075707
- Kleijer, G., 2002. Sélection des variétés de blé pour la qualité boulangère. *Revue suisse d'agriculture* 34, 253–259.
- Lande, R., Thompson, R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756.
- Lê, S., Josse, J., Husson, F., others, 2008. FactoMineR: an R package for multivariate analysis. *Journal of statistical software* 25, 1–18.
- Lecomte, C., 2005. L'évaluation expérimentale des innovations variétales Proposition d'outils d'analyse de l'interaction génotype - milieu adaptés à la diversité des besoins et des contraintes des acteurs de la filière semences.
- Lersten, N.R., 1987. Morphology and anatomy of the wheat plant. *Wheat and Wheat Improvement* 33–75.
- Lévy, Brabant, 2014. Facteurs influençant la teneur en protéines – résultats des projets de recherche d'Agroscope.
- Li, J., Ji, L., 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95, 221.
- Li, J., Wan, H.-S., Yang, W.-Y., 2014. Synthetic hexaploid wheat enhances variation and adaptive evolution of bread wheat in breeding processes. *Jnl of Systematics Evolution* 52, 735–742. doi:10.1111/jse.12110
- Li, Y., Song, Y., Zhou, R., Branlard, G., Jia, J., 2009. Detection of QTLs for bread-making quality in wheat using a recombinant inbred line population. *Plant Breeding* 128, 235–243. doi:10.1111/j.1439-0523.2008.01578.x
- Lindsay, M.P., Skerritt, J.H., 1999. The glutenin macropolymer of wheat flour doughs: structure–function perspectives. *Trends in Food Science & Technology* 10, 247–253.
- Liu, H., Zhou, H., Wu, Li, Zhao, Zuo, 2015. The Impact of Genetic Relationship and Linkage Disequilibrium on Genomic Selection. *PLoS One* 10, e0132379–e0132379. doi:10.1371/journal.pone.0132379
- Méléard, B., 2014. Le rôle des protéines en panification.
- Meuwissen, T.H., Hayes, B.J., Goddard, G., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Michel, S., Gallee, M., Löschenberger, F., Buerstmayr, H., Kummer, C., 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. *Journal of Cereal Science* 77, 24–34. doi:10.1016/j.jcs.2017.07.012
- Moreau, L., Nicolas Bardol, Fabio Valente, Alain Charcosset, 2011. De la SAM à la sélection génomique : exemple du maïs Le Sélectionneur Français, 55–67.

- Nedelkou, I.-P., Maurer, A., Schubert, A., Léon, J., Pillen, K., 2017. Exotic QTL improve grain quality in the tri-parental wheat population SW84. *PLOS ONE* 12, e0179851. doi:10.1371/journal.pone.0179851
- Osborne, T.B., 1907. *The proteins of the wheat kernel*. Washington, D.C.: Carnegie Institution of Washington.
- Oury, F.-X., Chiron, H., Faye, A., Gardet, O., Giraud, A., Heumez, E., Rolland, B., Rousset, M., Trottet, M., Charmet, G., Branlard, 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. doi:10.1007/s10681-009-9997-1
- Partners and Food in Emergency and Development Aid, 1997. *La filière pain*.
- Patterson, N., Price, A.L., Reich, D., 2006. Population Structure and Eigenanalysis. *PLoS Genet* 2. doi:10.1371/journal.pgen.0020190
- Paulic Minotiers, 2016. *Tarifs des analyses sur farine et blé tendre*.
- Plessis, A., Ravel, C., Bordes, J., Balfourier, F., Martre, P., 2013. Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. *Journal of Experimental Botany* 64, 3627–3644. doi:10.1093/jxb/ert188
- Rasheed, A., Safdar, T., Gul-Kazi, A., Mahmood, T., Akram, Z., Mujeeb-Kazi, A., 2012. Characterization of HMW-GS and evaluation of their diversity in morphologically elite synthetic hexaploid wheats. *Breeding Science* 62, 365–370. doi:10.1270/jsbbs.62.365
- Roldan, D.L., 2011. *Détection de QTL: interaction entre dispositif expérimental et méthodes statistiques*. Université de Toulouse, Université Toulouse III-Paul Sabatier.
- Roussel, P., Chiron, H., Della Valle, G., Ndiaye, A., 2010. *Glossaire terminologique appliqué aux pains français*. INRA CEPIA. Available from: <http://www4.inra.fr/cepia/Editions/glossaire-pains-francais> [June 2011].
- Rousset, M., 1978. *Amélioration du blé tendre pour sa valeur d'utilisation*.
- Sahin, S., Sumnu, S.G., 2007. *Physical Properties of Foods*. Springer Science & Business Media.
- Segura, 2017. *mlmm: An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations*. GMI: Gregor Mendel Institute of Molecular Plant Biology GmbH.
- 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. doi:10.1038/ng.2314
- Shewry, P.R., Halford, N.G., Belton, P.S., Tatham, A.S., 2002. The structure and properties of gluten: an elastic protein from wheat grain. *Philosophical Transactions of the Royal Society B: Biological Sciences* 357, 133–142. doi:10.1098/rstb.2001.1024
- Snyder, L.R., Kirkland, J.J., Glajch, J.L., 2012. *Practical HPLC Method Development*. John Wiley & Sons.
- Stevenson, K.R., 1987. The relationship between baking quality and protein content: some cultivar and environmental effects, in: *Proceedings of the Agronomy Society of New Zealand*. pp. 121–127.
- Tamura, T., Osawa, M., Kimura, R., Inaoka, Y., Tanaka, S., Satoh, F., Sato, I., 2013. Evaluation of the allele-sharing approach, known as the IBS method, in kinship analysis. *Journal of Forensic and Legal Medicine* 20, 112–116. doi:10.1016/j.jflm.2012.05.005

- Uthayakumaran, S., Newberry, M., Keentok, M., Stoddard, F.L., Bekes, F., 2000. Basic rheology of bread dough with modified protein content and glutenin-to-gliadin ratios. *Cereal Chemistry* 77, 744–749.
- Vieira, I.C., dos Santos, J.P.R., Pires, L.P.M., Lima, B.M., Gonçalves, F.M.A., Balestre, M., 2017. Assessing non-additive effects in GBLUP model. *Genetics and Molecular Research* 16. doi:10.4238/gmr16029632
- Voorrips, R.E., 2002. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J Hered* 93, 77–78. doi:10.1093/jhered/93.1.77
- Weegels, P.L., Hamer, R.J., Schofield, J. 1994, 1996. Functional properties of wheat glutenin. *Journal of Cereal Science* 23, 1–17.
- Wrigley, C.W., 1982. The use of genetics in understanding protein composition and grain quality in wheat. *Plant Foods for Human Nutrition* 31, 205–227. doi:10.1007/BF01108631
- Wrigley, C.W., Corke, H., Seetharaman, K., Faubion, J., 2015. *Encyclopedia of Food Grains*. Academic Press.
- Yara, 2015. Wheat quality: how to increase proteins? *Pure Nutrients Facts*.
- Yu, J., Pressoir, G, Briggs, W. H, Bi, V., Yamasaki, 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics* 38, 203–208.
- Zhao, F., Hawkesford, M., McGrath, S., 1999. Sulphur Assimilation and Effects on Yield and Quality of Wheat. *Journal of Cereal Science* 30, 1–17. doi:10.1006/jcrs.1998.0241
- Zhen, S., Han, C., Ma, C., Gu, A., Zhang, M., Shen, X., Li, X., Yan, Y., 2014. Deletion of the low-molecular-weight glutenin subunit allele Glu-A3a of wheat (*Triticum aestivum*L.) significantly reduces dough strength and breadmaking quality. *BMC Plant Biology* 14, 367. doi:10.1186/s12870-014-0367-3

ANNEXE

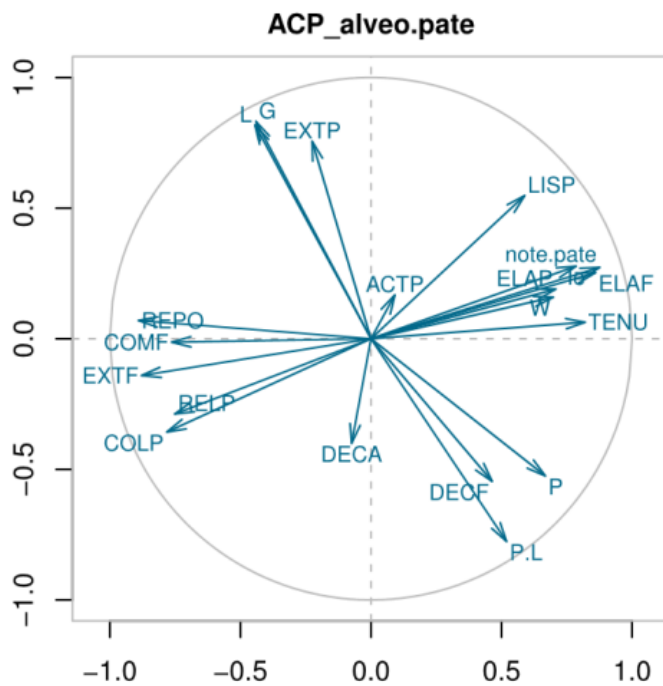
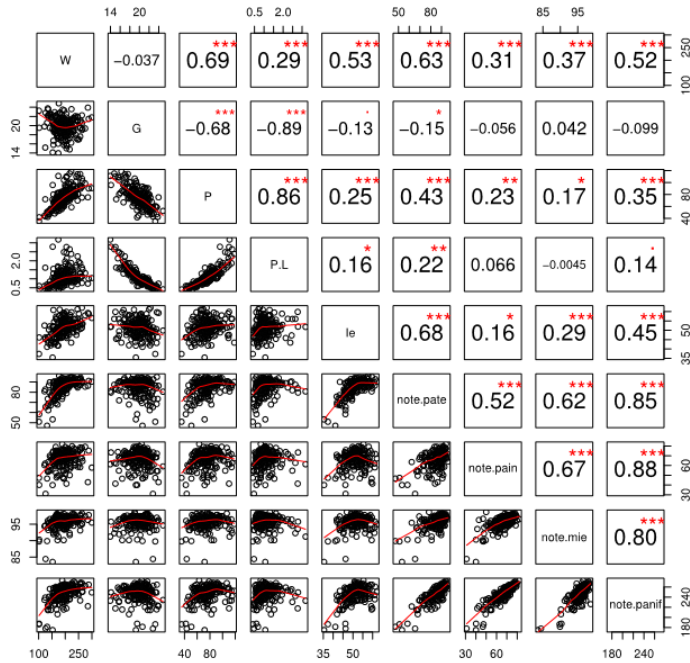
Annexe 1 : Grille de notation du test BIPEA selon la norme NF V03-716 (BIPEA, 2017)

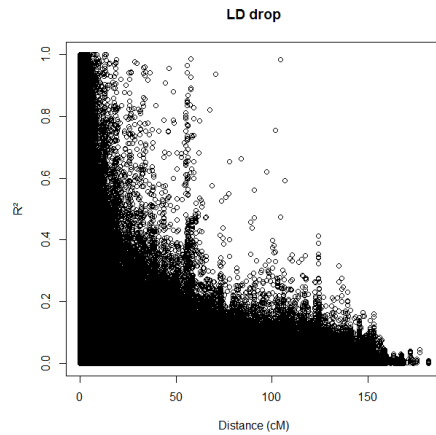
		Insuffisant	Excès						
		1	4	7	10	7	4	1	
PETRISSAGE									
LISSAGE									x 0,5 = #N/A
COLLANT DE LA PATE									x 0,5 = #N/A
CONSISTANCE									x 0,5 = #N/A
EXTENSIBILITE									x 0,5 = #N/A
ELASTICITE									x 0,5 = #N/A
RELACHEMENT									x 0,5 = #N/A
									#N/A x #N/A = #N/A /25
									TOTAL COEFFICIENT PETRISSAGE
POINTAGE									
DETENTE : RELACHEMENT									x 1 = #N/A /10
									POINTAGE
FACONNAGE									
ALLONGEMENT									x 0,5 = #N/A
DECHIREMENT									x 0,5 = #N/A
ELASTICITE									x 0,5 = #N/A
COLLANT DE LA PATE									x 1 = #N/A
									#N/A x #N/A = #N/A /25
									TOTAL COEFFICIENT FACONNAGE
									#N/A /100
									NOTE PATE
APPRET									
ACTIVITE FERMENTAIRE									x 0,5 = #N/A
PATE : DECHIREMENT									x 0,5 = #N/A
									#N/A /10
									APPRET
MISE AU FOUR									
COLLANT DE LA PATE									x 1 = #N/A
TENUE DE LA PATE									x 2 = #N/A
									#N/A x #N/A = #N/A /30
									TOTAL COEFFICIENT MISE AU FOUR
CARACTERISTIQUES DU PAIN									
SECTION									x 1 = #N/A
COULEUR									x 2 = #N/A
EPAISSEUR									x 0,5 = #N/A
CRUSTILLANT									x 0,5 = #N/A
									#N/A x #N/A = #N/A /70
									TOTAL COEFFICIENT ASPECT DU PAIN
COUPS DE LAME									
DEVELOPPEMENT									x 1 = #N/A
REGULARITE									x 1 = #N/A
DECHIREMENT									x 1 = #N/A
									#N/A /100
									NOTE PAIN
VOLUME									
VOLUME MOYEN	1er	2e	3e	4e					cm ³
MASSE MOYENNE									g
	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	0
									#DIV/0!
									Moyenne
									#DIV/0! cm ³
									#DIV/0! g
									DENSITE (g/cm3)
									#DIV/0! /30
									VOLUME
CARACTERISTIQUES DE LA MIE									
COULEUR									x 1 = #N/A
TEXTURE									x 3 = #N/A
PLASTICITE									x 1 = #N/A
COLLANT									x 1 = #N/A
									#N/A /100
									NOTE MIE
ALVEOLAGE									
REGULARITE									x 2 = #N/A
ODEUR									x 3 = #N/A
									#N/A /300
									NOTE TOTALE

NB : La grille doit être remplie avec la lettre 'x'. Ne mettre qu'une seule croix par ligne.

Annexe 2:

En haut : Corrélation entre les variables d'alvéographe et les notes de pâte, pain, mie et totale. En bas : Cercle de corrélation des variables d'alvéographe avec les composantes de la note de pâte. Les noms des caractères sont répertoriés dans le tableau 4.

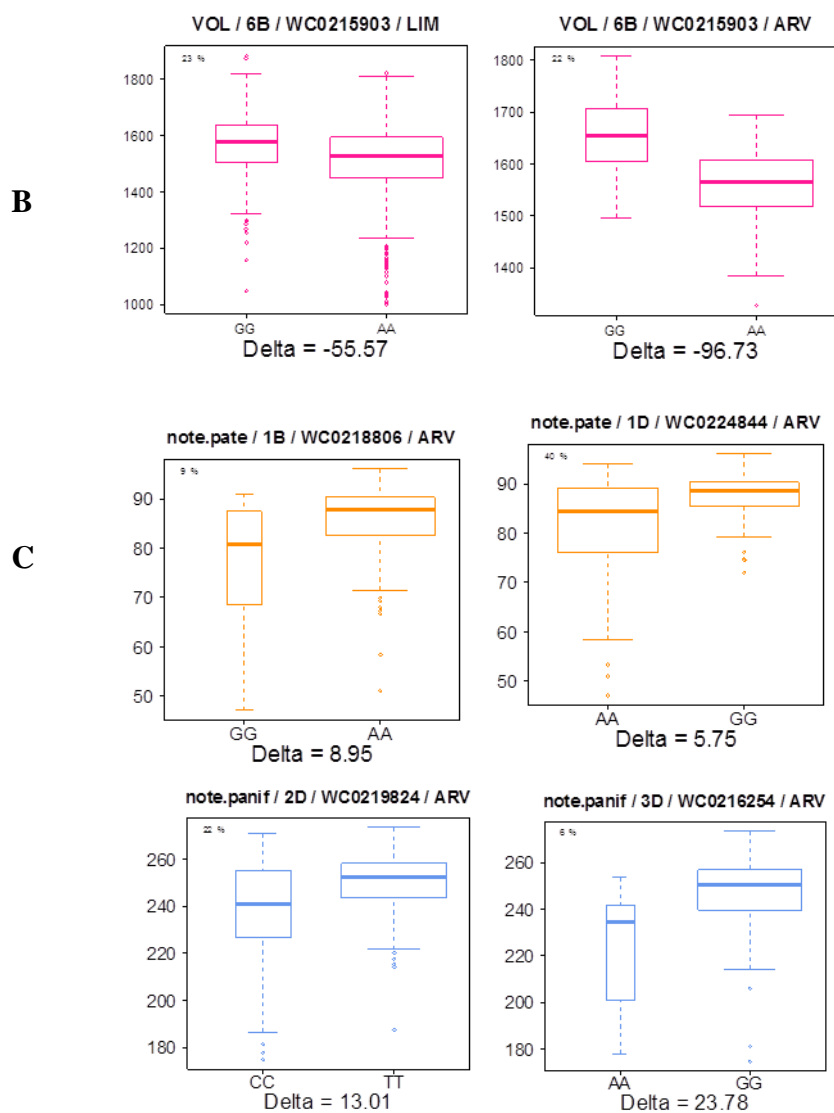




Annexe 3: Evolution du déséquilibre de liaison sur le panel Arvalis en fonction de la distance entre marqueurs. Le déséquilibre de liaison a été estimé par le calcul d'un coefficient de corrélation pour toutes les combinaisons de marqueurs deux à deux.

	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

Annexe 4 : Occurrence des variétés testées en fonction des années. Le chiffre indiqué représente le nombre de variétés communes entre les deux années (ligne, colonne). Les cases sur la diagonale représentent les variétés présentes dans l'année correspondante.



Annexe 5:

A : Distribution du volume du pain pour les deux classes alléliques au marqueur WC0215903 localisé sur le chromosome 6B pour le panel Limagrain (à gauche) et le panel Arvalis (à droite).

B : Distribution de la note de pâte pour les deux classes alléliques au marqueur WC0218806 localisé sur le chromosome 1B (à gauche) et au marqueur WC0224844 localisé sur le chromosome 1D (à droite) pour le panel Arvalis.

C : Distribution de la note de panification pour les deux classes alléliques au marqueur WC0219824 localisé sur le chromosome 2D (à gauche) et au marqueur WC0216254 localisé sur le chromosome 3D (à droite) pour le panel Arvalis.