

Les déterminismes génétiques de la tolérance au froid et de la précocité chez l'orge

Justine Flament

▶ **To cite this version:**

Justine Flament. Les déterminismes génétiques de la tolérance au froid et de la précocité chez l'orge. Sciences agricoles. 2014. dumas-01114775

HAL Id: dumas-01114775

https://dumas.ccsd.cnrs.fr/dumas-01114775

Submitted on 9 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AGROCAMPUS OUEST - CFR Rennes

65, rue de Saint Brieuc

CS 84215 - 35042 RENNES cedex

AGROSUP DIJON

26, boulevard Docteur Petitjean

BP 87999 - 2109 DIJON cedex

Limagrain Europe

Site LA GARENNE - Route d'Ennezat

63720 CHAPPES

Mémoire de Fin d'Etudes

Diplôme d'Ingénieur de l'Institut Supérieur des Sciences Agronomiques, Agroalimentaires, horticoles et du Paysage

Année universitaire : 2013 - 2014

Spécialité : Sciences et Productions Végétales

Spécialisation et Option : Amélioration des plantes

Les déterminismes génétiques de la tolérance au froid et de la précocité chez l'orge

Par : Justine FLAMENT

CONFIDENTIEL



Volet à renseigner par l'enseignant responsable de l'option ou son représentant*

Date : …./…/… Signature

Bon pour dépôt (version définitive) ☐

Autorisation de diffusion : Oui ☐ Non☐

Devant le jury *:*

Soutenu à Rennes le 12 septembre 2014

Sous la présidence de : Maria MANZANARES - DAULEUX

Maître de stage : Anne-Marie BOCHARD

Enseignant référent : Maria MANZANARES - DAULEUX

Rapporteur : Alain BARANGER

Examinateur : Mélanie JUBAULT

Autres membres du jury (Nom, Qualité):

« Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST »

Fiche de confidentialité et de diffusion du mémoire

Cadre lié à la confidentialité :

Aucune confidentialité ne sera prise en compte si la durée n'en est pas précisée.

Préciser les limites de la confidentialité [(2)] **:**

➢ Confidentialité absolue : ☐ oui ☐non

(ni consultation, ni prêt)

Si oui ☐1 an ☐5 ans ☐10 ans

Le maître de stage[(4)]**,**

Cadre lié à la diffusion du mémoire :

A l'issue de la période de confidentialité et/ou si le mémoire est validé diffusable sur la page de couverture, il sera diffusé sur les bases de données documentaires nationales et internationales selon les règles définies ci-dessous :

Diffusion de la version numérique **du mémoire** : ☐oui ☐non

Référence bibliographique diffusable[(3)] : ☐oui ☐non

Résumé diffusable : ☐oui ☐non

Mémoire papier consultable sur place : ☐oui ☐non

Reproduction autorisée du mémoire : ☐oui ☐non

Prêt autorisé du mémoire papier : ☐oui ☐non

……..

Diffusion de la version numérique **du résumé** : ☐oui ☐non

↳ *Si* oui, l'auteur[(1)] complète l'autorisation suivante :

Je soussigné(e) , propriétaire des droits de reproduction dudit résumé, autorise toutes les sources bibliographiques à le signaler et le publier.

Date : Signature :

Rennes/Angers, le

L'auteur[(1)]**,** **L'enseignant référent**

ou son représentant

(1) auteur = étudiant qui réalise son mémoire de fin d'études

(2) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.

(3) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé.

(4) Signature et cachet de l'organisme

Remerciements

Mes remerciements vont tout d'abord à Anne-Marie BOCHARD. Merci de m'avoir choisie pour cette mission. Merci pour ton encadrement hors pair que l'on ne peut que recommander. Merci pour la confiance et l'indépendance que tu m'as accordées. Merci pour l'attention que tu as portée à la correction de mon rapport. Merci pour ton soutien, ton aide et ta bonne humeur. Pour tout, merci.

À Constance LAVERGNE, ma très chère collègue, pour le nettoyage des données moléculaires que tu as réalisé et qui m'a été des plus utiles, mais surtout pour ton soutien, ton aide et ta bonne humeur au quotidien.

À Stephanie CHAUVET, pour sa patience et son aide précieuse dans l'utilisation du logiciel MAST et la production de données d'analyse. Merci également à Sébastien DUCROCQ et Alexandre MULLER pour leur aide dans le développement de scripts R.

À Jordi COMADRAN, pour son expertise et son aide dans la compréhension du fonctionnement des gènes de précocité.

À Agathe MINI et Mickaël THROUDE pour leur aide dans l'acquisition des compétences d'utilisation des logiciels Strudel et Biomercator.

À Bernard DUPERRIER et Fabien LECOUVIOUR pour leur accueil sur le site de Mons pour une journée de participation aux travaux de castration et de croisement sur du blé.

Enfin, je tiens à remercier toutes les personnes du site la Garenne qui ont participé à mon intégration au sein de l'entreprise. À tous, pour votre accueil et votre bonne humeur, merci.

Table des matières

Remerciements II
Liste des figures et tableaux VIII
Listes des abréviations XII
Introduction 2
Présentation générale de l'espèce 2
Les objectifs et critères de sélection 4
Présentation de la précocité et de la tolérance au froid 4
Problématique et démarche 14
Matériels et Méthodes 16
1- Germplasm 16
2- Données et outils de phénotypage 18
Les variables 18
Le nettoyage des données 18
Les dispositifs expérimentaux et l'ajustement des données phénotypiques 18
3- Données de génotypage 20
4- Cartographie des marqueurs 20
5- Imputation des données de marquage 20
6- Méthode d'analyses d'associations 22
7- Méthode de sélection génomique 24
Résultats 26
1- Germplasm 26
2- Données et outils de phénotypage 26
Rendement 26
Précocité 28
Résistance au froid 28
Corrélation entre les variables 28
Héritabilité 28
3- Cartographie des marqueurs 30
4- Imputation des données de marquage 32
Qualité d'imputation des données manquantes 32
Résultat sur l'efficacité de l'imputation 32
5- Résultats d'analyses d'associations 34

Impact de l'imputation 34
Résultats des études d'association concernant la précocité 34
Résultat des études d'association concernant le froid 34
Différentiations alléliques 36
6- Résultats de sélection génomique 36
Discussion 38
Conclusions et Perspectives 42
Bibliographie 46
Annexes 54

Liste des figures et tableaux

Figure 1 : Evolution de la température en France au cours de l'hiver 2011-2012. Diagnostic élaboré à partir d'un indicateur thermique constitué de la moyenne des températures quotidiennes de 30 stations métropolitaines. (Météo France, 2012) 7
Figure 2 : Proportion départementale des surfaces d'orge d'hiver détruites par le gel (ressemis indispensable) - % (Deswarte, 2012) 9
Figure 3 : Cartographie du chromosome de l'orge 5HL utilisant des populations "Nure"x"Tremois" HD et F2. (Francia, et al., 2007) 11
Figure 4 : Ensemble des gènes intervenant directement ou indirectement dans la tolérance au froid. Légende : ld = long day ; sd = short day ; v = vernalisation ; (Figure élaborée avec les travaux cités dans la partie « Présentation de la précocité et de la tolérance au froid ») 13
Figure 5 : Schéma explicatif de la méthode d'imputation des données manquantes 19
Figure 6 : Évaluation de regroupement par la méthode SSE. La rupture de pente matérialise le nombre de groupes. 25
Figure 7 : Répartition des individus selon leur note WKI1 (0 : résistant, 10 sensible) en fonction de la latitude (la latitude 2 est la latitude intermédiaire ; la latitude 3, la latitude méridionale). 27
Figure 8 : Carte consensus réalisée sur la base d'une carte Morex*Barke ainsi qu'une carte génétique et une carte physique 29
Figure 9 : Répartition des espaces entre marqueurs par chromosome (à gauche) et sur le génome entier (à droite) 30
Figure 10 : Distribution du Call Rate par individu avant (a) et après (b) imputation 31
Figure 11 : Visualisation du gain de données moléculaires en fonction du Call Rate des parents (pourcentage de données de marquages). Le Call Rate de référence (sans imputation) du parent 1 est matérialisé par un gradient de couleur allant du blanc (CR minimum de 3%) au rose (CR maximum de 98%) en passant par le jaune (CR de 50%). Le Call Rate de référence du parent 2 est matérialisé par un gradient de taille allant de 2.95% de CR pour les cercles les plus petits à 98% de CR pour les cercles les plus grands. 31
Figure 12 : Récapitulatif du nombre de QTLs en fonction des variables 33

Tableau 1 : Récapitulatif des seuils minimum et maximum choisis pour la détection de données aberrantes 18
Tableau 2 : Récapitulatif de la taille des jeux de données après la réalisation de la correspondance entre les sous-populations et les données de marquage moléculaire imputées (CR de 80%) . Les populations en rouge ne sont pas analysées car elles comptent moins de 100 individus. 22
Tableau 3 : Comparaison de l'héritabilité des variables de précocité suivant les types d'orges et les populations "type x climat" 25
Tableau 4: Répartition des individus des trois groupes selon les types variétaux printemps : S, hiver deux rangs : W2, hiver six rangs : W6. Les variétés hiver dont le nombre de rangs n'a pas été identifié sont placées dans le groupe WI. 26
Tableau 5 : Résumé des moyennes des héritabilités des variables étudiées 27

Tableau 6 : Résumé d'observations de BoxPlots "rendement VS pays" par type 28
Tableau 7 : Résumé du test de la qualité d'imputation en fonction du pourcentage de données manquantes. "-" pourcentage faible de données manquantes, "+" pourcentage élevé de données manquantes. 32
Tableau 8 : Récapitulatif des r² Cross Validés (r² CV), des héritabilités des modèles de moyennes ajustées (h²) et du nombre d'individus en fonction du type variétal 35
Tableau 9 : Présentation des majorités alléliques des loci FR-H3, Ab_Fro_5H_168, FR-H2 et PPD-H1 selon rdt+ et rdt-. S : sensible au froid, R : Résistant au froid, T : tardif, P : précoce, = : proportions égales allèle précoce et tardif. + : l'allèle majoritaire est présente à plus de 75 %. En vert : les loci ne montrant pas de différences entre lieux. En jaune : Les lieux montrant des différences entre lieux. 36

Listes des abréviations

BLUP : Best Linear Unbiased Prediction

CCA1 : Circadian Clock Associated 1

DOE1 : Date of Earing

EAR : Earing

EARC : Earing Code

FAOSTAT : Food and Agriculture Organization of the United Nations Statistics

FD : Une protéine de médiation des signaux

FT : Locus de Floraison T

GI : Gigantea

GNIS : Groupement National Interprofessionnel des Semences et Plants

GS : Sélection Génomique

HvCEN : Hordeum vulgare Centroradialis

IBGSC : International Barley Genome Sequencing Consortium

IGC : International Grains Council

Mat-a ou EAM8 : Praematurum – a

NGS : Next-generation sequencing

PPD-H (1 ou 2): Photoperiod-H (1 ou 2)

RIP : Ripening

SSE : Sum of Squares due to Error

TOC1 : Pseudo-Response Regulator

VRN-H (1,2 ou 3) : Vernalization-H (1,2 ou 3)

WKI1, WKI2 : Winter Kill 1 et 2

Introduction

Présentation générale de l'espèce

L'orge, *Hordeum vulgare*, est une céréale. Elle se place au 4ème rang mondial en terme de production derrière le maïs, le riz et le blé. C'est la céréale qui possède la zone de culture la plus étendue de par les latitudes, altitudes ou conditions pédoclimatiques qu'elle occupe (Briggs, 1978). Sa culture remonte aussi loin que les débuts de l'agriculture (Bonjean & Picard, 1990). Elle est originaire de Mésopotamie (Irak actuelle) où elle est apparue 7000 ans av. J.C. et descend de l'orge sauvage 2 rangs (*H.spontaneum*) (GNIS, 2006). En 2012, la production mondiale comptabilise 133 millions de tonnes pour une surface de 49.5 millions d'hectares. Les principaux producteurs d'orge sont l'Europe, la Russie et l'Australie. De manière générale c'est 20 % de la production qui circule chaque année (25.5 millions de tonnes) et les pays producteurs exportent entre autres vers l'Arabie Saoudite, la Chine et le Japon. Ces pays étaient les premiers importateurs en 2011 (FAOSTAT, 2013).

Dans le monde, l'orge est principalement utilisée pour l'alimentation animale, à hauteur de 73 %. Ensuite, 19 % de la production est utilisée en malterie, 6 % pour l'alimentation humaine et 2 % pour d'autres filières industrielles comme la filière bioéthanol (Reg & Chengdao, 2008) (Passion Céréales, 2011). Après une baisse de la production depuis le début des années 90 (Bonjean & Picard, 1990 ; Moule, 1980 ; USDA, 2013), un certain essor peut aujourd'hui être observé. En effet, l'International Grains Council prévoit une augmentation de la production de 11 % en 2014 et de 17 % d'ici 2018 (IGC, 2013).

L'orge est une plante monocotylédone de la famille des Poaceae, tribu des Triticeae, avec un génome diploïde à 2n=2x=14 chromosomes. C'est une plante annuelle avec un épi barbu qui peut mesurer entre 60 cm et 1.3m de haut. Elle est de type hiver ou printemps en fonction des dates de semis. Chaque type peut être dit 6 (*H. hexastichum L.*) ou 2 rangs (*H. distichum*). Le nombre de rangs est une description servant à caractériser le nombre d'épillets présents à chaque étage du rachis. Dans le cas d'une orge 6 rangs on y compte 3 épillets et seulement 1 pour une orge 2 rangs, les épillets latéraux étant stériles (Moule, 1980 ; GNIS, 2006). Le mode de reproduction est principalement l'autogamie avec une part d'allogamie résiduelle de 1% (GNIS, 2006). Le génome de l'orge, d'une taille d'environ 5.1 Gb et contenant entre 23000 et 56000 gènes (IBGSC, 2012) a été séquencé par un consortium, mais les séquences ne sont pas complétement ordonnées. En effet le séquençage a été effectué à l'aide de différentes techniques dont font parties les techniques de nouvelle génération de séquençage (NGS). Ces techniques génèrent des petits fragments de séquence (IBGSC, 2012) et le génome des céréales étant composé de 80 à 90 % de séquences d'ADN répétées, dérivées d'éléments transposables (Daron, 2011), l'assemblage des séquences est difficile. Cependant des progrès dans l'ordonnancement du génome sont régulièrement proposés : ainsi (Mascher, et al., 2013) ont créé la méthode POPSEQ (population sequencing). POPSEQ propose dans un premier temps de séquencer, grâce aux techniques deNGS, une population biparentale. Les fragments générés sont regroupés en fonction de leur profil de ségrégation et sont génétiquement ancrés sur une carte et regroupés par bin. Cette méthode permet d'ancrer 3 fois

plus de séquences du parent que l'on souhaite séquencer que les techniques de séquençage classiques seules pour un coût raisonnable (< $100 K) et en quelques mois seulement.

Les objectifs et critères de sélection

Le principal objectif de sélection, que ce soit pour les orges de printemps ou les orges d'hiver, est l'amélioration du rendement, mais les résistances aux maladies, la qualité brassicole et les caractères agronomiques tels que la verse, la hauteur ou la précocité sont également des critères importants (Doré & Varoquaux, 2006 ; Limagrain, 2013).

Les efforts de sélection ont conduit à des progrès génétiques non négligeables. Ces derniers seraient responsables de 50 % de l'augmentation du rendement des orges : 1.2 q/ha/an depuis 1970 (Doré & Varoquaux, 2006). La qualité de l'orge brassicole a pu être également améliorée grâce à l'étude d'enzymes-clés de dégradation de l'amidon dans le processus de maltage comme les β-amylases. Par exemple, la découverte de SNP prédictifs des 4 allèles de la thermostabilité de la β-amylase 1 ont permis de sélectionner sur ce critère des orges au stade plantule (Paris, et al., 2002) .

De nos jours, les conditions environnementales sont changeantes et les aléas climatiques se multiplient. L'équipe de sélection orge de Limagrain souhaite développer des variétés adaptées à ces nouvelles contraintes. Pour cela, il est nécessaire de mieux comprendre le déterminisme génétique de la tolérance aux stress abiotiques et la phénologie de l'orge. Le thème de ce stage propose d'étudier deux aspects importants pour l'adaptabilité recherchée et qui sont étroitement liés : la précocité et la tolérance au froid.

Présentation de la précocité et de la tolérance au froid[1]

Précocité

Les mécanismes moléculaires liés à la précocité sont nombreux. Leur organisation est complexe et à ce jour ils ne sont pas tous connus. Néanmoins c'est un sujet qui fait l'objet de nombreuses recherches et l'état des connaissances évolue rapidement.

La floraison est un phénomène dont la fonction est de permettre à la plante de se reproduire. Pour cela de nombreux mécanismes redondants coexistent lui permettant de capter les informations de son environnement et ainsi de fleurir mais surtout de fleurir au bon moment.

La base du mécanisme de floraison est relativement simple : Une protéine de médiation des signaux (FD) est couplée à une protéine issue d'un gène de la famille *FLOWERING LOCUS T (*FT*).* Ce couple se fixe sur la zone promotrice du gène VERNALIZATION-H1 (VRN-H1), et si le couple est activateur, VRN-H1 est transcrit, alors qu'il est inhibé si le couple FD-FT est

[1] Remarque : Les éléments bibliographiques cités dans cette partie sont rassemblés dans la Figure 4 p13.

répresseur (Abe, et al., 2005). C'est la transcription de VRN-H1 qui fait passer la plante de l'état végétatif à l'état reproducteur.

La famille FT est une très grande famille de gènes (Higgins, et al., 2010) dont les plus importants sont VRN-H3 (ou HvFT1), PHOTOPERIOD-H2 (PPD-H2), également appelé HvFT3 (Kikuchi, et al., 2009) et CENTRORADIALIS (HvCEN ou TFL1) (Comadran, et al., 2012). Les deux premiers sont exprimés dans les feuilles, le dernier au niveau de l'apex caulinaire (Higgins, et al., 2010). Tous les gènes de cette famille sont en compétition pour le même site de fixation.

C'est le niveau d'expression de ces gènes qui va activer ou non la floraison. Par exemple, si la quantité de transcrits des complexes FD-VRN-H3 ou FD-PPD-H2 est suffisante, ils activent la floraison via la transcription de VRN-H1. Les variations alléliques peuvent avoir différents effets sur ces quantités de transcrits. PPD-H2 présente un allèle hiver non fonctionnel et un allèle printemps fonctionnel qui active VRN-H1 et donc la floraison. Sur un autre principe, un gène peut être inhibiteur ou activateur. C'est le cas du gène HvCEN homologue du gène *Antirrhinum CENTRORADIALIS* d'Arabidopsis sous-jacent au locus EPS2. L'allèle hiver augmente la précocité alors que l'allèle printemps est antagoniste du premier (Comadran, et al., 2012). À cela s'ajoute la priorité de certains mécanismes sur d'autres. En effet, si l'allèle hiver de HvCEN favorise la floraison, l'action du gène ne sera visible qu'à la suite une période de froid suffisante car le mécanisme lié à la vernalisation prime sur lui.

L'expression de ces gènes FT est contrôlée de manière directe ou non par les différentes voies de régulations suivantes :

La vernalisation fait intervenir les gènes VRH-H1, VRN-H2 et VRN-H3. VRN-H2, qui code pour une protéine en doigt de zinc (Yan, et al., 2004), est un répresseur de VRN-H3 (Trevaskis, et al., 2007) : la floraison est bloquée quand VRN-H3 est inhibé. Un froid prolongé induit VRN-H1. La transcription de VRN-H1 réprime VRN-H2. VRN-H3 n'est alors plus réprimé par VRN-H2 et la transcription de VRN-H3 active la transcription de VRN-H1 dans des proportions plus importantes (Trevaskis, et al., 2006 ; von Zitzewitz, et al., 2005).

Ce mécanisme n'est présent que chez les orges d'hiver car VRN-H2 est rendu non fonctionnel par des délétions chez le type printemps (von Zitzewitz, et al., 2005)

La photopériode intervient à plusieurs endroits clés. Les jours longs activent l'expression de VRN-H2 (Trevaskis, et al., 2006). Les jours courts activent l'expression de PPD-H2 (Kikuchi, et al., 2009). Mais si PPD-H2 active VRN-H1 l'hiver (jours courts), il n'active pas nécessairement la floraison car dans le contexte d'une lignée de type hiver c'est l'inhibition par la vernalisation qui prime. Les transcrits de HvCO1, pivot entre l'horloge circadienne (décrite plus bas) et VRN-H3, ne sont stables que pendant les jours longs (Turck, et al., 2008) ce qui ne permet pas la floraison en jours courts. Et enfin, PPD-H1 est le gène qui a le plus

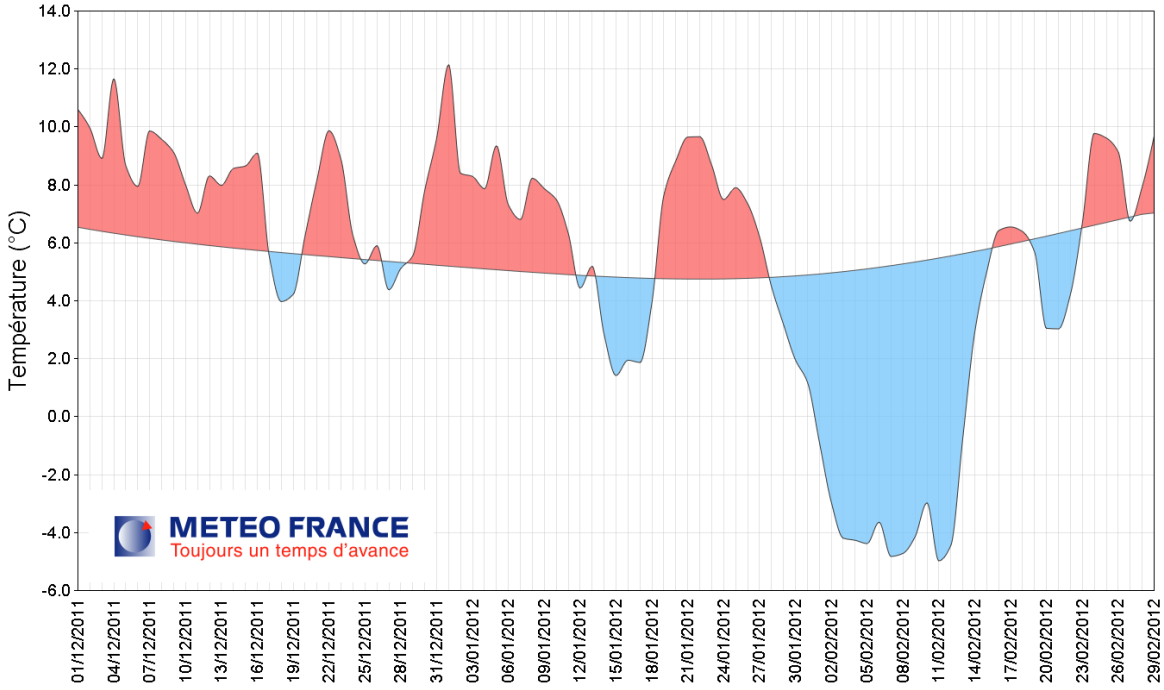


Figure 1 : Evolution de la température en France au cours de l'hiver 2011-2012. Diagnostic élaboré à partir d'un indicateur thermique constitué de la moyenne des températures quotidiennes de 30 stations métropolitaines. (Météo France, 2012)

d'effet sur le temps de floraison aussi bien en jours courts qu'en jours longs en activant la transcription de VRN-H3 (Laurie, et al., 1995).

L'horloge circadienne correspond à un rythme biologique d'environ 24h. Les gènes faisant parties de cette horloge répondent à une alternance jour-nuit qui se traduit par un phénomène d'expression-répression. Cette horloge se compose des boucles de régulations suivantes :

CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) est exprimé le matin. PSEUDO-RESPONSE REGULATOR (TOC1) est exprimé le soir. CCA1 réprime TOC1 alors que ce dernier induit l'expression de CCA1 (Higgins, et al., 2010). TOC1 lui réprime GIGANTEA (GI) alors que ce dernier induit l'expression de TOC1 (Locke, et al., 2006). Enfin, Mat-a réprime PPD-H1 et GI (Faure, et al., 2012), tout comme EAM10, gène sous-jacent à HvLUX1 (Campoli, et al., 2013). Au locus praematurum-a (Mat-a ou EAM8), le gène HvElf3 a été identifié comme étant le gène homologue chez l'orge du gène Elf3 d'Arabidopsis thaliana. (Zakhrabekova, et al., 2012).

Mat-a active aussi directement VRN-H3 par une voie indépendante de la voie de réponse à la photopériode. (Faure, et al., 2012). En jours longs, lorsque les transcrits HvCo1 sont stables, GI active VRN-H3 en passant par le gène HvCO1 (Turck, et al., 2008 ; Campoli, et al., 2012) (chromosome 7H) qui est un gène orthologue à CONSTANS (CO) d'Arabidopsis (Higgins, et al., 2010). PPD-H1 active également VRN-H3 indépendamment de HvCO1 en agissant en aval de celui-ci (Campoli, et al., 2012). PPD-H1 fait également partie des gènes dont l'expression varie selon l'horloge circadienne. (Turner, et al., 2005)

La voie autonome (*Arabidopsis thaliana*).
Il existe une voie de régulation autonome mettant en jeu les gènes *TEMPRANILLO* (TEM), notamment TEM1 et TEM2. Ils agissent comme un répresseur des locus FT. Ce mécanisme évite les floraisons prématurées. Il existe donc un équilibre entre les activités de CONSTANS et TEMPRANILLO (Castillejo & Pelaz, 2008). Les gènes TEM inhibent la production d'acide gibbérellique (Osnato, et al., 2012) qui elle active la floraison via le promoteur LEAFY (Blázquez, et al., 1998).

Tolérance au froid

Depuis 50 ans, de grands efforts de sélection ont été fournis pour améliorer le rendement et la résistance aux stress biotiques et abiotiques. Mais malgré cela, et au vu de l'augmentation de la population et des aléas climatiques qui fragilisent les récoltes, il faudrait encore doubler le rendement d'ici 2050 (Xu, 2012). Parmi les stress affectant les cultures, il est intéressant de se pencher sur le cas du froid. La vague de froid de l'hiver 2011-2012 a été particulièrement violente en Europe. Des événements de ce type restent assez rares mais cela représente une occasion de mieux connaitre ce sujet. Le germplasm utilisé en sélection est-il tolérant ou non au froid ? Quels sont les mécanismes liés à ce phénomène ?

Le froid de l'hiver 2012 se caractérise par un début d'hiver très doux en décembre et janvier, qui n'a pas permis l'endurcissement (décrit par la suite), suivi par un épisode de froid intense

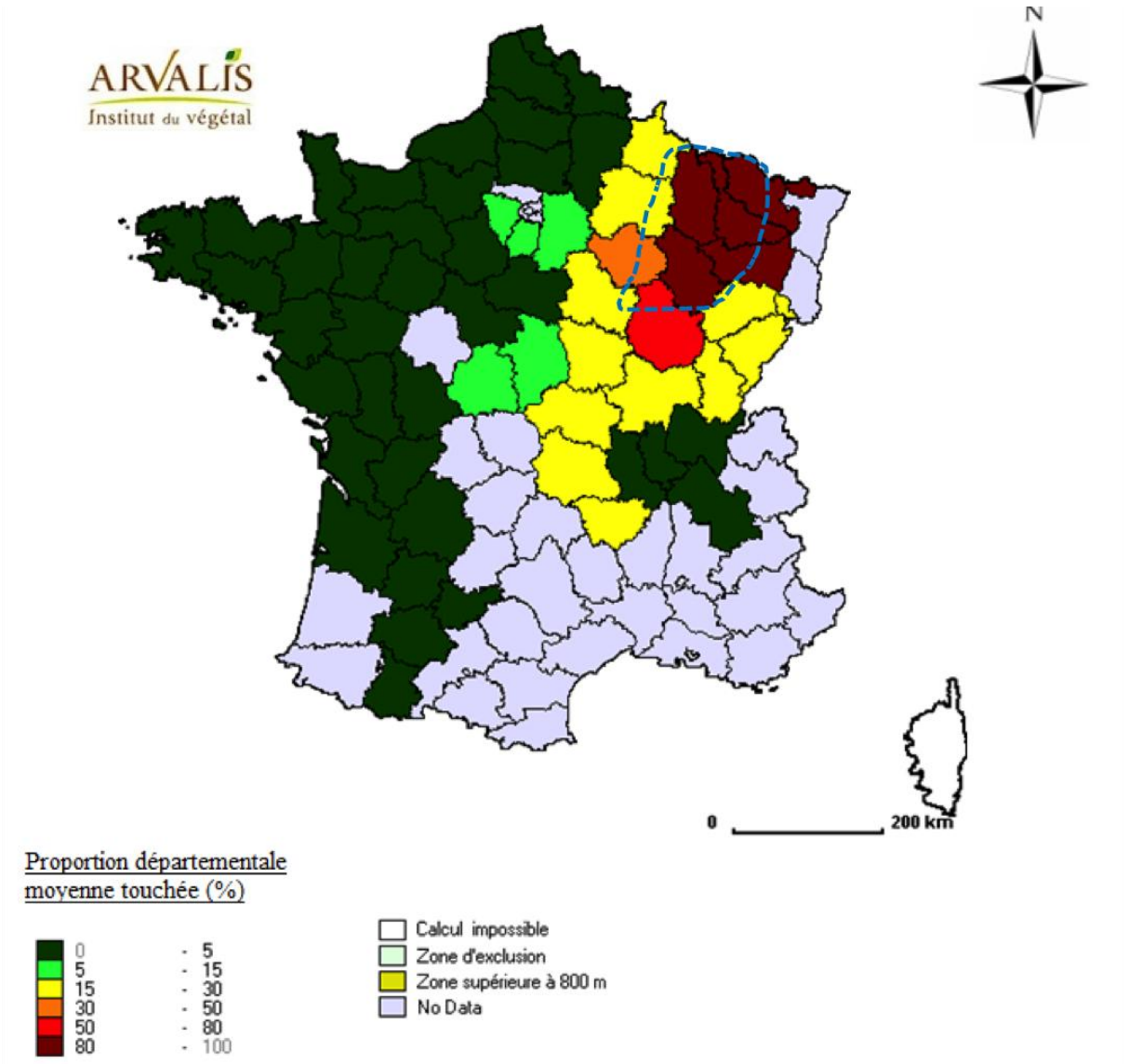


Figure 2 : Proportion départementale des surfaces d'orge d'hiver détruites par le gel (ressemis indispensable) - % (Deswarte, 2012)

(par rapport aux normales de saison : Figure 1). Des vents glacials venus de Sibérie ont soufflé pendant plus de deux semaines sur toute l'Europe (Météo France, 2012) et ont conduit à la destruction de nombreuses cultures. Pour l'orge d'hiver en France, 175 000 ha environ ont été retournés et emblavés de nouveau, soit 16 % des semis (I.E., 2012). Dans certaines régions, les surfaces détruites par le gel représentent plus de 80% de la proportion départementale (Figure 2) (Deswarte, 2012). Les dernières références de froid de cette intensité sont les hivers 1979 et 1985 (L'est Eclair, 2012). Ces évènements étant rares, il est difficile de réaliser une sélection efficace.

De nombreuses études ont été menées pour caractériser les mécanismes de la tolérance au froid. Cette tolérance peut se présenter sous différentes formes :

- L'échappement : si la plante fleurit après les périodes froides alors elle aura résisté au froid ;
- La résistance intrinsèque : la plante a la capacité de résister au froid
- L'endurcissement : la plante adapte progressivement son métabolisme à un environnement froid.

La tolérance au froid est contrôlée majoritairement par 3 QTLs FROST RESISTANCE (FR-H1, FR-H2 et FR-H3). Les deux loci les plus importants, FR-H1 et FR-H2, sont situés sur le bras long du chromosome 5H (Francia, et al., 2004).

FR-H1

Le gène VRN-H1 a un effet pléiotropique : il agit à la fois sur la précocité et la tolérance au froid (Dhillon, et al., 2010). La plante va perdre cette tolérance au froid au cours de sa transition vers son développement reproductif (Fowler & Limin, 2004). Le gène VRN-H1 induit une diminution plus ou moins importante de la tolérance au froid lorsqu'il est exprimé (Dhillon, et al., 2010). Les types hiver et alternatif portent l'allèle vrn-H1 ancestral, cet allèle retarde l'accumulation de transcrits, ce qui prolonge le stade végétatif et donc la tolérance au froid. Alors que les types printemps porte l'allèle Vrn-H1 qui accélère cette accumulation et provoque la floraison. Cette partie de la tolérance est donc liée au phénomène d'échappement. Si la plante fleurit après la vague de froid, elle y aura résisté. Les variations alléliques de VRN-H1 sont dues à une délétion dans le promoteur de celui-ci pour l'allèle printemps (Dhillon, et al., 2010 ; Rapacz, et al., 2010).

FR-H2

C'est à l'automne, en présence de températures modérément basses, pendant les stades de développement végétatif que l'orge est la plus résistante au froid (Hommo, 1994). Cette période correspond à l'acclimatation (ou endurcissement). (Fowler & Limin, 2004 ; Hommo, 1994). Plusieurs études réalisées sur des populations différentes ont montré l'importance du locus FR-H2 (Skinner, et al., 2006 ; Francia, et al., 2004) situé sur le chromosome 5H pour

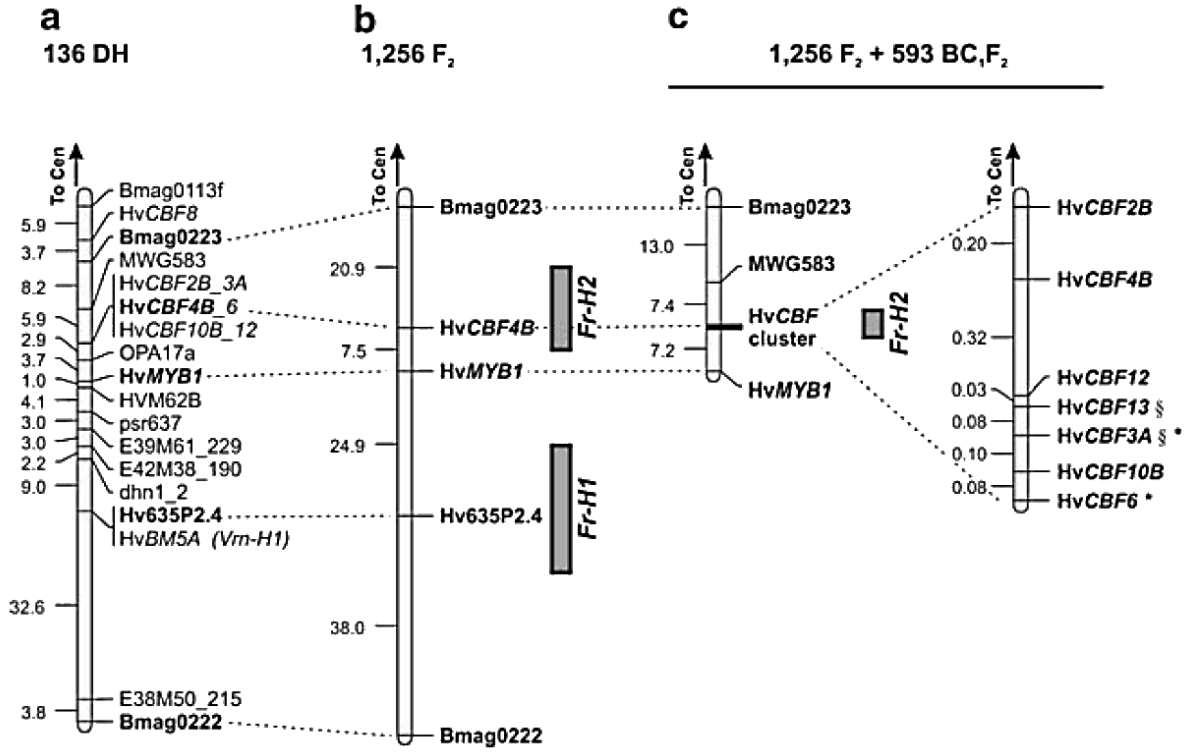


Figure 3 : Cartographie du chromosome de l'orge 5HL utilisant des populations "Nure"x"Tremois" HD et F2. (Francia, et al., 2007)

l'endurcissement. Ce locus, a été identifié comme étant à l'origine de cette tolérance au froid via la famille de gènes *Hordeum vulgare C-repeat binding factors* (HvCBFs) (Visioni, et al., 2013). Les gènes HvCBFs codent pour des facteurs de transcription qui déclenchent l'activation de gènes COR (Cold-Regulated) (Skinner, et al., 2006) qui augmentent la résistance au froid. On dénombre au moins 20 gènes HvCBFs différents répartis sur le génome. Les plus importants sont les gènes HvCBF 2A-B, 3,4A-B, 6, 9, 10A-B, 11, 12, 13 et 14. Les gènes cités sont tous situés au locus FR-H2 (5H). Ils peuvent expliquer jusqu'à 62% de la variation phénotypique liée à la tolérance aux basses températures (Skinner, et al., 2006). Une cartographie fine de 7 d'entre eux est disponible (Francia, et al., 2007) (Figure 3) et des études ont été menées pour comprendre les effets et le fonctionnement de certains de ces gènes. En l'occurrence, il a été montré qu'une induction par le froid active la transcription de ces gènes (Dhillon, et al., 2010) et qu'une surexpression de HvCBF2A accélère l'accumulation de transcrits des gènes COR (Jeknic, et al., 2013). Les variations alléliques, et donc le niveau d'expression de HvCBF2A, sont en partie dues à des variations en « nombre de copies ». Les gènes HvCBF2B et HvCBF4 sont contrôlés de la même manière : les orges de type hiver présentent un plus grand nombre de copies des gènes HvCBF2A-CBF2B et HvCBF4. Les variétés de type printemps ne présentent qu'une seule copie de ces gènes et un pseudogène HvCBF13. Il n'y a pas de différence au niveau de la séquence des transcripts par contre ils sont produits à de plus faibles niveaux chez le type printemps. Il semble que ce soit cela qui marque une partie de la tolérance au froid entre les types hiver et printemps. (Stockinger, et al., 2007 ; Knox, et al., 2010)**.**

FR-H3

Plusieurs études ont été menées sur la recherche d'un nouveau locus de tolérance aux basses températures. Une étude (Fisk, et al., 2012) a porté sur deux populations du Nebraska. L'une de ces populations est sensible à la vernalisation, l'autre ne l'est pas. Cette étude a permis d'identifier un locus situé sur le chromosome 1H. Il a été nommé FR-H3. Il explique 64% de la variation phénotypique dans le cas d'une expérimentation au champ. Dans cette étude, les valeurs de tolérance aux basses températures sont les mêmes dans les deux populations testées. Donc dans ce matériel génétique la tolérance au froid n'est pas couplée à la vernalisation. Pour donner suite à cette étude, une autre étude a été réalisée sur une population représentative du bassin méditerranéen. Dans cette population le locus FR-H3 n'a pas été identifié (Visioni, et al., 2013). Par contre, dans un panel européen d'orge de printemps 2 rangs, un marqueur SNP (12_10506) associé à la tolérance au froid a été identifié par (Tondelli, et al., 2014) et il est situé à 8 cM du marqueur 11_10764 qui a été identifié comme étant le plus proche de FR-H3 par (Fisk, et al., 2012).

On observe donc un locus lié à l'endurcissement, un autre semble intrinsèque à la plante et le dernier est lié à VRN-H1 et donc à un phénomène d'échappement.

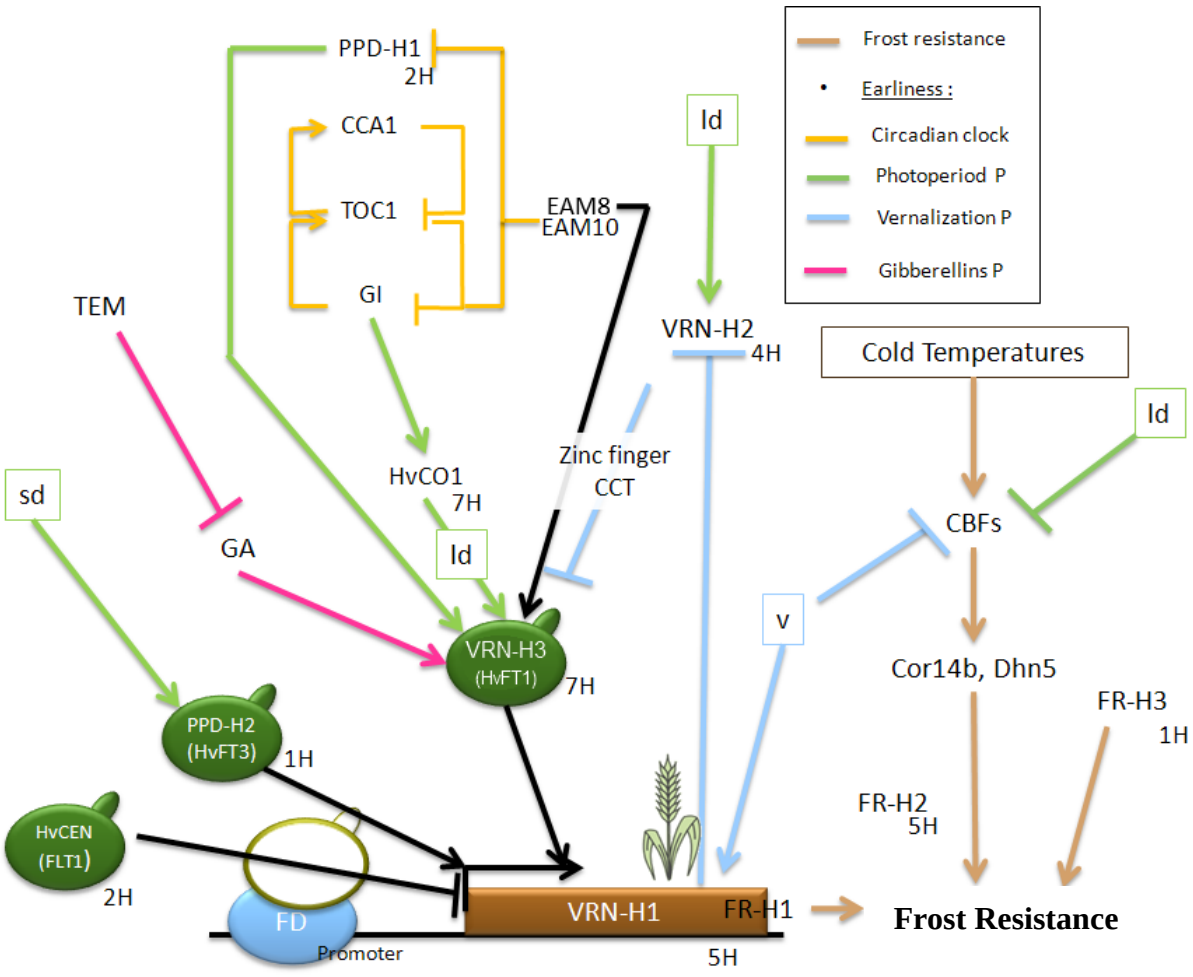


Figure 4 : Ensemble des gènes intervenant directement ou indirectement dans la tolérance au froid. Légende : ld = long day ; sd = short day ; v = vernalisation ; (Figure élaborée avec les travaux cités dans la partie « Présentation de la précocité et de la tolérance au froid »)

Problématique et démarche

La problématique est la suivante : Quelles sont les déterminismes génétiques de précocité et de tolérance au froid au sein du matériel des programmes de sélection orge de Limagrain et peut-on prédire le comportement des lignées?

Pour répondre à cette problématique, des tests d'association seront menés parallèlement à la construction de modèles de sélection génomique sur un panel constitué de toutes les lignées d'orge utilisées en sélection entre 2006 et 2013. La première démarche permettra de détecter les loci importants dans ces mécanismes et d'identifier les quelques marqueurs les plus importants et la seconde donnera les moyens de prédire le potentiel du matériel de manière globale pour un caractère mais avec un nombre important de marqueurs. Cette thématique permet de soulever d'autres questions : La sélection génomique (GS) permet de mieux prédire des caractères complexes (Meuwissen, et al., 2001) mais est-elle aussi efficace pour des locus spécifiques et identifiables ? La présence de ces nombreux marqueurs en GS n'ajoute-elle pas de bruit dans le calcul du modèle ?

Matériels et Méthodes

1- Germplasm

La population étudiée comprend toutes les lignées en cours de sélection développées par l'équipe Orge de Limagrain Europe et ayant atteint le stade des essais rendement entre 2006 et 2013. Cette population est nommée ALLDATA. Après avoir été épurée de ses données aberrantes, ce jeu de données comprend 74193 individus qui au final ont été testés de manière déséquilibrée sur 73 lieux différents en Europe. Il est composé de 34912 lignées printemps, 26136 lignées deux rangs hiver, 12975 lignées six rangs hiver et 169 lignées hiver pour lesquelles le nombre de rangs n'a pas pu être déterminé.

Le germplasm européen est structuré très distinctement selon le type variétal (Comadran, et al., 2012) (Annexe 1). Des sous-populations ont été créées selon ces types variétaux. La méthode Sum of Squares due to Error (SSE) a été utilisée pour estimer le nombre de groupes présents parmi les 7320 individus génotypés disponibles. En connaissant le nombre de groupes présents dans la population, un groupe est attribué à chaque individu en utilisant la méthode de variance minimum de Ward avec 10 boostraps (Ward, 1963).

D'autres sous-populations ont été créées en formant des groupes homogènes à partir de la population ALLDATA. Elles répartissent les individus selon le pays, la latitude ou le climat du lieu d'essais.

Ces répartitions ont conduit à la création de 56 sous-populations (Tableau 2) qui sont les combinaisons des groupes suivants:

- La population ALLDATA précédemment citée ;
- Les populations S, W2, W6 formées respectivement des variétés de type printemps, hiver deux rangs et hiver six rangs. La population WI regroupe les populations W2, W6 et les variétés de types hiver dont le nombre de rangs n'a pas été déterminé ;
- Les populations « pays » sont : la Belgique (BE), la République Tchèque (CZ), l'Allemagne (DE), le Danemark (DK), la France (FR), les Pays-Bas (NL) et le Royaume-Uni (UK).
- Les populations « latitudes » sont LAT1, LAT2 et LAT3, délimitées par les latitudes 50°N et 55°N (LAT1 étant la population la plus septentrional, Annexe 3).
- Les populations « climat » sont Terre et Mer. Elles regroupent les lieux qui respectivement présentent un climat plutôt continental ou plutôt océanique (Annexe 3).

2- Données et outils de phénotypage

Les variables

Les variables retenues pour la précocité sont :

- DOE1 : La date d'épiaison en jour,
- EAR : Une note de 0 à 10 relative à un témoin (0 : précoce, 10 : tardif),
- EARC : Une note correspondant à l'échelle de Zadoks (Zadoks, et al., 2006),
- RIP : Une note de 0 à 10 sur le temps d'arrivée à maturité depuis le semis (0 : précoce, 10 : tardif),

Les variables retenues pour l'impact du froid sont Winter Kill 1 et 2 (respectivement WKI1 et WKI2). Ces deux notes évaluent les dommages causés par le froid à deux dates différentes sur une échelle de 0 à 10 (0 : résistant, 10 : sensible)

Et enfin le rendement (YLD) est choisi comme indicateur de performance.

Le nettoyage des données

Des intervalles seuils ont été déterminés pour le nettoyage des données (Tableau 1) afin de déterminer la présence de données aberrantes. Mis à part certains cas particuliers, les valeurs ne se situant pas dans ces intervalles ont été éliminées. L'observation des données, lieu par lieu, a permis de supprimer la totalité des informations d'un lieu pour une variable si le pourcentage de données aberrantes sur ce lieu était trop élevé.

Tableau 1 : Récapitulatif des seuils minimum et maximum choisis pour la détection de données aberrantes

Variable	Minimum	Maximum
DOE1	60	170
EAR	1	10
EARC	40	69
RIP	1	10
YLD	20	150
WKI1	0	10
WKI2	0	10

Les dispositifs expérimentaux et l'ajustement des données phénotypiques

Les essais utilisés pour réaliser le phénotypage se présentent sous la forme de dispositifs expérimentaux comportant de une à trois répétitions. Les dispositifs non répétés comportent environ 10 à 20% de témoins randomisés au sein de l'essai alors que les essais répétés suivent une randomisation en « alpha plan ». Les unités expérimentales mesurent 6 à 12 m². Les différentes variables sont mesurées à l'échelle de la parcelle. Les données rendement sont mesurées sur toutes les répétitions alors que les autres variables ne sont mesurées que sur une seule répétition car elles présentent généralement une héritabilité forte. Les variables mesurées sont ajustées par lieu pour le rendement (les autres variables n'étant pas répétées) et ensuite au sein du dispositif expérimental grâce aux témoins.

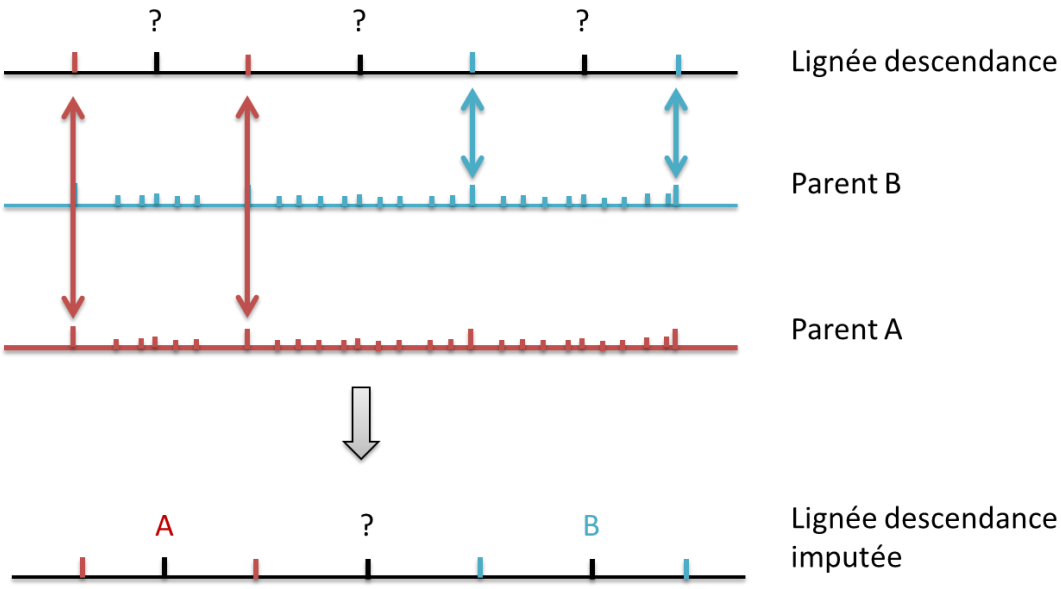


Figure 5 : Schéma explicatif de la méthode d'imputation des données manquantes

Les données étant très déséquilibrées, des moyennes ajustées ont été calculées grâce au modèle Best Linear Unbiased Prediction (BLUP) suivant :

$$Y = \text{Individu} + \text{Année}*\text{Lieu}*\text{Groupe d'essai},$$

l'interaction étant placée en effet aléatoire.

3- Données de génotypage

Pour cette étude 7320 individus génotypés sont disponibles avec un nombre de marqueurs allant de 232 à 11476. Les marqueurs utilisés sont des Single Nucleotide Polymorphism (SNP) et les données ont été générées avec les différentes technologies disponibles depuis 2006 dans le laboratoire de Limagrain. La variation du nombre de marqueurs traduit l'évolution des technologies. De ce fait tous les individus génotypés disponibles ne présentent par la même densité de marquage. Les données brutes présentent au total 46.44% de données manquantes. Des marqueurs placés sur les gènes d'intérêt sont disponibles mais ils n'ont pas été utilisés pour génotyper un grand nombre d'individus.

4- Cartographie des marqueurs

La première carte génétique utilisée pour effectuer les tests de cette étude est une carte de 7477 marqueurs SNP construite à partir de la carte de Comadran, et al. (2012).

Une première version de la carte physique ayant été disponible au cours de ce stage, une nouvelle carte consensus a été réalisée. Celle-ci se base sur une carte génétique Morex x Barke (Comadran, et al., 2012) sur laquelle ont été projetées des données d'autres cartes génétiques et physique grâce au logiciel Biomercator (Sosnowski, et al., 2012). Les inversions et erreurs de positionnement de marqueurs sont visualisés grâce au logiciel Strudel (Bayer, et al., 2009) et supprimés.

5- Imputation des données de marquage

En utilisant cette nouvelle carte et les informations de génotypage des parents disponibles, les données manquantes de lignées dont les parents avaient été densément génotypés ont été imputées. Si un marqueur n'est pas génotypé pour une lignée descendante mais a deux marqueurs flanquants portant l'allèle du parent A alors ce marqueur sera imputé avec l'allèle du parent A (Figure 5).

Pour juger de la qualité d'imputation des données de marquage moléculaire huit génotypes qui présentent peu de données manquantes sont choisis. Pour ces génotypes un certain

pourcentage de marqueurs, choisis aléatoirement, sont transformés en données manquantes (10%, 50% et 90%).

L'imputation est réitérée avec, à la place des 8 génotypes du départ, les génotypes tests nouvellement créés. La distance génétique entre génotype réel et génotype test est mesurée grâce à la distance de Rogers.

Suite à l'imputation, les individus présentant plus de 20% de données manquantes ont été éliminés (Call rate ou CR de 80%) (Figure 10b). La correspondance faite entre les données de génotypage et de phénotypage permet de répertorier le nombre d'individus dans chaque population étudiée (Tableau 2). Ce qui représente entre 8600 et 10800 marqueurs.

Tableau 2 : Récapitulatif de la taille des jeux de données après la réalisation de la correspondance entre les sous-populations et les données de marquage moléculaire imputées (CR de 80%) . Les populations en rouge ne sont pas analysées car elles comptent moins de 100 individus.

		ALLDATA	S	W	W2	W6
NOMBRES D'INDIVIDUS		3411	2817	583	297	296
	BE	291	133	164	1	161
	CZ	71	549	27	_	27
	DE	1281	1508	387	201	193
	DK	41	41	_	_	_
	FR	1496	1548	309	195	114
	LAT1	769	770	_	_	_
	LAT2	2007	2814	583	297	296
	LAT3	1575	1563	462	203	266
	NL	1075	806	254	31	230
	UK	2486	2189	301	281	23
	Terre	1383	1304	_	30	49
	Mer	1385	997	_	7	381

6- Méthode d'analyses d'associations

Les tests d'associations sont réalisés grâce au modèle EMMAX (Efficient Mixed-Model Association eXpedited) (Kang, et al., 2010) en éliminant les allèles rares (présents dans moins de 10 génotypes). Ce modèle se base sur le modèle EMMA (Kang, et al., 2008) pour lequel chaque marqueur explique une faible partie d'une variable complexe. Ceci permet d'éviter des calculs de variance qui seraient redondants et ainsi de diminuer le temps de calcul.

Dans un premier temps, les tests d'association sont donc réalisés sur les individus avec un CR supérieur à 80% et à 60% avec les données de marquage moléculaire non imputées (Figure 10a) et uniquement sur les populations comptant au moins 100 individus par souci de significativité des résultats.

45 populations remplissaient les deux conditions précédentes pour 7 variables.

Un marqueur est considéré comme associé quand le logPvalue > 3. Ce faible seuil est compensé par la superposition d'informations : Un locus ne sera validé que si plusieurs marqueurs associés co-localisent dans une population et/ou dans plusieurs populations. Un locus est considéré intéressant s'il explique au moins 10% de la variation phénotypique.

Après imputation des données moléculaires de nouveaux tests d'associations sont réalisés sur les individus avec un CR minimum de 80 % en EMMAX sur les populations présentées (Tableau 2).

Les différentiations alléliques (froid et précocité) entre lieux concernant le rendement, qui est l'objectif premier d'un programme de sélection, ont également été étudiées. Pour cela, les loci qui ont le plus d'effet pour la résistance au froid et pour la précocité sont sélectionnés et les fréquences alléliques sont calculées pour deux groupes de lignées : les 15% ayant les rendements les plus élevés (rdt+) et les 15% ayant les rendements les plus faibles (rdt-). Ces calculs sont effectués pour les lieux RORO, RSAD, VRVR qui sont des lieux distincts au niveau environnemental (Annexe 3). Les individus de types W2 sont utilisés pour étudier ce point.

7- Méthode de sélection génomique

L'objectif de la sélection génomique est de prédire la valeur génétique d'un individu grâce à l'ensemble des marqueurs positionnés sur le génome. Cette méthode attribue un effet à chaque marqueur utilisé contrairement aux analyses d'association qui ne permettent d'identifier qu'un petit nombre de QTLs et de ne considérer que l'effet de ces QTLs. Cette méthode nécessite donc un très grand nombre de marqueurs pour témoigner au mieux de l'effet d'une zone du génome. La sélection génomique comporte plusieurs étapes. Tout d'abord une population d'entrainement est utilisée pour créer un modèle dans lequel les effets des marqueurs sont calculés grâce aux notations phénotypiques des individus. Ensuite, ce modèle est utilisé pour calculer la valeur génétique de nouveaux individus qui ont été génotypés uniquement. (Robert-granié, et al., 2011). Il existe de nombreux modèles de calculs dont les plus fréquemment utilisés sont la méthode RR-BLUP qui considère que la variance des effets des marqueurs est constante et la méthode bayésienne qui au contraire considère que la variance des effets des marqueurs n'est pas constante (Gallais, 2011).

Une étude récente (Heslot, et al., 2012) montre une similarité de précision entre tous les modèles de sélection génomique disponibles actuellement. La méthode utilisée dans cette étude sera la Ridge Regression Blup (RR-Blup), suivant ses caractéristiques de rapidité, d'efficacité de calcul et de fiabilité sur un large éventail de critères phénotypiques et de jeux de données.

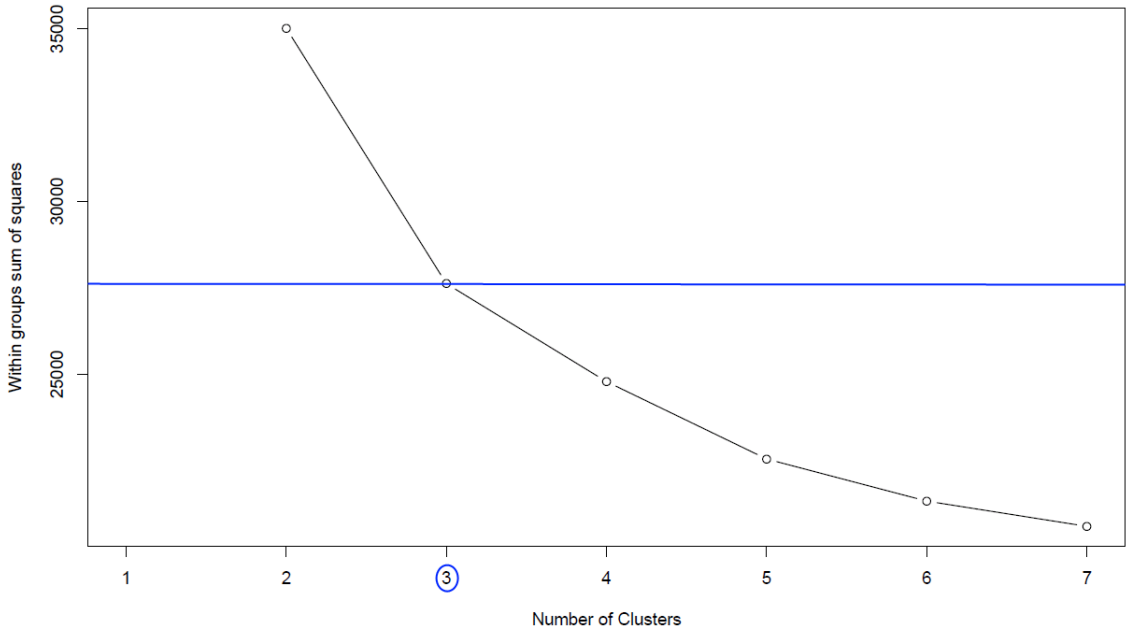


Figure 6 : Évaluation de regroupement par la méthode SSE. La rupture de pente matérialise le nombre de groupes.

Tableau 3 : Comparaison de l'héritabilité des variables de précocité suivant les types d'orges et les populations "type x climat"

	printemps			hiver		
Variable	alldata	mer	terre	alldata	mer	terre
XDOE1.PERSE.-.all	0.7	***0.58***	***0.62***	0.79	***0.59***	
XEAR.PERSE.-.all	0.78	0.81	***0.58***	0.79	0.8	0.75
XEARC.PERSE.-.all	0.84	0.83		0.71	0.69	
XDOE1.PERSE.+.all	0.61	***0.25***	0.67	0.74	0.69	
XEAR.PERSE.+.all	0.7	0.68	***0.62***	0.79	0.78	0.72
XEARC.PERSE.+.all	0.67	0.67	0.78	0.81	0.81	
	W2			W6		
	alldata	mer	terre	alldata	mer	terre
XDOE1.PERSE.-.all	0.8	0.77		0.82	***0.31***	
XEAR.PERSE.-.all	0.8	0.81	0.76	0.75	0.76	0.77
XEARC.PERSE.-.all	0.79	0.79		0.57	***0.52***	
XDOE1.PERSE.+.all	0.8	0.83	0.88	0.71	***0.44***	0.77
XEAR.PERSE.+.all	0.81	0.78	0.75	0.69	***0.61***	0.7
XEARC.PERSE.+.all				0.81	0.81	

Résultats

1- Germplasm

La méthode SSE, pour estimer le nombre de groupes présents sur les 7320 individus génotypés disponibles, montre une structuration en trois groupes (Figure 6). Chaque individu se voit attribuer un groupe en utilisant la méthode de variance minimum de Ward avec 10 boostraps (Ward, 1963). Cette répartition correspond bien aux trois types variétaux principaux de l'orge (Tableau 4) (Annexe 2).

Tableau 4: Répartition des individus des trois groupes selon les types variétaux printemps : S, hiver deux rangs : W2, hiver six rangs : W6. Les variétés hiver dont le nombre de rangs n'a pas été identifié sont placées dans le groupe WI.

Count of INDIVIDUAL	Group			
Growth Habit	**1**	**2**	**3**	**Total**
S	5154	37	2	5193
W2	9	1034	5	1048
W6	2	24	937	963
WI	1	60	55	116
Total	**5166**	**1155**	**999**	**7320**

L'héritabilité des caractères suivant la dichotomie « climat » montre une variabilité pour les caractères mesurant la précocité (Tableau 3), ce qui n'est pas normal car ces traits sont très héritables. De plus cette variabilité n'apparait pas selon la séparation « latitudes » (données non présentées).

Donc sur les 45 populations envisagées, celles qui répartissent les individus selon le caractère maritime ou terrestre du lieu d'essai ne sont pas retenues.

2- Données et outils de phénotypage

En observant les données phénotypage brutes, des tendances dans la répartition des individus apparaissent.

Rendement

- Le traitement n'a pas d'effet évident sur les variables mis à part sur le rendement par lieu et par latitude.
- Les lignées hiver ont un plus fort rendement que les lignées printemps.
- En comparant les pays, des tendances au niveau des performances existent :

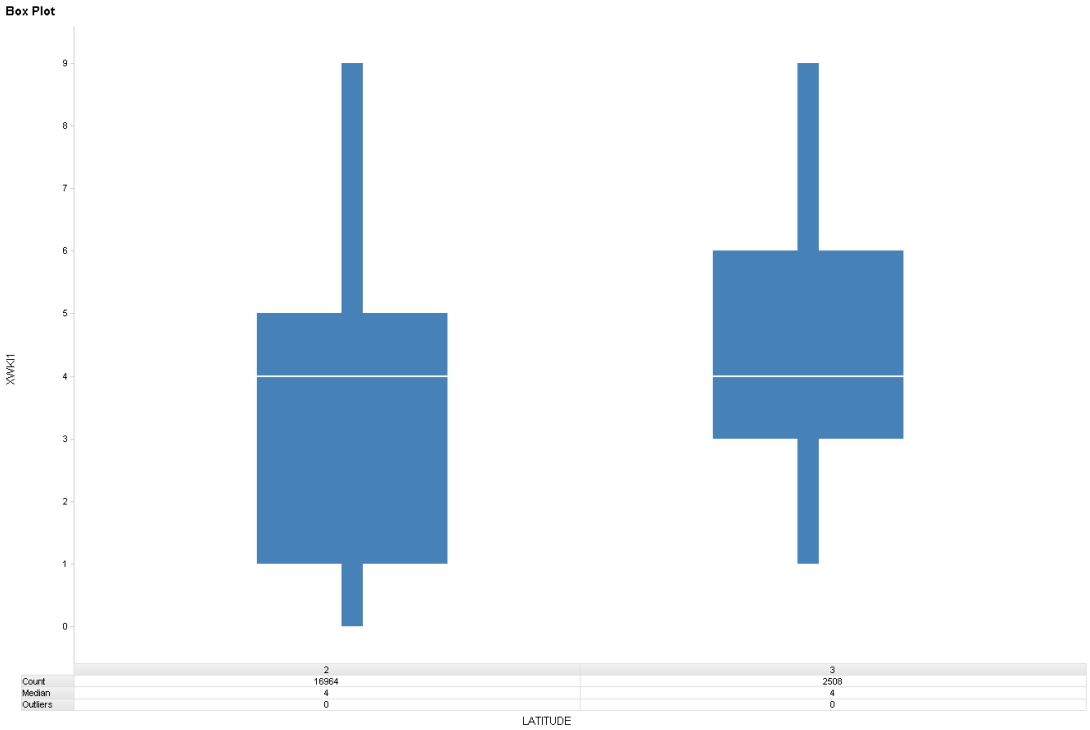


Figure 7 : Répartition des individus selon leur note WKI1 (0 : résistant, 10 sensible) en fonction de la latitude (la latitude 2 est la latitude intermédiaire ; la latitude 3, la latitude méridionale).

Tableau 5 : Résumé des moyennes des héritabilités des variables étudiées

Variables	Moyenne de H^2	Ecart-type
DOE1-	0.760	0.110
DOE1+	0.730	0.095
EAR-	0.758	0.115
EAR+	0.724	0.064
EARC-	0.751	0.110
EARC+	0.743	0.089
RIP-	0.584	0.130
RIP+	0.468	0.097
YLD-	0.504	0.149
YLD+	0.451	0.157
WKI1-	0.549	0.235
WKI1+	0.465	0.245
WKI2-	0.608	0.256
WKI2+	0.561	0.063

- Globalement, le Danemark et la République Tchèque présentent dans des médianes basses de rendement et la Belgique des médianes hautes (excepté pour le printemps) (Tableau 6).

Tableau 6 : Résumé d'observations de BoxPlots "rendement VS pays" par type

Type	Classement des pays en fonction de leur performance
Spring	DK < CZ = DE = FR < BE = UK < NL
Winter	CZ < FR = DE = NL = UK < BE
W2	FR < DE = NL = UK ≤ BE
W6	CZ < UK < DE = NL = FR ≤ BE

Précocité

- Les lignées hiver sont plus précoces que les lignées printemps.
- Les populations LAT3 semblent présenter des lignées plus précoces que les populations LAT2 suivant la variable DOE1.

Résistance au froid

- Il n'y a pas d'effet évident de différence de résistance en les types deux et les types six rangs pour les scores WKI1 et WKI2.
- Il semble aussi y avoir une plus grande résistance au froid en LAT2 pour le score WKI1 (Figure 7) qui traduit la capacité de la plante à se maintenir lors d'un épisode de froid intense.

Corrélation entre les variables

Il existe une forte corrélation entre les variables qui notent la précocité. DOE1, EAR et EARC présentent des coefficients de corrélation supérieure à 90% en valeur absolue.

Par contre, la corrélation entre WKI1 et WKI2 semble plus dépendante de l'environnement. En effet, si la corrélation de ces variables en essais non traité est de 88 %, elle chute à 53% pour les essais traités.

Héritabilité

Les variables exprimant la précocité (DOE1, EAR, EARC) ont une forte héritabilité, alors que les autres variables étudiées (RIP, YLD, WKI1, WKI2) ont une héritabilité moyenne (Tableau 5).

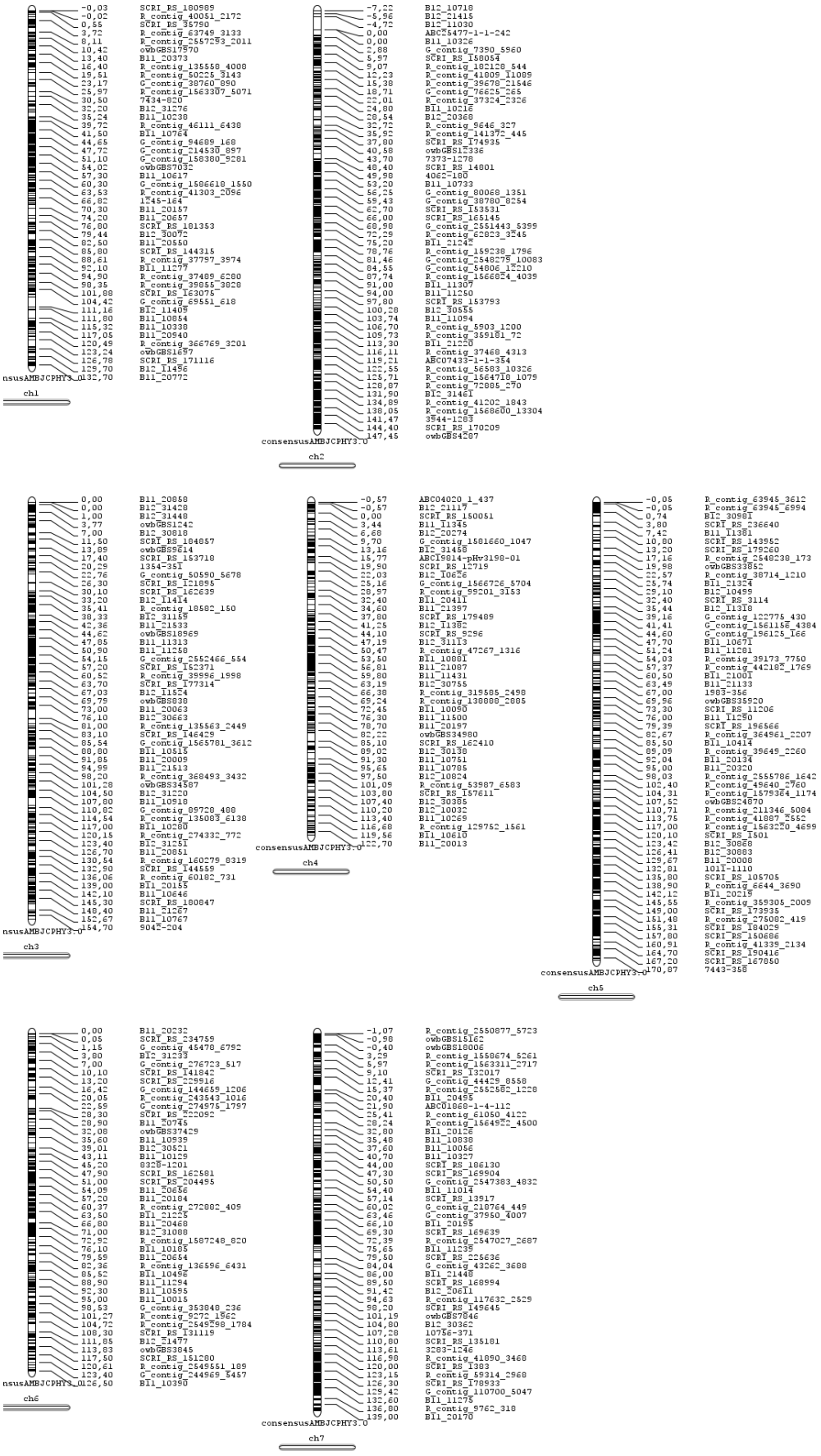


Figure 8 : Carte consensus réalisée sur la base d'une carte Morex*Barke ainsi qu'une carte génétique et une carte physique

3- Cartographie des marqueurs

La nouvelle carte se compose de 13618 marqueurs SNP (Figure 8).

Elle présente une densité moyenne de 0.185cM. En termes de distance génétique, 27 % de cette carte présente une densité supérieure à un marqueur tous les 0.2 centiMorgan. 75% de la carte est couverte au minimum avec une densité d'un marqueur tous les 0.5 cM. Le plus grand espace sans marqueur est de 4.70 cM et se situe au début du chromosome 2 (Figure 9).

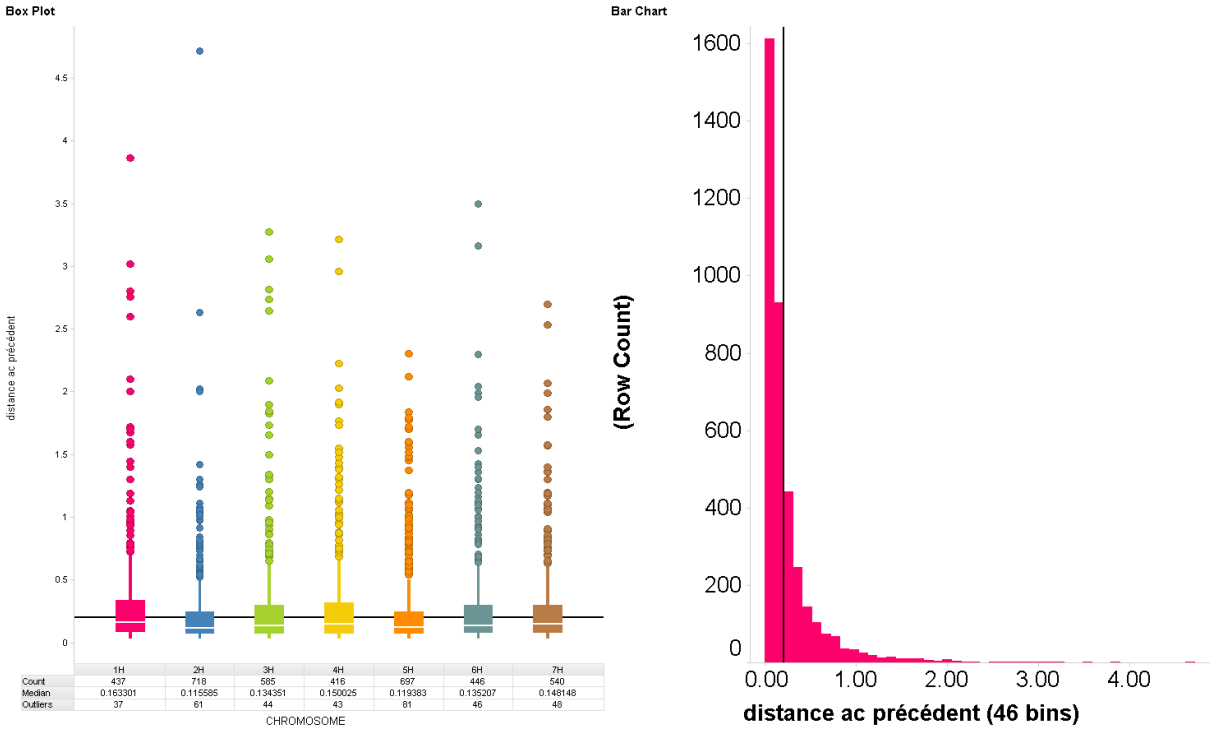


Figure 9 : Répartition des espaces entre marqueurs par chromosome (à gauche) et sur le génome entier (à droite)

Il est désormais possible de faire le lien entre carte physique et carte génétique (Annexe 4). Ce lien permet d'identifier pour chaque chromosome la place du centromère mais également des erreurs de positionnement de marqueurs.

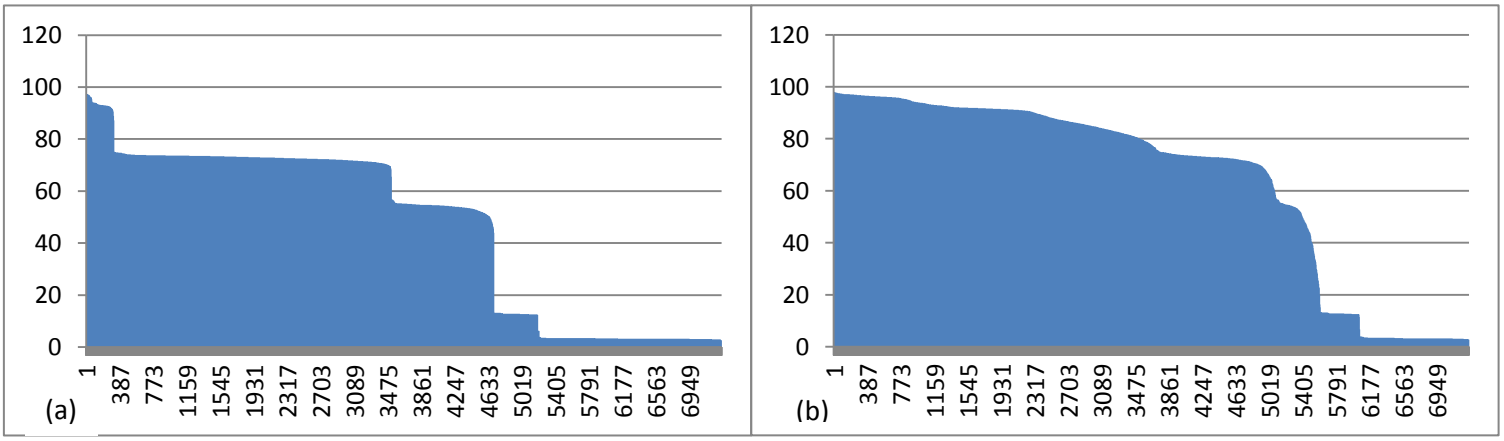


Figure 10 : Distribution du Call Rate par individu avant (a) et après (b) imputation

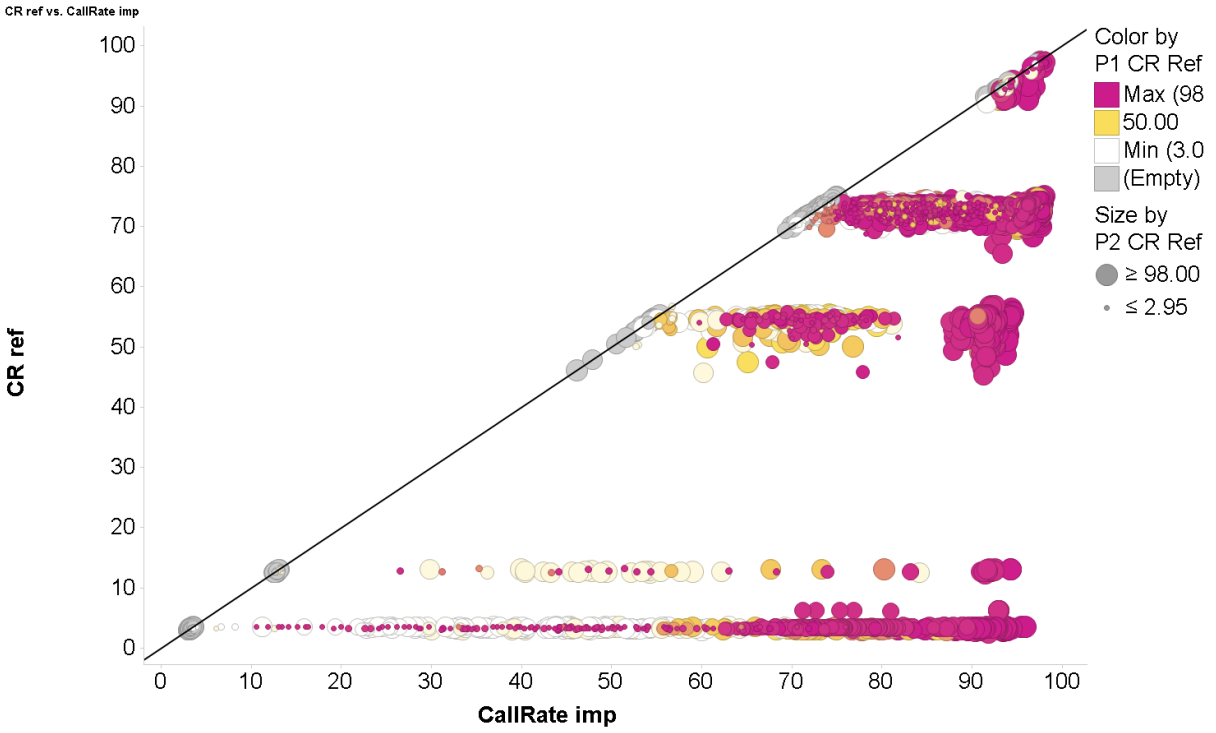


Figure 11 : Visualisation du gain de données moléculaires en fonction du Call Rate des parents (pourcentage de données de marquages). Le Call Rate de référence (sans imputation) du parent 1 est matérialisé par un gradient de couleur allant du blanc (CR minimum de 3%) au rose (CR maximum de 98%) en passant par le jaune (CR de 50%). Le Call Rate de référence du parent 2 est matérialisé par un gradient de taille allant de 2.95% de CR pour les cercles les plus petits à 98% de CR pour les cercles les plus grands.

4- Imputation des données de marquage

Qualité d'imputation des données manquantes

L'ensemble des données manquantes n'étant pas imputées, l'efficacité de l'imputation est mesurée sur le pourcentage de données manquantes imputées, ce qu'on appelle le gain. La qualité d'imputation des données manquantes est mesurée par la distance génétique entre le génotype témoin et celui imputé. Cette distance augmente en fonction du nombre de données manquantes appliquées. Si l'on considère qu'au-dessus d'une distance de Rogers (RD) supérieure à 0.02, l'identité entre deux génotypes n'est plus confirmée alors au-delà de 40% de gain, la qualité de l'imputation n'est plus optimale (Tableau 7).

$$\text{Distance de Rogers} = \sqrt{\frac{\sum_i (p_{x,i} - p_{y,i})^2}{2}},$$

où i est un locus, x et y sont les populations que l'on souhaite caractériser (Rogers, 1972).

Résultat sur l'efficacité de l'imputation

L'imputation des données permet de réduire de 17.6% le nombre de données manquantes pour l'ensemble des individus génotypés (Figure 10). Les génotypes imputés dont au moins un parent est genotypé avec peu de marqueurs montrent un taux d'imputation plus faible que les génotypes dont les deux parents sont génotypés avec de nombreux marqueurs (Figure 11).

Tableau 7 : Résumé du test de la qualité d'imputation en fonction du pourcentage de données manquantes. "-" pourcentage faible de données manquantes, "+" pourcentage élevé de données manquantes.

	Parent1	Parent2	**Gain (**10% DM)	RD (10%DM)	**Gain (**50% DM)	RD (50% DM)	**Gain (**90%DM)	RD (90 % DM)
CHRONICLE	-	-	6.2740	0.0002	22.4817	0.0041	3.9299	0.0103
CONCERTO	-	-	6.2304	0.0029	42.6891	0.0189	79.4179	0.0435
LN1147	-	+	4.4963	0.0004	26.6643	0.0071	44.2924	0.0364
CAVALIER	-	+	4.1129	0.0006	23.7888	0.0047	35.6396	0.0073
QUENCH	+	+	3.6860	0.0009	13.6807	0.0082	1.2461	0.0000
TESLA	+	+	3.3461	0.0006	14.5695	0.0111	2.5096	0.0261
WICKET	+	-	0.7930	0.0002	6.0997	0.0038	_	_
SEBASTIAN	+	-	0.4096	0.0000	3.1457	0.0017	5.6814	0.0082

	doe1	ear	earc	rip	wki1	wki2	TOTAL
alldata.be.imp				4	2		6
alldata.be.impHYB					1		1
alldata.de.imp					2	1	3
alldata.fr.imp					2		2
alldata.imp		4	4			2	10
alldata.lat2.imp						3	3
alldata.lat3.imp				1			1
alldata.nl.imp					1	2	3
alldata.uk.imp		1					1
s.cz.imp		1					1
s.de.imp	1						1
s.imp			1				1
s.lat2.imp	1	1	1				3
s.lat3.imp			1				1
s.nl.imp		1	1				2
s.uk.imp	1						1
w.be.imp				5	4		9
w.de.imp	1	1		2	3	1	8
w.fr.imp	1			1	2		4
w.imp	2	2		1	1	2	8
w.lat2.imp	2	1		1	1	2	7
w.lat3.imp	3	1		1		1	6
w.nl.imp		1	2	1		1	5
w.uk.imp		3					3
w2.de.imp		3		4	4	2	13
w2.fr.imp	4	3		4	1		12
w2.imp	4	3		4	3	4	18
w2.lat2.imp	2	3		3	5	3	16
w2.lat3.imp	3	3		4		1	11
w2.uk.imp		2					2
w6.be.imp				6	4		10
w6.de.imp	1	1		1	1		4
w6.fr.imp	1	1		2			4
w6.imp	1	1	2	4		1	9
w6.lat2.imp	2	1	2	1		1	7
w6.lat3.imp		2		1			3
w6.nl.imp		1	2	3			6
TOTAL	30	41	16	54	37	27	

Figure 12 : Récapitulatif du nombre de QTLs en fonction des variables

5- Résultats d'analyses d'associations

Dans ces résultats le terme « QTL» est utilisé comme un terme générique pour désigner les loci associés aux caractères étudiés.

Impact de l'imputation

De manière générale, les associations réalisées à partir de données imputées présentent moins de marqueurs associés aux caractères étudiés. Mais si le nombre global de marqueurs associés diminue, certains marqueurs sont associés de manière plus significative.

Les études menées à partir de données imputées montrant des associations plus significatives, donc présentant moins de chance d'être des faux positifs, seules ces études seront traitées dans la suite de ce rapport.

183 QTLs ont été détectés sur 85 zones différentes (Figure 12) (Annexe 5). Parmi ces 183 QTLs, seulement 9 sont issus de populations comportant des lignées de printemps.

Résultats des études d'association concernant la précocité

66 QTLs (r^2>0.1) sont liés à la précocité de floraison. Ils sont répartis sur 31 zones dans le génome.

55 QTls (r^2>0.1) sont liés à la précocité de maturité. Ils sont répartis sur 22 zones dans le génome.

Sur le chromosome 2H, PPD-H1 (22.2cM) et HvCEN (57cM) sont identifiés très nettement dans de nombreuses populations pour les caractères de précocité et de maturité mais uniquement chez des variétés de type hiver. Plus précisément, HvCEN n'est identifié que dans des populations de variétés deux rangs.

Le gène VRN-H2, situé sur le chromosome 4H colocalise avec 6 QTLs liés à la maturité. Le gène EAM8, situé sur le chromosome 1H, colocalise avec un QTL également lié à la maturité dans la population d'orges d'hiver 6 rangs.

Résultat des études d'association concernant le froid

62 QTLs (r^2>0.1) liés à la résistance au froid ont été détectés. Ils sont répartis sur 32 zones dans le génome.

Le gène FR-H3 (Fisk, et al., 2012), situé sur le chromosome 1H, colocalise avec deux QTLs, dans les populations hiver et alldata en essai aux Pays-Bas.

Sur le chromosome 5H, 6 QTLs sont situés aux alentours du locus FR-H2. Sur le même chromosome 5H, un QTL de la population d'orges six rangs en latitude 2 colocalise avec le QTL de résistance de froid FR-H1/VRN-H1.

Il existe aussi des différences entre variétés deux rangs et six rangs. Il y a beaucoup moins de zones détectées pour les orges des populations ne comportant que des cultivars six rangs.

Tableau 8 : Récapitulatif des r^2 Cross Validés (r^2 CV), des héritabilités des modèles de moyennes ajustées (h^2) et du nombre d'individus en fonction du type variétal

types	printemps			W6			W2		
variable	r^2CV	h^2	nb indiv	r^2CV	h^2	nb indiv	r^2CV	h^2	nb indiv
XDOE1.PERSE.-.all	0.3654	0.7	229	0.1098	c	60	0.0875	0.8	105
XDOE1.PERSE.+.all	0.2903	0.61	1068	0.2389	0.71	139	0.1066	0.8	197
XEAR.PERSE.-.all	0.3431	0.78	1151	0.2385	0.75	155	0.1659	0.8	132
XEAR.PERSE.+.all	0.3983	0.7	2608	0.1985	0.69	282	0.3156	0.81	291
XEARC.PERSE.-.all	0.4051	0.84	883	0.1596	0.57	283	0.1871	0.79	30
XEARC.PERSE.+.all	0.3076	0.67	452	0.2131	0.81	89	_	_	_
XRIP.PERSE.-.all	0.3041	0.6	1182	0.1303	0.5	284	0.0357	0.69	136
XRIP.PERSE.+.all	0.2168	0.45	1075	0.0732	0.37	282	0.2785	0.54	196
XWKI1.PERSE.-.all	_	_	_	0.2636	0.41	71	0.0121	0.44	76
XWKI1.PERSE.+.all	_	_	_	0.1638	0.41	269	0.0774	0.35	174
XWKI2.PERSE.-.all	_	_	_	0.0123	0.77	245	0.0005	0.58	41
XWKI2.PERSE.+.all	_	_	_	_	0.51	_	0.0687	0.56	103
XYLD.PERSE.-.all	0.2336	0.44	1669	0.2178	0.54	286	0.0699	0.46	212
XYLD.PERSE.+.all	0.3649	0.54	2809	0.1881	0.48	291	0.4177	0.59	296

Différentiations alléliques

Les loci choisis pour tester la différentiation entre lieux sont HvCEN, PPD-H1, Ag_Preco_5H_51, VRN_H1, Ag_Preco_4H_97 et Ag_Preco_5H_63 pour la précocité et Ab_Fro_1H_59, Ab_Fro_5H_100 (FR-H2), Ab_Fro_5H_168, Ab_Fro_6H_30 pour la résistance au froid. Ces loci sont ceux qui ont le plus d'effets sur la précocité et la résistance au froid. Les loci EAM8 pour la précocité et FR-H1 et FR-H3 pour la résistance au froid sont également suivis mais n'apparaissent pas dans les QTLs des populations W2.

EAM8 n'est pas discriminant car il est complétement fixé pour les W2 testés ici, qui portent l'allèle précoce.

Pour les autres loci, des comportements différents apparaissent ou non entre lieux (Tableau 9).

Tableau 9 : Présentation des majorités alléliques des loci FR-H3, Ab_Fro_5H_168, FR-H2 et PPD-H1 selon rdt+ et rdt-. S : sensible au froid, R : Résistant au froid, T : tardif, P : précoce, = : proportions égales allèle précoce et tardif. + : l'allèle majoritaire est présente à plus de 75 %. En vert : les loci ne montrant pas de différences entre lieux. En jaune : Les lieux montrant des différences entre lieux.

	Fro_1H_59			Fro_6H_30			FR-H1		
	RORO	RSAD	VRVR	RORO	RSAD	VRVR	RORO	RSAD	VRVR
Rdt+	R+	R+	R+	R+	R+	R+	S+	S+	S+
Rdt-	R	R	R	R	R	R	S+	S+	S+
	FR-H3			Ab_Fro_5H_168			FR-H2		
	RORO	RSAD	VRVR	RORO	RSAD	VRVR	RORO	RSAD	VRVR
Rdt+	S	S	S	S	S	S	R	R	R
Rdt-	R	R	S	S	S	R	R	R	S

	HvCEN			Preco_4H_97			VRN-H1		
	RORO	RSAD	VRVR	RORO	RSAD	VRVR	RORO	RSAD	VRVR
Rdt+	P+	P+	P+	P+	P+	P+	T	T	T
Rdt-	P+	P+	P+	P	P	P	T	T	T
	PPD-H1			Preco_5H_51			Preco_5H_63		
	RORO	RSAD	VRVR	RORO	RSAD	VRVR	RORO	RSAD	VRVR
Rdt+	=	=	P	P+	P+	P+	P+	P+	P+
Rdt-	T	=	=	P+	P+	P+	P+	P+	P+

Les fréquences alléliques des loci présentant des comportements différents sont présentées en annexe (Annexe 6).

6- Résultats de sélection génomique

Pour la sélection génomique, le nombre d'individus a une forte importance sur le calcul du modèle. Les résultats montrent qu'avec un faible nombre d'individu le modèle est mal établi. Ainsi les modèles sont peu prédictifs pour les populations W6 et W2 qui comportent moins de 1000 génotypes et semblent plus solides pour la population printemps (Tableau 8).

Les r^2 Cross Validés (r^2CV) des populations hiver pour la précocité sont en moyenne de 0.18 (160 individus en moyenne) alors que les populations printemps les r^2CV sont en moyenne de 0.35 (1065 individus en moyenne).

Discussion

La population mère ALLDATA comprend des lignées en cours de sélection. C'est un jeu de données qui n'a pas été créé pour des études d'association. Il est donc très déséquilibré et comporte de nombreuses données manquantes. Néanmoins les lignées qui forment ce jeu de données ont atteint le stade des essais rendement. C'est donc un matériel quasiment fixé qui présente une relative diversité génétique travaillée par les sélectionneurs. Les QTLs identifiés pourront être directement utilisés pour améliorer le matériel de sélection. Cette approche permet également de travailler des données de phénotypage disponibles dans les bases de données. De plus le modèle utilisé pour calculer les moyennes ajustées essaie de prendre en compte un maximum d'effets en considération (année, lieu, groupe d'essai).

Les sous-populations concernant le climat n'ont pas été retenues car l'héritabilité calculée n'était pas satisfaisante. Le modèle n'a pas réussi à prendre en compte correctement l'effet environnement. Pourtant cette dichotomie pourrait faire apparaitre des adaptations au climat. Il faudrait pouvoir utiliser des données environnementales plus détaillées en cofacteurs (pluviométrie, température, …) pour créer un modèle plus efficace.

L'observation des données brutes révèle des évidences (Les lignées hiver sont plus précoces et ont un rendement plus élevé que les lignées printemps, le traitement n'a d'effet que sur le rendement, etc.) qui montrent la cohérence des notations au fil des années. Le fait que les lignées testées en LAT3 soient plus précoces qu'en LAT2 est également un résultat cohérent avec le fait que l'accumulation de degré-jour nécessaire à la floraison de la plante soit plus rapide dans le sud et/ou avec le fait que les lignées qui sont testées sont choisies pour être précoces et échapper à d'éventuelles périodes de sécheresse et de chaleur en début d'été. De la même manière, les lignées testées sur les lieux LAT2 sont plus tolérantes au froid que les lignées testées en LAT3, car ces lieux LAT2 sont soumis à des stress de froid plus importants.

L'étude utilise trois variables notant la précocité de la floraison. Il serait intéressant d'harmoniser les notations et de n'en utiliser qu'une seule mais il est difficile de déterminer si une note est meilleur indicateur qu'une autre. Elles présentent chacune des avantages et des inconvénients. En effet, DOE1 nécessite de nombreux passages dans les essais pour déterminer quelle parcelle a épié mais c'est une note très simple à attribuer et assez objective. EAR est la note la plus subjective et peut contenir des effets notateurs assez importants mais elle ne nécessite qu'un seul passage dans les essais. Enfin, EARC est une note basée sur les stades physiologiques de la plante, c'est donc une note plutôt objective et qui ne nécessite qu'un seul passage mais elle nécessite un niveau d'expertise plus important des stades physiologiques de la plante. De plus ces trois variables présentent la même héritabilité, ce critère n'est pas non plus une clé de détermination.

La notation de la maturité est plus difficile, ce qui explique que l'héritabilité du caractère soit moyenne. De plus de nombreux loci intervenant dans la précocité de floraison apparaissent

dans les tests d'associations de la variable RIP. Il est intéressant de se demander si c'est bien la précocité de la maturité qui transparait ici ou plutôt la précocité de la floraison. Pour éviter cet éventuel problème des essais pourraient être mis en place où une nouvelle variable RIP' serait la différence entre la date de maturité et la date de floraison par exemple.

Enfin pour la caractérisation du froid, l'héritabilité est également moyenne. Après une période de froid importante qui laisse apparaitre une réaction différente entre les génotypes d'un essai le sélectionneur décide de noter la résistance des individus (WKI1) puis ensuite la reprise de végétation (WKI2). Il est difficile de caractériser le froid qui est appliqué aux essais, en effet d'une année sur l'autre celui-ci peut être très différent. Les caractères qui notent le froid sont donc peu héritables. Il pourrait ici aussi être intéressant de disposer de plus de données environnementales telles que la température ou le vent pour qu'ils soient pris en compte dans les calculs de moyennes ajustées. De plus ces événements sont plutôt rares et la quantité de données disponibles est encore assez limitée.

L'imputation des données a permis de réduire de 17.6% le nombre de données manquantes. Cette méthode est donc efficace. Mais il est probable qu'au-delà d'un certain « gain seuil » par individu l'imputation génère des erreurs. Il serait intéressant de tester dans quelle proportion les erreurs de génotypage ou les erreurs d'imputations liées aux crossing over entrainent des faux positifs dans les tests d'association et une réduction du r^2CV des modèles de sélection génomique. Créer un fichier de génotypage qui tiendrait compte de ce gain seuil pourrait être une amélioration. De plus, disposer des parents plus densément génotypés permettrait de réaliser une imputation plus efficace.

Le nombre important de QTLs obtenus avec un faible effet témoigne de la complexité des caractères et de la diversité génétique de ce matériel de sélection par rapport à des populations bi-parentales. Une partie des gènes d'intérêt présentés en introduction sont détectés. Néanmoins, certains tel qu'EAM8, VRN-H1 et VRN-H2 sont détectés uniquement avec la variable RIP. Cela pourrait traduire le fait que ces QTLs majeurs sont fixés pour les variables de précocité dans le matériel de sélection ou alors que peu de lignées ont été génotypé avec des marqueurs sur ces gènes. PPD-H2 n'apparait pas dans les résultats, ses allèles doivent être également fixés en fonction des populations étudiées. Ceci est une hypothèse à vérifier. TEM ou GI n'ont pas pu être placés précisément sur la nouvelle carte, il est impossible de conclure sur une éventuelle présence de ces gènes parmi les QTLs détectés.

Concernant les QTLs de résistance au froid il s'avère qu'ils sont moins nombreux pour les lignées W6. Ce résultat corrobore une observation formulée par l'équipe en étudiant les notes de résistance au froid des lignées commerciales : Les cultivars W2 semblent plus résistants que les cultivars W6. Le type 6 rangs est issu de types ancestraux 2 rangs et il est possible que des loci de résistance existant dans les 2 rangs aient été perdus au cours de ces sélections. De la même façon, le type printemps a été obtenu à partir du type hiver. Il a subi une pression de sélection forte et il semble que ce soit un matériel très fixé au vu du faible nombre de QTLs observé.

Concernant l'observation des différentiations de fréquences alléliques, il semble logique que ce soit les variétés présentant les allèles de résistance au froid qui aient un rendement plus

élevé. Par contre, il est surprenant que les variétés présentant les allèles précoces aient un meilleur rendement. Ce résultat devra être vérifié sur un plus grand nombre de lieux. Pour les réelles différences entre sites, aucune explication n'est trouvée pour l'instant.

Pour juger de la qualité des modèles de sélection génomique et les comparer de manière équitable, il faudrait avoir le même nombre de lignées hiver que de lignées printemps. Les modèles printemps semblent efficaces alors qu'il y a peu de loci à effet fort détecté. Cela correspond bien au modèle infinitésimal où tous les marqueurs jouent un rôle dans le caractère alors que les études d'association sont plus adaptées pour la détection de gènes à effet fort et permettent d'exploiter des populations plus petites. Des modèles plus adaptés aux gènes à effets fort restent à tester tel que le modèle BayesCπ (Heslot, et al., 2012).

Conclusions et Perspectives

Au cours de ce stage, les loci importants pour la précocité et la résistance au froid ont pu être identifiés. Une partie de ces loci colocalise avec des gènes importants des mécanismes connus dans la littérature. La variabilité existante des gènes devra être vérifiée dans les populations testées pour les loci qui ne se sont pas détectés. Il sera également intéressant d'utiliser les génotypes dont les données de phénotypage n'ont pas été prises en compte (les données de 2014 par exemple) pour comparer les méthodes de sélection génomique et d'analyses d'association pour voir qui des loci choisis grâce aux analyses d'association ou des modèles de sélection génomiques est meilleur pour prédire le comportement des lignées.

Les principales perspectives de ce stage visent à approfondir le traitement des données en prenant davantage en compte des variables environnementales qui seraient nécessaires à la création des populations « climat » mais également pour mieux décrire le froid spécifique à un lieu et une année.

Des voies d'amélioration sont envisagées concernant le test de l'impact des erreurs des données moléculaires imputées sur les résultats d'associations et de sélection génomique mais également dans la création d'une variable RIP' qui caractériserait la précocité de maturité de manière plus précise. Enfin, réaliser des classifications des individus en fonction des allèles aux loci d'intérêts permettra une meilleure connaissance du matériel travaillé.

Les mécanismes de précocité sont décrits dans de nombreux articles. Mais la complexité de ce caractère requiert encore davantage d'études. En effet, même s'il est question d'un caractère très héritable et très travaillé, les informations disponibles aujourd'hui ne permettent pas de mettre en lumière toutes les zones d'ombre qui persistent encore dans la compréhension des mécanismes de ce caractère Des mécanismes épistatiques peuvent également intervenir. De nombreuses combinaisons d'allèles sont possibles et la modélisation de l'interaction des gènes n'est pas simple.

Un plus grand nombre de données de froid et de meilleure qualité pourrait apporter plus de réponse. Mais cela est difficile à mettre en place à l'échelle d'un sélectionneur. Ainsi Limagrain participe à une étude de grande échelle dans laquelle publics et privées collaborent. La problématique « Low Temperature Tolerance » du projet Triticeae CAP en partenariat avec l'université de l'Oregon a pour objectif de tester un grand nombre variété dans des environnements qui présente des périodes de froid intense chaque année.

L'exploitation de la précocité et de la tolérance au froid en sélection assistée par marqueurs n'est pas aisée : des gènes majeurs sont fixés et les autres gènes ont des effets faibles difficiles à détecter et à comprendre. A terme, l'objectif est double. Tout d'abord, il faut comprendre les mécanismes et la diversité disponible afin de suivre les gènes d'intérêt au cours de croisement et d'introgression grâce aux tests d'association. Le deuxième objectif est de prédire efficacement la précocité et la tolérance au froid grâce à la sélection génomique. Ce travail s'inscrit dans ces objectifs : Les loci détectés sur les populations de lignées hiver et les modèles de sélection génomique calculés pour les populations de lignées printemps seront directement utilisés dans les programme de sélection.

Bibliographie

Abe, M. et al., 2005. FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. *SCIENCE*, pp. 1052-1056.

Bayer, M., Milne, I. & Marshall, D., 2009. *Strudel 1.12.03.20.* s.l.:Plant Bioinformatics Group, SCRI.

Blázquez, M. A. et al., 1998. Gibberellins Promote Flowering of Arabidopsis by Activating the LEAFY Promoter. *The Plant Cell*, pp. 791-800.

Bonjean, A. & Picard, E., 1990. *Les céréales à paille : Origine – Histoire – Economie – Sélection.* s.l.:Softword et Groupe ITM - Groupe Verneuil.

Briggs, D., 1978. *Barley.* London: Chapman and Hall.

Campoli, C. et al., 2012. Functional characterisation of HvCO1, the barley (Hordeum vulgare) flowering time ortholog of CONSTANS. *The Plant Journal*, pp. 868-880.

Campoli, C. et al., 2013. HvLUX1 is a candidate gene underlying the early maturity 10 locus in barley: phylogeny, diversity, and interactions with the circadian clock and photoperiodic pathways. *New Phytologist*, pp. 1045-1059.

Castillejo, C. & Pelaz, S., 2008. The Balance between CONSTANS and TEMPRANILLO Activities Determines FT Expression to Trigger Flowering. *Current Biology*, pp. 1338-1343.

Comadran, J. et al., 2012. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics*, 18 November.

Daron, J., 2011. *Caractérisation de gènes de blé non synténiques avec les génomes des espèces apparentées.*. [En ligne]
Available at:
http://www4.clermont.inra.fr/var/internet_clermont_umr1095/storage/htmlarea/genome/PDF/ReportM2R_Josquin_Daron_10-06-2011.pdf
[Accès le 4 Janvier 2014].

Deswarte, J.-C., 2012. *Cartographie : Froid sur céréales d'hiver : où sont localisés les dégâts en France ?.* [En ligne]
Available at: http://www.arvalis-infos.fr/view.jspz?obj=arvarticle&id=9547&syndtype=null&hasCookie=false&hasRedirected=true&region=4,5,6,13,83,84
[Accès le 13 Aout 2014].

Dhillon, T. et al., 2010. Regulation of Freezing Tolerance and Flowering in Temperate Cereals: The VRN-1 Connection. *Plant Physiology*, pp. 1846-1858.

Doré, C. & Varoquaux, F., 2006. *Histoire et amélioration de cinquante plantes cultivées.* collection Savoir-Faire éd. s.l.:Editions Inra-Quae.

FAOSTAT, 2013. *FAOSTAT.* [En ligne]
Available at: http://faostat3.fao.org/faostat-gateway/go/to/home/E
[Accès le 12 Mai 2013].

Faure, S. et al., 2012. Mutation at the circadian clock gene EARLY MATURITY 8 adapts domesticated barley (Hordeum vulgare) to short growing seasons. *Proceedings of the National Academy of Sciences*, pp. 8328-8333.

Fisk, S. P. et al., 2012. FR-H3: a new QTL to assist in the development of fall-sown. *barley with superior low temperature tolerance*, pp. 335-347.

Fowler, D. & Limin, A., 2004. Interactions among Factors Regulating Phenological Development and Acclimation Rate Determine Low-temperature Tolerance in Wheat. *Annals of Botany*, pp. 717-724.

Francia, E. et al., 2007. Fine mapping of a HvCBF gene cluster at the frost resistance locus Fr-H2 in barley. *Theorical Applied Genetics*, pp. 1083-1091.

Francia, E. et al., 2004. Two loci on chromosome 5H determine low-temperature tolerance in a "Nure" (winter) x"Tremois" (spring) barley map. *Theorical Applied Genetics*, pp. 670-680.

Gallais, A., 2011. *Méthodes de création de variétés en amélioration des plantes.* Versailles: Quae.

GNIS, 2006. *L'amélioration des plantes autogames : blé tendre, blé dur, orge et pois protéagineux,* Paris: GNIS.

Heslot, N., Yang, H.-P., Sorrels, M. E. & Jannink, J.-L., 2012. Genomic Selection in Plant Breeding : A comparison of Models. *Crop Science*.

Higgins, J. A., Bailey, P. C. & Laurie, D. A., 2010. Comparative Genomics of Flowering Time Pathways Using Brachypodium distachyon as a Model for the Temperate Grasses. *PLOS ONE*, April, Volume 5.

Hommo, L. M., 1994. Hardening of Some Winter Wheat (Triticum aestivum L.), Rye {Secale cereale L.), Triticale (X Triticosecale Wittmack) and Winter Barley {Hordeum vulgare L.) Cultivars During Autumn and the Final Winter Survival in Finland. *Plant Breding*, Volume 112, pp. 285-293.

I.E., 2012. *Céréales/Dégâts de gel : Chute de 5,4 % des surfaces de blé tendre comparé à 2011 (ministère et FranceAgriMer).* [En ligne] Available at: http://www.lafranceagricole.fr/actualite-agricole/cereales-degats-de-gel-chute-de-5-4-des-surfaces-de-ble-tendre-compare-a-2011-ministere-et-franceagrimer-56091.html [Accès le 13 Aout 2014].

IBGSC, 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature*, Volume 491, pp. 711-717.

IGC, 2013. *Five-year global supply and demand projections.* [En ligne] Available at: www.igc.int/fr/downloads/grainsupdate/igc_5yrprojections.pdf [Accès le 17 Décembre 2013].

Jeknic, Z. et al., 2013. Hv-CBF2A overexpression in barley accelerates COR gene transcript accumulation and acquisition of freezing tolerance during cold acclimation. *Plant Molecular Biology*, 15 Aout, pp. 67-82.

Kang, H. M. et al., 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, Issue 42, pp. 348-354.

Kang, M. et al., 2008. Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics*, 178(3), pp. 1709-1723.

Kikuchi, R. et al., 2009. Molecular and Functional Characterization of PEBP Genes in Barley Reveal the Diversification of TheirRoles in Flowering. *Plant Physiology*, pp. 1341-1353.

Knox, A. K. et al., 2010. CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theorical Applied Genetics*, pp. 21-35.

Laurie, D., Pratchett, N., Bezant, J. & Snape, J., 1995. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter x spring barley (Hodeum vulgare L.) cross. *Genome*, 12 February, pp. 575-585.

L'est Eclair, 2012. *Le gel ravage céréales d'hiver, pois et colza.* [En ligne] Available at: http://www.lest-eclair.fr/article/a-la-une/le-gel-ravage-cereales-dhiver-pois-et-colza [Accès le 13 Aout 2014].

Limagrain, 2013. *Le guide des semences : céréales à paille et protéagineux.* s.l.:s.n.

Locke, J. C. et al., 2006. Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. *Molecular Systems Biology*, 14 Août.

Mascher, M. et al., 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *the plant journal*, 10 October, pp. 718-727.

Météo France, 2012. *Bilan de l'hiver 2011-2012.* [En ligne] Available at: http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/bilan-2012/bilan-de-lhiver-2011-2012
[Accès le 14 Mai 2014].

Meuwissen, Hayes & Goddard, 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics,* 157(4), pp. 1819-1829.

Moule, C., 1980. *Céréales.* Paris: La maison rustique.

Osnato, M., Castillejo, C., Mathias-Hernandez, L. & Pelaz, S., 2012. TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in Arabidopsis. *NATURE COMMUNICATIONS*, pp. 808-808.

Paris, M., Jones, M. G. & Eglinton, J. K., 2002. Genotyping Single Nucleotide Polymorphisms for Selection of Barley p-amylase Alleles. *Plallt Molecl/lar Biology Reporter*, Juin, pp. 149-159.

Passion Céréales, 2011. *Des chiffres et des céréales, l'essentiel de la filière.* 2010-2011 éd. s.l.:s.n.

Rapacz, M., Tyrka, M., Gut, M. & Mikulski, W., 2010. Associations of PCR markers with freezing tolerance and photosynthetic acclimation to cold in winter barley. *Euphytica*, pp. 293-301.

Reg, L. & Chengdao, L., 2008. *Barley Breeding – Past, Present & Future.* [En ligne] Available at: http://www.ioa.uwa.edu.au/__data/assets/pdf_file/0006/1147686/Barley-improvement-in-WA.pdf
[Accès le 13 Mai 2014].

Robert-granié, c., Legarra, A. & Ducrocq, V., 2011. Principe de base de la sélection génomique. *INRA Productions Animales,* 4(24), pp. 31-340.

Rogers, J., 1972. Measures of genetic similarity ans genetic distances. *Univ. of Texas Publ.*.

Skinner, J. S. et al., 2006. Mapping of barley homologs to genes that regulate low temperature tolerance in Arabidopsis. *Theorical Applied Genetics*, pp. 832-842.

Sosnowski, O. et al., 2012. *BioMercator V4.2.* s.l.:INRA-UMRGV du Moulon.

Stockinger, E. J. et al., 2007. Expression levels of barley Cbf genes at the Frost resistance-H2 locus are dependent upon alleles at Fr-H1 and Fr-H2. *The Plant journal*, pp. 308-321.

Tondelli, A. et al., 2014. Allelic variation at Fr-H1/Vrn-H1 and Fr-H2 loci is the main determinant of frost tolerance in spring barley. *Environmental and experimental Botany,* Volume 106, pp. 148-155.

Trevaskis, B., Hemming, M. N., Dennis, E. S. & Peacock, W. J., 2007. The molecular basis of vernalization-induced flowering in cereals. *TRENDS in Plant Science*, pp. 1360-1385.

Trevaskis, B., Hemming, M. N., Peacock, W. J. & Dennis, E. S., 2006. HvVRN2 Responds to Daylength, whereas HvVRN1 Is Regulated by Vernalization and Developmental Status. *Plant Physiology*, pp. 1397-1405.

Turck, F., Fornara, F. & Coupland, G., 2008. Regulation and Identity of Florigen: FLOWERING LOCUS T Moves Center Stage. *Annual Review of Plant Biology*, pp. 573-594.

Turner, A. et al., 2005. The Pseudo-Response Regulator Ppd-H1 Provides Adaptation to Photoperiod in Barley. *SCIENCE*, pp. 1031-1034.

USDA, 2013. *Table 05 Barley Area, Yield, and Production. Production, Supply and Distribution Online Reports.* [En ligne]
Available at:
http://www.fas.usda.gov/psdonline/psdReport.aspx?hidReportRetrievalName=Table+05+Barley+Area%2c+Yield%2c+and+Production++&hidReportRetrievalID=885&hidReportRetrievalTemplateID=1
[Accès le 13 Mai 2014].

Visioni, A. et al., 2013. Genome-wide association mapping of frost tolerance in barley (Hordeum vulgare L.). *BMC Genomics*.

von Zitzewitz, J. et al., 2005. Molecular and structural characterization of barley vernalization genes. *Plant Molecular biology*, 28 June, pp. 449-467.

Ward, J., 1963. Hierachical grouping to optimize an objective function. *Journal of the american statistical association,* Issue 58, pp. 236-244.

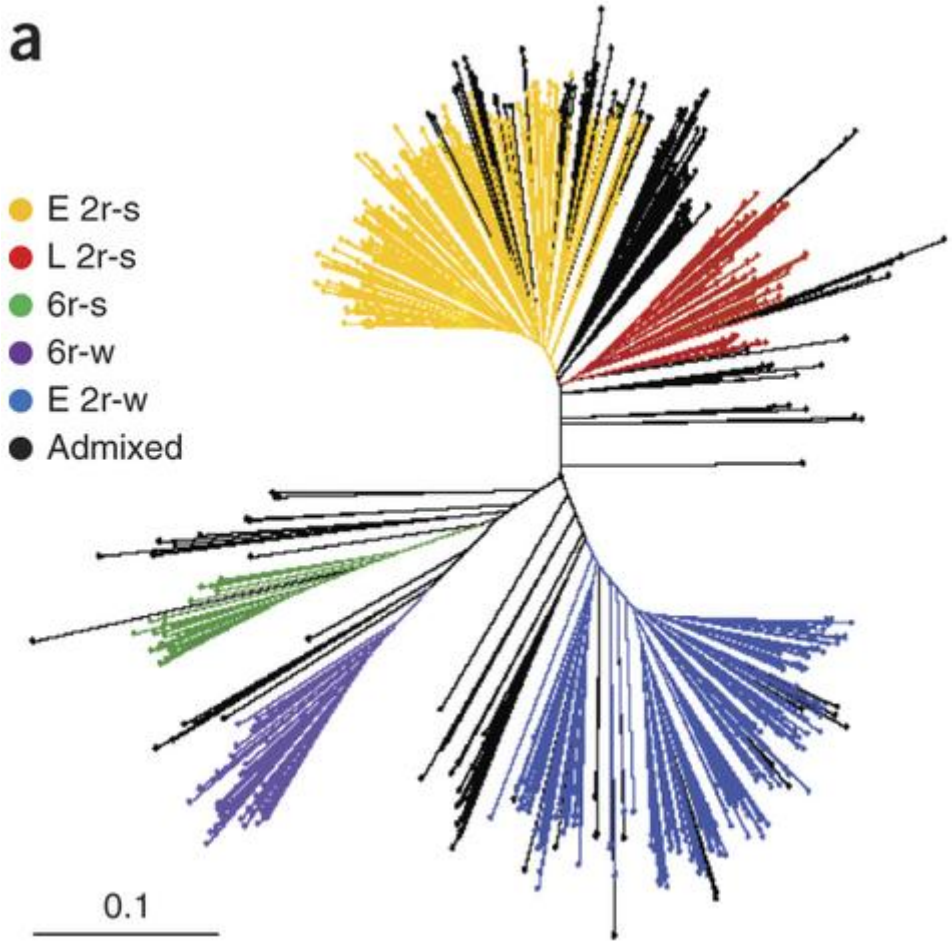
Yan, L. et al., 2004. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. *SCIENCE*, pp. 1640-1644.

Zadoks, J., Chang, T. & Konzak, C., 2006. A Decimal Code for the Growth Stages of Cereals. *Weed Research*, 28 juillet, pp. 415-421.

Zakhrabekova, S. et al., 2012. Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. *Proceedings of the National Academy of Sciences*, pp. 4326-4331.

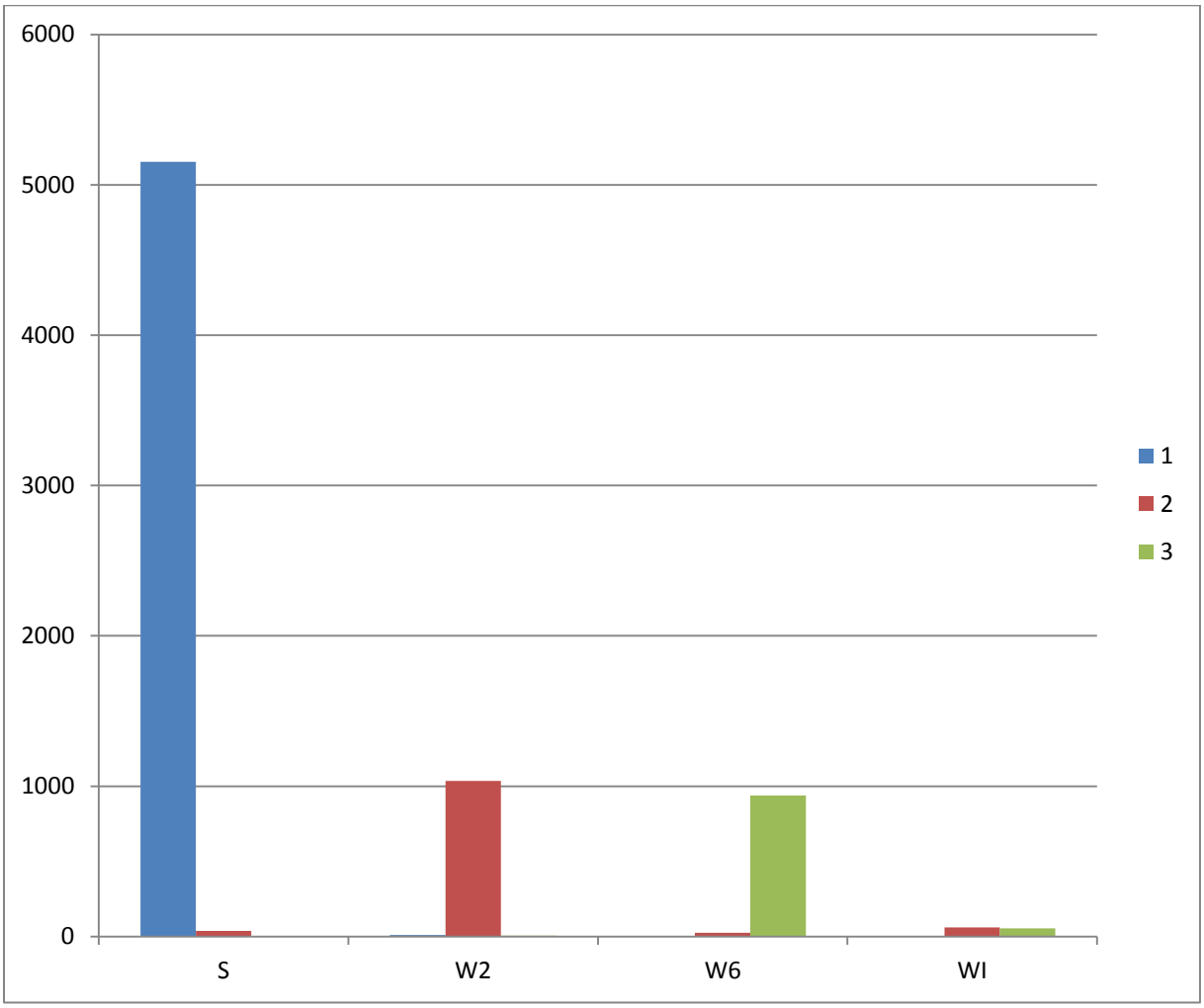
Annexes

Annexe 1 : Genetic structure and population differention in European barley germplasm

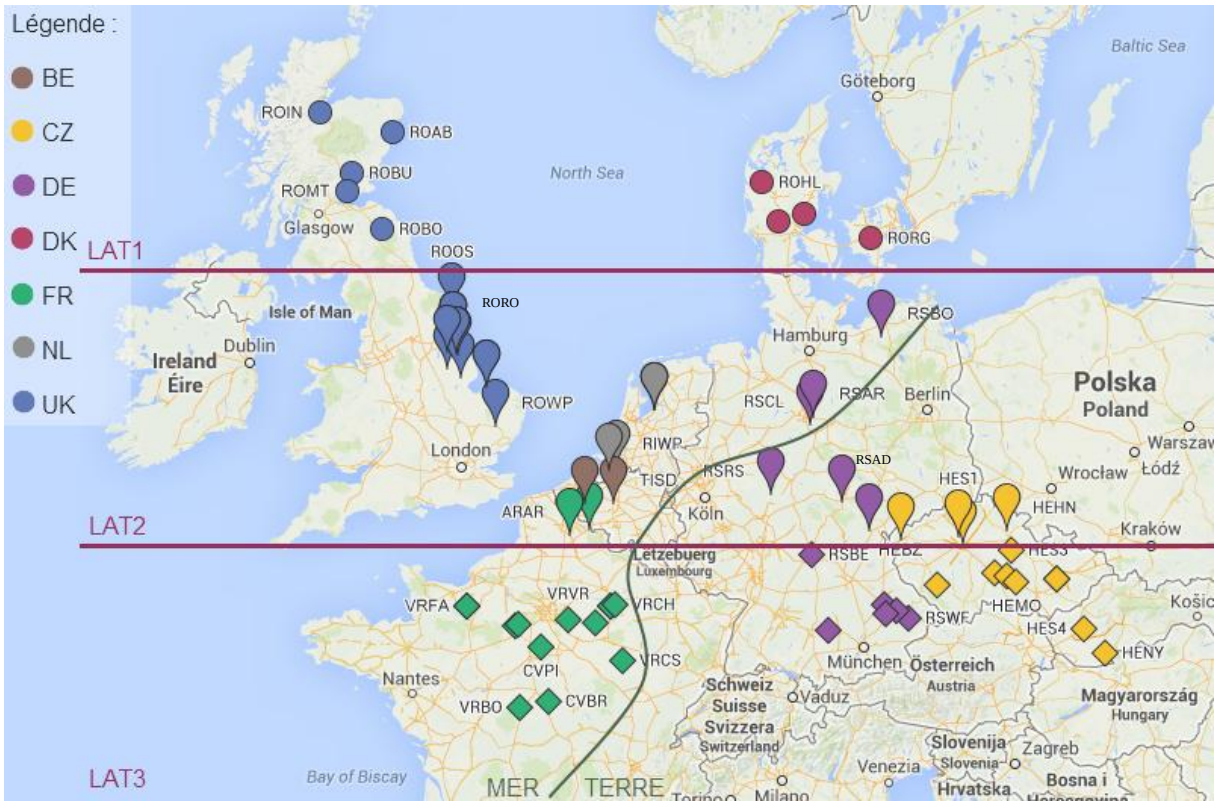


(a) Neighbor-joining tree of 804 spring and winter barley cultivars constructed from simple matching distances of all SNPs. E 2r-s, elite European two-rowed springs; L 2r-s, old land race European two-rowed springs; 6r-s, European six-rowed springs; 6r-w, European six-rowed winters; E 2r-w, elite European two-rowed winters. Units for the scale bar are simple matching distances by nonmetric multidimensional scaling analysis.

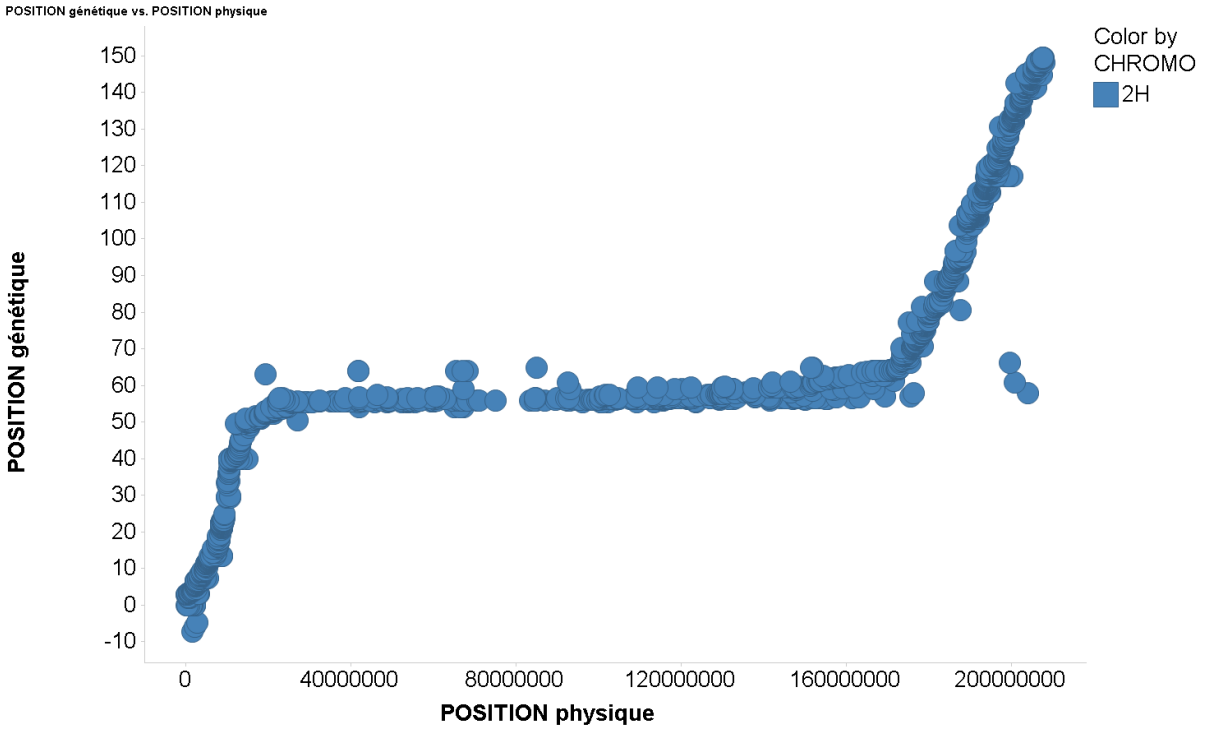
Annexe 2 : Répartition des individus selon leur groupe par la méthode de Ward par rapport au type variétaux



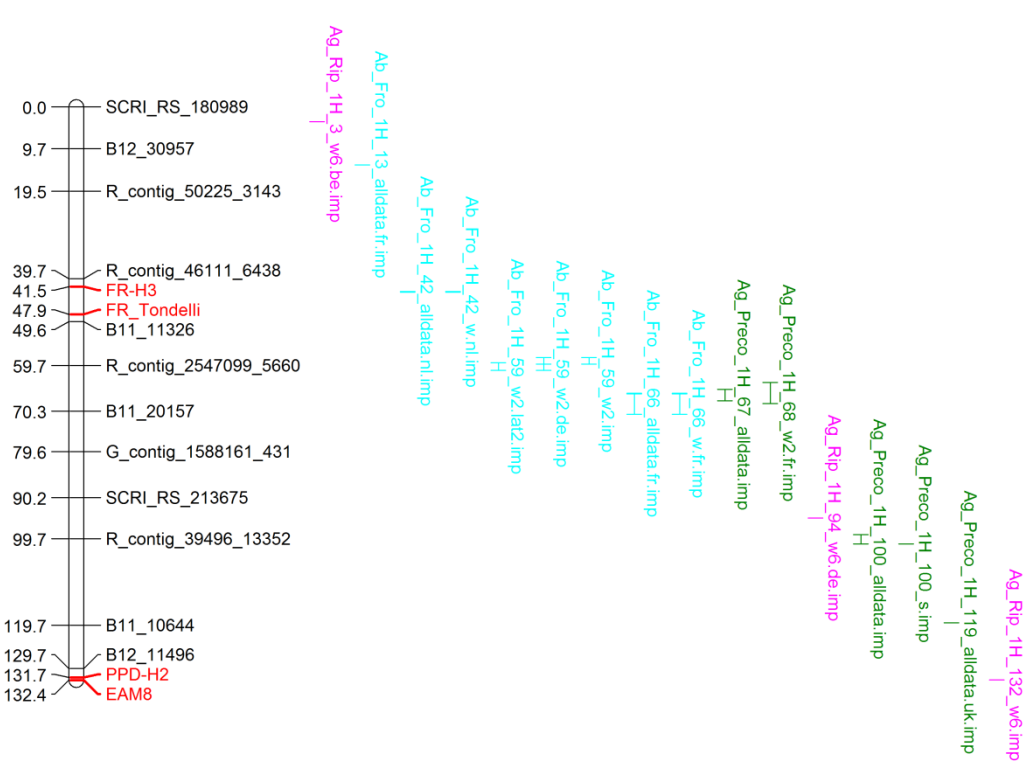
Annexe 3 : Situations des stations d'expérimentation orge utilisées dans cette étude

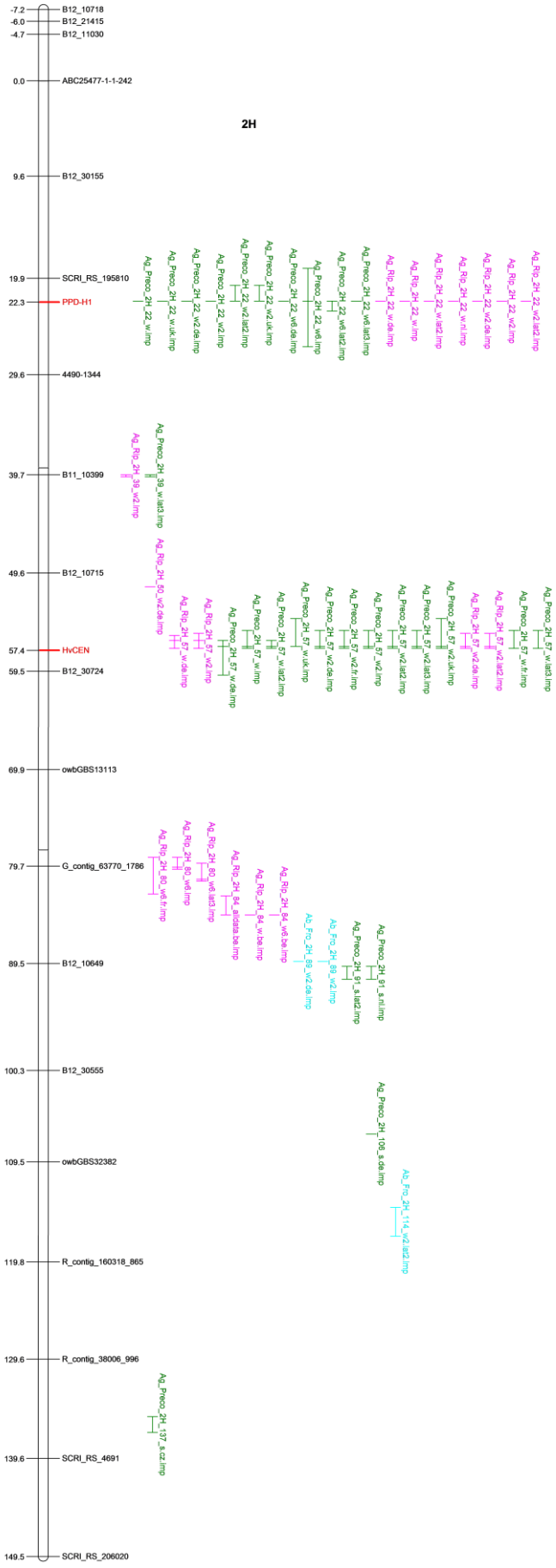


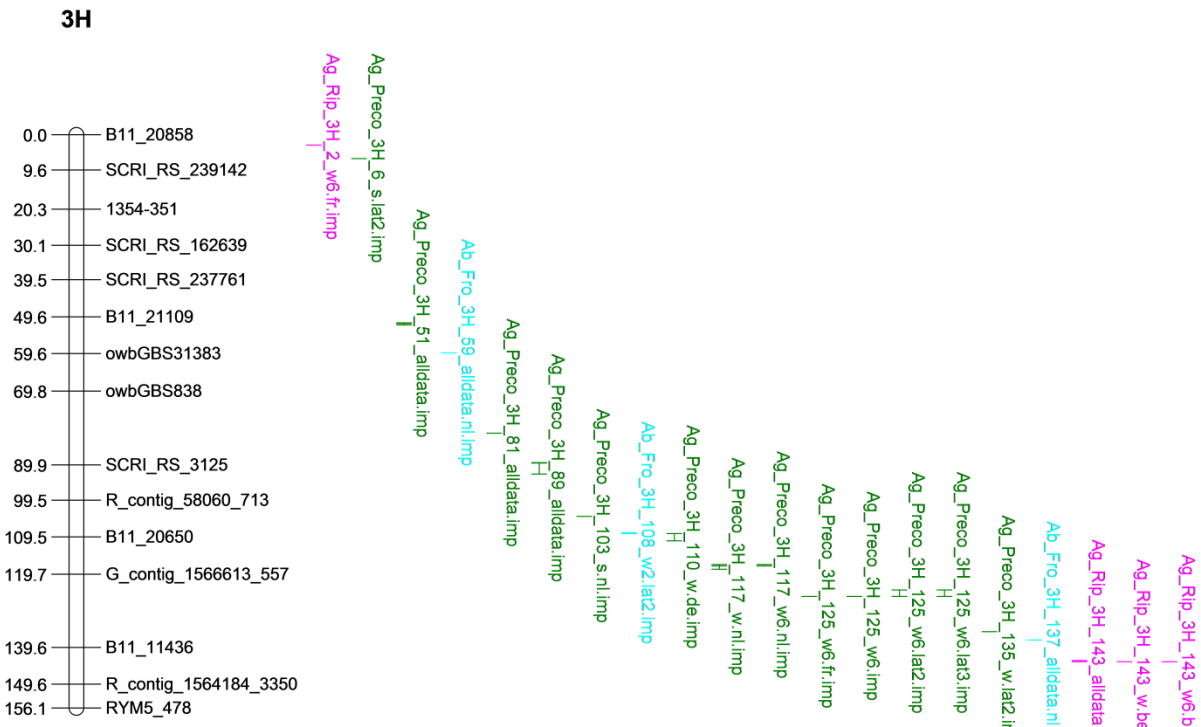
Annexe 4 : Comparaison de la nouvelle carte génétique avec la carte physique disponible

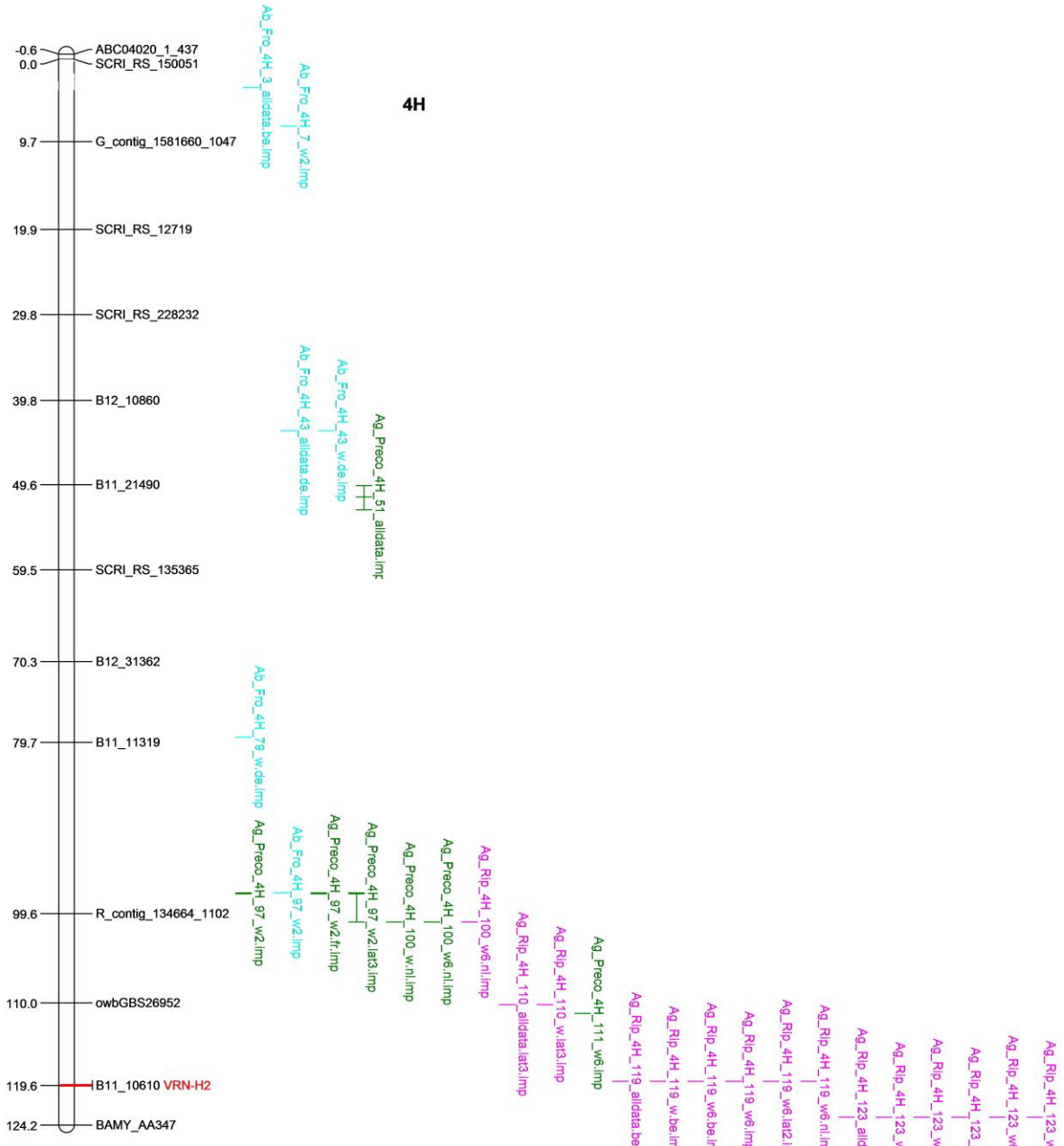


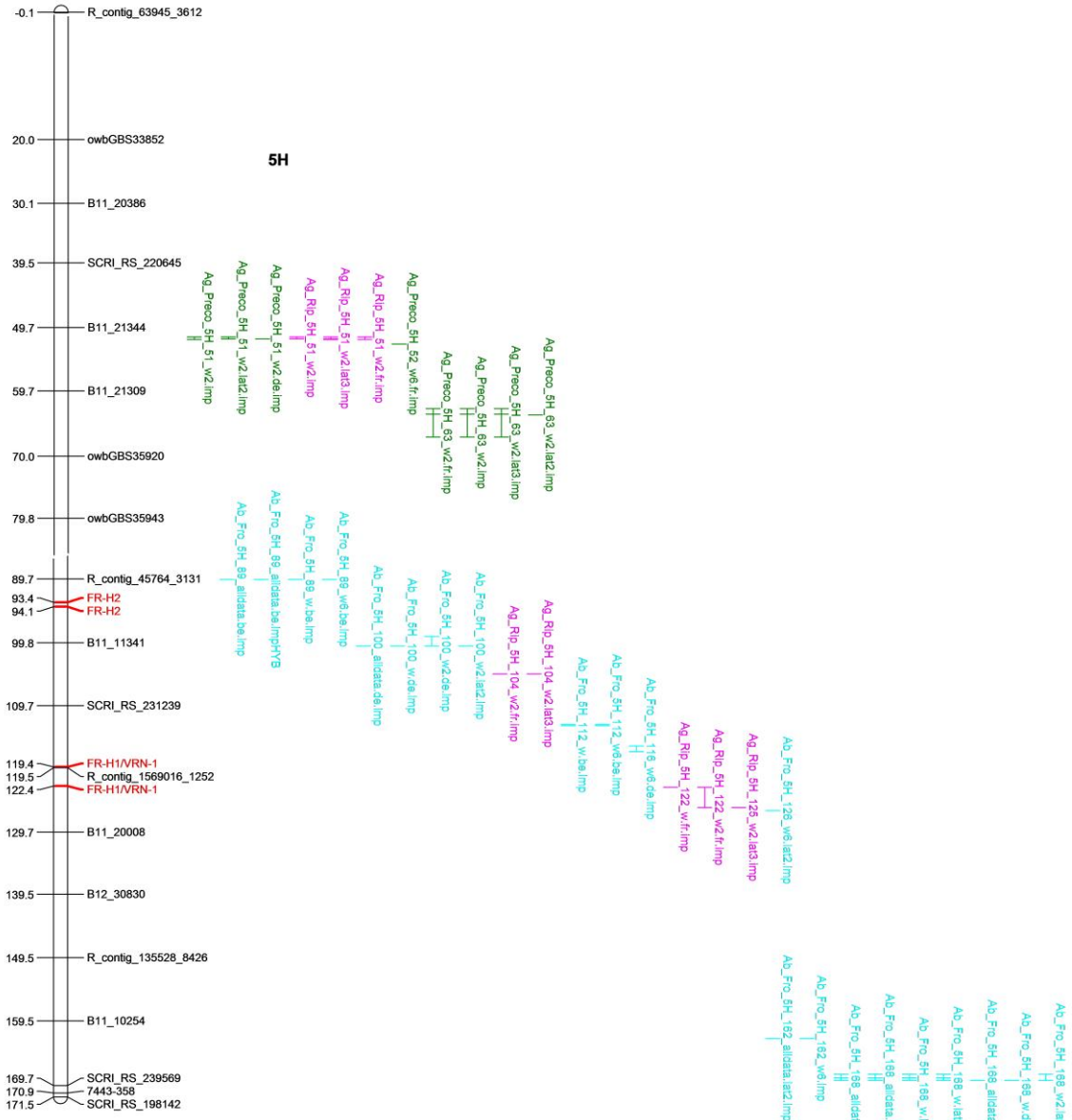
Annexe 5 : Les Loci détectés lors des études d'associations (en rose les QTLs Rip, en bleu les QTLs de froid et en vert les QTLs de précocité), les repères rouges sur les chromosomes sont les gènes connus dans la littérature

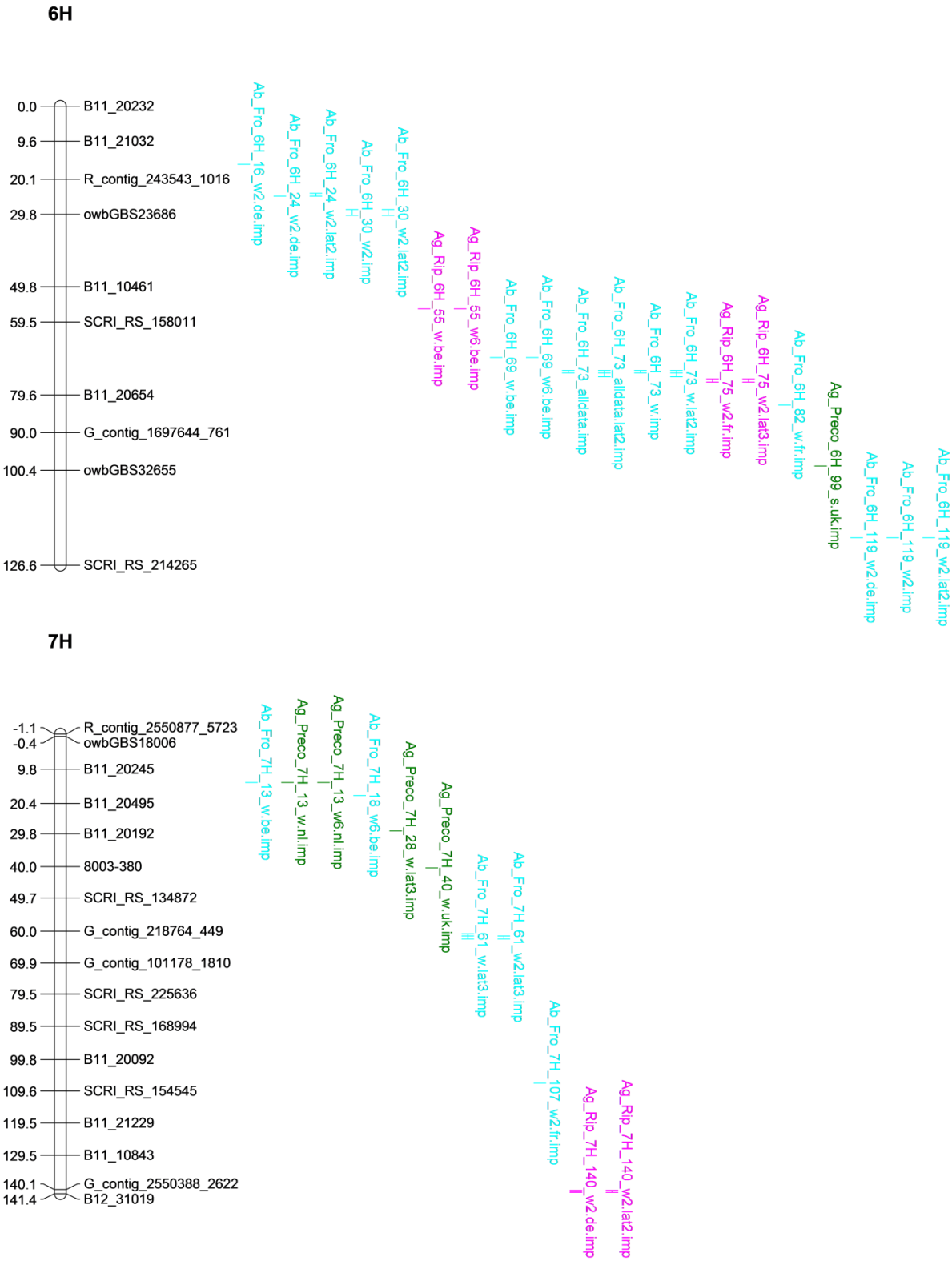



2H
-7.2 B12_10718
-6.0 B12_21415
-4.7 B12_11030
0.0 ABC25477-1-1-242
9.6 B12_30155
19.9 SCRI_RS_195810
22.3 PPD-H1
29.6 4490-1344
39.7 B11_10399
49.6 B12_10715
57.4 HvCEN
59.5 B12_30724
69.9 owbGBS13113
79.7 G_contig_63770_1786
89.5 B12_10649
100.3 B12_30555
109.5 owbGBS32382
119.8 R_contig_160318_865
129.6 R_contig_38006_996
139.6 SCRI_RS_4691
149.5 SCRI_RS_206020
Ag_Rlp_2H_22_w2.lat2.imp
Ag_Rlp_2H_22_w2.imp
Ag_Rlp_2H_22_w2.de.imp
Ag_Rlp_2H_22_w.nl.imp
Ag_Rlp_2H_22_w.lat2.imp
Ag_Rlp_2H_22_w.uk.imp
Ag_Rlp_2H_22_w.de.imp
Ag_Preco_2H_22_w6.lat3.imp
Ag_Preco_2H_22_w6.lat2.imp
Ag_Preco_2H_22_w6.de.imp
Ag_Preco_2H_22_w6.imp
Ag_Preco_2H_22_w2.uk.imp
Ag_Preco_2H_22_w2.lat2.imp
Ag_Preco_2H_22_w2.imp
Ag_Preco_2H_22_w2.de.imp
Ag_Preco_2H_22_w.uk.imp
Ag_Preco_2H_22_w.imp
Ag_Rlp_2H_39_w2.imp
Ag_Preco_2H_39_w.lat3.imp
Ag_Rlp_2H_80_w2.de.imp
Ag_Preco_2H_57_w.lat3.imp
Ag_Preco_2H_57_w.fr.imp
Ag_Preco_2H_57_w2.uk.imp
Ag_Rlp_2H_57_w2.lat2.imp
Ag_Rlp_2H_57_w2.de.imp
Ag_Preco_2H_57_w2.lat3.imp
Ag_Preco_2H_57_w2.lat2.imp
Ag_Preco_2H_57_w2.de.imp
Ag_Preco_2H_57_w2.imp
Ag_Preco_2H_57_w2.fr.imp
Ag_Preco_2H_57_w.uk.imp
Ag_Preco_2H_57_w.lat2.imp
Ag_Preco_2H_57_w.imp
Ag_Preco_2H_57_w.de.imp
Ag_Rlp_2H_57_w2.imp
Ag_Rlp_2H_57_w.de.imp
Ag_Rlp_2H_80_w6.fr.imp
Ag_Rlp_2H_80_w6.imp
Ag_Rlp_2H_80_w6.lat3.imp
Ag_Rlp_2H_84_allidata.be.imp
Ag_Rlp_2H_84_w.be.imp
Ag_Rlp_2H_84_w6.be.imp
Ab_Frq_2H_89_w2.de.imp
Ab_Frq_2H_89_w2.imp
Ag_Preco_2H_91_s.lat2.imp
Ag_Preco_2H_91_s.nl.imp
Ag_Preco_2H_108_s.de.imp
Ab_Frq_2H_114_w2.lat2.imp
Ag_Preco_2H_137_s.cz.imp

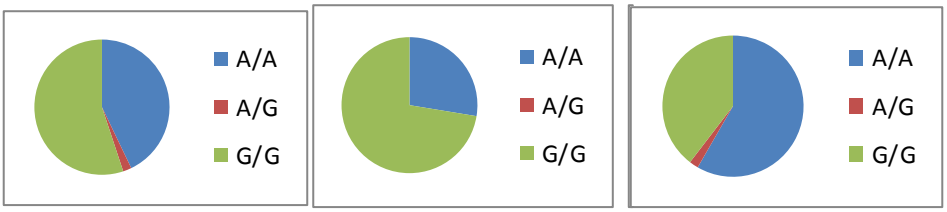
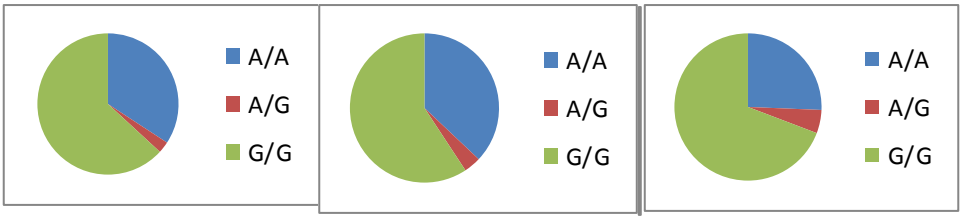
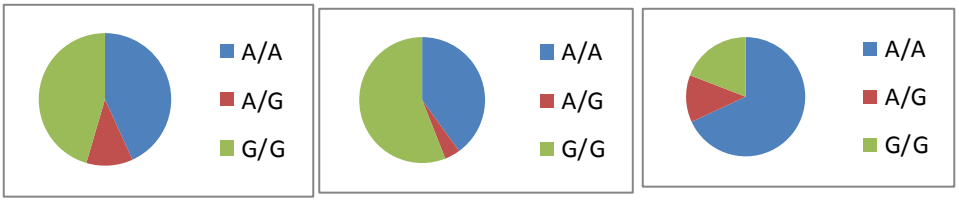
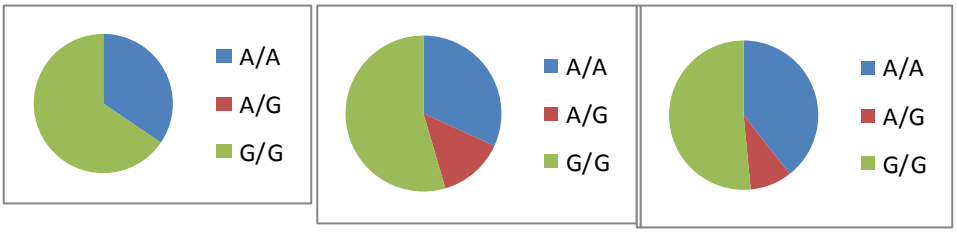
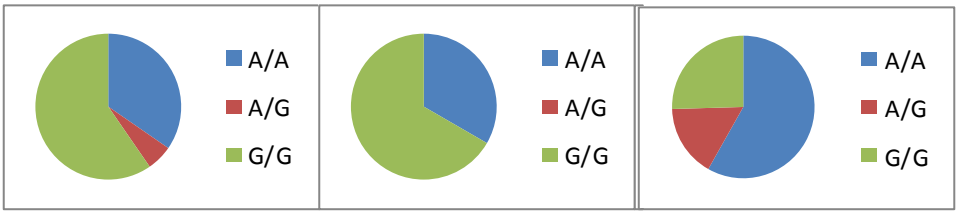
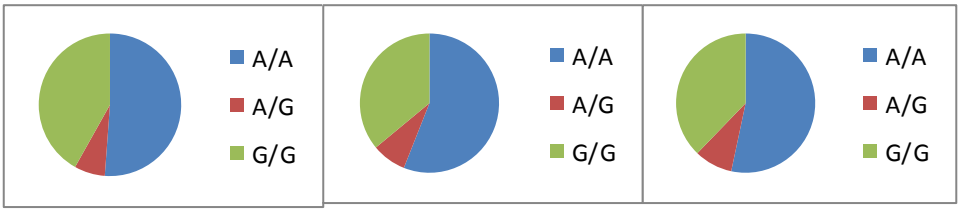

3H
0.0
9.6
20.3
30.1
39.5
49.6
59.6
69.8
89.9
99.5
109.5
119.7
139.6
149.6
156.1
B11_20858
SCRI_RS_239142
1354-351
SCRI_RS_162639
SCRI_RS_237761
B11_21109
owbGBS31383
owbGBS838
SCRI_RS_3125
R_contig_58060_713
B11_20650
G_contig_1566613_557
B11_11436
R_contig_1564184_3350
RYM5_478
Ag_Rip_3H_2_w6.fr.imp
Ag_Preco_3H_6_s.lat2.imp
Ag_Preco_3H_51_alldata.imp
Ab_Fro_3H_59_alldata.nl.imp
Ag_Preco_3H_81_alldata.imp
Ag_Preco_3H_89_alldata.imp
Ag_Preco_3H_103_s.nl.imp
Ab_Fro_3H_108_w2.lat2.imp
Ag_Preco_3H_110_w.de.imp
Ag_Preco_3H_117_w.nl.imp
Ag_Preco_3H_117_w6.nl.imp
Ag_Preco_3H_125_w6.fr.imp
Ag_Preco_3H_125_w6.imp
Ag_Preco_3H_125_w6.lat2.imp
Ag_Preco_3H_125_w6.lat3.imp
Ag_Preco_3H_135_w.lat2.imp
Ab_Fro_3H_137_alldata.nl.imp
Ag_Rip_3H_143_alldata.be.imp
Ag_Rip_3H_143_w.be.imp
Ag_Rip_3H_143_w6.be.imp


4H
-0.6
0.0
ABC04020_1_437
SCRI_RS_150051
9.7
G_contig_1581660_1047
19.9
SCRI_RS_12719
29.8
SCRI_RS_228232
39.8
B12_10860
49.6
B11_21490
59.5
SCRI_RS_135365
70.3
B12_31362
79.7
B11_11319
99.6
R_contig_134664_1102
110.0
owbGBS26952
119.6
B11_10610 VRN-H2
124.2
BAMY_AA347
Ab_Fro_4H_3_alldata.be.imp
Ab_Fro_4H_7_w2.imp
Ab_Fro_4H_43_alldata.de.imp
Ab_Fro_4H_43_w.de.imp
Ag_Preco_4H_51_alldata.imp
Ab_Fro_4H_79_w.de.imp
Ag_Preco_4H_97_w2.imp
Ab_Fro_4H_97_w2.imp
Ag_Preco_4H_97_w2.fr.imp
Ag_Preco_4H_97_w2.lat3.imp
Ag_Preco_4H_100_w.nl.imp
Ag_Preco_4H_100_w6.nl.imp
Ag_Rip_4H_100_w6.nl.imp
Ag_Rip_4H_110_alldata.lat3.imp
Ag_Rip_4H_110_w.lat3.imp
Ag_Preco_4H_111_w6.imp
Ag_Rip_4H_119_alldata.be.imp
Ag_Rip_4H_119_w.be.imp
Ag_Rip_4H_119_w6.be.imp
Ag_Rip_4H_119_w6.imp
Ag_Rip_4H_119_w6.lat2.imp
Ag_Rip_4H_119_w6.nl.imp
Ag_Rip_4H_123_alldata.be.imp
Ag_Rip_4H_123_w.be.imp
Ag_Rip_4H_123_w6.be.imp
Ag_Rip_4H_123_w6.imp
Ag_Rip_4H_123_w6.lat2.imp
Ag_Rip_4H_123_w6.nl.imp


5H
-0.1 R_contig_63945_3612
20.0 owbGBS33852
30.1 B11_20386
39.5 SCRI_RS_220645
49.7 B11_21344
59.7 B11_21309
70.0 owbGBS35920
79.8 owbGBS35943
89.7 R_contig_45764_3131
93.4 FR-H2
94.1 FR-H2
99.8 B11_11341
109.7 SCRI_RS_231239
119.4 FR-H1/VRN-1
119.5 R_contig_1569016_1252
122.4 FR-H1/VRN-1
129.7 B11_20008
139.5 B12_30830
149.5 R_contig_135528_8426
159.5 B11_10254
169.7 SCRI_RS_239569
170.9 7443-358
171.5 SCRI_RS_198142
Ag_Preco_5H_51_w2.imp
Ag_Preco_5H_51_w2.lat2.imp
Ag_Preco_5H_51_w2.de.imp
Ag_Rip_5H_51_w2.imp
Ag_Rip_5H_51_w2.lat3.imp
Ag_Rip_5H_51_w2.fr.imp
Ag_Preco_5H_52_w6.fr.imp
Ag_Preco_5H_63_w2.fr.imp
Ag_Preco_5H_63_w2.imp
Ag_Preco_5H_63_w2.lat3.imp
Ag_Preco_5H_63_w2.lat2.imp
Ab_Fro_5H_89_alldata.be.imp
Ab_Fro_5H_89_alldata.be.imp.HYB
Ab_Fro_5H_89_w.be.imp
Ab_Fro_5H_89_w6.be.imp
Ab_Fro_5H_100_alldata.de.imp
Ab_Fro_5H_100_w.de.imp
Ab_Fro_5H_100_w2.de.imp
Ab_Fro_5H_100_w2.lat2.imp
Ag_Rip_5H_104_w2.fr.imp
Ag_Rip_5H_104_w2.lat3.imp
Ab_Fro_5H_112_w.be.imp
Ab_Fro_5H_112_w6.be.imp
Ab_Fro_5H_116_w6.de.imp
Ag_Rip_5H_122_w.fr.imp
Ag_Rip_5H_122_w2.fr.imp
Ag_Rip_5H_125_w2.lat3.imp
Ab_Fro_5H_128_w6.lat2.imp
Ab_Fro_5H_162_alldata.lat2.imp
Ab_Fro_5H_162_w6.imp
Ab_Fro_5H_168_alldata.imp
Ab_Fro_5H_168_alldata.lat2.imp
Ab_Fro_5H_168_w.lat2.imp
Ab_Fro_5H_168_w.imp
Ab_Fro_5H_168_alldata.de.imp
Ab_Fro_5H_168_w.de.imp
Ab_Fro_5H_168_w2.lat2.imp


6H
0.0
9.6
20.1
29.8
49.8
59.5
79.6
90.0
100.4
126.6
B11_20232
B11_21032
R_contig_243543_1016
owbGBS23686
B11_10461
SCRI_RS_158011
B11_20654
G_contig_1697644_761
owbGBS32655
SCRI_RS_214265
Ab_Fro_6H_16_w2.de.imp
Ab_Fro_6H_24_w2.de.imp
Ab_Fro_6H_24_w2.lat2.imp
Ab_Fro_6H_30_w2.imp
Ab_Fro_6H_30_w2.lat2.imp
Ag_Rip_6H_55_w.be.imp
Ag_Rip_6H_55_w6.be.imp
Ab_Fro_6H_69_w.be.imp
Ab_Fro_6H_69_w6.be.imp
Ab_Fro_6H_73_alldata.imp
Ab_Fro_6H_73_alldata.lat2.imp
Ab_Fro_6H_73_w.imp
Ab_Fro_6H_73_w.lat2.imp
Ag_Rip_6H_75_w2.fr.imp
Ag_Rip_6H_75_w2.lat3.imp
Ab_Fro_6H_82_w.fr.imp
Ag_Preco_6H_99_s.uk.imp
Ab_Fro_6H_119_w2.de.imp
Ab_Fro_6H_119_w2.imp
Ab_Fro_6H_119_w2.lat2.imp
7H
-1.1
-0.4
9.8
20.4
29.8
40.0
49.7
60.0
69.9
79.5
89.5
99.8
109.6
119.5
129.5
140.1
141.4
R_contig_2550877_5723
owbGBS18006
B11_20245
B11_20495
B11_20192
8003-380
SCRI_RS_134872
G_contig_218764_449
G_contig_101178_1810
SCRI_RS_225636
SCRI_RS_168994
B11_20092
SCRI_RS_154545
B11_21229
B11_10843
G_contig_2550388_2622
B12_31019
Ab_Fro_7H_13_w.be.imp
Ag_Preco_7H_13_w.nl.imp
Ag_Preco_7H_13_w6.nl.imp
Ab_Fro_7H_18_w6.be.imp
Ag_Preco_7H_28_w.lat3.imp
Ag_Preco_7H_40_w.uk.imp
Ab_Fro_7H_61_w.lat3.imp
Ab_Fro_7H_61_w2.lat3.imp
Ab_Fro_7H_107_w2.fr.imp
Ag_Rip_7H_140_w2.de.imp
Ag_Rip_7H_140_w2.lat2.imp

Annexe 6 : Résumé des fréquences alléliques des loci présentant une différentiation par rapport au lieu en rdt+ et rdt-

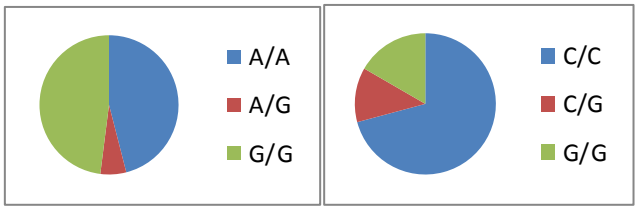
	FR-H2	Ab_Fro_168	FR-H3
RORO	R_contig_85326_844	SCRI_RS_232575	B11_10275
rdt +			
rdt-			
RSAD	R_contig_85326_844	SCRI_RS_232575	B11_10275
RDT+			
RDT-			
VRVR	R_contig_85326_844	SCRI_RS_232575	B11_10275
rdt +			
rdt -			

PPD-H1 Ag_Preco_5H_51

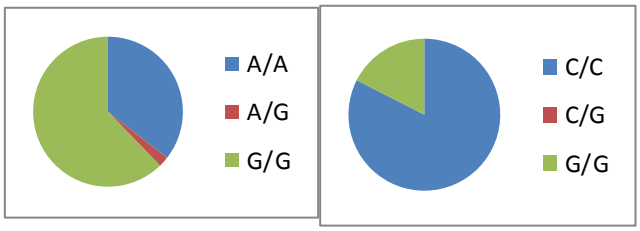
B12_30872 B11_20441

RORO

Rdt+



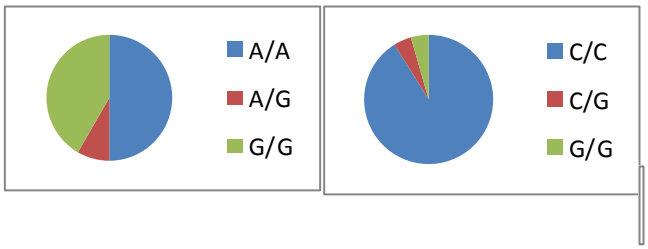
Rdt-



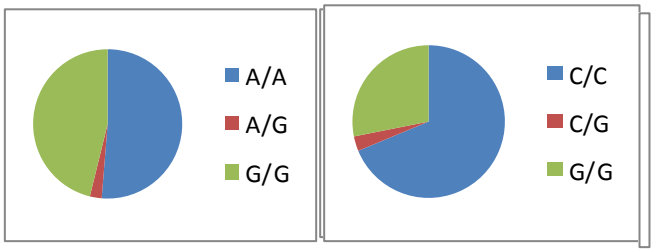
B12_30872 B11_20441

RSAD

Rdt+



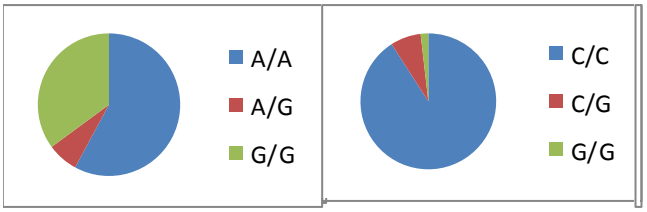
Rdt-



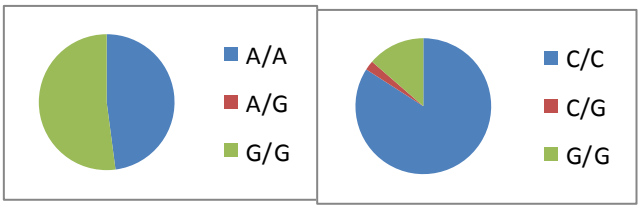
B12_30872 B11_20441

VRVR

Rdt+



Rdt-



Diplôme : Ingénieur de l'Institut Supérieur des Sciences Agronomiques, Agroalimentaires, Horticoles et du Paysage

Spécialité : Sciences et Productions Végétales

Spécialisation / option : Amélioration des Plantes

Enseignant référent : MANZANARES-DAULEUX Maria

Auteur(s) : FLAMENT Justine

Date de naissance : 24/10/1990

Nb pages : 74 Annexe(s) : 22

Année de soutenance : 2014

Organisme d'accueil : Limagrain Europe

Adresse : Site LA GARENNE

Route d'Ennezat

63720 CHAPPES

Maître de stage : BOCHARD Anne-Marie

Titre français : Etude des déterminismes génétiques de la tolérance au froid et de la précocité chez l'orge

English Title : Study of the genetic determinants of cold tolerance and earliness in barley

Résumé : L'orge (*Hordeum vulgare*) est la quatrième céréale mondiale. De nos jours, les conditions environnementales sont changeantes et les aléas climatiques se multiplient. L'équipe de sélection orge de Limagrain souhaite développer des variétés adaptées à ces nouvelles contraintes. Pour cela, il est nécessaire de mieux comprendre le déterminisme génétique de la tolérance aux stress abiotiques et la phénologie de l'orge. Le thème de ce stage propose d'étudier deux aspects importants pour l'adaptabilité recherchée et qui sont étroitement liés : la précocité et la tolérance au froid. Afin de mieux connaître les déterminants génétiques de la précocité et de la tolérance au froid, des études d'association et de sélection génomique ont été réalisées. Au total, 183 QTLs ont été détectés dans 40 populations de lignées en cours de sélection et pour un ensemble de 6 variables. Une partie de ces loci colocalise avec des gènes importants des mécanismes de précocité et de tolérance au froid. Les QTLs identifiés pourront être directement utilisés pour améliorer le matériel dans lequel ils ont étaient détectés. Un approfondissement du traitement des données prenant davantage en compte des variables environnementales est envisagé ainsi qu'une comparaison entre sélection génomique et analyse d'association.

Abstract: Barley (*Hordeum vulgare*) is the fourth cereal in the world. In our days, environmental conditions are changing and climatic hazards multiply. The Limagrain barley breeding team wants to develop adapted varieties to these new constraints. For this, it is necessary to understand the genetic determinism of abiotic stress tolerance and barley phenology. The theme of this internship proposes to explore two important aspects for adaptability and which are closely related: Earliness and frost tolerance. To better understand the genetic determinants of earliness and frost tolerance, genome wide association studies (GWAS) and genomic selection were performed. 183 QTLs were detected in 40 populations of inbred lines and for a set of six variables. Some of these loci colocalize with genes important in earliness and frost tolerance mechanisms. The identified QTLs will be directly used to improve the material from which they were detected. Data analysis will be improved by taking into account environmental variables and a comparison between genomic selection and association analysis.

Mots-clés : orge, précocité, tolérance au froid, analyse d'association, sélection génomique, cartographie

Key Words : barley, frost tolerance, GWAS, genomic selection, map making