Investigation and validation of QTL for yield and yield components in winter barley

Guillaume Barral Baron

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Newcastle University

School of Agriculture, Food and Rural Development

June 2016

Abstract

The rapid development of biotechnologies in crop genetics has increased the prospects for more efficient crop improvement. In barley breeding programmes, marker assisted selection (MAS) approaches for quantitative trait loci (QTL) for yield and yield components is still developing as it requires a thorough understanding of the genetic architecture of complex traits. This project reports an investigation of QTL for yield and yield components in two-row winter barley using three QTL mapping experiments. First, a bi-parental mapping population from an elite cross identified 23 genetic factors involved in the control of complex traits, including a strong grain weight QTL on the short arm of chromosome 2H. Second, two genome wide association studies (GWAS) were used to explore the genetic diversity for agronomic traits in European variety panels used in the NUE-CROPS and the AGOUEB projects. The integration of QTL mapping results revealed clustering of significant effects as potential targets for MAS. A major QTL cluster identified on 2H suggests that the centromeric HvCEN candidate gene is strongly involved in controlling the phenology and number of grains per ear in two-row winter barley and has additional pleiotropic effects on several agronomic traits. Some QTL effects were further confirmed by a QTL validation experiment using near isogenic lines (NILs) developed from advanced breeding material alongside the mapping experiments. Most of the QTL clusters involving different yield components showed that allele effects mirrored phenotypic correlations and a few QTL clusters were identified that had unidirectional increasing effects on all traits, such as an important tillering locus on 4HL. The exploitation of comparative genomics with rice revealed that SNP haplotypes could be used for candidate gene discovery at barley QTL clusters.

The complexity of the QTL clusters associated with yield and yield components highlight the challenges in identifying relevant targets for marker assisted breeding when accounting for pleiotropic effects of loci controlling phenology and correlated traits. The study provides insights into the genetic architecture of complex traits in small grain cereals and for the implementation of associated QTL in commercial barley breeding activities.

Acknowledgement

First and foremost I would like to thank my main supervisor Dr Peter Werner. His contributions of time, guidance, ideas and generous support have motivated me during the whole course of the PhD project. His enthusiasm and passion for plant breeding, genetics and applied research have been a source of inspiration for this work.

I am more than grateful to my supervisor Dr Luke Ramsay at the James Hutton Institute for his support and input during the course of this project. His expertise on barley genetics and biotechnology was greatly appreciated for the in-depth analysis of my results.

I would like to thank my supervisor Dr Julia Cooper at Newcastle University, for the extremely valuable input during the course of this project especially for the writing up stages. Her advices, suggestions and encouragements have been kept me motivated during this project. I am also very thankful for all the administrative support she gave me during the duration of that project.

I am indebted to Dr Bill Thomas for all the time and work he devoted to this work. His extensive knowledge on barley breeding, quantitative genetics, statistics and all the constructive discussion we had for both NUE-CROPS and the PhD have contributed immensely to this study.

I would like to express my gratitude to Carlo Leifert at Newcastle University to establish the EU-FP7 NUE-CROPS project and providing very positive support and encouragement for the establishment of a company funded PhD within this framework. I am also indebted to Professor Chris Tapsell at KWS-UK Ltd for supporting this PhD project and KWS-UK Ltd for providing the main funding of this project.

I acknowledge the contribution of excellent students Kim Goodger and Morgane Roth who have helped greatly during the busy harvests. The completion and quality of field measurements and phenotypes would not have been possible without them.

This work would not have been possible without the much appreciated contribution of Joanne Russell, Allan Booth and Jordi Comadran from the James Hutton Institute for the excellent trainings on genetics, markers technology, laboratory work and GWAS. I owe a very special thank you to my colleagues at KWS and more especially the trials team for providing excellent trial management and support, to Claire Fremann and Debra Manning for keeping me motivated, to David Harrap for his expertise in barley breeding, to Mark Dodds my manager for his support and giving me some time and to Jacob Lage for his

review of my thesis. I am grateful to Stephen Wilcockson and Lois Bell at Newcastle University for their comments and help.

I would like to thank my family for the support and encouragements during this project.

The authors gratefully acknowledge funding from the European Community financial participation under the Seventh Framework Programme for Research, Technological Development and Demonstration Activities, for the Integrated Project NUE-CROPS FP7-CP-IP 222645. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the information contained herein.

Table of contents

List of figures

List of tables

List of appendix

List of abbreviations

ANOVA	Analysis of Variance
AGOUEB	Association Genetics Of UK Elite Barley
BLUEs	Best Linear Unbiaised Estimates
BLUPs	Best Linear Unbiaised Predictors
BOPA	Barley Oligo Pool Assay
DUS	Distinctness Uniformity and Stability
GC	Genomic Control
GWAS	Genome Wide Association Study
HIFs	Heterogeneous Inbred Family
JHI	The James Hutton Institute
LLHS	Lower Leaf Hairy Sheath
MAB	Marker Assisted Breeding
MAGIC	Multi-parent Advanced Generation Inter-Cross
MAS	Marker Assisted Selection
MTA	Marker Trait Association
MLM	Mixed Linear Model
NAM	Nested Association Mapping
NILs	Near Isogenic Lines
NUE	Nitrogen Use Efficiency
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
QTL	Quantitative Trait Locus
RAPD	Random Amplified Polymorphic DNA
AFLP	Amplified Fragment Length Polymorphism
REML	Residual Maximum Likelihood
RILs	Recombinant Inbred Lines
SA	Structured Association
SNP	Single Nucleotide Polymorphism
UPOV	Union for the Protection Of new Varieties of plant

Introduction

Agriculture has been central to human evolution and the expansion of farming based cultures and is now at the heart of food supply. Amongst main agriculture activities, the selection of crops has been a continuous process to adapt our food production to face changes in environmental conditions and in supply requirements across evolutionary times. About 10,000 years ago in the fertile crescent, the first barley crops were domesticated from a wild relative *Hordeum spontaneum* (Badr et al., 2000). Generations later, plant breeding is still the key activity to improve yield and agronomic traits of that widely grown cereal. This was illustrated by the research developments leading to remarkable step-change in the farm environment during the green revolution of the 1960s. The progress achieved unveiled the potential of plant genetics to significantly increase yields of cereal crops worldwide and address sustainability and stability of crop production. Subsequently, the development of biotechnology in modern plant breeding has been aimed at harnessing the potential of genes for crop improvement. Currently, plant breeders are faced with the challenge of exploiting and translating the genetic variation in crops to meet the goals of future global food security.

This rapid development of biotechnology has given rise to new perspectives in plant breeding by providing access to an unprecedented depth of understanding of the plant genetic information contained in the DNA. The wider use of genetic markers especially the single nucleotide polymorphisms (SNP), enables characterisation of the genetic variation over entire genome with a high coverage to ultimately associate variation in genetic sequences with phenotypic variations. More importantly, the use of genetic markers coupled with statistical analysis provides an insight into understanding of the genetic architecture of traits controlled by multiple genes, commonly referred to as quantitative traits. To do so, the different genetic mapping methods of quantitative traits locus (QTL) work under the hypothesis of detecting significant correlations between segregating genetic alleles in mapping populations and phenotypic variation (Collard et al., 2005). QTL mapping experiments in plants have successfully been able to identify loci and genes involved in the control of qualitative and quantitative traits (Bernardo, 2008). In barley, QTL for distinctness, uniformity and stability (DUS) traits, disease resistance and complex traits of grain quality and yield components have been

published (Aghnoum et al., 2009; Bezant et al., 1997; Cockram et al., 2010; Comadran et al., 2011b). Recently, the completion of the genome wide association study in the AGOUEB project (www.agoueb.org) has demonstrated the potential of using wider genetic diversity in barley QTL mapping studies to investigate trait variation available to breeders (Waugh et al., 2009). Plant breeders now have the ability to better understand the genetics of complex traits, and access to large amounts of genetic information, both of which should result in breeding progress. Nevertheless, it should be noted that the knowledge generated from research is only slowly implemented in breeding programmes (Bernardo, 2008; Rae et al., 2007) as plant breeders require validated and reliable effects at QTL targets.

Grain yield and yield stability (i.e. constant yield over years and environments) are the major breeding targets to improve the sustainable production of small grains cereals such as barley and wheat. At a genetic level, yield is the most complex trait and the result of the whole crop cycle. Therefore breeding for high yield requires a comprehensive knowledge of the genetic architecture of traits and environmental factors responsible for variation (Abeledo et al., 2003; Mackay et al., 2011; Mackay et al., 2009). To address yield architecture, barley yield can be divided into the sum of components consisting of the number of fertile tillers per plant, the number grains per ear and the grain weight. The genetic control of these traits results from the expression of additive effects and interact with the environment during the whole plant cycle (Xue et al., 2010). In particular, the plant adaptation to the environment that reflects physiology such as resource use efficiency and phenology are determinant factors for yield and yield component variation (Abeledo et al., 2003; Cockram et al., 2007). Therefore, breeding progress in yield is no more than the selection for the best combinations of alleles for a given environment. A better understanding of the genetic architecture can identify the alleles that are consistently favourable, thus increasing the possibility of generating these optimal allele combinations. Furthermore, once associations of genetic markers with trait have been validated, the markers can be used in marker assisted selection (MAS) strategies to increase the efficiency of selection in order to compile positive effects in improved varieties (Collard and Mackill, 2008).

This research project focuses on the potential of using genetic marker applications to improve yield in two-row winter barley in an applied and commercial breeding programme. The main objective is to increase our understanding about the genetic

architecture of economically and agronomically relevant quantitative traits in two-row winter barley, and furthermore evaluate the potential for breeding applications. The project is an extension of the NUE-CROPs project which aimed at investigating the genetic potential in four main crops species to maintain current yields while reducing the environmental impact of industrial agriculture (http://research.ncl.ac.uk/nefg/nuecrops/page.php).

The scope of the project is reviewed in Chapter 1 with an emphasis on the prospects of strategies for genetic progress of complex traits in a breeding context. The foundation of this study consists of three QTL mapping experiments using a range of genetic resources to identify consistent genetic factors (QTL) for yield and yield components in winter barley. Chapter 2 reports QTL mapping exercise using a bi-parental double-haploid (DH) population from an elite cross between Saffron and Retriever. Chapter 3 presents the genome wide association studies (GWAS) for yield and yield components and a range of agronomic traits including disease resistance and nitrogen use efficiency. A panel made of 226 European two-row winter barley varieties was used. The varieties were also included in the AGOUEB and NUE-CROPS projects where extensive genotype information from 9K SNP markers was generated. The evaluation and in-depth analysis of the QTL mapping results integrated on a common genetic map are presented in Chapter 4. This enabled the identification of QTL clusters and a more comprehensible interpretation of the genetic architecture of yield, especially the influence of adaptive loci on yield variation. This allow for identification of the precise position of useful molecular markers to locate the genetic origin controlling trait variations. The genome collinearity with rice can then be used to propose candidate genes at relevant loci with the prospects of identifying diagnostic markers for precision breeding. Based on the QTL results obtained in previous chapters, Chapter 5 describes the implementation of a QTL validation strategy using Near Isogenic Lines (NILs) developed from advanced generation breeding material. Finally, the relevance of the loci of interests and prospects for marker applications for the selection of yield and yield components are discussed in Chapter 6 of this thesis.

The work carried out during this project involved the acknowledged input of my supervisors and the help for a range of persons and institutes (see ii). The project work was divided as follow: All the trial management at KWS was carried out by the KWS-UK trials team. The phenotyping the DH biparental mapping study (Chapter 2) was carried out at KWS-UK by me with the help of a temporary student. I carried out the

statistical and QTL analysis in the chapter. I managed the NUE-CROPs trials and phenotyping of those trials at both KWS-UK and location in Germany with the help temporary students. The James Hutton Institute (JHI) carried out the phenotyping of the NUE-CROPs trials in Dundee (Chapter 3). The JHI provided the genotype information for the NUE-CROPs panel of varieties. I carried out the statistical analysis on the phenotypes for the results reported in Chapter 3. The phenotype data relative to the AGOUEB project were collected by the AGOUEB consortium before this PhD project started. It was accessed by KWS-UK as a partner in the project. The JHI provided the genotype information for the AGOUEB panel of varieties. I carried out the GWAS for both the NUE-CROPs and AGOUEB dataset presented in that study. I carried out the work and analysis reported in Chapter 4 and Chapter 5. The JHI provided carried out the sequencing of the Ppd-H1 gene presented in Appendix 5. 4.

Chapter 1
Literature review

Plant breeding and crop science have been under continuous development over many years to improve crop production and agronomic performance amongst which yield, disease and pest resistances and production quality are the major targets (Heisey et al., 2002; Muurinen et al., 2006; Rae et al., 2007; Zhang et al.). Amongst small grain cereals, barley has benefited from state-of-art breeding methods and technologies based on advanced genetic knowledge. The breeding community needs to encourage the use available resources and research tools in order to improve crop varieties, and more particularly the investigation of the genetic control of quantitative traits. This will be a leading theme throughout this PhD project. Chapter 1 sets the context of the project by referring to the recent advances in barley breeding, crop sciences and genetics that are leading to thriving developments in crop improvement.

1.1 The cereal crop of winter barley.

1.1.1 Biology and production of barley.

Domesticated barley, *Hordeum vulgare* L. is a self-pollinating small grain cereal member of the Poaceae family. The genus *Hordeum* is composed of 32 species amongst which *H. vulgare* itself is divided into the domesticated and commonly cultivated *H. vulgare ssps. vulgare*, and a wild ancestor *H. vulgare ssps. Spontaneum* (Bastergue et al., 2006). The species domestication events are believed to originate from the Fertile Crescent around 10 000 years ago while further diversification occurred in Himalayan areas (Badr et al., 2000). The main trait which allowed domestication was the selection of non-shattering spikes to avoid seed dispersal at maturity detrimental for harvest. The barley spike is composed of an alternate succession of spikelets arranged in triplets along the rachis node. The two-row barley is characterized by having a single central spikelet fertile and side spikelets reduced or absent. The restored fertility of all spikelets in six-row barley is controlled by a multi allelic gene *vrs1* (Komatsuda and Mano, 2002; Komatsuda et al., 2007). The three types of barley growth habits are winter, factultative and spring barleys. These habits characterize the physiological traits involved in sensitivity to vernalization, photoperiod, and cold tolerance. Unlike spring barley, the winter barley type requires to be exposed for a period under low temperature (8°C)

referred to as vernalization in order to initiate flowering (Szucs et al., 2007). Facultative barley is vernalization-insensitive and more tolerant to cold temperatures compared to spring barley. The sensing of photoperiod sensitivity affect plant response and flowering initiation under variable day length and participate to the geographical distribution and adaptation of different barley types across growing areas (Cockram et al., 2007; Laurie, 1997).

In 2012, the world production of barley amounted to 132 Mt ranking as the fourth most produced cereal after maize, rice and wheat (faostat.fao.org). The main producers of barley are the Russian federation (13.9Mt), France (11.3Mt) and Germany (10.4Mt). Barley production in the United Kingdom was 5.5 Mt in 2012, the 2nd most important cereal crop behind wheat (13.2 Mt) making the UK the 10th barley producing country in the world.

The use of barley depends on the grain quality required by end-user processes which would classify it for either food or feed product categories. In the quality barley, grains are used for the production of malt for the brewing and distilling industries and to a lesser extent for food. Alternatively barley unsuited for malt production due to higher protein content and lower grain quality is used as feed. In this category, the breeding efforts are entirely targeted to maximise the output of grain per land area, i.e. the grain yield, while maintaining and improving beneficial agronomic characteristics such as disease resistance.

1.1.2 Barley genetics.

Barley is a diploid species composed of 2n=14 chromosomes for a size of 5.1 Gb. The international barley genome sequencing project initiated by a North American and EU collaboration led to the assembly of barley genome sequence resources into a physical and genetic map framework in 2012 (Mayer et al., 2012). The barley genome is highly repetitive with 84% of the sequence data attributable to mobile elements or other repeat structures. The number of high confidence genes is estimated as 26,159 genes from the 79,379 transcript clusters found (Mayer et al., 2012). This is half the number of predicted rice genes (Yu et al., 2002) and 3 times smaller than the wheat genome (Brenchley et al., 2012). By comparison, the human genome contains about 25,000 genes for 2.9 Gb (IHGSC, 2004). International efforts of the barley research community have generated substantial genetic and genomic resources with thousands of genetic

markers available on genotyping platforms. These dynamic developments in plant genetics have allowed ambitious initiatives such as AGOUEB and Barley CAP to investigate the genetic diversity in the worldwide barley germplasm (Waugh et al., 2009). Thousands of genetic markers are now available and routinely implemented in studies aiming at understanding the genetic control of complex traits ranging from disease resistances to quality and agronomic traits such as yield.

The relationship of synteny between species based on genetic markers has been well described in the literature depicting the conservation of genome blocks across species (Gale and Devos, 1998; Salse and Feuillet, 2009). Rice and Brachypodium have initially been in the foreground of grass studies but the genome co-linearity of barley with other grass species is also an advantage of the species for small grains cereals research. Barley combines the advantages of diploid genetics, a short life cycle and a close genomic proximity with wheat (*Triticum aestivum* L.) to become a leading model crop species (Distelfeld et al., 2008). Already, the exploitation of the syntenic relationship between barley and maize has led to successful gene discovery and validation (Ramsay et al., 2011). Barley is a key crop to help understanding the genetic control of traits underpinning cereal crops performance in different environments.

1.1.3 Barley breeding.

Since early domestication of the species, barley breeders have used a range of methods to create improved varieties. The self-pollinating nature of the species implies that the combinations of favorable traits are created by manual crossing of carefully selected parental lines. The conventional breeding methods of pedigree breeding and single seed descent are commonly used for barley breeding. The more sophisticated method of double haploid production offers breeders the possibility to accelerate the breeding progress by reducing the time to reach complete homozygosity of the lines. Because of its spontaneous chromosome doubling barley lends itself well to this type of production (Touraev et al., 2009). More recently, the better understanding of the genetic control of cytoplasmic male sterility and restoration of fertility has opened up new possibilities for the development of hybrid barley varieties (Mühleisen et al., 2013). These continuous efforts in barley breeding have facilitated constant and sustained breeding progress resulting in a yield increase of 1% per annum from 1983 to 2004 for both winter and spring types in the UK (Rae et al., 2007). Amongst the recent high yielding UK feed

barley, varieties Saffron and Retriever have demonstrated a significant step forward in feed barley yields as presented by the HGCA recommended list 2010/11 (Table 1.1).

The desirable traits under selection depend on the location and the strategy of the breeding programmes. They include yield, quality (feed or malting), disease and insect resistance, abiotic stress resistance and other agronomic traits. These multi-trait selections have largely contributed to the yield increases in the UK and proved to be necessary to overcome historical environmental changes and disease resistance breakdown (Mackay et al., 2011). Specific traits are maintained to meet the market demands. For example, the two-row and six-row types are conserved by selecting for specific allelic variant at the *vrs1* and *int-c* genes involved in the control of fertility of lateral spikelets on the ears (Komatsuda et al., 2007). In the UK, the winter barley varieties are mainly grown for feed purposes (NIABtag, 2016) whereas the better grain quality obtained in spring barley (e.g. protein content) is aimed at satisfying malting, brewing and distilling markers demands. These different phenotypes have led commercial breeders to breed within separate gene pool although pre-breeding efforts are used to generate novel allelic combinations by crossing between them. This strategy is reinforced by the complexity of traits involved in malting quality for the brewing and distilling industries that requires breeders to work within adapted germplasm.

Nowadays, changes in barley breeding are instigated by rapid developments in genetic research bringing with it innovative tools and methods. The genetic markers and high throughput genotyping platform have encouraged initiatives to associate genetic polymorphisms with trait variation with the view of marker assisted selection (MAS) strategies (Close et al., 2009; Collard and Mackill, 2008; Waugh et al., 2009). More recently, the genomic selection (GS) prospects combining biotechnologies and biostatistics have been realised in barley breeding programmes and for other cereals (Reynolds et al., 2011; Varshney et al., 2005). With all these tools available to breeders, a practical understanding of the methods and their applications is required in order to realise their full potential and convert scientific progress into advances in yield, disease resistance, nutrient use efficiency or any traits with added value.

Table 1.1 Two-row feed winter barley varieties in HGCA recommended list 2010/11.

Table of agronomic performances and recommendations of winter barley varieties adapted from HGCA recommended list of varieties 2010/11.

Source: HGCA (http://cereals.ahdb.org.uk/varieties/ahdb-recommended-lists.aspx)

	Retriever	KWS-Cassia	Saffron	Suzuka	Carat	Average LSD (5%)
Variety type	**Two-row feed**					
Scope of recommendation	UK	UK	UK	UK	Sp	
UK treated yield as % control (8.6 t/ha)	107	107	103	101	97	2.6
Main market options						
Year first listed	07	10	05	07	02	
Grain quality						
Specific weight (kg/hl)	66.6	70.8	70.3	69.5	71.6	0.9
Screenings % through 2.25 mm	4.8	1.9	1.9	2.4	0.8	2.0
Screenings % through 2.5 mm	16.8	6.2	7.4	7.3	2.6	5.7
Grain yield as % treated control						
UK with fungicide (8.6 t/ha)	107.1	106.9	102.5	101.1	96.8	2.6
Dry (East) region with fungicide (8.8 t/ha)	106	108	103	101	97	3.6
North region with fungicide (8.7 t/ha)	111	[106]	101	102	95	4.1
Wet (West) region with fungicide (8.2 t/ha)	106	[106]	103	101	98	4.6
Untreated grain yield (% treated control in comparable trials)						
UK without fungicide	83	87	83	85	79	3.5
Agronomic features						
Resistance to lodging	6.4	7.8	7.8	7.5	7.8	-
Straw height (cm)	84	89	88	90	82	3.0
Ripening (+/- Pearl, -ve = earlier)	-1	-1	0	-2	-2	1.0
Winter hardiness #	6	-	5	6	6	1.0
Disease resistance						
Mildew	6.0	4.6	2.8	5.7	6.6	1.1
Yellow rust	7.9	5.6	6.3	8.0	7.1	1.4
Brown rust	5.8	7.0	7.0	7.2	4.8	1.0
Rhynchosporium	8.0	4.2	4.0	8.1	5.3	1.7
Net blotch	6.3	7.5	7.9	7.0	5.9	0
BaYMV	R	R	-	R	R	
Annual yields as % treated control						
2005 treated yield (8.5 t/ha)	109	-	101	101	98	3.7
2006 treated yield (8.7 t/ha)	106	-	102	101	95	3.0
2007 treated yield (8.3 t/ha)	106	108	103	101	96	4.7
2008 treated yield (8.8 t/ha)	107	106	102	101	99	5.1
2009 treated yield (8.8 t/ha)	109	107	103	101	95	3.7
Soil type (about 50% of trials are medium soils)						
Light soils (8.3 t/ha)	109	107	102	102	96	4.3
Heavy soils (8.7 t/ha)	107	105	103	102	96	4.7
Agronomic characteristics						
Lodging % without PGR	4	0	1	1	0	
Lodging % with PGR	4	2	2	2	2	

1.2 The current understanding of complex traits in barley.

1.2.1 The genetic architecture of complex traits.

Quantitative traits and Quantitative Trait Locus.

The majority of traits under selection in crops, and barley in particular are quantitative. The phenotypic variation for quantitative traits follows the curve of a normal distribution in which the phenotype measured quantitatively can take any continuous value showing no clear discontinuity (e.g. yield, height). This quantitative distribution is the observed result of multiple gene effects that segregate in a Mendelian manner and can be resolved genetically in mapping experiments of segregating in populations (Lander and Botstein, 1989). Biometrical genetics aims at identifying and locating those genes responsible for part of the trait variation in order to gain a better understanding of the genetic control of the quantitative trait. It mainly revolves around the identification and analysis of quantitative trait locus which describe the association of a chromosome segment containing polymorphic markers with the phenotypic variation (Collard et al., 2005). The developments of efficient genotyping technologies (SNP, multiplexed genotyping array) have made QTL mapping an accessible and reliable tool for plant genetic research and crop science (Bernardo, 2008; Mackay et al., 2009). A detailed description of QTL analysis will be made in paragraph 1.3. Although a better insight of the genetic basis of quantitative traits gained from QTL studies, most of the genetic architecture underpinning quantitative traits in barley, is yet to be uncovered (Bernardo, 2008).

Genetic architecture of complex traits.

QTL mapping studies have enabled a greater insight into the genetic architecture of quantitative traits by identifying genes responsible for phenotypic variation, their number, location, allelic effects and interactions in particular genetic backgrounds and environmental conditions. The genetic complexity underlying quantitative traits revealed the strong influence of non-additive effects including epistasis, genotype by environment interactions and pleiotropy on phenotypic variation (Cooper et al., 2009). Such networks of interactions reduce the rate at which QTL results can be exploited and incorporated in breeding programmes, as in most cases the desired alleles at a QTL will be specific to a genetic background and environmental conditions. In a recent review, Mackay et al. (2009) pointed out that QTL studies have been successful in identifying a large number of gene polymorphisms with small effects on phenotype. The pleiotropic

effects, which refer to a particular allele (or gene) affecting more than one phenotypic trait, have been found to occur between traits not known to be functionally related. The role of epistatic interactions with background loci also needs to be considered in the genetic architecture of traits as they affect the detection and estimation of the QTL effects and increase the difficulty of selecting best alleles or combination of alleles in complex gene networks (Cheverud and Routman, 1995; von Korff et al., 2010). These QTL interactions and pleiotropic effects suggest that the genetic control of quantitative traits is substantially more complex than a simple additive model, and the attempts at QTL modelling, integration of QTL results and trait dissection, can provide additional insights on the genetic architecture of traits (Cooper et al., 2009; Emebiri, 2013; van Eeuwijk et al., 2010). Studies using high marker densities, accurate phenotypes and advanced QTL mapping methods also offer the prospects of describing genetic architecture and predicting the future evolution and variation of phenotypes (Mackay et al., 2009; van Eeuwijk et al., 2010). For example the association mapping method using a large population size has the ability to detect a large number of loci with small effects QTL that can help in fine tuning QTL position and give a better picture of the complexity of genetic control (Ingvarsson and Street, 2011).

A first reduction of the level of complexity of the trait consists of looking at its underlying components expected to be under simpler genetic controls. In barley this approach showed that QTL for yield were often associated with the QTL of the components of yield: tillering, grains per ear and grain weight (Yin et al., 2002). The yield component QTL however identified additional locations on the genome that had not been associated with yield suggesting that trait dissection revealed increased complexity of yield control. Other strategies to link genes with phenotype involve the characterisation of mutagenized populations (mutant plants) or TILLING populations to screen for specific phenotypes generated by mutations have been developed in barley (Druka et al., 2010; Rossini et al., 2006). The extent of understanding of the genetic architecture of traits is limited by the ability to identify and delve into their multiple components and interactions. The physiological models and genotype by environment by management interactions have to be considered to evaluate QTL in a particular system (Cooper et al., 2009; van Eeuwijk et al., 2010).

1.2.2 Yield and yield components.

The main trait driving genetic progress in barley and other cereals is the harvested grain yield, a key trait in selection. This complex trait is the result of agronomical, structural

and physiological factors that will define the yield potential of the plant based on source and sink traits (Bingham et al., 2007b; Reynolds et al., 2011). In a more simplistic manner, yield can be described as the product of yield components which are the number of ears per land area, the final number of grains per ear and the average weight of a grain (Yin et al., 2002).

QTL mapping studies for yield, yield components and agronomic traits are well documented in barley, using mapping populations made from selected parents differing in the traits of interest. An example is the double haploid (DH) population derived from the two-row/six-row barley cross Harrington/Morex that revealed the importance of the genes *vrs1* and *int-c* in the control of inflorescence row type and other agronomic traits (Ayoub et al., 2002; Marquez-Cedillo et al., 2001). Bezant et al., (1997) reported QTL for agronomic and quality traits as well as plant responses to biotic and abiotic stresses in a study using a DH population derived from two two-row spring barleys (Blenheim x Kym). Other QTL studies for yield and yield components have been reported for multiple effects across the barley genome (Inostroza et al., 2009; Rae et al., 2007).

The environmental conditions have a strong influence on the mapping precision and the estimation of QTL effects and need to be considered in the phase of interpretation (Cooper et al., 2009). For the yield and yield components, environmental variation can explain the low replicability of the QTL results between studies (Li et al., 2005; Saal et al., 2011; Schnaithmann and Pillen, 2013; Xue et al., 2010). Better consistency between QTL mapping results was found when effects were detected at major genes such as those controlling phenology and morphological traits, also reported as strong candidate QTL for yield (Comadran et al., 2011b; Cuesta-Marcos et al., 2009; Kraakman et al., 2004). Nevertheless, very few reports of QTL for yield and yield components on elite material grown in optimal conditions have been presented and may not be reported by commercial breeding research. These QTL have a genuine value to breeders aiming at rapid genetic progress in environmental conditions similar to those of current crop production.

Tillering.

Tillering in small grain cereals is an essential yield component that affects the variation of number of ears/m^2 and can provides compensatory mechanisms to overcome poor plant establishment. The trait is affected by tillering ability and tillering survival. The genetic control of branching in grass plants is made up of a complex network of regulatory pathways, hormones and structural genes that affect the activity of apical and

axillary meristems to set the basis of vegetative architecture (Doust, 2007; Kebrom et al., 2013). Studies on mutant barley lines have identified genes that could affect axillary meristem development that produce low tillering phenotypes: *low number of tillers1 (lnt1), absent lower laterals1 (als1), intermedium-b (int-b), uniculm2 (cul2), uniculm4 (cul4),* and *uzu* (Hussien et al., 2014). Transcriptome analysis suggests that *Als1* expression is required for secondary tiller development, linked to stress response and function independently to *Lnt1* (Dabbert et al., 2009; Dabbert et al., 2010). Babb and Muehlbauer (2003) showed that the *cul2* gene on barley chromosome 6H is necessary to initiate the development of axillary meristems. Other mutations repressing axillary meristem development have also been described and include *densonidosum6 (den6), granum-a (gra-a), intermedium spike-m (int-m),* and *many noded dwarf1 (mnd1)* (Dabbert et al., 2010). The location of tillering QTL also suggests the presence of strong pleiotropic effects from major genes acting on compensatory mechanisms such as changes in plant architecture (Comadran et al., 2011b; Hussien et al., 2014). Additionally, the genetic regulation of the duration of pre-anthesis developmental phases can modify the development of meristems and final tiller number (Borràs-Gelonch et al., 2011). As a main yield component, tillering shows strong underlying complexity influenced by multiple interacting factors, both genetic and environmental.

Grain number per ear.

Despite the evident contribution of grain number per ear in the overall yield figure, especially in the six-row/two-row types, this yield component has received limited attention from the research community. The main factor affecting the grain number per ear in barley is the fertility of the lateral spikelets on an ear. The recessive allele at the gene *vrs1* has been identified as responsible for the six-row spike morphology (Komatsuda et al., 2007). Detailed molecular analysis of the gene has revealed that *vrs1* is a HD-ZIP I-class homeobox gene located on chromosome 2H and expressed in spikelet primordia. Loss of function and homozygozity for the *vrs1* gene is sufficient to give a complete six-row spike that was selected during domestication. Recently Ramsay et al., (2011) showed that allelic variation at *int-c* (*intermedium-c*), an ortholog of the maize domestication gene *teosinte branched 1*, could modify the phenotype of *vrs1* by affecting lateral grain plumpness and fertility gradients. Many QTL mapping studies have been able to locate the strong effects of these two genes (Comadran et al., 2011b; Kjaer and Jensen, 1996). The loci for kernel number per spike showed strong pleiotropic effect on most agronomic traits in a population from a cross between two-

row and six-row parents (Marquez-Cedillo et al., 2001). Other genes have been described with modification effects on the positioning and extent of fertility in the spikelets (Koppolu et al., 2013).

Additional controls of inflorescence architecture has been observed in cereals where some of the control of meristem differentiations and branching structure are shared between species (Tanaka et al., 2014). These may be relevant to understand the genetic control of grain number per ear independently from genes involved in lateral floret fertility, by instead looking at the spike elongation within the two-row and six-row groups. Although this trait is less documented in barley, a higher number of grains per ear would come from the initiation of additional spikelets along the rachis. Such variation could correspond to QTL with smaller effects found in mapping analysis of grains per ear. It could be expected that the genetic control of meristem fate and development time would affect the spikelet number. In addition, comparable allele effects may be observed between two-row and six-row types leaving aside the effect of the genes for lateral spikelet fertility.

Thousand grain weight.

A third yield component and major contributor to yield and yield variation is the average weight of a grain that is often measured as thousand grain weight (TGW). It is a desirable trait in breeding in order to increase yield and seed viability. TGW is relatively simple to measure and can be further described by grain length, width and thickness which are under control of different genetic and physiological factors (Breseghello and Sorrells, 2006). The grain is a sink organ in the plant and high TGW potential depends also on the plant's ability to achieve optimal grain shape and starch accumulation. Grain filling in small grain cereals occurs after anthesis when the starch granules accumulate within the protein matrix of the grain. In rice, QTL mapping studies have identified the loci controlling grain size traits. Fan et al. (2006) reported the locus for *GS3* on chromosome 3 with a strong effect on grain length and weight. *GW2* on chromosome 2 was found to associate with significant variation in TGW (Oh et al., 2010). Polymorphisms in the wheat ortholog candidate *TaGW2* on chromosome 6A was associated with significantly wider grains and TGW (Su et al., 2011). Barley TGW QTL have been identified in various mapping populations (Li et al., 2005; Saal et al., 2011). Schmalenbach et al (2009) validated QTL for TGW on chromosomes 2H, 4H and 6H using a population of lines containing introgressed chromosome segments of wild barley. Understanding the gene effects on the traits will help in differentiating the

genuine alleles impacting on grain weight and the extent of pleiotropic effects from other loci. For instance, many association mapping studies have highlighted significant effects at the genes controlling ear morphology and phenology on TGW (Comadran et al., 2011b; Pasam et al., 2012). Phenology genes may affect the duration of photosynthetic activity and therefore the quantity of photosynthetates available for sink organs. In addition, the greater number of grains in six-row barley may increase the sink size. However the controlling gene *vrs1* on 2H was associated with a reduction in grain size and TGW as lateral grains are generally smaller grains (Ayoub et al., 2002). Additionally, the control of the duration of the grain filling period can impact on TGW and consequently yield (Laurie, 1997). Delayed senescence has been put forward as a trait to maintain longer photosynthetic activity in source organs and lengthen the remobilisation period during grain filling (Gregersen et al., 2008; Parrott et al., 2010; Verma et al., 2004). Therefore, the investigation of candidates genes underlying TGW and other yield component QTL should consider physiological aspects such as photosynthesis, radiation use efficiency and nutrient use efficiency especially in (Reynolds et al., 2011).

1.2.3 Genetic control of winter habit

The control of growth habit and flowing time in barley provides plant adaptation and distribution across environments and farming practises (Cockram et al., 2007). The winter growth habit takes advantage of the autumn season to establish before winter and requires a period of cold (vernalization) to initiate the reproductive growth. The genetic control of growth habit in barley can be attributed to loci of major effects on photoperiod and vernalization responses (Laurie, 1997), some of them being homologous to other cereals (Faure et al., 2007; Griffiths et al., 2003).The winter growth habit is determined by the vernalization pathway involving three loci *vrn-H1* (5HL), *vrn-H2* (4HL) and *vrn-H3* (7H). The *vrn-H1* gene is a MADS-box transcription factor that promotes the transition from the vegetative state to the reproductive state (Yan et al., 2003). *Vrn-H2* gene is a zinc-finger CONSTANS, CO-like and TOC1 (CCT)-domain protein that represses flowering in plants that have not been vernalized (Yan et al., 2004). The determination of vernalization sensitivity involve the gene x gene interaction (epistatic interaction) between *vrn-H1* and *vrn-H2* that is influenced by intron length variation in *vrn-H1* (Szucs et al., 2007).

The winter vernalization sensitive varieties are often associated with the sensitivity to photoperiod and the detection of long days that enable to adjust plant phenology. The

ppd-H1 gene (2HS) is a pseudo-response regulator 7 (prr7) that encode for a component of the circadian clock with a dominant allele that promotes flowering in both winter-and spring-sown plants (Turner et al., 2005). The gene is located at a chromosome segment homologous to a junction of rice chromosomes (Dunford et al., 2002). The *ppd-H2* gene (1HL) is homologous to Arabidopsis FT and paralogue to *vrn-H3* (Yan et al., 2006). It encodes for PEBP and promotes flowering under short days. The interconnection of both vernalization and photoperiod pathways has been shown as the *Vrn-H2* is a repressor of *ppd-H2* (Casao et al., 2011). Other loci involved in the genetic control of heading date have been identified in QTL mapping studies (Comadran et al., 2012; Ren et al., 2012) suggesting that the photoperiod genes *ppd-H1* and *ppd-H2* are not the only factors involved in providing environmental adaptation to winter barley. The genetic control of phenology is especially relevant as photoperiod and vernalization have frequently been associated to QTL for agronomic traits of yield and yield components (Cuesta-Marcos et al., 2009; Li et al., 2005; Schmalenbach et al., 2009; Wang et al., 2010).

1.2.4 Biotic and abiotic stresses.

Specific environmental conditions can cause plant stresses that lead to a strongly negative impact on yield potential. However, the allelic make up of a variety can confer adaptations to environments and enable it to maximise the yield potential under conditions otherwise stressful for less adapted germplasm. Both biotic and abiotic stresses generate plant responses through the triggering of multiple genes acting in different metabolic pathways and interactions (Atkinson and Urwin, 2012). Many of these genes may well be influential yield-related genes in some environments.

The negative impact of living organisms such as insects, fungi, viruses and bacteria on plants are referred to as biotic stress. In barley in the UK, biotic stresses are mainly observed in the context of diseases from fungal organisms that damage organs of the plant and impact resource capture. A number of resistance genes have been identified for powdery-mildew amongst which *mlo* and *Mlg* mediate plant resistances in a complete gene network (Aghnoum et al., 2009; Miklis et al., 2007). The allelic diversity found in landraces and wild barley can be exploited to introduce resistances to leaf rusts such as rynchosporium (Ellis et al., 2000; von Korff et al., 2005). Stein et al. (2005) identified diagnostic SNP polymorphisms in *Hv-eIF4E* that confer the resistances rym4 and rym5 to barley yellow mosaic virus. These resistances are present and selected in

UK germplasm. Alternatively mutated material offers another source for identifying resistances that can be introgressed in breeding programmes (Druka et al., 2010).

Crop performance is also affected by abiotic stresses such as nutrient stresses, waterlogging, drought, and harsh or toxic growing environments (Long et al., 2013; Passarella et al., 2005). Within species, genotypes vary in their ability to cope with or avoid abiotic stresses suggesting that allelic diversity is available for adaptation to environmental modifications such as the increase frequency in hotter and drier summers. For instance, the adjustment of plant phenology was shown to be an adaptive advantage in environments prone to drought (Faure et al., 2007; Foulkes et al., 2004) such as the *eam6* locus on 2H which provide environmental adaptation to mediterranean growing conditions and has pleiotropic effects on yield and yield components (Comadran et al., 2011b). Breeding for improved agronomics in stressful environments is possible as salt tolerance in barley was associated with a major effect QTL on the centromere region of 6H where a number of candidate genes involved in physiological pathways have been proposed (Long et al., 2013). Despite the rare occurrence of extreme stresses in the UK environment, the genetic progress for those traits should not be disregarded by breeding programmes and is needed for sustainable production of barley with maximal and consistent yields.

1.2.5 Resource use efficiency.

The challenge for breeding improved crop varieties is to maintain or increase current production while minimizing the impact on the environment. Although the specific analysis of traits for resource use efficiency is marginal in this project, it is a recurrent theme in studies on yield and yield components. Resource use efficiency, and in particular nitrogen (N) efficiency, has been identified as a main attribute of sustainable and high yielding crop production systems (Good et al., 2004; Raun and Johnson, 1999). The limited availability of these nutrient resources can cause abiotic stresses and interact with the yield and yield components (see 1.2.4). Efficient management of resources in agricultural production is essential to maximize farm profitability and minimise environmental damage (Hirel et al., 2007; Raun et al., 2002; Sylvester-Bradley and Kindred, 2009). For example, excessive levels of N can lead to ammonia volatilisation, denitrification, leaching, ammonium fixation, immobilisation and runoff which is detrimental to soil, air and water quality, while also increasing lodging which can reduce yield in cereal production. Sylvester-Bradley and Kindred (2009) proposed that the economic justification for nitrogen input is set by the value of an optimal N

input (N-opt) that takes into consideration the benefit for a grower of yield in an economic context of fertiliser costs and income from grain production. The study also highlighted that improving nitrogen use efficiency (NUE) in crops should be done with the aim of maintaining current yield levels and acknowledged that plant breeding is a method to deliver enhanced NUE in crops (Hirel et al., 2007; Muurinen et al., 2006). Therefore, the optimum use of fertiliser on cereal crops requires a rigorous management strategy determined by agronomic and economic factors. In the UK, the RB209 manual aims at guiding farmers for optimal crop fertilisation practices (DEFRA, 2010).

NUE-CROPS project.

With increasing concerns about issues related to food production and agricultural sustainability, nitrogen economy of crops has become a major field of investigation for academic and private institutions. A large scale European project NUE-CROPS was launched to investigate crop response to fertilization and genotypic variation in NUE, and promote a better transfer of research knowledge into commercial applications. Led by Newcastle University, this five year project was part of the FP-7 European framework and finished in 2014. It aimed at investigating NUE in five major cultivated crops: barley, wheat, maize, oilseed rape and potatoes. The project used cutting-edge technologies and worked directly with the crop breeding industry to identify tools to improve breeding for NUE and agronomic strategies to further enhance system-level NUE (http://research.ncl.ac.uk/nefg/nuecrops/page.php).

This PhD project was set up in the context of NUE-CROPS Work package 1 that dealt with NUE in barley and wheat with a specific focus on barley. Barley was used as a model crop in an association genetics experiment carried out over two years, in five geographical locations and at three rates of nitrogen application. The project aimed at quantifying and describing the allelic variation in European germplasm underpinning NUE components, yield and yield components. Further details on that work package can be found in project reports (Thomas et al., 2013).

Nitrogen use efficiency.

NUE is a quantitative trait of increased interest in plant breeding. In a study on maize, Moll et al, (1982) identified two main components for NUE, the nitrogen capture or uptake efficiency (NupE= Ncrop/Na) and nitrogen conversion or utilization efficiency (NutE= Yld/Ncrop); where Yld is the total marketable crop yield (kg ha^{-1}), Ncrop (kg ha^{-1}) is the total crop nitrogen uptake at harvest and Na (kg ha^{-1}) is the available

nitrogen during growth period. Additional components and adjustment can account for plant nitrogen loss, plant nitrogen content at different growth stages and nitrogen remobilization during the senescing period (Gregersen et al., 2008). Other approaches for describing NUE have been proposed excluding the productivity factor or uptake efficiency factor e.g. NUE= Ncrop-Nsoil/(Nfert) (Hirel et al., 2001; Raun et al., 2002). A possible consensus is the productive value of nitrogen defined by NUE= Yld/Na, which is also a scaling of yield.

Genetic progress in crop NUE has been achieved by the exploitation of genetic variation for NupE and NutE within species (Abeledo et al., 2008; Muurinen et al., 2006). A large number of genes involved in regulatory mechanisms in nitrogen metabolism can play a role in the observed variation in NUE (Good et al., 2004; Kant et al., 2010). For example, variable expression of different glutamine synthetase isoenzymes was shown to affect kernel size and number in maize (Gallais and Hirel, 2004; Hirel et al., 2007). Differences in NUE, NupE and NutE have also been reported in barley (Beatty et al., 2010; Le Gouis et al., 1999; Sylvester-Bradley and Kindred, 2009) with significant QTL × nitrogen interactions affecting yield and yield components (Saal et al., 2011). In a study on spring barley, Beatty et al., (2010) showed that the variation in NUE was mostly accounted for by variation in NutE suggesting that varieties differed in their ability to transfer leaf and stem nitrogen to the grain. In addition to the physiological aspects of NUE, it can be assumed that the variation in roots and canopy architecture play a leading role in the trait variation (Hirel et al., 2007). A breeding hypothesis suggests that the winter barley varieties Saffron and Retriever have contrasting abilities to cope with nitrogen stress which could be due to differences in both rooting abilities and physiology (David Harrap pers. comm.). Therefore, an improved genetic understanding of NUE and an estimation of the existing diversity would benefit breeding to tailor elite varieties to specific fertilisation regimes.

1.3 The genetic mapping of quantitative traits.

Recent advances in genetics and biotechnology have drastically increased progress in crops sciences and plant breeding methods. The development of genetic markers and high throughput genotyping platforms are now routinely used to characterise breeding material. One main application is the study of traits using QTL mapping to focus on

genomic regions with underlying genes and polymorphism in order to identify and track down any desired allelic variant for a faster varietal improvement.

1.3.1 Molecular markers and genotyping technologies.

Genetic markers.

Genetic markers have become a major tool for plant sciences and breeding. A genetic marker refers to a polymorphism in the DNA sequence of individuals in the form of a different nucleotide or sequence of nucleotides that can be linked to a trait variation and is unaffected by the environment (Collard et al., 2005). The description of individual genotypes by marker alleles enables their fingerprinting. All genetic markers can be tested for linkage relationships and recombination frequency between alleles in segregating populations in order to create genetic maps using mapping functions (Zhao and Speed, 1996).

The advantages and disadvantages of genetic markers depend on their proprieties that reflect high polymorphism, abundance in the genome, dominant or co-dominant inheritance, cost and the flexibility of assessments and assays (Collard et al., 2005). The types of genetic markers are described in reference to their method of detection and the sequence plolymorphism. Most genetic markers such as AFLP (amplified fragment length polymorphism), SSR (single sequence repeat or microsatellites) and SNP (single nucleotide polymorphism) are based on polymerase chain reaction (PCR). A SNP marker corresponds to a single base-pair change in the DNA sequence which generates two alleles. The extreme abundance of SNP in the genome and the possibility for multiplexing on high throughput arrays has driven the cost reduction and attractiveness of genotyping. SNP have greater scope to describe and locate allelic variation with QTL mapping (Close et al., 2009; Waugh et al., 2009), track further the desired alleles with MAS (Collard and Mackill, 2008) and develop advanced breeding methods of genomic selection (Jannink et al., 2010).

Barley genotyping technology.

The SNP multiplexing technology pioneered by Illumina in collaboration with the barley research community has enabled the development of genotyping platforms for barley that cover the genome with a high density of markers in a cost effective manner (Close et al., 2009). The Illumina beadXpress that contained 384 SNP from all seven barley chromosomes (Bx384) was commonly used in the breeding industry. The increased marker density was achieved with OPA1 and OPA2 SNP assays using a larger

scale platform with each chip containing 1536 SNP that covered the barley genome with an average resolution of 1 SNP per cM based on OPA consensus maps (Close et al., 2009). More recently, the Illumina infinium genotyping assay comprises 9000 SNP of which 2832 are covered by OPA assays. This platform has already been used in barley mapping studies (Comadran et al., 2012) and was available as a genotyping resource for the NUE-CROPs project. The broader implementation of genotyping by sequencing technology (GBS) is also being considered for barley in order to combine low cost genotypes and high marker density (Mascher et al., 2013a) that can then be referenced to the barley physical and genetic map framework (Mayer et al., 2012). Other advantages for SNP markers are found in modern genotyping arrays and technologies. For example, the KBiosciences' KASPar genotyping platform can be designed to target any individual SNP such as those available on other platforms or from DNA sequence information (www.lgcgenomics.com). This technology is now used routinely in barley breeding for MAS. The barley genetic resources generated from genotyping are accessible via public databases such as Germinate (ics.hutton.ac.uk/germinate), Graingenes (wheat.pw.usda.gov/GG2/index.shtml) and also Ensemble plants (plants.ensembl.org).

1.3.2 Linkage QTL mapping.

The mapping of QTL is a core activity used for understanding the genetic basis of quantitative phenotypic variation (Rae et al., 2007). The basic principle behind QTL mapping is a regression between phenotype and explanatory variables with the assumption that traits can be understood by linear additive models (Bernardo, 2008). It is a test for significant associations between the genotype marker classes and the phenotypic variation between groups of individuals partitioned according to marker classes (alleles) and the proportion of the variance accounted for by those classes (Collard et al., 2005). The underlying hypothesis is that a trait is controlled by multiple genes in linkage with classes of genetic predictors. The significant QTL is positioned on the chromosome at its highest result of association with a confidence interval assigned (support interval).

The type of population used for mapping purposes varies depending on the trait's complexity and the strategy adopted to investigate the trait. In cereals, and barley in particular, the large majority of studies have used DH populations derived from a bi-parental cross of lines with contrasting phenotypes (Bezant et al., 1997; Borràs-Gelonch et al., 2011; Hayes et al., 1993; Xue et al., 2010). In most studies, the DH population

size varies between 100 and 200 individuals. Alternatively, advanced backcross QTL mapping (AB-QTL mapping) has mainly been done to investigate the effects of chromosome introgressions from landraces and exotic material into elite lines (von Korff et al., 2010). The segregation obtained in recombinant inbred line (RIL) populations derived from the segregating lines is also used for mapping (Liu et al., 2010).

In order to position the association results on a genetic map, preliminary analysis of the recombination frequency between markers is used to identify linkage groups of genetic predictors (markers). The first QTL mapping methods were based on single marker analysis (SMA) to test the association of single marker alleles with the phenotypic distribution using statistical methods such as *t*-test, ANOVA, and linear regression (Collard et al., 2005). More elaborate mapping algorithms have been proposed to exploit linkage maps and the interval between markers pairs. Lander and Bostein (1989) introduced the simple interval mapping (SIM) method using RFLP markers and estimated genetic predictors in-between mapped markers. Composite interval mapping (CIM) combines the linear regression approach of interval mapping and inclusion of additional cofactors (markers) in the model which adds more precision in positioning the QTL and tests with adequacy for residual effects across the rest of the genome (Zeng, 1994). Both SIM and CIM are implemented in statistical packages such as R/qtl (www.rqtl.org/) and GenStat 14th (Payne *et al.*, 2009).

It needs to be born in mind that QTL mapping based on a bi-parental population also has limitations such as the set number of alleles segregating in the population (Collard et al., 2005). The alleles segregating in the populations will be those of the crossing parents and may only represent a small proportion of the allelic diversity in an elite crossing programme. This diversity is also minimal in comparison to the range of alleles that can be accessed by breeders in the wider barley germplasm (Comadran et al., 2009; Ellis et al., 2000). The number of recombinations between chromosomes is another limiting factor associated with bi-parental mapping, especially for identifying precise QTL positions. An increase in population size (i.e. the amount of recombination events sampled) can help to narrow the QTL support interval. The cloning of *vrs1* involved 9,831 gametes to identify appropriate recombination events around the gene (Komatsuda et al., 2007). The downside of low QTL resolution is that markers in association with the trait span a significantly large chromosome segment that can contain numerous polymorphic gene candidates and can contribute to a linkage drag of undesired alleles with negative effects on crop performance. Nevertheless the bi-

parental QTL mapping approach still has a value when considered in the context of large scale breeding programmes where large numbers of segregating families are generated from sets of common parents closely related (Würschum, 2012).

1.3.3 Genome wide association mapping.

The key principle in genome wide association studies (GWAS) is to use the linkage disequilibrium (LD) between a trait and the alleles of a genetic marker to map the significant effects of the trait along the chromosomes (Gupta et al., 2005). LD is the non-random association of alleles at separate loci on a chromosome that can be caused by structure in populations, genetic drift, relatedness between individuals and the selection process. Rapid LD decay is observed when there is only a short distance between pairs of markers in LD suggesting frequent recombination events in the population. In barley, the LD was shown to extend up to 10cM (Kraakman et al., 2004) but generally declined rapidly after 2.6 cM (Zhang et al., 2009). The breeding processes aim at maintaining favourable alleles in the germplasm and populations exploited by selecting positive alleles in LD with linked marker loci. This genotypic information can be used to gain mapping resolution compared to simple linkage analysis (1.3.2). Three main advantages of GWAS have been advanced: an increased resolution of mapping, the reduced research time and the greater diversity investigated (Zhu et al., 2008). With high throughput genotyping technologies and efficient statistical modelling, GWAS is a key method for studying genetic architecture of quantitative traits and identifying valuable polymorphisms in a panel of diverse individuals.

One main obstacle to GWAS is describing and accounting for confounding effects caused by population structure present in the panel used. The geographical origins and relatedness of lines that compose the panel of individuals (e.g. varieties) tested in GWAS create stratification in the population that can be captured by genetic markers (Comadran et al., 2009). In a panel of 329 lines restricted to six-row winter barley, the hull and hulless traits were clearly separated using marker information (Berger et al., 2012). This complex population structure causes Type-II errors (false negatives) in association tests which lead to significant marker-trait associations due to shared pedigree rather than true genetic linkage. To assess the population structure and adjust for it in GWAS scans, a range of statistical methods have been developed that exploit the random background genotypic information (Price et al., 2010). Structured association (SA), genomic control (GC), principal components (PC) and mixed linear

models (MLM) can be used to account for population structure and test for appropriate marker-trait (i.e on the residuals of phenotypes after structure correction).

Structured association (SA) consists of inferring a population structure from the genotype information and a clustering statistical model implemented in the software STRUCTURE (Pritchard et al., 2000). This approach assigns individuals to populations (*k* parameter) by minimizing disequilibrium within them. The main output of STRUCTURE is a matrix Q with *k* vectors reporting the estimates of population membership for each individual. SA can be time consuming and inconsistent in some cases (Cockram et al., 2008). Devlin and Roeder (1999) proposed a population structure correction by genomic control (GC) in adjusting uniformly across the genome for the inflation of statistics caused by this structure. This method assumes that the structure has the same effect on all loci and tends to decrease the power of detection of associations. The principal component (PC) or EIGENSTRAT analysis is also used for estimating structure in GWAS. In general, the PCs are estimated from the genetic markers and the loadings of PCs included as covariate in the GWA models. These loadings can be interpreted as a proportion of subpopulation membership and tend to reflect family relatedness, long range LD and assay artefacts (Price et al., 2010). The flexibility found with MLM allows adequate accounting for multiple levels of relatedness in population structure in plants (Yu et al., 2006). In MLM, the genetic similarities between individuals can be included in the model as both fixed and random effects and the different combinations evaluated to better account for the levels of relatedness in mapping panels (Wang et al., 2012). A kinship matrix K computed from the marker data reports the degrees of covariance between pairs of individuals interpreted as population structure. The combination of different cofactors accounting for structure (e.g. Q+K) was shown to improve mapping power (Yu et al., 2006). In the case of autogamous species like wheat and barley, the appropriate K matrix may yield very acceptable results (Stich et al., 2008; Wang et al., 2012). MLM are nowadays computationally accessible thought web resources such as TASSEL (Bradbury et al., 2007) and EMMAX (Kang et al., 2010).

Barley has been a model species for GWAS in small grain cereal and was used in major projects such as AGOUEB and Barley-CAP (Waugh et al., 2009). Large diversity panels comprising commercial varieties and landraces have been assembled and exploited in GWAS (Close et al., 2009; Comadran et al., 2011a; Haseneyer et al., 2010). Cockram et al. (2008) showed that after correction for population structure, association mapping was able to detect the partitioning loci of vernalization genes *vrn-H1* and *vrn-*

H2 (Cockram et al., 2008). The interactions between genes involved in the control of phenology was also reported by GWAS (Stracke et al., 2009). Other barley GWAS reports have investigated agronomic traits involved in yield and yield stability (Comadran et al., 2011b; Kraakman et al., 2004; Rostoks et al., 2006), stress tolerance (Long et al., 2013; Visioni et al., 2013) and simplier morphological traits (Cockram et al., 2010). These positive results have shown the potential of GWAS to exploit genetic information for progress in breeding.

1.4 Achieving genetic progress

1.4.1 Phenotypes

As genotypic information because more accessible, the current challenge for QTL discovery is the bottleneck of achieving sufficient and accurate phenotype information (Furbank and Tester, 2011). The large panel of accessions required for GWAS can result in phenotyping becoming a costly and time consuming process. This step cannot be overlooked since it has a much greater effect on the power of detection of associations than the number of data points from genetic markers (Ingvarsson and Street, 2011). Since the aim of GWAS is to investigate the genetic control of a trait, the ideal phenotyping approach should aim at maximising the proportion of the trait variance due to the genetic component while reducing experimental error and environmental variation. One main strategy is to include replication of accessions to better partition the component of variance in the phenotype in order to get a measure of the error variance and an accurate estimate or prediction of the mean value for the accessions. This is routinely implemented in plant breeding and variety testing with replicated trial networks under varied field conditions. The attractiveness of phenomic tools may also offer an alternative option for plant scientists to narrow down the distance between phenotypes and genotypes (Furbank and Tester, 2011). Furthermore, the adequate statistical analysis of phenotypic data collected in replicated trials and phenotyping experiments is essential to predict robust breeding values. Often, these are obtained using mixed model procedures which assign appropriate values to the components of variance (Piepho et al., 2008). Breeding values can then be used for selection purposes and included in the statistical models to test for marker-trait associations in a two-stage mapping approach (Stich et al., 2008) and is predominantly applied in barley and plant GWAS.

1.4.2 Marker assisted breeding

Marker assisted selection

The intention of marker assisted selection (MAS) is to use genetic markers as a tool in conventional selection in order to screen for alleles associated with a QTL for a trait of interest. In other words, MAS aims to increase the efficiency and effectiveness of breeding over conventional methods. Amongst its advantages, MAS allows the screening of a large panel of individuals in the early stages of the selection process, whilst reducing resources allocated to phenotyping activities (Collard and Mackill, 2008). MAS is also unaffected by environmental conditions, reproducible over a range of material and years and has high heritability. Last but not least, it can be used to apply marker technology in order to describe genetic background. This is particularly useful in strategies of marker assisted backcrossing to select individuals containing both an introgressed allele at a QTL target and the maximal genetic background of the recurrent parent (Kandemir et al., 2000).

It needs to be born in mind that the factors affecting QTL mapping accuracy such as the mapping methods, population size, nature of genetic markers and GxE interactions will affect the extent of the realized utility of the linkage between the markers and QTL (Asíns, 2002). These factors need to be considered for MAS applications and a marker development pipeline should be implemented from the QTL mapping experiment to the validation of QTL effects (Collard and Mackill, 2008). Indeed, the validation of marker-trait associations is necessary to exclude rare events of double recombination and discrepancy between genetic and physical map distances. The QTL effects should also be confirmed in replicated experiments and both QTL and markers validated in relevant germplasm. A QTL in different genetic backgrounds may be subject to epistatic interactions. In some cases, the conversion of a marker on a different platform will speed up the screening process (e.g. KASPar technology). These steps downstream from the "QTL mapping" are necessary in order to integrate and exploit the results of molecular genetic research into conventional breeding by effective MAS.

In breeding programmes, MAS can be applied to increase the breeding value of the lines that are continued and screen larger panels with more stringency in order to increase the frequency of desired alleles in the following generations. Pairs of flanking markers can be used to track chromosome segments containing the relevant QTL alleles. A combination of MAS and phenotype screening can increase the success rate of having the favourable alleles and also identify useful cases of recombination between marker

and phenotype. The ideal situation for MAS is a “perfect marker” or “diagnostic marker” that describes the sequence polymorphism responsible for the trait variation. Marker screenings for disease resistance alleles are perfect examples for the rarely reported MAS applications in plants, often because those are diagnostic markers (Miklis et al., 2007; Stein et al., 2005). This is the case for barley yellow mosaic virus for which the actual change in nucleotide sequence in the gene that causes the resistance is used as a marker (Rae et al., 2007). A sequence deletion in *ant-2* can also be used as a marker to check the barley pigmentation (Cockram et al., 2010). In rice, some markers have been tagged to genes directly involved in yield and yield components variation (Yan et al., 2009). In barley, QTL for complex traits are yet to be characterised and efforts are being made to find and locate the polymorphisms strongly associated with markers that can be implemented in MAS.

QTL validation with near isogenic lines

The validation of a QTL effect is required if one wants to exclude potential mapping errors and use markers for selection. The main validation step involves a test of the consistency of significant QTL effects in different environments and genetic backgrounds in the plant material which is expected to carry the desired alleles of the QTL.

Near isogenic lines (NILs) have been advanced as a tool for QTL validation. NILs enable the reduction of the phenotypic variation caused by the environment and the genetic background so that the effect observed is principally due to the QTL (Kandemir et al., 2000; Navara and Smith, 2013). Ideally, NILs share an identical genome (i.e. identical alleles and genes) with the exception of a specific segment located at the putative QTL. When tested under the same conditions, phenotypic differences can provide strong evidence for a genuine QTL effect. NILs can also help to locate genomic regions involved in the control of agronomic traits without any assumptions about QTL (Venuprasad et al., 2011). Generally NILs can be obtained by backcrossing an allele of a QTL into a recurrent parent from the mapping study or into a conventional variety (Kandemir et al., 2000; Kongprakhon et al., 2009).

The validation of a QTL using the original mapping population is also possible (Yun et al., 2006) although this approach limits the transferability of the QTL effect to a wider germplasm pool. Alternatively, the development of heterogeneous inbred families (HIF) can be used to test and validate QTL (Tuinstra et al., 1997). The material in the higher generations of breeding programmes has low heterozygosity but the residual

segregation at a QTL location can be exploited to identify lines from which HIF and NILs can be derived. This strategy has the double advantage of validating QTL and alleles in a genetic pool of direct relevance to commercial breeding purposes. Along with QTL validation, the markers segregating in the NILs in close linkage to the QTL can also be used for MAS. The need to describe the causal gene is less essential for that process. Ultimately, the better understanding of the genetic control of traits can only be achieved if the polymorphic genes are described and diagnostic markers identified.

1.4.3 Finding the genes

Even though the scientific approaches to characterise diagnostic markers seem easily accessible, the actual number of this type of markers used by breeders remains low. The diagnostic markers are essential for accurate MAS as they eliminate the risks of recombination between markers and the causal polymorphism. In addition, the description of the polymorphic gene by its gene sequence can help to elucidate both genetic and physiological pathways involved in the control of the trait. Only a few barley genes involved in the control of phenology, disease resistance and major morphological changes have been described at the sequence level (Miklis et al., 2007; Ramsay et al., 2011; Turner et al., 2005). Variation in gene sequence found in mutagenized populations and causing extreme phenotypes is also used to investigate the genetic architecture of traits and identify genes (Druka et al., 2010; Rossini et al., 2006). However, the description of genes involved in the genetic control of agronomic quantitative traits is rare due to variable effects observed in different backgrounds and the presence of QTL x environment interactions (Collard and Mackill, 2008). For example, the pleiotropic effects of the photoperiod gene *Ppd-H1* may affect the final TGW indirectly by changing the adaptability of a plant to an environment, hence its potential to fill the grain (Kandemir et al., 2000; Wang et al., 2010).

The identification of candidate genes is not simple as hundreds of genes can be present in the few centiMorgans delimited by the QTL support interval on a barley chromosome. Meta-analysis of QTL and studies across species can help reduce this interval and identify plausible biological pathways that may be involved (Swamy et al., 2011). In addition, the synteny across species can also help to narrow down the number of candidate genes underlying a QTL (Mayer et al., 2011) and the sequencing information can also be used to find and clone a candidate gene (Cockram et al., 2010).

Comparative genomics

Comparative genomics, also referred to as synteny, is the study of co-linearity of loci between genomes of related species. The genome sizes of grass species are very variable but the linear order of genes has remained well conserved over million years of evolution (Gale and Devos, 1998). Plant common ancestry helps to understand genome dynamics by revealing both ancient and recent segmental insertion and replication, polyploidy events, and genome altering processes that have translated into different genome structures, functions and biological effects (Bennetzen and Chen, 2008). Therefore comparative genomics is a powerful tool to exploit the genomes of fully sequenced model species such as rice or *Brachypodium* and infer candidate gene positions in related species such as barley and wheat. The 'genome zipper' between cereal species has helped to establish a linear order of genes for the barley genome (Mayer et al., 2011) and online resources such as Gramene (www.gramene.org) enable the visualization of the co-linearities of genomes. Comparative genomics can also be used in marker development to investigate the sequence of flanking putative genes to saturate a chromosome region with new markers.

The macro-co-linearity of marker order at the genetic map level is distinguished to micro-co-linearity at the genomic sequence level (Bennetzen and Chen, 2008; Muehlbauer et al., 2009). Both can give better insight of genomic regions to understand the gene arrangement, function and sequence. The micro-co-linearity can be used to check for conservation of sequences at orthologous positions and can reflect the conservation of a gene with similar function across species. The barley flowering time gene *Ppd-H1* was shown to be located at a position orthologous to a junction between chromosomes 4 and 7 of rice (Dunford et al., 2002; Turner et al., 2005). There is strong evidence for conserved full or partial control of some plant architecture traits and molecular pathways. For example, the gene involved in the variation of lateral spikelet fertility in barley *Int-c* is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1 (TB1) (Ramsay et al., 2011). TB1 controls the fate of axillary meristems in maize and the development and expression of fertility in lateral spikelets of barley. Other gene candidates have been investigated using comparative genomics between rice, wheat and barley (Distelfeld et al., 2008; Muehlbauer et al., 2009; Stein et al., 2005). In some cases, the co-linearity between genomes is insufficient to resolve candidate genes as traits are not conserved across species or different pathways are involved (Muehlbauer et al., 2009; Ramsay et al., 2011). Griffiths et al. (2003) did not

confirm the QTL for rice flowering time at CONSTANS genes in orthologous barley genes. Nevertheless, the potential of comparative genomics in the search and validation of candidate genes is highly valued.

Sequence information and diagnostic markers

Once confidence has been established on the genetic location of the QTL, there is a benefit in acquiring the detailed genetic sequence information to potentially identify a diagnostic marker useful for MAS. Although this project will not delve into extensive sequencing experiments, it is necessary to mention the role of DNA sequencing in the characterisation of polymorphism at the genetic level as the sequence polymorphisms can be used to develop better or diagnostic markers for MAS. The sequencing of gene plays a key role to understand the changes in the DNA sequence at the origin of a modification in the function of a protein that can affect a whole physiological pathway and phenotype (Cockram et al., 2010; Szucs et al., 2007; Zitzewitz et al., 2005). A deletion in the *ant-2* gene sequence was shown to be responsible for the expression of anthocyanin pigmentation in barley (Cockram et al., 2010). This extra level of resolution may be helped by the recent release of the partial barley genome sequence which identified 26,159 'high-confidence' genes with homology support from other plant genomes (Mayer et al., 2012). The benefits are noteworthy as sequence information and across crops synteny could be exploited to identify candidate genes from QTL mapping the alleles to use in MAS.

1.5 Research objectives

The remarkable advances in crop science and barley research have placed the breeding industry in a position to diversify and improve their methods for improving crops. The QTL mapping studies are the initial stage to find associations between trait variation and genetic polymorphism. However these studies are the tip of the iceberg in understanding the genetic architecture of traits and identifying useful markers for MAS with validated effects (**Chapter 1**). Subsequent approaches of comparative genomics, QTL validation using NILs and characterisation of causal polymorphism at the DNA sequence level are all necessary steps to help defining and refining the most convincing targets (Collard and Mackill, 2008). Amongst the complex traits of interest for breeders, yield and yield components would greatly benefit from genomic research and increased marker applications. However, the current understanding of the genetic control of the

traits and their association with genetic markers as reviewed in Chapter 1 suggests that further exploration is needed to efficiently exploit genetic resources.

This PhD project has been elaborated at the cutting edge of the current breeding and research activities in barley to investigate the genetic architecture of the highly valuable but not so well understood traits of yield and yield components.
This project intends to deliver both the breeding industry and research community with enhanced knowledge on the genetic factors controlling complex agronomic traits which are targeted in elite winter barley breeding programmes, as well as exploitable results to enable a step-change in genetic progress.

The main objectives of the project provide the framework for the different chapters in this thesis and each chapter is structured based on specific objectives as follow:

- Carry out QTL mapping for yield and yield components using the bi-parental DH population from a cross between Saffron and Retriever (**Chapter 2**)
    - Collect and analyse phenotypes for a range of agronomic traits on the population grown in different environments and seasons.
    - Genotype the DH population and create the genetic map.
    - Carry out QTL x Environment analysis.
    - Identify genetic regions and QTL of interest to provide targets for further investigation of underlying polymorphic genes and alleles.

- Carry out the Genome Wide Associations Studies using the two-row winter barley panel of varieties and phenotypes from the NUE-CROPs and AGOUEB projects (**Chapter 3**).
    - Collect the phenotypes for a range of agronomic traits including yield and yield components on the NUE-CROPS trials managed by KWS–UK.
    - Carry out the statistical analysis of phenotype data collected across five locations using mixed modelling.
    - Gather the phenotypes for the varieties of the AGOUEB panel.
    - Gather and analyse the genotypes on the 9000 SNP Infinium genotyping chip for all varieties included in the panels.
    - Analyse the population structure of the panels

    - Carry out GWAS using appropriate statistical models (correction for population structure) and report on QTL discovery.

- Combine the results from the three mapping experiments to enable direct comparison across studies and built up confidence on QTL targets (**Chapter 4**).
    - Establish a consensus map using common SNP across the three mapping studies described previously.
    - Position the QTL on the consensus map to understand the genetic architecture of traits and identify genetic factors involved in the control of traits within and across studies.
    - Inspect the pattern of significance and allele effects for the range of traits for the SNP at relevant clusters to suggest targets for MAS.
    - Use comparative genomics and synteny between rice and barley genomes to propose candidate genes involved in the control of traits for some promising genetic factor.

- Validate QTL for agronomic traits by developing and testing Near Isogenic lines for agronomic traits (**Chapter 5**).
    - Identify breeding material segregating at QTL targets found in previous chapters to develop HIFs and NILs.
    - Use a panel of genome wide SNP to select and develop NILs with minimal background heterozygosity.
    - Carry out field trial testing and phenotyping for a range of agronomic traits on the NILs.
    - Interpret the effects and report on QTL validation between NILs.

- Discuss the results and expand on the knowledge generated during the project in the context of a commercial barley breeding programme focusing on improving crop yield. (**Chapter 6**)

Chapter 2

QTL mapping for yield and yield components in a bi-parental DH population of an elite winter barley cross.

2.1 Introduction

The challenges of sustainable food production are a main concern for agriculture which has to maintain high levels of production while reducing inputs as a key target for reaching sustainability. Plant breeding has had a major impact on food production. Improved varieties have been released to deliver greater benefits from producer to consumers, enhancing yields, quality requirements and sustainability (Abeledo et al., 2008; Fufa et al., 2005; Le Gouis et al., 2000; Rae et al., 2007; Sanchez-Garcia et al., 2013). The majority of traits of interest that are expressed quantitatively result from multiple factors or components controlled by a range of genes across the genome (Bernardo, 2008). In barley, the grain yield can be dissected into the yield components of tillering, grains per ear and grain weight (see 1.2.2). Therefore, the overall yield improvement in this crop comes from the combination of favourable genes and alleles controlling each one of these yield components. In order to select for positive alleles, plant breeders require a thorough understanding of the genetic architecture of yield and yield components so that the optimal allele combinations can be generated and maintained in segregating progenies.

Genetic mapping of QTL helps researchers to understand the genetic control of traits and by associating genetic markers with phenotypic variation that can be exploited for marker assisted selection (MAS) (Collard et al., 2005). In barley, the double haploid (DH) populations are a valuable tool to exploit the segregation of alleles and investigate agronomic traits (see 1.3.2). QTL for yield and yield components have been described in mapping studies of bi-parental crosses (Backes et al., 1995; Bezant et al., 1997; Hayes et al., 1993; Yin et al., 2002). The recombinant inbred line (RIL) population from a cross between spring barley and wild barley revealed that both elite and wild parents carried positives alleles for yield, tillering and TGW with changes in magnitude of the effects attributable to QTL x E interactions (von Korff et al., 2006). Yield QTL with strong effects were associated with the known loci *vrs1* and *int-c* controlling inflorescence structure in a mapping population from a cross between a two-row and six-row barley (Marquez-Cedillo et al., 2001). Similarly, loci involved in the control of

vernalization and photoperiod in barley were significantly associated with yield differences in crosses between spring and winter types (Cuesta-Marcos et al., 2009). In barley, different pool of germplasm can be identified based on genetic markers and clusters of varieties monomorphic for the alleles at loci controlling row type, vernalization and photoperiod (Comadran et al., 2009; Zhang et al., 2009). These pools are rarely crossed between each other in elite breeding programmes in order to avoid extreme segregation in progenies (David Harrap pers. comm.). Therefore the yield variation the within each pool originates another set of segregating alleles and genes.

Bi-parental populations are continuously produced in the process of plant breeding and can be used for research purposes. An elite bi-parental DH population was generated at KWS-UK from the cross between two-row winter barley varieties Saffron and Retriever. This population provides an opportunity to understand the genetic architecture of yield and yield components in elite two-row winter barley material. The complementarity of the varieties in terms of agronomic characteristics had been spotted by the breeder willing to exploit their contrasting genetics (David Harrap pers. comm. and Table 1.1). Saffron and Retriever are known to differ in tillering ability and grain weight and have different yield responses under first and second cereal conditions affecting their yields and yield components. However, the magnitude of genetic main effect, pleiotropic effects and genotype x environment interactions for these traits and their interaction with other traits remains unclear. Barley was shown to have different yield response to nitrogen supply (Abeledo et al., 2003) and varying root architecture (Hargreaves et al., 2009). It is possible that Saffron and Retriever have contrasting soil scavenging abilities and responses to early nitrogen availability. The nitrogen economy may also impact on plant development and fate of above ground material associated to yield performance (Gregersen et al., 2008).

A preliminary QTL mapping was carried out at KWS in 2009 using a single marker analysis (SMA) approach implemented as a routine programme in Excel with the raw phenotypes adjusted to the best fitting linear model. The SMA method identified a number of marker–trait associations for yield and yield components across the genome but did not estimate QTL × E interaction. Nevertheless, preliminary breeding trial results indicated that the S×R (Saffron × Retriever) DH population was a valuable tool to study the genetic control of agronomic and phenotypic traits in two-row winter barley (Cockram et al., 2010). Although the SMA results were encouraging for the number and magnitude of significant associations, a re-analysis of the data using composite interval mapping (CIM) (Zeng, 1994) was suggested.

This PhD project investigates the genetic architecture of yield and its underpinning yield components to provide breeders with relevant genetic targets for selection. Chapter 2 aims at providing a complete QTL analysis of the bi-parental DH population S×R in order to better understand the genetic architecture of yield for this elite two-row barley cross. The objectives of this study are: 1) to carry out QTL mapping using CIM on adjusted phenotypes collected on the DH population grown in yield trials at two sites, 2) to investigate and confirm the genetic control of traits using phenotypes from an additional experiment under untreated growing conditions, 3) to identify chromosome regions with candidate QTL involved in the control of agronomic traits that could potentially be targets for MAS approaches and investigated further to identify candidate genes. This study aims at generating a solid knowledge for the genetic control of yield and yield components in a cross of elite two-row winter barley and a working base for interpreting additional QTL mapping studies.

2.2 Methods

2.2.1 Plant material

A double haploid two-row feed winter barley population of 530 DH lines from a cross between elite varieties Saffron and Retriever was developed in 2007 at KWS UK for both breeding and research purposes.

Saffron is a KWS-UK bred variety and had the leading market share in the UK during the early 2000s. The variety exhibits very high yield potential in the high fertilisation regime of first cereal conditions where it develops large grains and numerous tillers. Retriever is a high yielding variety from Limagrain (Nickerson) with strong yield potential in second cereal conditions where it maintains its ability to tiller well and its grain filling. Breeder's observations suggest that the variety is unable to benefit from higher fertilisation regimes under which it tends to over tiller and produces thin grains (David Harrap personal commu.). At maturity, the variety is prone to collapse of the straw above the last node which can also affect harvest conditions.

2.2.2 Phenotypic evaluation

2009 first and second cereal sites

The population was grown in yield plots (6 m^2, spacing between rows: 13.6 cm) at the sites of Fowlmere (Fowl09) and Elmdon (Elm09) in east Hertfordshire-UK. Fowl09

was characterised as a first cereal site with high residual soil nitrogen and high yielding potential where barley is grown at the beginning of the crop rotation. Elm09 was a second cereal site, with lower soil residual nitrogen as winter barley is sown following a first crop of winter wheat. At each site, a subset of the population (211 DH lines) was grown as single replicates in a block containing 7 to 8 plots of each of the control varieties: Saffron, Retriever, and Cassia. Phenotype data (Table 2.1) was collected for a range of traits on each plot at both sites. Field topography was also recorded as a potential factor in a covariate analysis on measured phenotypes (Appendix 2. 2).

Table 2.1 List of phenotypes measured on the S×R DH population

Trait measured	Abbreviations	Trait description
Ear emergence	**EE**	Date of 50% of the ears above flag leaf in a plot
Height	**Ht**	Height of the plot
Tillering measured	**Til-mes**	Tillering measured from plot drilled rows after harvest
Tillering calculated	**Til-cal**	Tillering calculated from yield components
Grains per ear	**GE**	Average number of grains per ear
Thousand grain weight	**TGW**	Weight of a thousand grains
TGW from grab sample	**TGW-GS**	Weight of a thousand grains from the grab samples
Yield	**Yld**	Marketable grain yield of the plot
Lodging	**Ldg**	Proportion of lodging in a plot, visual score
Hectoliter weight	**HLW**	Weight of an hectoliter of combine harvested grains
Grain protein	**GP**	Percentage of protein in ground grain
Grain sugars	**GS**	Percentage of soluble sugars in ground grain
Mildew	**Mil**	Susceptibility to mildew, visual score
Brown rust	**BR**	Susceptibility to brown rust, visual score
Stay green	**SG**	Proportion of healthy plant tissues at ripening stage, visual score
Straw collapse	**SC**	Plant collapsing on itself, visual score
Straw degradation	**SD**	Degradation of the straw by disease at ripening stages, visual score
Ear glaucosity	**E_Glau**	Presence or absence of wax layer on the ear
Antocyanin colour	**Antho**	Colour of awn tips and grains
Aleurone colour	**Aleu**	Colour of aleurone layer in grain

EE was recorded at growth stage 59 and Ht was recorded at growth stage 81 (Zadoks et al., 1974). The percentage of lodging (Ldg) was scored before harvest. Grab samples consisting of 30 random ears within a plot were taken at maturity (GS81). Ears were dried at 40ºC for 48 hours, threshed and weighed to calculate the thousand grain weight from the grab sample (TGW-GS) and the grains per ear (GE). The raw yield was obtained as the grain weight of a plot (Yld) and used to calculate fertile tillers (Til-cal) derived from Yld, TGW-GS and GE (Til-cal = Yld/TGW-GS/GE). An independent measure of TGW was made using a seed sample from the combined plot for which

smaller grains should have been sieved out by the combine during harvest. This trait can be used as a validation trait for TGW-GS. In order to get an independent measure of fertile tillering, the cut tillers in stubble were counted on four segments of 50 cm in each plot and converted to an area (Til-mes). Bulked grains from combined yield plots were then analysed for HLW. In 2011, grains from each site were ground to flour to pass through a 0.5 mm mesh and analysed by NIR (Foss-5000 instrument) with 2 technical reps. Grain nitrogen and grain sugars (GS) were estimated using KWS-UK in-house NIR calibrations (Aunir-group 10). The grain proteins (GP) was calculated using the standard 6.25 nitrogen to protein conversion factor (Mariotti et al., 2008).

2012 Population maintenance and additional phenotyping

In 2012, the complete DH population of 530 DH lines was grown as single rep in a large block in untreated conditions at the Fowlmere site (Fowl12). The parental lines Saffron and Retriever were grown in 19 replicates each randomised within the population. All DH lines were grown as a pair of 1 meter long rows containing 25 plants each. A purity check was done between the seed stock from 2008 used for genotyping of the population and the seeds from the grab samples of Fowl09. The rogueing for off-types and identification of mixed seed lots was done by comparing phenotypes and genetic marker information (e.g. anthocyanin pigmentation). At GS51, one plant of each DH line was bagged to avoid cross pollination and harvested to compose the definitive population. The entire population was phenotyped for EE and Ht, mildew (Mil) and brown rust (BR) during the grain filling stages. Grab samples of 25 ears were taken on each of the 530 DH lines and replicated controls and dried at 40ºC for 48 hours. The grains from threshed ears were counted to record GE and TGW-GS. Grains were milled to pass through a 0.8 mm mesh and the flour was analysed by NIR using the same calibration as for 2009 experiments. Straw characteristics of stay green (SG), straw degradation (SD) and straw collapse (SC) were scored at the start of ripening stage (GS89) on a 1 to 5 scale, a high score indicating that the character is visible to a large extent. In this study, SG was scored using a scale estimating the remaining areas of healthy straw and leaves. SC is a measure of the extent of weakness of the straw visible by straw twisting and bending on the last 2 internodes. This trait observed in the variety Retriever differs from a weak straw at the plant base which induces the lodging. SD was defined as an indicator of straw degradation due to senescence and diseases and is associated with SG. The variety Saffron tends to be good at SG while retriever is prone to SC and SD.

Estimation of adjusted means

Experimental trials at Fowl09, Elm09 and Fowl12 produced highly unbalanced data sets due to single replication of DH lines and only two or three replicated control varieties. Therefore, adjusted means for all phenotypes were obtained by linear mixed model analysis implemented in REML algorithms in GenStat 14th Edition (Payne *et al.*, 2009). Raw phenotypes were initially screened to identify outliers at an arbitrary cut-off value of +/- 3 standard deviations in association with breeder's notes on the quality of the plots. The genotypes were analysed as fixed term in the model. To account for environmental variation, blocking structures (row, column) and covariates (topography, lodging) were handled as random terms in the models (Ve). For each trait and DH line, the Best Linear Unbiased Estimates (BLUEs) were obtained by fitting the model minimising residual error (Appendix 2. 1). The final BLUEs were used for phenotypic and QTL mapping analysis in this study.

An estimate of the heritability (h2) of traits was calculated from the replicated controls using the best REML models identified for each trait respectively. The variance components for each term of the models were obtained by setting all models terms as random effects so that h2=Vg/(Vg+Ve) where Vg represents the variance component of the genotype and Ve is the sum of variance components for environment, blocking and residual error variance.

2.2.3 Genotyping information

211 DH lines of the population were genotyped in 2009 using a set of 1536 SNP markers contained within the BOPA1 array, Illumina Golden gate technology on a Beadstation. 173 additional lines were genotyped using custom subset of 384 SNPs using the Illumina BeadXpress platform. SNP marker names were standardised using the BOPA_C nomenclature (e.g. 11_10022) and replacing "11_" by the letter A (e.g. A10022). Each SNP was associated with their expected chromosome from barley OPA2009 consensus map available at Graingenes 2.0 (wheat.pw.usda.gov/) to facilitate the linkage map construction (Close et al., 2009) (Supplementary data 1). Genotypes were transformed into ABH codes with A-Saffron, B-Retriever and H-Heterozygotes calls. Genetic map distances between markers were estimated using Mapdisto software v1.7. (Lorieux, 2012) with a LOD score of 3 and the Kosambi mapping function used to define initial linkage groups. The relevant linkages groups were associated to chromosomes based on the consensus chromosome position. Markers showing segregation distortion were kept throughout the process. Bootstrap and ripple order

functions were used in Mapdisto (Lorieux, 2012) along with published consensus map position to determine optimal marker order and distances. A skimmed map consisting of one unique SNP marker for each of the individual positions was kept for the subsequent QTL mapping analysis (Appendix 2. 3).

2.2.4 QTL analysis

Flapjack format files were used to carry out QTL mapping using GenStat14th Edition (Milne et al., 2010; Payne et al., 2009). A simple interval mapping (SIM) procedure was first carried out and candidate QTL were defined as cofactors in a mixed model based Composite interval mapping (CIM) using a minimum cofactor proximity window of 30 cM. A LOD score of 3 was defined as the threshold for detection of significant marker trait associations and putative QTL. For traits with no QTL detected initially, the threshold was lowered to a 2.9 LOD. QTL support intervals were defined as the distance corresponding to a decrease of 2 LOD scores from the QTL peak position. The traits measured in more than one site/year combination were analysed in a multi-environment QTL analysis. The traits measured only in 2009 were analysed in QTL × Env analysis using environments of Fowl09 and Elm09. Traits measured at Fowl09, Elm09 and Fowl12 were analysed in a QTL × Env analysis using the three environments. A single environment QTL analysis was done for traits measured only at Fowl12.

Genetic predictors which are genotypic covariates that reflect the genotypic composition of a genotype at a specific chromosome location were set at every 2cM (Lynch & Walsh 1998). The best variance covariance model for multi-environment QTL analysis was estimated for each trait based on Schwarz Information Criterion (SIC). Genome-wide QTL scans on multi-environment trials data was carried out in GenStat 14th Edition (Payne et al., 2009) by fitting statistical model incorporated in the QMESTIMATE procedure (Malosetti et al., 2004). The detection model assumes environment as a fixed term and genotypes as a random term.

2.3 Results

2.3.1 Phenotypic analysis

2009 experiment

Phenotypic variation was observed for all traits measured. The adjusted means for each DH line estimated as BLUEs were obtained by taking into account the effects of different environmental factors in the prediction models (Appendix 2. 1). Although not always significant, environmental factors included in the model made changes to predicted values and reduced the standard errors of estimates. At Fowl09, the field topography affected soil moisture and influenced the distribution of some traits. (Appendix 2. 2). Positive variance components were found for topography when estimating traits linked to plant development and productivity. The lodging affected the traits of Yld and the GE measured from combine samples, suggesting losses of grain during the combining process. At Elm09, the topography was less variable across the trial and only row or columns factors significantly captured environmental effects.

Significant location effects were found for all the traits in 2009 (Table 2.2). Fowl09 produced on average higher yields compared to Elm09. These higher yields were associated with higher tillering and grains per ear despite a lower TGW. Yld at both sites was positively correlated with both tillering and TGW although those two traits were negatively correlated with each other (Table 2.3). At Elm09 and Fowl09, Til-cal was negatively correlated with TGW and grains per ear (GE). However, the independent measure of tillering Til-mes only confirmed that negative correlation with TGW at Fowl09 (-0.38), highlighting the value of independent measurements of traits. GE was significantly positively correlated to Yld at Elm09 only. At both sites late EE appeared to have negative impact on Yld and yield components mainly associated with a drastic reduction of TGW suggesting that early lines produced bigger grains but lower yields.

Table 2.2 ANOVA of DH lines grown at three sites

One way ANOVA on BLUE of the DH lines in the population grown at 3 sites in 2009 and 2012. Traits have been abbreviated as follow: Yield (Yld), Thousand Grain weight (TGW), Thousand Grain weight from grab samples (TGW-GS), Tillering calculated (Til-cal), Tillering measured (Til-mes), Grain per ear (GE), Hectoliter weight (HLW), Ear emergence (EE), Height (Ht), Grain proteins (GP) and Grain sugars(GS).

	One way ANOVA				Site Means (s.e.)				
Trait	**Source of variation**	**d.f**	**m.s.**	**F pr.**	**Elm09**	**Fowl09**	**Fowl12**	**average s.e.d**	**LSD (5%)**
Yld	Env	1	861	**<.001**	8.1 (0.1)	11.1 (0.1)		0.07	0.15
	Residual	396	0.6						
TGW	Env	1	709	**<.001**	58.1 (0.2)	55.4 (0.2)		0.34	0.68
	Residual	408	12.2						
TGW-GS	Env	2	17158	**<.001**	60.0 (0.3)	56.5 (0.3)	42.7 (0.3)	0.40	0.79
	Residual	610	16.6						
Til-cal	Env	1	2781588	**<.001**	540 (4)	706 (4)		6.24	12.27
	Residual	401	3925						
Til-mes	Env	1	4620741	**<.001**	656.64 (4.95)	869 (5)		7.01	13.78
	Residual	407	5027						
GE	Env	2	293	**<.001**	26.2 (0.1)	27.6 (0.1)	28.6(0.1)	0.19	0.38
	Residual	609	3.8						
HLW	Env	1	1660	**<.001**	68.6 (0.1)	64.6 (0.1)		0.12	0.23
	Residual	408	1.4						
EE	Env	2	2041	**<.001**	19.9 (0.2)	16.7 (0.2)	23.1 (0.2)	0.28	0.54
	Residual	610	7.8						
Ht	Env	2	43073	**<.001**	81 (0.4)	86 (0.4)	109 (0.4)	0.52	1.03
	Residual	610	27.8						
GP	Env	2	16.5	**<.001**	10.3 (0.1)	9.8 (0.1)	10.9 (0.1)	0.06	0.11
	Residual	609	0.3						
GS	Env	2	4.4	**<.001**	1.66 (0.02)	1.74 (0.02)	1.46 (0.02)	0.03	0.06
	Residual	610	0.1						

Table 2.3 Correlation between traits at each experimental site

Pearson correlation coefficient between different traits at a) Elmdon 2009, b) Fowlmere2009 and c) Fowlmere 2012. Coefficients are significantly different at * p<0.05, ** p<0.01 and *** p<0.001.

a)

Elm 09	EE	Ht	Til-cal	Til-mes	GE	TGW	TGW-GS	Yld	HLW
Ht	0.12								
Til-cal	**-0.15***	**-0.24*****							
Til-mes	0.01	0.03	**0.28*****						
GE	**0.26*****	0.23**	**-0.42*****	-0.12					
TGW	**-0.57*****	0.05	**-0.3*****	-0.08	-0.05				
TGW-GS	**-0.54*****	0.04	**-0.17***	-0.06	-0.09	**0.77*****			
Yld	**-0.31*****	-0.07	**0.46*****	0.13	**0.17***	**0.36*****	**0.27*****		
HLW	-0.09	-0.05	-0.03	-0.1	-0.09	**0.3*****	**0.34*****	0.11	
GP	**0.18***	0.02	-0.03	-0.09	-0.09	**-0.22****	**-0.31*****	**-0.22****	**-0.17***

b)

Fowl 09	EE	Ht	Til-cal	Til-mes	GE	TGW	TGW-GS	Yld	HLW
Ht	**0.34*****								
Til-cal	0.06	**0.3*****							
Til-mes	0.05	**0.17***	**0.54*****						
GE	0.11	**0.15***	**-0.23****	-0.08					
TGW	**-0.51*****	**-0.2****	**-0.41*****	**-0.38*****	**-0.18***				
TGW-GS	**-0.56*****	**-0.2****	**-0.25*****	**-0.27*****	**-0.18***	**0.78*****			
Yld	**-0.29*****	**0.15***	**0.57*****	0.14	0.07	**0.3*****	**0.32*****		
HLW	-0.05	-0.01	0.01	-0.11	**-0.2****	**0.38*****	**0.33*****	**0.33*****	
GP	**0.4*****	**0.55*****	**0.29*****	**0.21****	**0.24*****	**-0.47*****	**-0.39*****	-0.11	**-0.16***

c)

Fowl 12	EE	Ht	GE	TGW-GS	GP	Ldg	Mil	BR	SG	SC
Ht	**-0.24*****									
GE	0.09	0.08								
TGW-GS	**-0.42*****	**0.31*****	**0.18***							
GP	-0.04	**0.16***	0.08	-0.12						
Ldg	**-0.25*****	**0.5*****	-0.06	**0.15***	-0.05					
Mil	0.02	0.12	**0.24*****	**0.26*****	0.08	0.09				
BR	**-0.22****	-0.09	0.1	**0.19****	-0.03	-0.04	**0.28*****			
SG	**0.29*****	-0.05	**0.17***	0	**0.2****	**-0.27*****	0.1	0.05		
SC	**-0.24*****	**0.27*****	**-0.18***	-0.08	-0.12	**0.5*****	0.01	0.01	**0.44*****	
SD	**-0.21****	**0.29*****	-0.1	-0.07	**-0.14***	**0.52*****	0.01	0.02	**0.46*****	**0.7*****

Table 2.4 Summary statistics on eleven agronomic traits

Distribution of the DH population BLUEs for agronomic traits at both sites in 2009. Control means, control F probability, least significant difference (LSD) and heritability (h2) are reported from the REML analysis using the replicated control only.

	Yld (t/ha)	TGW-GS (g)	TGW (g)	Til-cal (Tiller/m2)	Til-mes (Tiller/m2)	GE	HLW (kg/hl)	EE (days of May)	Ht (cm)	GP (%)	GS (%)
Fowl09											
Min	8.1	47.1	43.7	501	671	23.4	60.5	9.4	71	8.2	1.0
1st Qu	10.6	53.9	53.3	658	816	26.4	65.8	14.5	83	9.3	1.5
Median	11.1	57.0	55.5	711	862	27.5	66.6	16.5	86	9.7	1.7
Mean	11.1	56.5	55.6	705	868	27.6	66.5	16.7	86	9.8	1.7
3rd Qu	11.7	59.5	58.4	758	934	28.6	67.4	18.9	90	10.2	1.9
Max	13.2	65.8	65.0	898	1118	32.0	69.2	24.2	99	11.9	2.4
Missing	9	-	1	4	2	1	8	-	-	1	-
Control means											
Cassia	11.70	59.62	59.28	731.40	869.60	26.33	66.66	16.02	91.64	10.47	2.04
Retriever	10.86	53.21	53.23	752.20	911.50	27.15	62.91	14.75	84.63	9.75	1.79
Saffron	11.42	58.37	57.91	688.20	861.80	27.45	66.33	18.06	90.03	10.38	2.04
av sed	0.19	1.32	0.93	17.83	28.44	0.57	0.34	0.47	2.19	0.27	0.09
F.pr	**0.002**	**<0.001**	**<0.001**	**0.008**	0.186	0.17	**<0.001**	**<0.001**	**0.013**	**0.024**	**0.009**
LSD (5%)	0.40	2.85	2.09	38.47	59.62	1.19	0.72	1.05	4.64	0.55	0.18
h2	0.18	0.37	0.23	0.17	0.03	0.05	0.40	0.22	0.18	0.28	0.36
Elmd09											
Min	6.4	51.8	50.2	421	517	21.5	65.4	14.3	71	9.2	0.6
1st Qu	7.8	57.6	55.7	512	621	25.0	68.1	18.0	79	10.1	1.5
Median	8.2	60.0	58.6	538	660	26.0	68.7	19.1	82	10.3	1.7
Mean	8.1	60.0	58.1	540	657	26.1	68.6	19.9	81	10.4	1.7
3rd Qu	8.6	62.2	60.3	568	689	27.0	69.3	22.1	84	10.6	1.8
Max	9.5	68.0	65.1	669	809	30.5	70.6	27.0	90	11.4	2.3
Missing	3	2	-	3	-	-	-	-	1	-	-
Control means											
Cassia	7.98	62.21	61.16	535.90	680.70	24.60	69.42	19.14	83.57	10.74	1.96
Retriever	8.42	62.37	61.55	529.00	681.60	25.76	68.54	16.94	80.33	9.70	1.67
Saffron	7.98	59.86	58.95	527.20	656.90	25.46	69.28	20.04	83.02	10.96	1.84
av sed	0.13	0.72	0.45	10.07	17.62	0.27	0.16	0.40	1.07	0.08	0.07
F.pr	**0.024**	**<0.001**	**0.004**	0.666	0.297	**0.002**	**<0.001**	**<0.001**	**0.018**	**<0.001**	**0.003**
LSD (5%)	0.34	1.52	1.00	21.93	37.02	0.62	0.35	0.84	2.26	0.18	0.15
h2	0.15	0.38	0.55	-	0.03	0.30	0.48	0.61	0.21	0.87	0.32

The replicated control varieties showed a response to site similar to the rest of the DH population (Table 2.4). The population distribution indicated transgressive segregation at both sites for all traits. Retriever had the lowest yield amongst the control varieties in the first cereal growing conditions at Fowl09 (10.8 t/ha) and achieved the highest control yield at the second cereal site of Elm09 (8.1 t/ha). Both Saffron and Cassia (pedigree: Saffron×(Eden×Carat)) had a reduction in yield at Elm09 but Cassia maintained significantly higher TGW compared to Saffron at that site. These yield variations observed between the controls also illustrate the correlations between yield components (Table 2.3) and confirm the breeder's comment on the adaptation of both varieties to different fertility. Retriever has a substantial increase in TGW at Elm09 where it matches Cassia's. However, the variety has low TGW (53.2g) and HLW (62.9 kg/hl) the first cereal site Fowl09 suggesting a varietal interaction with nitrogen availability. Retriever reaches GS61 four days earlier than Saffron and has high tillering

and low TGW at Fowl09 with a GE equivalent to other controls. This trait balance suggests an increase in sink organs in high fertility conditions (Fowl09) that could affect the overall grain filling and TGW as suggested by the negative correlation between tillering and TGW in Table 2.3. Saffron and Cassia show a similar behaviour in tillering ability and yield at each site but Cassia is slightly earlier and maintains higher TGW in second cereal conditions.

Significant differences between the control varieties were also observed for grain quality traits. Retriever had the lowest GP (9.7%) and GS at both sites. Saffron and Cassia maintained a higher GP even despite a substantial increase in Yld at Fowl09 which could have led to nitrogen dilution effect in the grain. These results suggest different responses to nitrogen fertilisation between the two varieties Saffron and Retriever. GS is a measure of soluble sugar in ethanol in the ground grain. The value of this trait has not yet been established in KWS breeding programme but it seems that Saffron and Retriever are significantly different and show segregation in the DH population. These soluble sugars could potentially relate to potential malting qualities and enzyme activity under genetic control that it may be worth analysing.

The heritability of the traits calculated for the controls ranged from 0.03 (tillering Eml09) to 0.87 (grain proteins Elm09). Low heritability values indicate that a large amount of the trait variation is attributable to error variance suggesting that interpretation of the results for these traits should be made with caution. Additionally, the large error variance can be due to the statistical models unable to account for undescribed environmental variation.

2012 experiment

After the verification of 552 DH lines for a match between genotype with phenotypes, 515 lines were conserved for purity to complete the entire S×R DH population.

The BLUEs approach was favoured for the analysis of phenotypes due to a limited number of controls and the non-replicated DH lines. Environmental effects of blocking structures were found significant for most traits (Appendix 2. 1). The GE, Ht and GP were higher at Fowl12 compared to Fowl09 and Elm09 whereas TGW was reduced (Table 2.2) suggesting different growing conditions. It needs to be borne in mind that due to the larger space between both the rows (23 cm) and the plants at Fowl12, plant tillering was increased in comparison to a yield plot which may affect the balance between yield components.

EE at Fowl12 was negatively correlated to Ht whereas these two traits were positively correlated in 2009. The correlation of EE and GP was not significant in 2012 (Table 2.3). These observations support the differences between both sites and contrasting growing conditions of 2009 and 2012 (extremely wet season) but also the genuine differences between varieties. Similar trends were observed for EE and TGW confirming that early lines have significantly higher TGW. Untreated conditions at Fowl12 captured additional differences in straw characteristics and disease susceptibility. SD and SC were significantly positively correlated with Ht and negatively with EE and GP. Surprisingly, the susceptibility to mildew was significantly positively correlated with GE and TGW although an increase in disease is not expected to increase any of those two traits. A close proximity between the genetic controls of those traits may be possible.

The significant differences between sites presented in Table 2.2 are illustrated by the phenotypic distributions for the traits measured across multiple sites (Table 2.4 Table 2.5). Fowl12 growing conditions produced larger differences between the controls for TGW, EE and Ht than either 2009 sites. Retriever was always the shortest with lower TGW compared to Saffron and was earlier to reach the flowering stage (GS61). GE was measured with lower precision from the grab samples at Fowl12 (heritability of 0.15) and Saffron had significantly lower GE than Retriever ($P<0.04$), not observed at Fowl09. GP was very stable and highly heritable for each variety across sites confirming the low protein values for the grain of Retriever.

The disease resistance scores collected for the 515 lines suggested also that the genetic control of resistance was segregating within the population. Saffron was slightly more susceptible to brown rust (+1.1) and by +3.6 point more susceptible to mildew than Retriever. The straw characteristics of Retriever indicated a high susceptibility to collapse (SC) associated with lower SG and higher SD and Ldg.

Table 2.5 Summary statistics on 12 phenotypes collected in 2012

Distribution of the phenotypic means (BLUEs) of the DH population for agronomic traits at Fowl12. Control means, Controls F probability, LSD and heritability are reported from the REML analysis using only replicated controls. For the following traits, the scoring scale was: stay green (SG): 1 green, 5 senesced; Brown rust (BR): 1 resistant, 9 susceptible; Mildew (Mil): 1 resistant, 9 susceptible, Straw degradation (SD): 1 healthy, 5 highly degrading; Straw collapse (SC): 1 standing, 5 collapsed, Aleurone colour (Aleu): 1 white, 2 blue; Anthocyanin: 1 white, 2 red; Ear glaucosity (E_Glau): 1 non glaucous or waxless, 2 glaucous or waxy.

	TGW_GS (g)	GE	EE (days of May)	Ht (cm)	GP (%)	GS (%)	SG	BR	Mil	SD	SC	Ldg	Aleu	Antho	E_Glau
Fowl12															
Min	32.0	17.7	18	92	8.7	0.7	1.0	1.5	1.0	0.9	0.6	-0.9	1.0	1.0	1.0
1st Quartile	39.1	27.2	22	104	9.7	1.2	1.0	3.9	3.2	1.0	1.1	2.6			
Average	42.6	28.5	23	109	10.1	1.5	3.2	5.1	4.6	2.6	3.0	4.3			
Median	43.1	28.6	23	108	10.1	1.4	3.0	5.1	4.6	3.0	3.1	4.2			
3rd	46.0	30.1	25	112	10.5	1.7	5.0	6.4	6.0	4.0	4.8	6.0			
Max	53.5	34.1	28	122	12.1	2.5	5.0	9.1	8.5	5.1	5.1	9.9	2.0	2.0	2.0
Control means															
Retriever	43.8	31.1	21	111	9.7	1.7	4.8	4.7	2.6	4.9	4.9	6.4	1.0	2.0	2.0
Saffron	48.0	30.5	24	117	10.6	1.7	1.1	5.8	6.4	1.3	1.5	3.6	2.0	1.0	1.0
sed	1.1	0.3	0.3	1	0.1	0.1	0.3	0.4	0.3	0.2	0.2	0.3			
Saffron effect	4.2	-0.6	4	6	0.8	0.0	-3.6	1.1	3.8	-3.6	-3.4	-2.8			
F statistic	13.4	4.7	111.6	55.0	56.9	0.0	172.1	7.7	195.8	310.6	209.2	110.7			
d.d.f.	25	24	28	22	23	19	22	21	15	22	21	15			
F pr	**0.001**	**0.04**	**<0.001**	**<0.001**	**<0.001**	0.864	**<0.001**	**0.012**	**<0.001**	**<0.001**	**<0.001**	**<0.001**			
LSD (5%)	2.3	0.6	0.7	1.7	0.2	0.1	0.6	0.8	0.6	0.4	0.5	0.6			
h2	0.36	0.15	0.86	0.68	0.67	-	0.91	0.23	0.87	0.95	0.92	0.60	1.00	1.00	1.00

The two years of experiments clearly exposed the varietal differences between Saffron and Retriever for agronomic traits, especially yield components. Retriever performs better than Saffron in second cereal conditions (Elm09). Its development is less adapted to high fertility first cereal conditions where it tends to develop large numbers of tillers with low TGW. It is therefore apparent that the S×R DH population will be a useful research tool for the study of genetic control of yield and yield components in different growing conditions. The contrasting environment responses for each of the varieties are especially relevant in the context of cereal breeding and crop performance in rotation. The stability of the GP content across sites is remarkable and highlights strong genetic differences in nitrogen metabolisms between the varieties that could provide research leads in future studies.

2.3.2 Genotypic analysis

S×R genetic map

The BOPA1 genotyping identified 309 polymorphic SNPs between the two parents. A total of 211 lines were used to create the genetic map including lines 139 and 192 despite their large number of heterozygotes called markers. Genotype information for

Saffron and Retriever from different sources was used to impute missing SNP alleles wrongly called as either missing or heterozygote. A preliminary genetic map was created using all 309 polymorphic markers (mapdisto 1.7.5b4 2007-all markers). The default settings produced nine linkage groups including more than three markers. A21141_5H did not show any significant linkage with any of the linkage groups and was excluded from the mapping set. The optimisation of the marker order using information from consensus map and the optimisation functions of Mapdisto produced a final genetic map of seven distinct linkage groups corresponding to the seven barley chromosomes (Appendix 2. 3).

The total length of the final SR map is 896.4cM with chromosome length ranging from 49.8cM (5H) to 192.6 cM (3H) (Appendix 2. 3). Lower recombination rates were found at centromeric regions with more co-segregating marker loci. The exclusion of co-segregating marker loci produced a map with 174 different map positions (Figure 2.1). Although these co-segregating marker loci may have different consensus map positions, they did not show any recombination in the S×R population.

For the analysis, the locus file created for QTL mapping analysis included all DH lines in trials with phenotypes. The lines with genotype information but no phenotypes were removed (14,100,177,206,207,216). 208 valid DH lines were used for QTL mapping in GenStat 14th Edition (Payne et al., 2009).

Genetic polymorphisms

The S×R population genetic map showed unequal distribution of SNP polymorphisms and varied chromosome length (Appendix 2. 3). Chromosome 1H and 5H showed the lower polymorphism with only 9 polymorphic SNPs each. A higher number of polymorphic markers was observed on 3H (98 SNP) and 4H (61 SNP). The marker order on the genetic map mostly agreed with the order observed in consensus map OPA 2009 (Graingene). The SNP array detected four large monomorphic chromosome segments between the two varieties on chromosomes 1H, 2H, 5H and 7H spanning respectively 91.7cM, 27.3cM, 111.6cM and 33.3cM based on consensus map distances (Figure 2.1). Those regions on 1H and 7H include the chromosome centromeres whereas the quasi entire length of chromosome 5H long arm appears to be monomorphic based on the marker results. Genotyping of the parental lines with increased numbers of marker such as the 9K SNP array may help to confirm these observations.

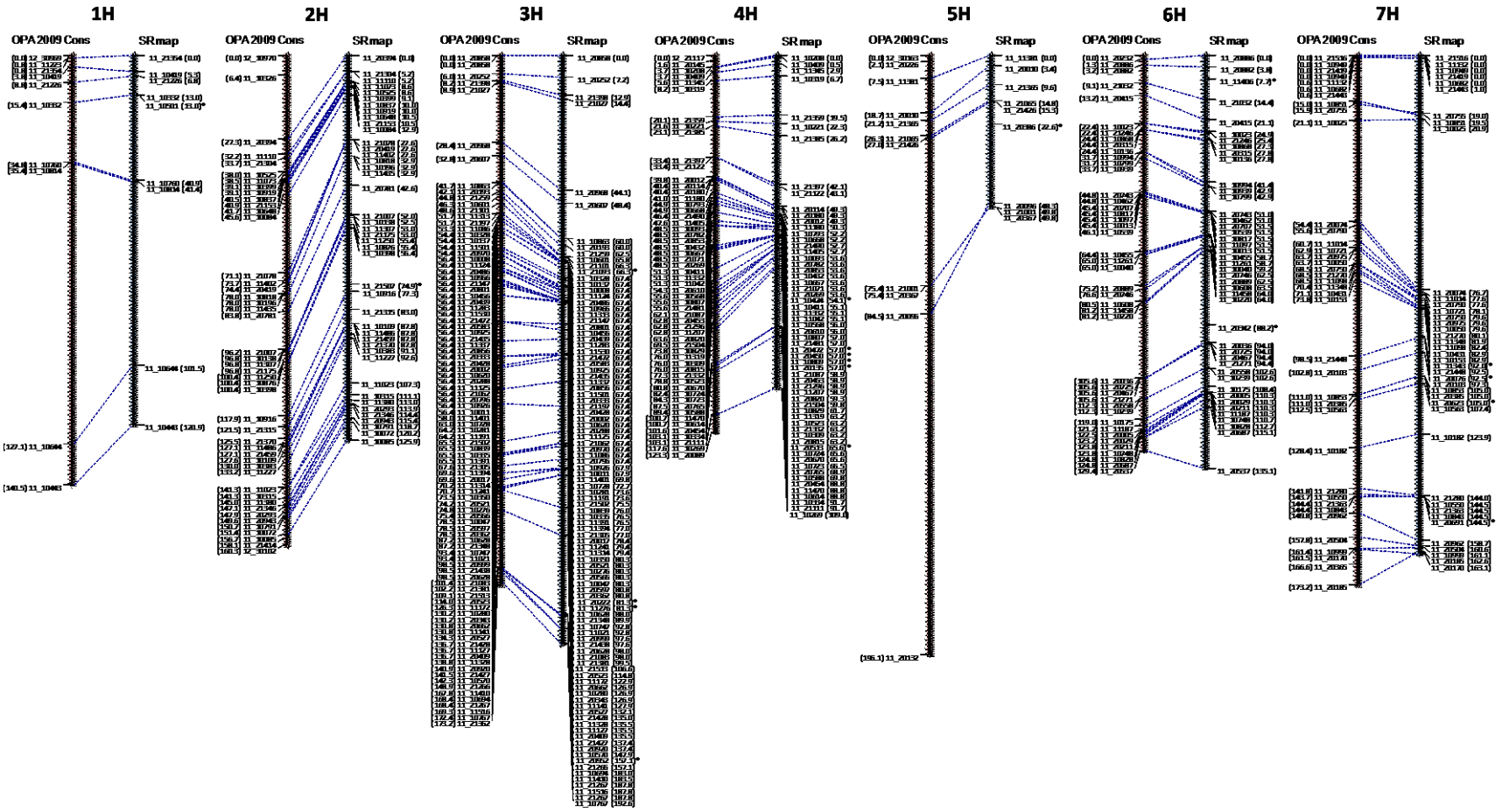


Figure 2.1: Alignment of S×R SNP genetic map with the OPA consensus 2009 distances.
Representation of the map distortion and repartition of polymorphic SNP markers between the S×R DH population genetic map (Blue) (Appendix 2. 3) and barley OPA consensus map 2009 (red) (Close et al., 2009).

2.3.3 Quantitative Trait Loci

Mapping results of QTL analyses for the traits are presented for individual experimental years in Table 2.6 (2009) and Table 2.7 (2012). The multi-environment QTL analysis for traits measured at all three sites is reported in Table 2.8.

QTL were mapped on all the seven chromosomes and for all traits measured in 2009 and 2012. For each trait, the QTL mapping identified from one to five QTL. Only a few QTL by Environment interactions (QTLxE) were detected showing different magnitude of the effects. However, in both Table 2.6 and Table 2.8, the majority of the QTL showed similar effect size between the sites. The QTL results had an average support interval for the mapped QTL was 22.6cM providing generally well defined QTL positions on the S×R genetic map although relatively large for targeting genes.

Ear emergence

Four different QTL locations were identified for EE across analysis with three of them showing QTLxE interactions (Table 2.6, Table 2.8). QTL EE.2_1 (3H) was only significant at Elm09 (Table 2.6) and was not detected in the three site analysis while QTL EE.3_3 (7H) was only significant at Fowl12 (Table 2.8). EE.3_1 and EE.3_2 were mapped in all sites with the Saffron allele associated with late ear emergence. EE.3_2 had the highest effect on EE of 1.6 days difference (Elm09 and Fowl09) but its magnitude was reduced at Fowl12 (EE.1_2) to a non-significant effect of 1 day. Remarkably, EE.3_2 was mapped on the monomorphic chromosome segment of 5H. Although this QTL explains up to 32% of the trait variation at Elm09, the monomorphic segment was confirmed by additional genotyping. EE QTL on 5H was considered as an artefact of CIM method. A mapping of EE excluding 5H produced a unique QTL on chromosome 4HL corresponding to EE.3_1.

Table 2.6 QTL results of S×R mapping at Fowl09 and Elm09.

Table of QTL detected in the Saffron×Retriever DH population in the 2009 experiments at Fowl09 and Elm09. The QTL distance (Dist) on the chromosome (Chr) and maximum likelihood scores (LOD) are reported from the multi-environment mapping analysis. A negative (-) effect corresponds to an increase of the traits by the Saffron allele and a positive (+) by the Retriever allele. Effects in italic are non-significant. A SNP marker with "*" indicate that the closest marker reported is not located within the QTL support interval. The QTL are named as "Δ.2_▲" where Δ is the trait abbreviation and ▲ the QTL number for that trait.

Trait	QTL Name	Chr	Dist	LOD	closest marker (cM)		-2 LOD	-1 LOD	+1 LOD	+2 LOD	F stat	d.d.f.		F prob	Effect Elm09 (S.e)	Effect Fowl09 (S.e.)	%Expl. Elm09	%Expl. Fowl09	Effect
GE	GE.2_1	3H	132.1	10.58	A20527	132.1	127.9	129.3	133.7	133.8	50.7	201		<0.001	0.62 (0.09)	0.62 (0.09)	12.6	13.2	Main
	GE.2_2	4H	91.7	3.95	A10334	91.65	76.5	77.1	100.9	105.3	15.6	201		<0.001	-0.35 (0.09)	-0.35 (0.09)	4.1	4.3	Main
	GE.2_3	6H	115.1	3.16	A20687	115.1	95.2	110.6	126.6	134.3	8.8	201		0.003	0.26 (0.09)	0.26 (0.09)	2.3	2.4	Main
TGW	TGW.2_1	2H	3.5	9.03	A21304	5.23	0.0	0.0	9.2	10.2	41.2	201		<0.001	-1.23 (0.19)	-1.23 (0.19)	15.7	10.4	Main
	TGW.2_2	3H	151.6	3.07	A10570	148	137.4	142.4	163.4	164.3	8.4	201.6		<0.001	*-0.36 (0.2)*	-0.95 (0.25)	1.3	6.1	QTLxE
	TGW.2_3	5H	36.5	5.18	*A20096	48.33	20.4	24.8	46.7	48.7	23.2	201		<0.001	1.04 (0.22)	1.04 (0.22)	11.1	7.3	Main
TGW_GS	TGW_GS.2_1	2H	3.5	13.74	A21304	5.23	0.0	0.6	7.4	9.6	68.3	200		<0.001	-1.53 (0.19)	-1.53 (0.19)	19.7	15.7	Main
	TGW_GS.2_2	3H	155.3	3.76	A20952	157.1	140.3	143.4	163.7	168.5	18.7	200		<0.001	-0.64 (0.19)	-0.64 (0.19)	3.4	2.7	Main
	TGW_GS.2_3	5H	36.5	5.40	*A20096	48.33	20.4	25.2	45.9	48.6	23.4	200		<0.001	1.07 (0.22)	1.07 (0.22)	9.6	7.6	Main
	TGW_GS.2_4	7H	54.8	4.42	*A20074	76.68	40.2	41.7	69.0	69.6	15.4	200		<0.001	1.12 (0.29)	1.12 (0.29)	10.5	8.3	Main
Til-cal	Til_cal2_1	2H	6.9	3.17	A11073	8.57	0.0	0.0	11.6	39.2	19.1	198.7		<0.001	9 (3)	9 (3)	4.5	1.4	Main
	Til_cal2_2	2H	111.1	4.05	A10315	111.1	95.6	96.4	112.9	123.7	13.0	198.5		<0.001	10 (3)	10 (3)	5.1	1.6	Main
	Til_cal2_3	3H	132.1	6.63	A20527	132.1	120.7	128.9	135.9	140.4	28.6	198.4		<0.001	-13 (3)	-13 (3)	9.2	2.9	Main
	Til_cal2_4	7H	50.8	4.13	*A20074	76.68	34.2	35.8	62.9	63.5	17.4	199.3		<0.001	-17 (4)	-17 (4)	15.4	4.9	Main
Til-mes	Til_mes.2_1	2H	3.5	5.38	A21304	5.23	0.0	0.0	8.5	10.1	12.7	200.8		<0.001	9 (4)	28 (6)	2.9	10.6	QTLxE
	Til_mes.2_2	2H	94.4	3.50	A11227	92.56	82.0	82.7	104.5	111.7	16.1	201.3		<0.001	13 (3)	13 (3)	6.1	2.4	Main
Yld	Yld.2_1	3H	165.1	3.09	A20952	157.1	150.7	151.5	179.1	180.2	7.1	197.4		0.001	*0 (0)*	-0.3 (0.1)	0.1	8.9	QTLxE
EE	EE.2_1	3H	98.0	3.00	A20628	98.03	90.4	94.2	105.1	114.6	7.6	201.1		<0.001	*0.2 (0.2)*	-0.5 (0.2)	0.3	2.6	QTLxE
	EE.2_2	4H	93.6	4.26	A10334	91.65	82.5	88.8	103.6	109.0	10.3	201		0.002	-0.7 (0.2)	-0.7 (0.2)	6	4.6	Main
	EE.2_3	5H	34.5	14.43	*A20386	22.58	27.7	30.3	40.5	43.1	66.4	201		<0.001	-1.6 (0.2)	-1.6 (0.2)	32.4	24.7	Main
Ht	Ht.2_1	4H	97.4	4.36	A10334	91.65	82.5	85.0	107.2	109.0	10.0	202		<0.001	*-0.5 (0.3)*	-2 (0.4)	1.8	11	QTLxE
HLW	HLW.2_1	2H	22.1	4.89	A20419	27.56	8.8	9.0	30.4	33.4	28.9	199		<0.001	-0.3 (0.1)	-0.3 (0.1)	8.5	3.1	Main
	HLW.2_2	2H	91.1	3.72	A10383	91.14	52.3	79.6	97.4	102.3	8.0	199.6		<0.001	0.2 (0.1)	*-0.1 (0.1)*	4.2	0.9	QTLxE
	HLW.2_3	3H	76.5	4.42	A10335	76.5	61.1	67.7	86.8	90.4	21.6	199		<0.001	0.2 (0.1)	0.2 (0.1)	7.5	2.8	Main
	HLW.2_4	3H	135.5	11.80	A11328	135.5	133.6	134.4	136.6	141.8	64.3	199		<0.001	-0.3 (0)	-0.3 (0)	15.5	5.7	Main
	HLW.2_5	7H	153.2	4.69	A20962	158.7	139.9	144.7	159.8	163.1	10.8	200.6		<0.001	*-0.1 (0.1)*	-0.4 (0.1)	2.5	8.6	QTLxE
GP	GP.2_1	2H	48.3	9.52	A20781	42.61	31.0	34.1	48.5	48.7	18.2	199.9		<0.001	0.14 (0.02)	0.14 (0.02)	11.7	3.8	Main
	GP.2_2	2H	117.1	18.49	A10791	118.7	113.3	114.6	118.7	119.6	99.4	200		<0.001	-0.19 (0.02)	-0.19 (0.02)	23.1	7.4	Main
	GP.2_3	6H	102.6	3.87	A20558	102.6	96.2	99.2	116.3	117.2	16.7	200		<0.001	-0.07 (0.02)	-0.07 (0.02)	3.5	1.1	Main
	GP.2_4	7H	111.1	3.97	A10563	107.4	102.4	105.9	119.0	123.9	59.0	200		<0.001	-0.16 (0.02)	-0.16 (0.02)	16.6	5.3	Main
GS	GS.2_1	1H	5.3	4.65	A10419	5.31	0.0	1.7	8.4	11.0	22.1	200		<0.001	-0.06 (0.01)	-0.06 (0.01)	4.4	4.2	Main
	GS.2_2	4H	2.9	4.49	A11345	2.87	0.0	0.0	11.9	16.5	24.5	200		<0.001	0.06 (0.01)	0.06 (0.01)	4.4	4.1	Main
	GS.2_3	6H	1.9	11.62	A20886	0	0.0	0.0	4.2	5.8	57.0	200		<0.001	-0.1 (0.01)	-0.1 (0.01)	12.4	11.6	Main
	GS.2_4	7H	160.6	3.41	A20504	160.6	145.3	148.6	163.1	163.1	15.6	200		<0.001	-0.05 (0.01)	-0.05 (0.01)	3.4	3.2	Main

Table 2.7 QTL results of S×R mapping at Fowl12.

Table of QTL detected in the S×R DH population in the untreated 2012 experiment at Fowl12. The QTL distance (Dist) on the chromosome (Chr) and maximum likelihood scores (LOD) are reported from the multi-environment mapping analysis. A negative (-) effect corresponds to an increase of the traits by the Saffron allele and a positive (+) by the Retriever allele. A SNP marker with "*" indicate that the closest marker reported is not located within the QTL support interval. The QTL mapped at Fowl12 only are named as "Δ.1_▲" where Δ is the trait abbreviation and ▲ the QTL number for that trait.

Trait	QTL Name	Chr	Dist	LOD	closest marker	(cM)	-2 LOD	-1 LOD	+1 LOD	+2 LOD	F stat	d.d.f.	F prob	%Expl. Var.	effect (S.e.)	Positive allele
GE	GE.1_1	2H	20.2	6.95	A10084	12.9	4.9	5.2	33.2	35.0	28.7	199	<0.001	15.2	-0.91 (0.17)	Saf
	GE.1_2	4H	83.1	6.51	A20454	88.8	73.2	76.9	91.4	97.4	28.1	199	<0.001	14.0	-0.87 (0.17)	Saf
TGW_GS	TGW-GS.1_1	2H	0.0	5.11	A20394	0.0	0.0	0.0	10.3	14.7	20.5	198	<0.001	8.5	-1.39 (0.3)	Saf
	TGW-GS.1_2	3H	14.4	4.65	A21027	14.4	5.4	10.5	27.5	28.7	25.0	198	<0.001	7.1	-1.28 (0.29)	Saf
	TGW-GS.1_3	3H	157.1	3.04	A20952	157.1	141.7	142.6	166.2	172.1	14.2	198	<0.001	4.3	-0.99 (0.3)	Saf
	TGW-GS.1_4	5H	36.5	3.58	A20096	48.3	21.9	23.0	49.8	49.8	13.8	198	<0.001	7.3	1.3 (0.35)	Ret
EE	EE.1_1	4H	85.0	8.32	A20454	88.8	75.2	78.0	97.0	97.7	32.5	199	<0.001	16.1	-0.9 (0.1)	Saf
	EE.1_2	5H	34.5	8.67	*A20386	22.6	22.0	26.8	41.2	45.2	38.9	199	<0.001	19.7	-1 (0.2)	Saf
	EE.1_3	7H	19.0	3.36	A20755	19.0	4.1	7.8	33.2	33.9	12.8	199	<0.001	4.6	0.5 (0.1)	Ret
Ht	Ht.1_1	4H	95.5	2.91	A10334	91.7	78.1	89.0	109.0	109.0	10.7	201	0.001	6.2	-1.5 (0.4)	Saf
GP	GP.1_1	1H	1.8	4.35	A21354	0.0	0.0	0.0	15.0	15.6	16.6	200	<0.001	7.9	-0.17 (0.04)	Saf
	GP.1_1	7H	94.1	5.20	A21448	92.5	84.3	87.4	99.8	103.6	21.5	200	<0.001	9.5	-0.19 (0.04)	Saf
GS	GS.1_1	1H	6.8	5.06	A21226	6.8	0.0	0.0	9.1	11.1	20.8	201	<0.001	9.6	-0.1 (0.02)	Saf
BR	BR.1_1	4H	55.1	3.88	A10411	55.1	41.2	45.7	57.2	64.4	15.2	201	<0.001	7.4	-0.4 (0.1)	Saf
Mil	Mil.1_1	1H	11.4	4.03	A10332	13.0	0.0	2.1	24.2	27.5	14.3	198	<0.001	5.0	-0.4 (0.1)	Saf
	Mil.1_2	3H	89.9	4.50	A21348	89.9	81.7	86.2	94.8	104.9	6.8	198	0.01	6.4	0.4 (0.1)	Ret
	Mil.1_3	4H	59.3	20.04	A20820	59.3	50.5	52.0	62.1	62.4	115.4	198	<0.001	34.2	-1 (0.1)	Saf
	Mil.1_4	5H	9.6	2.94	A21365	9.6	0.0	0.0	19.5	43.6	10.9	198	0.001	3.2	0.3 (0.1)	Ret
SD	SD.1_1	1H	1.8	9.58	A21354	0.0	0.0	0.0	5.7	10.2	40.1	199	<0.001	17.4	0.7 (0.1)	Ret
	SD.1_2	2H	98.1	5.36	A11227	92.6	84.4	88.8	113.0	113.2	22.8	199	<0.001	9.6	0.5 (0.1)	Ret
	SD.1_3	6H	41.4	3.23	A10994	41.4	26.8	30.5	50.4	55.8	12.2	199	<0.001	4.5	0.3 (0.1)	Ret
SC	SC.1_1	1H	1.8	5.92	A21354	0.0	0.0	0.0	12.8	25.1	20.8	198	<0.001	9.0	0.5 (0.1)	Ret
	SC.1_2	2H	103.6	13.25	A11023	107.3	97.2	99.1	109.3	112.9	67.7	198	<0.001	24.7	0.8 (0.1)	Ret
	SC.1_3	6H	42.9	4.09	A10939	42.9	18.9	19.4	47.9	48.4	16.2	198	<0.001	5.4	0.4 (0.1)	Ret
SG	SG.1_1	1H	1.8	13.06	A21354	0.0	0.0	0.0	4.9	11.8	61.0	199	<0.001	23.9	0.8 (0.1)	Ret
	SG.1_2	2H	115.4	3.67	A20943	115.4	101.0	101.6	118.8	124.1	14.3	199	<0.001	4.9	0.4 (0.1)	Ret
	SG.1_3	6H	53.3	4.54	A20707	51.5	42.3	45.5	59.6	63.7	18.3	199	<0.001	6.7	0.4 (0.1)	Ret
Ldg	Ldg.1_1	5H	49.8	3.46	*A21001	49.8	35.7	38.3	35.6	54.9	13.2	201	<0.001	6.2	0.5 (0.1)	Ret
E_Glau	E_Glau.1_1	1H	0.0	88.10	A21354	0.0	0.0	0.0	0.5	1.0	1254.0	203	<0.001	88.6	0.47 (0.01)	Ret
Antho	Antho.1_1	2H	52.5	53.41	A10138	52.5	51.4	52.2	52.7	52.9	461.0	203	<0.001	70.4	0.91 (0.04)	Ret
Aleu	Aleu.1_1	4H	59.3	46.61	A20820	59.3	58.9	59.1	59.5	59.8	397.6	179	<0.001	70.4	0.74 (0.04)	Saf

Height

The genetic variation of height in that cross appeared to be significantly associated with a maximum of two controlling genetic factors. The QTL HT.3_1 on 4H was recurrently detected in all experiments and showed QTLxE interactions. The support interval of that QTL was also consistent between the different mapping experiments (84-109 cM). Ht.3_1 is located is the same region as EE.3_1 suggesting a common genetic control for the traits located in the distal region of 4HL. Retriever's early type is also associated with a shorter straw length. Ht.3_2 was only observed in the multi-environment analysis but not in specific years separately (Table 2.8).

Yield and yield components

- Yield

A unique QTL was found for yield on 3HL at 165 cM (Table 2.6). Yld.2_1 was detected with a LOD score of 3.09 and a significant positive effect from Saffron allele at Fowl09. Yld.2_1 did not overlap any other QTL on chromosome 3HL (Figure 2.2) despite the significant correlations to yield components and EE (Table 2.3). An attempt at QTL mapping of yield using a lower threshold of LOD=2 produced additional QTL: a QTL on 3H at 104.9cM with a LOD of 3.28 (Retriever main effect = 0.13), a QTL on 3H at 165cM significant at Fowl09 only and corresponding to Yld.2_1 (LOD=3.30; Saffron positive effect 0.26), a QTL on 5H at 49.8 cM (LOD= 2.81; Retriever main effect of 1.11) and a QTL on 6H at 5.8 cM (LOD= 2.88; Retriever main effect of 0.12). Although these QTL are not entirely comparable to the QTL detected at a LOD3 threshold, they could constitute putative Yld QTL and may be of interest for the interpretation of the genetic architecture of yield if confirmed by other components. For example, the Retriever yield QTL on 3H at 104.9cM is associated with the Retriever earliness effect of EE.2_1.

- Thousand grain weight

Five different QTL locations were significantly associated with TGW variation in the multisite analysis and explained up to 37.4% of the variance at Elm09 (Table 2.8). TGW-GS.3_1, TGW-GS.3_3 TGW-GS.3_4 were mapped in all experiments. TGW-GS.3_2 was only significant at Fowl12. TGW-GS.3_5 was not detected at Fowl12 but significant as main effect in the multi-site analysis. Analysis of TGW combine sample (TGW) confirmed the grab sample QTL results (TGW-GS) on 2H, 3HL and 5H but did not capture any significant associations on 3HS and 7H. A major QTL TGW-GS.3_1 on

2HS explained from 10 to 20% of the trait variance depending on the site, Saffron giving a consistent 1.5 g advantage across sites. The strongest association is with the first polymorphic marker on 2H (A20394) located at 0cM on S×R genetic map and at 27cM on the OPA1 consensus map (Figure 2.1). On chromosome 3H, TGW-GS.3_2 accounted for 8% of the variance at Fowl12 only suggesting a strong site effect on that locus while all other loci showed reduced influence on the trait (Table 2.8). TGW-GS.3_3 on 3HL overlapped with Yld.3_1 with both QTL showing Saffron as positive allele. However the QTLxE observed for Yld.3_1 was not observed at TGW-GS.3_3. Both TGW.2_3 and TGW-GS.3_4 were mapped in the monomorphic region of 5H at 36.5 cM. These mapped QTL will be considered as spurious as was EE.3_2 in this study despite being the only position where these two highly correlated traits have overlapping QTL. TGW-GS.2_4 was mapped at 55cM on 7H in 2009 (Table 2.6) and a nearby hit of TGW-GS.3_5 was found at 83cM across all sites (Table 2.8). Similar effects, direction and overlapping support intervals for the two QTL suggest the that they are identifying the same genetic factor. The peak of TGW-GS.3_5 corresponds to the position of Ht.3_2 where Retriever carries the positive allele for both traits.

The strong environment interactions detected for TGW-GS.3_2 suggest that the condition of 2012 benefited the grain filling for lines carrying the Saffron allele (Table 2.5, Table 2.8). In addition, the independent measure of TGW by grab sample suggested that this phenotyping method captured additional variation in the trait (TGW-GS.3_5) which may relate to the small grains blown away during mechanical harvest. Both Retriever and Saffron had alleles contributing to increases in TGW and interact with sites conditions.

- Grains per ear

  Five different QTL where found in the QTL analyses for GE (Table 2.6, Table 2.8). Saffron has a positive effect at GE.3_1 and GE.3_3. QTL GE.3_3 on 4H 83 cM showed QTLxE interactions in the multi-site analysis and explained a large proportion of the variance at Fowl12, less at Fowl9 and was not significant at Elm09. Despite this interaction, GE.3_3 was found as the unique common QTL for grains per ear between the site specific mapping analysis (GE.1_2, GE.2_2). It is a strong QTL candidate associated with EE and TGW QTL and despite the low heritability of the trait at Fowl12 (0.15). GE.3_2 on 3H has the highest LOD score of 11.9. The Retriever allele contributes to an increase in 0.56 grains per ear and accounts for more than 10% of the variance of the trait at Fowl09 and Elm09 (Table 2.8). The negative correlation and

relationship of GE with Til-cal is also supported by the co-location and antagonist effect of GE.3_2 with Til-cal.2_3 QTL in the 2009 analysis. GE.2_3 was mapped in the 2009 analysis (Table 2.6) whereas GE.3_4 in the multisite analysis only (Table 2.8). This approach points out the benefits of making both single site and multisite analysis in order to detect all putative QTL for a trait, as it seems to reveal QTL with small effects and lower contribution to the variance. None of the grain per ear QTL was co-located with a TGW or yield QTL in this study(Figure 2.2).

- Tillering

  The tillering calculated from the yield components allowed the detection of four putative QTL whereas only two were found with the independent measure of tillering (Table 2.6). It needs to be borne in mind that tillering had the lowest heritability amongst the yield components. Til_mes.2_1 has a significant QTLxE interaction which could be related to the phenotypic variation observed between sites (Table 2.5). The overlapping support intervals of Til_mes.2_1 with Til_cal.2_1 on 2HS as well as Til_mes.2_2 with Til_cal.2_2 on 2HL suggests that these QTL represent the same two genetic factors controlling tillering on 2H (Figure 2.2). Til_cal.2_4 on 7H explained a maximum of 15.4% of variation at Elm09 and Til_mes.2_1 explained 10.6% of the variation at Fowl09. Til_mes.2_1 and Til_cal.2_1 are in the same location of TGW-GS.3_1 and TGW.2_1 on the short arm of 2H where the Saffron allele gives high TGW while the Retriever allele increase tillers per m^2. Despite the low heritability of tillering, the QTL mapping identified significant effects in genetic regions that carrying QTL for strongly negatively correlated traits. This association reinforces the importance of the 2HS region as a candidate carrying alleles with a strong influence on yield components. This inter-relationship between the two traits is supported by the reduced effect size of Til_mes.2_1 at Elm09 showing an increasing TGW effect of Retriever haplotype. Til_mes.2_2 is in a region where QTL for straw collapse (SC.1_2) straw degradation (SD.1_2) and maturity (SG.1_2) have been identified in 2012 suggesting that Retriever's higher tillering ability is associated with low straw quality. Increasing alleles for tillering from Saffron were only captured with the yield derived tillering measurements at Til_cal.2_3 and Til_cal.2_4. These QTL overlap with significant associations for GE and TGW-GS, two additional components of yield. On 3H, GE.3_2 and Til_cal.2_3 have opposite positive haplotypes. Similarly, the suspicious Til_cal.2_4 is mapped on a monomorphic chromosome segment of 7HS close to TGW-GS.2_4 (Table 2.6) and could correspond to TGW-GS.3_5 (Table 2.8). Although different

interpretations of the co-location of QTL are possible, all tillering QTL are associated with QTL for other yield components as expected from the phenotypic correlations between those traits.

- HLW

Five QTL were found for HLW in the 2009 data set (Table 2.6). HLW.2_1, HLW.2_3 and HLW.2_4 are main effect QTL whereas HLW.2_2 and HLW.2_5 are subject to QTLxE interaction. For HLW.2_2, the negative effect associated with the Saffron allele at Elm09 but a shift to a positive effect at Fowl09 albeit non-significant. This observation is supported by the QTL overlapping Til_mes.2_2 and Til_cal.2_2 for which Retriever has an increased tillering ability that could penalise the HLW under high fertility conditions at Fowl09. The possible relationship between grain quality QTL and HLW.2_3 is more difficult to establish. HLW.2_4 is the QTL accounting for the most variance of the trait with a high LOD and narrow interval that overlaps with GE.2_1. At this locus, the Saffron allele associates with an increase in HLW reduction in GE.

Grain protein and sugar content

Four QTL were mapped for the highly heritable trait of GP. At Elm09, the QTL described almost 50% of the trait variance highlighting with in particular locus GP.3_2 on 2HL. Saffron alleles contributed to higher GP at GP.3_2, GP.3_3 and GP.3_4. The increasing effect on GP from Retriever at QTL GP.3_1 overlaps with Saffron increasing effect at GE.3_1 suggesting a potential dilution effect of nitrogen due to higher grain number. GP.3_2 was also mapped in a region of Til_cal.2_2 and SG.1_2. Two QTL for grain proteins were found on chromosome segment of 6H and 7H that contain no overlapping QTL of other traits. Although both GP.3_1 and GP.3_2 are main effect QTL significant in multisite analysis (Table 2.8), these loci were not detected by the 2012 mapping alone. Instead, GP.1_1 was mapped on 1H at a locus of straw quality and grain soluble sugars QTL. Grain sugar was found with four small QTL of equivalent size effects. GS.3_4 mapped next to HLW.2_5 on 7H, GS.3_2 and GS.3_3 at the distal regions of 4HS and 6HS respectively had no overlap with other QTL. The trait may be associated with other aspects of the grain quality.

Diseases and straw characteristics

Four QTL were mapped for mildew resistance and one for brown rust. Mil.1_3 accounts for 24% of the trait variance and is located on 4H at around 59cM where the unique

brown rust QTL BR.1_1 was also mapped. The susceptible variety Saffron carries susceptibility alleles for Mil.1_3, Mil.1_1 and BR.1_1. Despite being more resistant, Retriever has susceptibility alleles at Mil.1_2 on 1H and Mil.1_4 on 5H, in a region where QTL effects have previously been described (Comadran et al., 2009).

The strong correlations between straw aspect, degradation and stay green traits are also depicted by similar QTL positions on 1H, 4H and 6H. The bright and green straw of Saffron at maturity is largely controlled by the main locus SG.1_1 on 1H (23.9% of the variance explained). The same locus also describes 17.4% of the variance of SD (with an increasing allele from Retriever) and matches the QTL positions of mildew resistance (Mil.1_1) and low GP (GP.1_1) (Figure 2.2). The unique QTL for the DUS trait of ear glaucosity E_Glau.1_1 is also precisely mapped with high LOD at the extremity of 1HS and seems to coincide with the 1HS QTL cluster. The high tillering QTL on 2HL are associated with poorer straw characteristics namely SC (SC.1_2), and could potentially influence GP content at GP.3_2. A small effect QTL for lodging was also observed in the monomorphic segment on 5H with significant EE and TGW QTL. The lower replication of the Fowl12 experiment and its design may have impacted strongly on some scorings of straw health and collapse. Nevertheless, the genetic associations found for those traits suggest that they have to be considered for the interpretation of the genetic architecture of other quantitative traits such as grain quality.

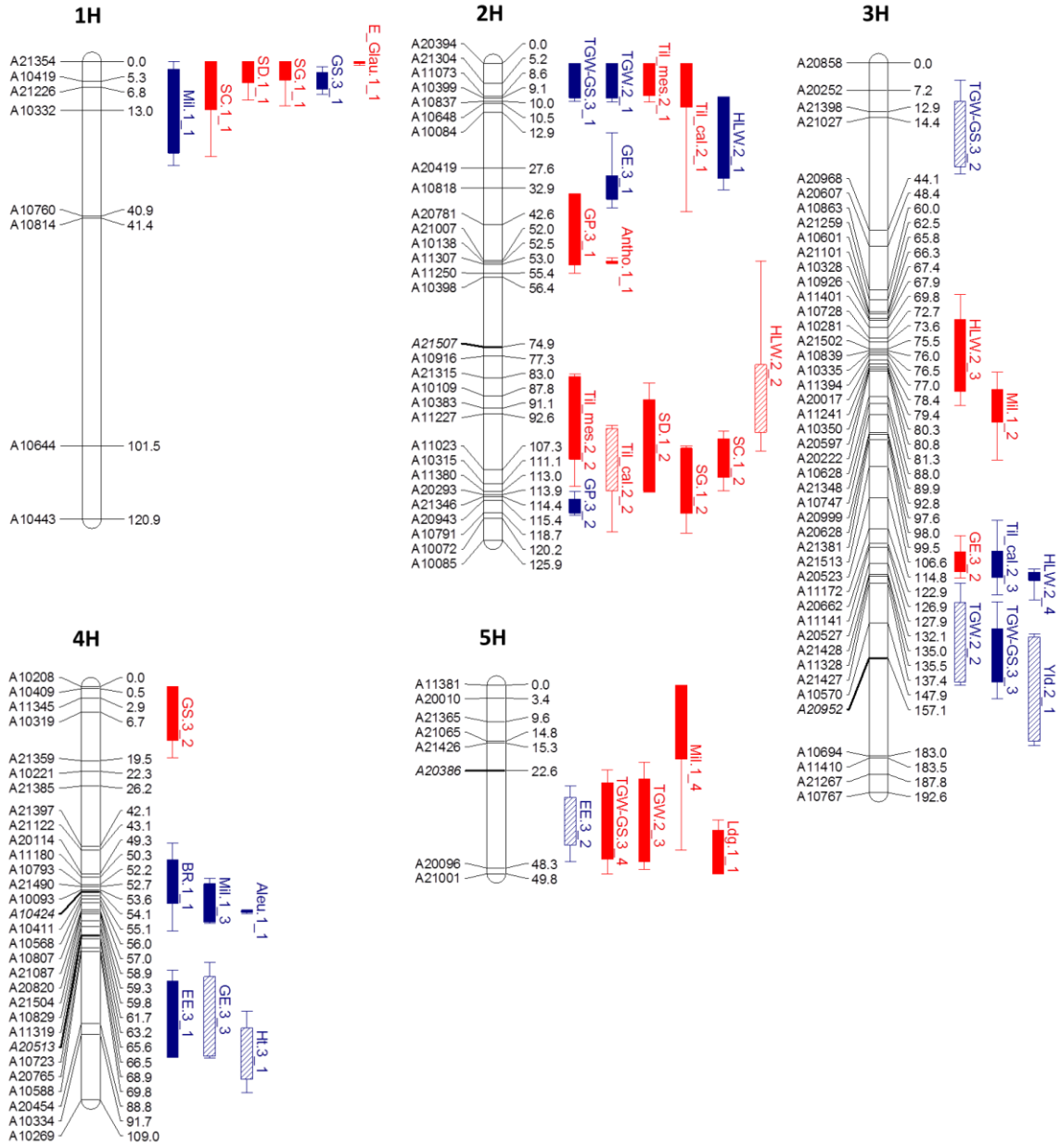


Figure 2.2 QTL location in the S×R population.

S×R QTL location for segregating traits collected from the multi-environment mapping experiments carried out in 2009 and 2012. The QTL bars represent 1 LOD and 2 LOD score decrease from the peak position. QTL with a positive additive effects from Saffron are in Blue and positive additive effect from Retriever in Red. Hatched QTL bars represent a QTL x E interaction with the colour given by the parent allele having the strongest significant effect. QTL from traits mapped in Elm09and Fowl09 are named by "Δ.2_▲" where Δ is the trait and ▲ the QTL number for that trait. QTL mapped at Fowl12 only are named by "Δ.1_▲" and the QTL from traits measured in the three sites (Fowl09 Elm09 and Fowl12) are named by "Δ.3_▲".Refer to Table 2.6, Table 2.7 and Table 2.8, for additional details on QTL.

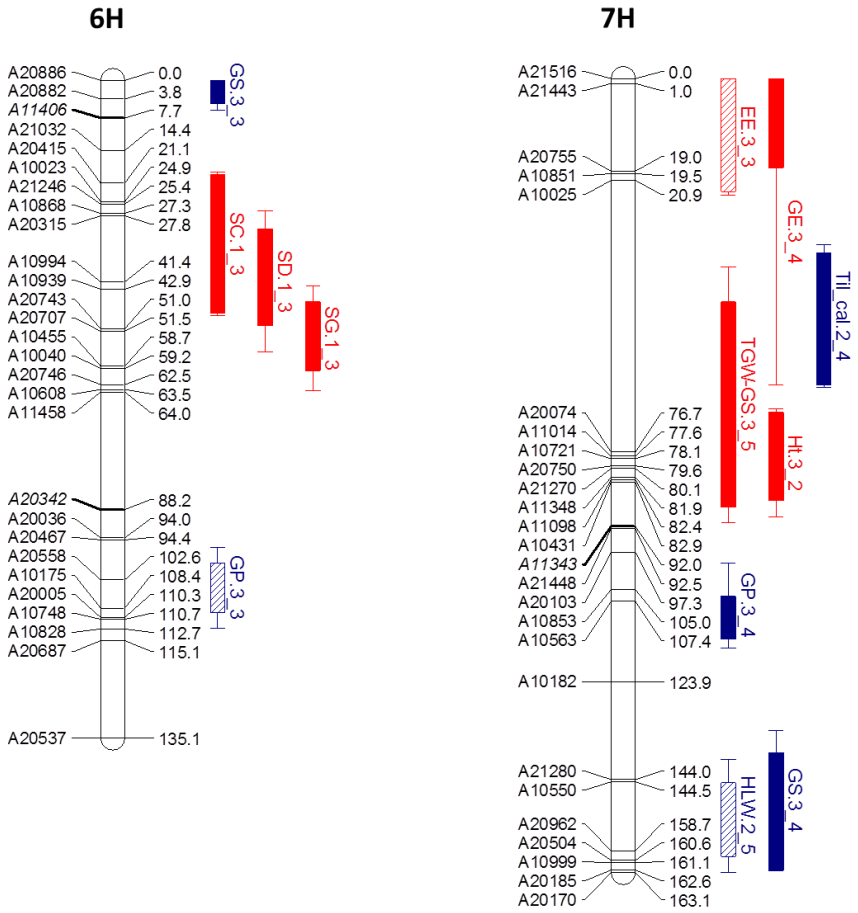


Figure 2.2 Cont

Table 2.8 QTL of S×R population for traits measured in 2009 and 2012

QTL x Environment results on phenotypes measured in 3 environments at Elm09, Fowl09 and Fowl12. The QTL from traits measured in the three sites (Fowl09 Elm09 and Fowl12) are named by "Δ.3_▲" where Δ is the trait and ▲ the QTL number for that trait. QTL effects are phased on the retriever alleles: a positive effect indicates the increasing allele is from Retriever. Effects in *italic* are non-significant with *pvalue* >0.05. A SNP marker with * indicates that the closest marker reported in the table is not located in the confidence interval of 1 LOD decrease.

							Confidence interval							Effect	Effect	Effect	%Expl.	%Expl.	%Expl.	
Trait	**QTL Name**	**Chrom**	**Pos**	**LOD**	**closest marker**	**(cM)**	**-2 LOD**	**-1 LOD**	**+1 LOD**	**+2 LOD**	**F stat**	**d.d.f.**	**F prob**	**Elm09 (S.e.)**	**Fowl09 (S.e.)**	**Fowl12 (S.e.)**	**Elm09**	**Fowl09**	**Fowl12**	**QTLxE**
GE	GE.3_1	2H	32.9	7.80	A10818_2H	32.92	18.3	29.6	35.9	38.1	39.0	198.2	<0.001	-0.45 (0.08)	-0.45 (0.08)	-0.45 (0.08)	6.6	6.9	3.7	Main
	GE.3_2	3H	132.1	11.89	A20527_3H	132.14	124.8	129.0	134.3	136.0	57.4	198.3	<0.001	0.56 (0.07)	0.56 (0.07)	0.56 (0.07)	10.5	10.9	5.8	Main
	GE.3_3	4H	83.1	6.41	A10334_4H	91.65	72.8	76.5	97.5	97.9	11.6	269.8	<0.001	*-0.18 (0.12)*	-0.43 (0.12)	-0.79 (0.17)	1.1	6.3	11.5	QTLxE
	GE.3_4	7H	6.4	3.15	A21443_7H	0.96	0.0	0.0	18.3	63.0	13.6	198.4	<0.001	0.3 (0.08)	0.3 (0.08)	0.3 (0.08)	3	3.1	1.7	Main
TGW_GS	TGW_GS.3_1	2H	1.7	14.27	A20394_2H	0	0.0	0.0	9.1	10.0	76.3	198.5	<0.001	-1.55 (0.18)	-1.55 (0.18)	-1.55 (0.18)	20.2	16.1	10.4	Main
	TGW_GS.3_2	3H	14.4	3.58	A21027_3H	14.35	4.5	10.2	27.4	29.2	8.6	271.9	<0.001	*-0.2 (0.2)*	*-0.15 (0.23)*	-1.36 (0.29)	0.3	0.1	8	QTLxE
	TGW_GS.3_3	3H	157.1	4.15	A20952_U	157.1	142.3	149.3	163.5	167.8	20.3	198.5	<0.001	-0.63 (0.17)	-0.63 (0.17)	-0.63 (0.17)	3.3	2.7	1.7	Main
	TGW_GS.3_4	5H	36.5	5.11	*A20096_5H	48.33	22.3	25.7	46.0	49.8	26.3	198.5	<0.001	1.02 (0.2)	1.02 (0.2)	1.02 (0.2)	8.7	6.9	4.5	Main
	TGW_GS.3_5	7H	82.9	4.97	A10431_7H	82.9	38.7	45.9	88.0	91.3	20.2	198.5	<0.001	0.76 (0.17)	0.76 (0.17)	0.76 (0.17)	4.9	3.9	2.5	Main
EE	EE.3_1	4H	83.1	7.79	A20454_4H	88.78	74.8	77.7	97.6	97.8	29.6	200.1	<0.001	-0.8 (0.1)	-0.8 (0.1)	-0.8 (0.1)	9	6.9	14.6	Main
	EE.3_2	5H	36.5	9.93	*A20096_5H	48.33	26.6	29.6	42.3	46.7	21.9	198.7	<0.001	-1.6 (0.2)	-1.6 (0.2)	*-1 (0.2)*	32	24.2	19.3	QTLxE
	EE.3_3	7H	11.8	3.57	A20755_7H	19.02	0.0	0.0	23.2	24.0	5.8	198.6	<0.001	*0.2 (0.2)*	*-0.1 (0.2)*	0.5 (0.1)	0.4	0.2	4.7	QTLxE
HT	Ht.3_1	4H	97.4	6.92	A10334_4H	91.65	85.6	90.1	103.6	107.2	10.2	273.1	<0.001	*-0.6 (0.3)*	-2 (0.4)	-1.6 (0.4)	2.4	12	7.2	QTLxE
	Ht.3_2	7H	82.9	3.52	A10431_7H	82.9	67.8	68.6	86.7	90.1	14.2	200.8	<0.001	0.8 (0.2)	0.8 (0.2)	0.8 (0.2)	4.2	1.7	1.7	Main
GP	GP.3_1	2H	36.8	8.26	A20781_2H	42.61	34.4	34.6	53.2	55.4	22.5	199.6	<0.001	0.13 (0.02)	0.13 (0.02)	0.13 (0.02)	10.2	3.3	4.4	Main
	GP.3_2	2H	117.1	18.59	A10791_2H	118.73	113.1	115.0	118.8	119.4	95.9	199.6	<0.001	-0.18 (0.02)	-0.18 (0.02)	-0.18 (0.02)	20.9	6.7	8.9	Main
	GP.3_3	6H	102.6	4.00	A20558_6H	102.6	96.0	99.1	109.4	112.6	8.1	270.5	<0.001	-0.09 (0.02)	*-0.01 (0.05)*	*0.05 (0.04)*	4.9	0	0.7	QTLxE
	GP.3_4	7H	111.1	12.32	A10563_7H	107.43	99.6	106.4	115.1	117.0	51.7	199.6	<0.001	-0.15 (0.02)	-0.15 (0.02)	-0.15 (0.02)	13.6	4.4	5.8	Main
GS	GS.3_1	1H	5.3	8.90	A10419_1H	5.31	1.4	2.9	7.4	8.7	40.8	199.3	<0.001	-0.07 (0.01)	-0.07 (0.01)	-0.07 (0.01)	6.9	6.5	5.1	Main
	GS.3_2	4H	4.8	4.73	A11345_4H	2.87	0.0	0.0	14.1	18.8	23.4	199.3	<0.001	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)	4.3	4	3.1	Main
	GS.3_3	6H	1.9	10.90	A20886_6H	0	0.0	0.0	4.7	6.1	48.4	199.3	<0.001	-0.08 (0.01)	-0.08 (0.01)	-0.08 (0.01)	8.5	8	6.2	Main
	GS.3_4	7H	149.5	4.22	A21280_7H	144.01	133.9	138.5	162.6	162.8	21.5	199.3	<0.001	-0.06 (0.01)	-0.06 (0.01)	-0.06 (0.01)	4.4	4.1	3.2	Main

2.3.4 Identification of genetic factors

The analysis of the QTL mapping results clearly shows that some QTL cluster at specific locations in the genome (Figure 2.2). In order to ease interpretation of the genetic control of traits, the QTL groups were identified as individual bins on chromosome segments. The sizes of the bins are delimited by the most extreme positions of the support intervals of the QTL associated with the cluster (Table 2.6, Table 2.7, Table 2.8). Each of these bin encompass a major QTL peak characterised by high LOD score and strong effect while other traits can be included in the cluster. However, some QTL intervals needed to be considered in more than one bin due to overlapping of support intervals. In total, 23 bins carrying putative genetic factors involved in the control of the traits measured in that study were identified across the genome and included QTL for one to six different traits (Table 2.9)

Table 2.9 S×R Genetic factors location and their associated QTL.
Location of the 23 bins containing putative genetic factors involved in the control of QTL identified in the S×R DH population in 2009 and 2012. The bin size and location is reported using S×R genetic map distances.

Genetic factor Bin	chr	Bin interval (cM)	Clustered SxR QTL									
1	1H	0.0 - 27.5	GP.1_1	GP.1_1	GS.1_1	GS.2_1	GS.3_1	Mil.1_1	SC.1_1	SD.1_1	SG.1_1	E_Glau.1_1
2	2H	0.0 - 39.2	HLW.2_1	TGW.2_1	TGW_GS.2_1	TGW_GS.3_1	TGW-GS.1_1	Til_cal.2_1	Til_mes.2_1			
3	2H	0.0 - 48.7	GE.3_1	GP.2_1	HLW.2_1	Til_cal.2_1	TGW-GS.1_1					
4	2H	31.0 - 102.3	GP.2_1	GP.3_1	HLW.2_2	Antho.1_1						
5	2H	52.3 - 124.1	HLW.2_2	SC.1_2	SD.1_2	SG.1_2	Til_cal.2_2	Til_mes.2_2				
6	2H	95.6 - 124.1	GP.2_2	GP.3_2	SG.1_2	Til_cal.2_2	Til_mes.2_2					
7	3H	4.5 - 29.2	TGW_GS.3_2	TGW-GS.1_2								
8	3H	61.1 - 90.4	HLW.2_3									
9	3H	81.7 - 114.6	EE.2_1	Mil.1_2	HLW.2_3							
10	3H	120.7 - 141.8	GE.3_2	GE.2_1	HLW.2_4	Til_cal.2_3						
11	3H	120.7 - 172.1	HLW.2_4	TGW.2_2	TGW_GS.2_2	TGW_GS.3_3	TGW-GS.1_3	Til_cal.2_3	Yld.2_1			
12	4H	0.0 - 18.8	GS.2_2	GS.3_2								
13	4H	41.2 - 64.4	BR.1_1	Mil.1_3	Aleu.1_1							
14	4H	72.8 - 109.0	EE.1_1	EE.2_2	EE.3_1	GE.2_2	GE.2_2	GE.3_3	Ht.1_1	Ht.2_1	Ht.3_1	
15	5H	0.0 - 43.6	Mil.1_4									
16	5H	20.4 - 54.9	EE.1_2	Ldg.1_1	TGW.2_3	TGW_GS.2_3	TGW_GS.3_4	TGW-GS.1_4	(Mil.1_4)			
17	6H	0.0 - 6.1	GS.2_3	GS.3_3								
18	6H	18.9 - 63.7	SC.1_3	SD.1_3	SG.1_3							
19	6H	95.2 - 134.3	GE.2_3	GP.2_3	GP.3_3							
20	7H	0.0 - 63.0	EE.1_3	EE.3_3	GE.3_4							
21	7H	38.7 - 91.3	Ht.3_2	TGW_GS.3_5								
22	7H	99.6 - 123.9	GP.2_4	GP.3_4								
23	7H	133.9 - 163.1	GS.2_4	GS.3_4	HLW.2_5							

The strong effects of TGW and tillering QTL detected on the short arm of chromosome 2H (2HS) were associated with the same genetic factor on bin 2. Bins 2, 11 and 16 contain the major genetic factors involved in the genetic control of yield components in the DH population. The closely mapped QTL EE.3_1, Ht.3_1 and GE.3_3 were all associated with bin 14 supporting the hypothesis of a common genetic control. The clusters of QTL tend to group traits phenotypically related or correlated such as yield components, disease susceptibility and grain and flour composition. Bins 13 and 18 correspond to genetic factors that seem to affect disease resistance and plant health

whereas bins 4, 6, 12 and 22 should be interpreted more in terms of the genetic control of grain quality. Some bins appear to make biological sense by grouping QTL for related traits that can help in the interpretation of the study. However, it needs to be born in mind that the large support intervals at QTL cluster describe only broadly the targets for investigating the genetic architecture of yield that can be used in association to additional mapping studies. (e.g. Chapter 4 on association genetics). In addition, the bins can be used to define the optimal contrasts of alleles and haplotypes that can be tested for validation of their effects using near isogenic lines.

2.4 Discussion

This QTL study uses phenotypes collected over different site and seasons that had a significant effect on the trait variation. The sites of higher fertility of Fowl09 and Fowl12 have been associated with increased tillering and grains sites per ear leading to larger sinks and poorer grain filling. In addition, the poor weather and untreated conditions in Fowl12 led to a significant increase in plant height, grains per ear, a drop of 14 g in TGW and delayed ear emergence. Therefore it was expected that the QTL mapping results would identify the different genetic controls that enable the parents Saffron and Retriever to better cope with their optimal environment.

2.4.1 Saffron × Retriever DH population

The genetic architecture of yield in current elite breeding material has been investigated in this project using a bi-parental mapping population of high breeding potential. The cross between Saffron and Retriever has produced the recommended line KWS-Discovery (Renamed KWS-Tower) and progenies of the latter have already been entered in the National List Trial system. The same population was used to validate functional polymorphisms of the *ant-2* gene involved in pigmentation of barley tissues that was detected initially in a mapping study of DUS characters (Cockram et al., 2010). These results are encouraging for the study of the yield components that may be controlled by alleles retained by selection in more recent varieties.

Although the initial seed quantity available was a limiting factor to achieve optimal replication, more than 200 DH lines were used in the study in both experimental years to produce robust phenotypes. Despite visible environmental variation, high heritabilities were obtained for the traits measured. The population revealed that higher yields were mainly achieved through higher TGW but the strong correlation between

yield components confirmed the compensation mechanisms occurring at a plot and plant scale. Interestingly, ear emergence was highly negatively correlated with TGW in the population suggesting that the timing of flowering influences the grain filling period and TGW. Earliness was beneficial for high TGW in both growing seasons. However this negative correlation was not supported by the behaviour of the parental lines as Saffron, a late emerging and maturing parent, appears to achieve higher TGW under first cereal conditions. Retriever proves to be the better variety under second cereal conditions where resources at establishment are limited. The variety produced a high yield at Elm09 but suffered from lodging and straw issues under high fertilisation and untreated conditions. These observations suggest that disease resistance coupled with stiff straw and early maturity may have been beneficial for barley yields in the conditions of 2009 and 2012 experiments. Transgressive segregation for all measured traits was encouraging for further QTL mapping on yield components, phenology, grain and straw characteristics and demonstrates the high potential of the population for understanding the genetics of these traits. The different responses of the parents to the site growing conditions and fertilisation regimes also indicates that the population can be used to further investigate environment interactions.

The extensive genotypic information collected on the population using the BOPA 1536 SNP markers added real value to this study and will allow comparison of the results with larger scale projects using overlapping genetic marker sets on similar platforms. Despite relatively few polymorphic markers, the S×R genetic map generated covers all barley chromosomes and the marker distribution agrees with published consensus maps (Close et al., 2009). The monomorphic segments on chromosomes 1H, 2H, 5H and 7H, confirmed with the genotyping of the parents with the higher density 9K iSselect SNP chip suggested close common ancestry. All BOPA markers are covered by the 9K platform which represents an additional source of markers to potentially increase marker density on the population. This is particularly interesting in the context of fine mapping and QTL validation in order to look for more recombinants in chromosome regions where initial marker density from BOPA1 was low.

2.4.2 Heading date and height

The genetic control of heading date and height in the population appeared to be related to two loci, excluding the spurious QTL on 5H. The control of heading date is well studied in barley and is controlled by genes affecting photoperiod, vernalization and earliness. The genes *Ppd-H1* (*eam1*, 2HS) and *ppdH2* (*eam8* or *HvFT3*, 1HL) control

the plant responses to photoperiod (Dunford et al., 2002; Faure et al., 2007; Griffiths et al., 2003). Three main loci *vrn-H1* (*sgh2*, 5HL) and v*rn-H2* (*Sgh1*, 4HL) and *vrn-H3* (*HvFT1*, 7H) determine vernalization response and the major differences between spring and winter types. *Vrn-H2* is a ZCCT transcription factor located on chromosome 4HL and is present in winter types (Dubcovsky et al., 2005). v*rn-H2* represses flowering by reducing expression of *vrn-H1* and *ppdH2* under long days affecting the timing of transition of the apical meristem from vegetative to reproductive state (Casao et al., 2011; Zitzewitz et al., 2005). QTL EE.3_1, Ht.3_1 and GE.3_3 suggest that the alleles of Saffron at that locus associate with a late ear emergence, taller plants and increased number of grains per ear that could be interpreted as longer ears. However additional QTL are involved in the control of grains per ear with Retriever being the positive parent. QTL for heading date and grain weight have been previously reported in that distal region of 4H (Backes et al., 1995; Bezant et al., 1997; Faure et al., 2007; Hayes et al., 1993; Tinker et al., 1996). Despite the two varieties being winter types, an allelic variation at *vrn-H2* locus could cause a change in onset of plant growth leading to variable growth rates and straw and ear lengths. However allele screening on a larger panel of winter varieties (NUE CROPS) revealed polymorphisms at v*rn-H2* within the winter barley varieties but none between Saffron and Retriever (Allan Booth, pers. comm.). Moreover the positioning of *vrn-H2* on GrainGenes consensus map next to 12_20760 at 118.34 cM (Close et al., 2009) does not overlap the QTL EE.3_1, Ht.3_1 and GE.3_3 significant SNPs A20454 and A10334, both mapped respectively at 101.62 cM and 103.11 cM. Faure et al (2007) reported an FT-like gene *HvFT5* co-segregating with the marker scsnp20989, close to A10611 but more than 15cM from A20454. Therefore although the biological interpretation of this particular cluster of QTL on 4H would strongly support a genetic factor involved in the control of phenology like *vrn-H2* and *HvFT5*, the QTL effect reported here appears to be distant enough to that locus to be associated with a novel genetic factor. This hypothesis will be investigated more closely by additional mapping and with increased marker density in Chapter 3 and Chapter 4.

2.4.3 Yield and yield components

Yield is the most important trait in barley breeding and results from the optimal expression of yield components in a given environment. The segregation for yield and yield components was partially explained by putative QTL detected on chromosomes 2H, 3H, 4H, and 7H. The initial expectation was to find associations of yield component

QTL with yield QTL that would ease interpretation and make biological sense as well as underpin the environmental yield response of the parents. This expectation was fulfilled in most cases with overlapping support intervals of yield component QTL at loci with significant effects on more than one trait (Figure 2.2), which have been identified as genetic factors in bins (Table 2.9).

Saffron yield potential on first cereal sites was supported by the unique yield QTL found on 3HL (Bin 11). This locus however does not include QTL for EE which has a strong genetic relationship with yield (Cuesta-Marcos et al., 2009; Li et al., 2005). Instead this positive yield effect appears to associate with a positive TGW effect. The SNP with the highest LOD for Yld.2_1 (A20952) maps 36cM away from the candidate TGW QTL mapped by Pansam et al (2012) (Barley OPA 2009 consensus (Close et al., 2009)). However this QTL for Yld.2_1 mapped in a chromosome region of the dwarfing gene *denso* where yield QTL have been mapped in spring barley (Li et al., 2005; von Korff et al., 2006). Here the bin 11 did not contain any QTL for plant height suggesting that the *denso* gene may not be a strong candidate for that genetic factor.

Unlike yield, the yield components were controlled by multiple loci with both parents contributing positive effects on these traits. At each genetic factor location, the direction of the QTL effect for the traits represented was in agreement with the correlations between the traits supporting a more plausible biological interdependence. Despite the poor growing conditions of 2012, the QTL results of 2009 were confirmed showing the consistency of the genetic control of those traits. The strong effect TGW QTL was mapped for the three mapping experiment on 2HS around 26cM on consensus OPA maps (Figure 2.1). At that locus, Retriever haplotype provides higher tillering but lower grain weight. This particular locus is located in the vicinity of the *Ppd-H1* gene, a candidate gene that has been associated with QTL for yield and yield components coupled with heading date (Schmalenbach et al., 2009; von Korff et al., 2006). However, no QTL for heading date was identified at that particular locus in any of the sites. *Ppd-H1* may still be a valid candidate gene as it is not excluded that a specific photoperiod response may affect more physiological pathways controlling tillering and TGW rather than phenology (Boden et al., 2015). For example, modifications of the tiller development in the early stages of growth and even grain filling in a later stage are plausible underlying mechanisms. The QTL candidate associated with genetic factor 2 will be subject to further investigation in a later chapter.

The Retriever haplotype alleles at the tillering QTL on 2HL around 100cM on S×R genetic map are associated with a strong increasing effect for straw collapse (SC.1_2)

and weaker positive effects on straw disease and stay green. A QTL for plant height has been described in that region in mapping studies (Hayes et al., 1993; Pasam et al., 2012) and Comadran et al., (2011b) detected a significant increase in heading date (A20366) associated with an increasing effect on harvest index at that location (A10376). However, no significant effects on plant height and heading date were reported in the present study. It is possible therefore that high tillering plants with weaker stems from the Retriever haplotype may well be more prone to straw collapse. The slight overlap with the confidence interval for the grain proteins QTL GP.3_2 brings only some evidence for associating these QTL in a single genetic factor hence the separation between 5 and 6 (Table 2.9**).**

It might however be possible to break the linkage between these two genetic factors.

Additional improvement of Retriever could be achieved by targeting the Saffron haplotypes on 3HL that combines QTL for increased TGW and tillering at the expense of grains per ear. Indeed, this tillering effect might result from a strong grain per ear QTL, a trait used to calculate tillering (Til_cal). Comadran et al., (2011b) mapped a QTL for harvest index at marker A10918 around 10cM proximal from GE.3_2 (A20527) based on consensus map distances. The yield QTL Yld.2_1 is too distal on 3HL to be directly associated with that marker and remains a valuable target for targeted breeding. The gene *denso* located in the vicinity could be considered as candidate although no significant plant height effect was observed. Other genes such as *int1* and *vrs4*, *als* (Dabbert et al., 2009; Dabbert et al., 2010; Koppolu et al., 2013) that are also located on the 3HL chromosome may also be considered as candidates underlying these yield components QTL but the effects detected on 3HL appear to be novel.

Although chromosome 7H contains QTL for GE and TGW, the low magnitude of their effects and large support interval suggest caution in interpreting the results and selecting a particular haplotype. The GE QTL GE.3_4 of 7H was associated with a QTLxE interaction for heading date in the region of HvFT1, a gene candidate for flowering time in barley orthologous to the rice gene OsFTL10 that has significant effects on heading date (Pasam et al., 2012). Low TGW effects were also detected in the centromeric region (TGW_GS.3_5). Chromosome 7H carries QTL with larger support intervals which impede a precise understanding of the location of the effects and the associated effects of the genetic factors. The fine mapping approach through association genetics could provide clearer patterns of the genetic control of those traits on 7H.

It becomes apparent that delivering sensible markers for selection of yield and yield components by the single method of bi-parental mapping is possible but not optimal.

The study shows that QTL with strong effects and narrow intervals would be more reliable and so the putative QTL for TGW and tillering found on 2H should be prioritized for validation over those on 7H. Yield and yield component mapping tends to capture effects associated with genes relevant to environmental adaptations such as vernalization and photoperiod genes (Cuesta-Marcos et al., 2009) or major ear morphology differences affecting row number and therefore yield. Few studies have reported QTL for yield and yield component in elite material and associated with a candidate gene that does not affect phenology or inflorescence type. The bi-parental mapping of yield and yield components complemented by fine mapping experiments therefore should be a novel opportunity to identify new genes and alleles controlling those traits exclusively.

2.4.4 QTL for grain quality

This study showed that Retriever's ability to produce high yields under low nitrogen regimes such as Elm09 was also associated with very low grain nitrogen compared to Saffron. The good mapping resolution on grain proteins QTL obtained and the small overlap with yield component QTL on 2HL favour the hypothesis of an independent control of the trait in that cross rather than the result of protein dilution effects in the grain. Strong genetic control of grain proteins was found on 2H (GP.2_1, GP.2_2, GP.3_1, GP.3_2) and 7H (GP.2_4, GP.3_4). GP.3_1 on 2HL mapped in the region of QTL6_CPC (Pasam et al., 2012) and QPc.nab-2H.1 (Marquez-Cedillo et al., 2001). Pasam et al (2012) also identified QTL8_CPC and QTL22_CPC at similar positions to GP.3_2 and GP.3_4 respectively based on barley OPA consensus map (Close et al., 2009). GP.3_2 is in a region of straw quality QTL that may well be affecting mechanisms such as nitrogen transfer in senescing growth stages. However, straw characteristics were only measured in 2012 and additional experiments may be required to confirm that hypothesis. The smallest effect QTL GP.3_3 is mapped 50 cM away from the barley homologue of the wheat gene *Gpc-B1* (Uauy et al., 2006) positioned on barley chromosome 6HS (Distelfeld et al., 2008; Jukanti and Fischer, 2008) and neither Mickelson et al., (2003) or Pasam et al. (2012) reported QTL effects this region. Retriever appears to have a rather unusual nitrogen metabolism by maintaining low grain proteins that could result from atypical nitrogen uptake or utilisation efficiency metabolism. The variety's response under lower nitrogen conditions and its high tillering abilities suggest that it is able to benefit in early establishment stages by having better uptake efficiency and better scavenging capabilities. Differences in barley root

structures have been observed between varieties (Hargreaves et al., 2009) that can directly impact on resource uptake and grain yield (Bertholdsson and Kolodinska Brantestam, 2009; Chloupek et al., 2006; Svačina et al., 2014). The study suggests that underground differences may exist between Saffron and Retriever and may relate to the above ground performance of traits that are segregating in this population. In addition, the low nitrogen content of the grain in a crop with a comparable yield to Saffron suggests a lower nitrogen transfer and a low remobilisation to the grain which could be seen as better nitrogen utilisation. Therefore the QTL for grain proteins found in this population potentially represent an important insight in the context of studies on barley nitrogen use efficiency.

2.4.5 QTL for disease resistance and straw characteristics

The ability of the plant to complete its life cycle healthily by overcoming biotic and abiotic stress such as diseases and adverse environmental conditions is a key to maximising yield potential. Breeders strongly select for plant aspect and health at grain filling and senescence stages as these traits affect grain filling and the marketability of the variety (David Harrap pers. comm.). This mapping study suggest that Saffron bright and stiff straw is controlled by several genetic factors involved in disease resistance and straw quality traits scored in 2012. An important QTL cluster (Bin 1) distal on chromosome 1HS indicated that the Retriever alleles had a negative impact on plant health during the ripening stage. That QTL cluster contains the major effect for the DUS trait of ear glaucosity, mapped at SNP A21354 (E_Glau1_1, LOD of 88.1). The composition and structure of culticular wax has been advanced as a mechanisms for plants to impede the establishment of diseases (Serrano et al., 2014). In barley, the cuticle properties associated to wax crystals was shown to play a key role at the prepenetration stage of mildew spores (Zabka et al., 2008). The results of S×R strongly suggest that the wax crystals of Saffron provide a natural barrier to prevent the disease to establish well and therefore benefit the plant health until ripening. In addition, the extremely high significance of the QTL suggest that SNP A21354 is close the causal polymorphism and can be used as a selection tool in breeding programs.

The straw collapse seen with Retriever as a twisting and bending of straw above the last node, has a negative impact on the crop harvest. The principal genetic control of straw collapse was found on 2HL in association with a tillering QTL. A higher tiller number may lead to a weaker straw structure also more inclined to twist as no association of straw collapse with height was found.

Because the 2009 yield trials were grown with fungicides, the disease QTL from the 2012 phenotyping rarely match yield and yield components QTL. Mildew resistance is under the genetic control of a large number of genes and positive allelic variation has been observed in barley varieties of diverse geographical locations (Berger et al., 2012; Comadran et al., 2009; Dreiseitl and Krianová, 2012; Dreiseitl and Platz, 2012). Of the four QTL detected for mildew, two mapped in regions of known resistance genes. Berger *et al,* (2012) identified the *Mla* resistance cluster on chromosome 1HS where the SNP A21226 is associated with Mil.1_1. However, the co-location of Mil.1_1 with E_Glau.1_1 strongly suggests that the wax cuticule affect the plant escape mechanism to mildew in that cross. The main resistance locus of Retriever on 4H described by Mil.1_3 and BR.1_3, is mapped at the position of the known resistance gene *Mlg* (Aghnoum et al., 2009; Kurth et al., 2001) also linked to QTL for brown rust susceptibility (von Korff et al., 2005). Common resistance mechanisms for the two pathogens have yet to be established but it is possible that susceptibility to mildew may have affected plant response to rust pressure. This study identified a convincing mildew restisance QTL Mil.1_4 on 5HS. Comadran *et al,* (2009) detected a QTL for mildew resistance on the short arm of 5H in a similar region although no specific resistance genes supported that novel resistance locus.

2.4.6 Taking the population forward

The interesting QTL mapping results obtained for yield components and other agronomic traits revealed the potential of the population in understanding yield architecture. Despite frequent QTL publication, few studies report implementation of MAS for yield component QTL (Collard and Mackill, 2008). One reason might be the constraints of bi-parental mapping, limiting the mapping resolution and the evaluation of a range of alleles. For MAS, the identification of small haplotype blocks is an advantageous in order to focus on the genetic factor and avoid negative linkage drag. A more in depth analysis of the QTL is therefore necessary to validate reliable targets that could be used for MAS. The clustering of QTL delimited by bins of genetic factors has the potential to identify the precise location for a potential target. The bins, associated with SNP markers, are transferable to other mapping studies using the same SNPs. In order to increase the understanding on the QTL and the genetic architecture of yield in elite barley and also to provide confidence on QTL for marker assisted breeding, further research will be implemented to exploit the results of Chapter 2. The comparison of S×R QTL mapping experiment with GWAS using two diversity panels mapping made

of two-row winter barley varieties will be the focus of Chapter 4. A QTL validation experiment using near isogenic lines for a subset of genetic factors identified in S×R QTL mapping will be presented in Chapter 5.

Chapter 3

Association genetics in European two-row winter barley.

3.1 Introduction

Plant breeding contributes to the genetic improvement of crops by creating new varieties with advantageous traits and characteristics to satisfy growers and user requirements. The release of varieties producing higher yields is the main objective of competitive commercial breeding programmes of feed barley, while other agronomic and quality traits are considered secondary. Breeding progress in UK yield of barley has been estimated at 1% per decade (Rae et al., 2007), a trend mirrored by TGW increase. However, the breeding efforts aim to maintain the positive contribution of other yield components on yield as well. The genetic variation for tillering, grain number and grain weight in barley germplasm and their response to environment contribute to the differences in grain yield between varieties. Therefore the understanding of the genetic control of these yield components can offer additional application and targets for marker assisted selection (MAS) strategies aimed at yield improvement and stability. In this study, the dissection of yield by its yield components is presented as an approach to obtain increased resolution in the architecture of the trait and potentially target its genetic control.

Despite the increase in molecular marker and QTL studies in crops, MAS of quantitative traits has not had the expected success in commercial breeding programmes (Rae et al., 2007). One of the plausible reasons for this low technology transfer is the limitation of the common bi-parental mapping approach. With this method, allelic variation is restricted to the alleles of the two parents (Chapter 2) so that QTL effects are estimated in a set genetic background which does not describe the scale of diversity often handled in breeding programmes. In addition, the resolution of the mapping is limited by the number of lines, the marker density and particularly by the number of recombination events. The population size of most published barley mapping studies is around 200 genotypes (see 1.3.2). While these are ideal for demonstrating genetic variation and providing robust tests of the approximate map location of effects, the increased resolution can only be achieved by screening extremely large populations in order to assay a larger number of recombination events (Komatsuda et al., 2007). The

average support interval of QTL in the S×R population was 22.6 cM (Chapter 2). Despite identifying convincing QTL with strong effects over two years, such a genetic distance encompasses a chromosome segment containing many genes that could be undesired for selection, even if the positive allele in a nearby gene is retained. The scope of bi-parental mapping is therefore limited and refined mapping strategies are needed to achieve the resolution required to provide an effective breeding tool.

Genome wide association studies (GWAS) have emerged in barley in parallel with the development of marker arrays containing the thousands of SNP necessary for high resolution mapping studies (Waugh et al., 2009). The method relies on the accumulated recombination events occurring over multiple rounds of meiosis during the previous generations of population development and selection. Linkage disequilibrium (LD) is the measure of the level of non-random assortment of polymorphisms for pairs of loci across the genome (Gupta et al., 2005). LD can be exploited to genetically map trait polymorphisms in association with genetic marker alleles. In the closely related population of barley cultivars, LD figures are quite high (about 10cM) whereas for the wider population of all UK material collected over the last 50 years the average LD is 1.2 (Cockram et al., 2010). Comadran et al., (2011a) showed that almost 90 % of the illumine 9K chip makers could be mapped by LD within 5cM of their expected consensus position verifying the potential of LD for fine mapping of simple traits and QTL of strong additive effects. Compared to bi-parental mapping, GWAS offers the advantage of investigating allele effects over a wide genetic diversity by using panels of varieties with different origins. Panels commonly encountered in barley GWAS include different geographical origins, morphology and growth habit which have been selected apart over years to optimise environment adaptation and product quality. These relationships between varieties are responsible for population structure in GWAS mapping panels (Pasam et al., 2012; Wang et al., 2012) and the relatedness between lines can lead to false or misleading associations due to pedigree rather than true linkage (Price et al., 2010). The distinction of stratification due to major genes such as *ppd-H1* in spring barley can also provide a better control of population structure and increase mapping resolution (Alqudah et al., 2014). Generally, genetic markers can be used to capture some of the population structure in mapping models and be included in statistical models to account for it (see 1.3.3). Therefore, the selection of a panel of two-row winter barley varieties that aims to limit population stratification is a strategy to

reduce extreme genetic variation due to major genes and focus on a closely related sub-group of the crop.

GWAS have been used to investigate yield as well as standard agronomical and morphological traits in barley (Comadran et al., 2011b; Inostroza et al., 2009). Recently, international resources have been pulled together to develop the base studies in barley association genetics with the creation of the AGOUEB (www.agoueb.org/) and Barley-CAP projects (Waugh et al., 2009). In barley-CAP, 10 US breeding programmes have shared genotypic and phenotypic information to carry out association mapping studies (Wang et al., 2012). Berger et al, (2012) analysed the six-row winter barley breeding programme of Virginia Tech to identify marker trait associations with agronomic traits of interest. GWAS was successfully implemented in European barley material with the AGOUEB project (http:// www.agoueb.org/) that focused on analysis of genotypic and phenotypic data collected during variety registration procedures on UK material (Waugh et al., 2009). The project delivered substantial QTL mapping results to the barley community with genotypic and phenotypic data made available to breeders through the Germinate database. In AGOUEB, morphological DUS traits data has been used for tight mapping, leading to the characterisation of functional polymorphism in the gene *Ant*-2 responsible for differences in anthocyanin pigmentation in barley (Cockram et al., 2010). The project was used to help the characterisation of the locus for *Int-c* discovery (Ramsay et al., 2011). This successful targeting of functional genes makes the method attractive for the dissection of the genetic control of complex traits.

Initiated in 2010, the NUE-CROPS project aimed at improving Nutrient efficiency on wheat, oilseed rape, potatoes and maize (see 1.2.5). It uses winter barley as a model crop for small grain cereal and the GWAS approach to identify candidate chromosome regions involved in the genetic control of important agronomic traits in wheat. The panel comprises 166 winter barley varieties selected to represent allelic diversity in European germplasm. The panel was grown in field conditions under different nitrogen regimes that included an optimal feed regime corresponding to a typical fertilisation regime for UK grown barley. The project produced robust estimates of standard agronomic traits of interest for breeders to be used in a GWAS (Thomas et al., 2013). This project provides a concrete opportunity to dissect yield related traits in order to investigate the genetic architecture and variation under optimal field growth conditions. In addition, the NUE-CROPS project takes advantage of the previous experience and

knowledge acquired during the AGOUEB project. Both projects share a core set of varieties as well as project specific two-row barley varieties from the range of European germplasm. More importantly, the project takes advantage from the recent advances in genotyping technology and the elaboration of Illumina's 9K chip to increase marker density and mapping resolution (Comadran et al., 2012). This array contains the SNP of the Oligo Pools Assays (OPAs) (Close et al., 2009) used in AGOUEB and Barley-CAP, and for the genotyping platform in Chapter 2 so that common SNP can be used to bridge across multiple mapping studies.

It has become apparent that MTA results from GWA studies represent an opportunity to overcome some of the limitations from bi-parental mapping by simultaneously providing increased mapping resolution and screening wider genetic variation. The focus on two-row winter barley suggests that the strong population structure commonly described in association mapping studies can be controlled voluntarily by designing adapted panels with reduced stratification. This chapter presents the second QTL mapping experiment carried out in the project for the investigation of genetic architecture of yield and yield components in winter barley. It focuses on two-row winter barley varieties from the NUE-CROPS and AGOUEB projects to carry out individual GWAS scans for a range of over 20 agronomic traits including yield and yield components.

3.2 Material and methods

3.2.1 NUE-CROPS project

Germplasm

The NUE-CROPS winter barley variety panel was composed of 166 including two-row and six-row types including most of the European winter barley diversity. 144 varieties were used in a first year of trial and 22 varieties were swapped for trials in the second year. Although the NUE-CROPS project collectively analysed all the 166 varieties as the association panel, this analysis will be focusing exclusively on the two-row barley subset of 126 varieties (Appendix 3. 1). The date of registration for all varieties recorded by the International Union for the Protection of New Varieties of Plants was collected and used as a trait in mapping (UPOV).

Genotyping

All varieties were genotyped by the James Hutton Institute with the Illumina 9K SNP chip (Comadran et al., 2012) using seed sources from JHI and KWS. 9K markers were renamed so that SNP present in the BOPA1 and BOPA2 assay had a prefix letter A and B respectively while 9K specific marker had the prefix I (Supplementary data 1). Genotypes of NUE and AGOUEB sets were compared using additional genotyping on 384 SNP for the 144 varieties grown in 2010. A dendrogram using all sources of genotypes available for the varieties was drawn using PAST (Hammer et al., 2001) and analysed so that mismatches between names and genotypes due to multiplication issues were corrected. The variety Hanna (NUE line 61) did not have available 9K genotype and was excluded. The genotype data of 166 varieties was processed by excluding markers with more than 50 % of missing values (4327 SNP remaining) and minor allele frequencies below 10% to produce genotypic information usable for GWAS scans (4255 SNP remaining). Heterozygotes and missing values were marked as "NA". The exclusion of the six-row barley genotypes resulted in a final cleaned matrix G_{NUE} for the two-row panel made of 125 varieties and 4041 SNP. Genotypes were coded at each SNP with their corresponding alleles for use in TASSEL and numerically transformed in a binary format 1 and 0 for association genetics scripts available in the statistical package R. The marker positions of the 9K SNP were provided by the JHI and based on map positions in the Morex × Barke RIL population together with LD mapping for additional SNP allowing allocation to map bin (Comadran et al., 2012).

Phenotyping

Phenotyping was carried out as part of the NUE-CROPS project. At each of the five trial sites, 144 varieties were grown in an incomplete split plot design with 48 genotypes replicated at each of the three fertilization regimes: "N1" had no nitrogen applied, "N3" was the optimal nitrogen rate for the site according to the fertiliser manual (RB209) (DEFRA, 2010) and "N2" corresponded to an intermediate rate for which the available N was set to one third of the optimal nitrogen rate (Sylvester-Bradley and Kindred, 2009). The sites were classified as second cereal sites to minimise residual nitrogen and the nitrogen supply (N supply) for each site was calculated as the sum of residual soil nitrogen + added nitrogen. Trials were carried out in 2010 and 2011 at the JHI in Dundee (JHI10, JHI11) and KWS UK in Fowlmere (KWS10, KWS11) with an additional site in central Germany (Seligenstad) in 2011 (LOC11). Phenotyping of agronomic and yield data was collected (Thomas et al., 2013).

A measure of heading date (Hd)[1] and height (Ht), the presence/absence of sterile spikelets and anthocyanin colour of the variety (Antho) were taken from the plots. Grab samples were taken on each plot at flowering (GS61) and maturity (GS87). At GS 61, a random sample of 10 to 20 plants was taken from the centre of each plot. The number of fertile tillers per plant was measured and the straw dried at 80°C for 48 hours and milled using a 0.3mm sieve for further NIR analysis. At GS87, plants were pulled out from 60cm of a row and counted. The number of plants and fertile tillers per plants were counted to provide independent measures of tillering as tillers per plants (Stems) and tillers per meter square of plot (Till.GS)[1]. The number of grain per ear (Grains)[1] and thousand grain weight (TGW-GS) were recorded from the grab sample. Straw and grains were dried at 80°C and 40°C for 48 hours respectively and weighted. Harvest index was calculated (HI). Grains were milled using a 0.5mm sieve and flour was analysed by NIR for nitrogen content (GrainN)[1] using a Group10 calibration from Aunir (www.aunir.co.uk). The dried straw from GS61 was analysed by NIR to measure stem nitrogen (StemN). At maturity, the plots were harvested for yield (Yld) and straw yield (SYld) and Biomass yield (BYld) = Yld/HI calculated. The TGW was measured on the combine sample (TGW) and used with Grains trait to calculate the tiller number (Till.Yld[1]). Nitrogen economy traits were calculated: Nitrogen Use Efficiency (NUE) = Yld / N supply; Grain nitrogen yield (GNYld) = GrainN × Yld; Straw Nitrogen Yield (SNYld) = StemN × SYld; Total nitrogen yield (TotNYld)= SNYld + GNYld; Nitrogen Uptake efficiency (NUpE) = TotNYld / N supply; Nitrogen Utilisation efficiency total (NUtEt) = BYld / N supply and Nitrogen Utilisation efficiency in grain (NUtEg) = Yld / TotNYld.

Although extensive phenotyping was carried out during the NUE-CROPS project, this study focuses particularly on phenotypes measured at the highest fertiliser rate (N3). This intend to provide a QTL study with results transferable to the farming practice and growing conditions of cereals in the UK. Indeed, the high fertilisation regime following guidelines of the fertiliser guide RB209 (DEFRA, 2010) is common practice for crop growing conditions in UK farming industry and in KWS-UK breeding trials . In addition, phenotypes collected on the higher fertilisation regime will best match the growing conditions utilised in the AGOUEB data set and the S×R QTL mapping study presented in Chapter 2.

[1] Different abbreviations for traits were used in this chapter and correspond to abbreviations in Chapter 2 as follow: Grains = GE; Till.Yld = Til.cal; Till.GS = Til.mes; Hd= EE; GrainN = GP

Statistics

All sites were analysed individually using the spatial REML analysis algorithms in GenStat 14th Edition (Payne et al., 2009). Variety means were predicted at each site as Best Linear Unbiaised predictors (BLUPs) and presented in a project report by Thomas et al., (2013). A multisite REML analysis accounting for the most parsimonious spatial model at each site was carried out to calculate overall variety BLUPs and variety by fertiliser BLUPs. The variance components were calculated in a mixed model fitting fertiliser level as a fixed effect and all other components as random. The estimate of heritability was obtained as a percentage for all traits using variance components of random factors (Thomas et al., 2013). For the reasons stated previously, only the phenotypic means (BLUPs) calculated for the N3 fertilisation rate from the 5 experimental sites were used in the association mapping exercise.

3.2.2 AGOUEB two-row winter barley analysis

Germplasm and genotyping

The barley collection used in the AGOUEB project includes spring and winter barley lines entered in national list trials in the UK. 201 two- and six-row winter barley varieties were genotyped using DNA from seeds grown during the AGOUEB project and Illumina 9K chip (Comadran et al., 2012).

The subset of two-row winter barley varieties from the larger AGOUEB panel was based on phenotypic DUS records and confirmed by the expected haplotypes at the *vrs1* and *int-c* loci involved in the control the two- and six-row phenotype. Four lines were excluded due to a mismatch between haplotype pattern and phenotypes recorded in AGOUEB dataset (Hurricane, Oleron, Sarah and Askanova). The final subset of 179 two-row winter barley is referred as the AGOUEB panel in this project (Appendix 3. 1). The genotypes of varieties in the AGOUEB panel were processed using the procedure presented in paragraph 3.2.1, resulting in a matrix G_{AG} of 3989 SNP.

Phenotypes

The phenotypes for agronomic traits and DUS for AGOUEB panel were downloaded from the Germinate database accessible to KWS-UK as an ex-consortium member (http://ics.hutton.ac.uk/germinate/). Agronomic traits recorded in national trialling system procedures were available as BLUPs from statistical models reported in AGOUEB. Traits available from treated trials were heading date (Hd); Straw length

(SL)[2]; straw strength (SS)[2]; Hectolitre weight (HLW); Thousand grain weight (TGW) (only yield component available); Yield (Yld); Grain Nitrogen content (GrainN); Hot water extract (HWE); Winter Hardiness (WintH); Brown rust (*Puccinia hordei*) (BR); Mildew (*Blumeria graminis f. sp. hordei.*) (Mil), Net blotch (*Pyrenophora teres*) (NB); Rynchosporium (*R. secalis*) (Ryncho); colour of aleurone layer (Aleu); Anthocyanin pigmentation (Antho); ear glaucosity (Ear-G)[2]; Ventral furrow hairs (VFH); lower leaf hairy leaf sheath (LLHLS); Sterile spikelets (StS). Some phenotypes were available from untreated trial data (identified by U). Since data used in the project was collected through the official national testing system it is unbalanced with the number of years for each variety varying with its success. In the statistical analysis for AGOUEB, BLUPs accounted for most of the environmental variation associated with sites and year effects (Wang et al., 2012) and were directly used as phenotypic information in the association mapping.

3.2.3 Association mapping

Population structure

Both panels described in paragraph 3.2.1 and 3.2.2 were analysed independently to study their underlying population stratification using genotypic information. Let g_{NUE} and g_{AG} be the matrices genotypes characterized as 0, 1, NA for the two panels. Element g_{NUEij} and g_{AGij} being the genotype of variety i at marker j, where i = 1 to N (the number of varieties in the panel N_{AG} and N_{NUE}) and j = 1 to M (the number of informative markers associated with each panel determined by MAF>0.1, missing value < 50% and constrained for pairs of markers with $r^2<0.9$; M_{NUE} = 1639 and M_{AG} = 1599). Dendrograms showing genetic distances were produced from g matrices using the cluster analysis function PAST with a paired group algorithm and Hamming similarity measures. Additional matrices X_{AG} and X_{NUE} of standardised genotypes were created by subtracting the column mean before dividing by its standard deviation for each variety in column j. The average relatedness of the lines or Kinship matrices K_{AG} and K_{NUE} of size N_{AG} and N_{NUE}, were computed from X_{AG} and X_{NUE} using the correlation function in the R statistical package (Team, 2013). Further Eigen decomposition of K_{AG} and K_{NUE} was performed with R (eigen function). The first 10 eigenvectors Q_{AG} and Q_{NUE} were regarded as 10 axis of variation capturing population structure to be used in PCA corrected association mapping models (Price et al., 2006). To analyse the overall

[2] Different abbreviations for traits were used in this chapter and correspond to abbreviations the following abbreviations in this document: SL = Ht; SS = Ldg; Ear_G = E_Glau

diversity for the 226 different varieties used in the study, the 1284 SNP in common between g_{AG} and g_{NUE} were pooled in matrix $g_{AG + NUE}$ in order to calculate the correlation matrix between the whole set of varieties. The latter was analysed by PCA using PAST software (Hammer et al., 2001) to identify distribution of varieties according to their panels or specific SNP allele.

Statistical models comparison

To carry out the GWAS, a collection of models and software calculating association tests accounting for different measures of population structure were investigated for the NUE-CROPS data.

A generalised linear model (GLM) script in R (Team, 2013) was used to carry out a naïve approach (R-GLM_naive) using tests for association between raw phenotype and each individual marker. The correction by PCA was also implemented in GLM (R-GLM_Q_{AG} or Q_{NUE}). The script predicts phenotypes and genotypes using the product of the Q matrices with the raw phenotype and normalised genotype matrices respectively. The predicted phenotypes and genotypes are then subtracted from the raw phenotypes and normalised genotypes respectively so that population structure has been accounted for. An *F*-test investigating the null hypothesis of no association between the phenotype residuals and marker residuals is carried out and –log10 (*p-values*) reported. The package EMMA (Efficient Mixed Model Association, http://mouse.cs.ucla.edu/emma/news.html) in R (Team, 2013) was also evaluated for a mixed linear model approach (MLM) to account for population relatedness (Yu et al., 2006). The functions EMMA.ML.LRT and EMMA.REML.t were used to test associations on a naïve (R-MLM_naive) and Kinship correction for population structure (R-MLM_K) respectively.

Additional analyses were performed with TASSEL 3.0 (http://www.maizegenetics.net), which offered GLM and MLM algorithms, together with quicker computing time, flexibility of data manipulation and the additional information on the test carried out (Bradbury et al., 2007). The naïve and PCA models using Q matrices were investigated (Tassel-GLM_naive; Tassel-GLM_Q). Additionally MLM was carried out with naïve (Tassel-MLM_naive), a Kinship (Tassel-MLM_K) and PCA and Kinship correction (Tassel-MLM_Q+K). Distributions of *p*-value obtained from the different models were compared to the expected distribution in a quantile-quantile plot (QQ-plot) to identify the best fitting model to minimize spurious QTL identification.

QTL significance threshold

A preliminary strategy to identify a suitable QTL significance threshold was used by calculating the False Discovery Rate (FDR) to control for the expected proportion of type I errors based on the number of independent tests (i.e. function of the number of independent markers). Similar to Pasam et al., (2012), the False discovery rate (FDR) calculated for each trait was found to be highly stringent and potentially could reject biologically real effects (Peter Werner pers. comm.). A more liberal approach could be taken by considering the bottom 0.1 percentile of the distribution of p-values. Limitations of this approach were identified when GWAS scans were carried out on the non-reduced genotypes matrices G_{NUE} and G_{AG} that included a large number of markers with identical *p*-values caused by high pairwise r^2 and an identical or close mapping position. Pasam et al., (2012) used the distribution of p-value to identify a relevant threshold of –log10(0.03)=1.5 to declare MTA. In this study, a compromise was made to retain a moderate number of associations mindful of the risks of false positives while providing higher stringency to Pasam et al., (2012). For both NUE-CROPs and AGOUEB GWAS the significant MTA are defined at a threshold of -log10(0.003)=2.523 which maintains a high stringency in defining a QTL.

An arbitrary QTL support interval size was fixed at +/- 5cM from the highest significant marker (Bill Thomas pers. comm.).

3.3 Results

3.3.1 Phenotypes

NUE-CROPS

The detailed statistical analysis of the complete NUE-CROPS winter barley panel including six-row and two-row types is available in the publishable report of NUE-CROPS WP1.1 (Thomas et al., 2013). Only the phenotypes (BLUP) of the two-row winter barley subset grown at high fertilisation regime were used for this study. The summarized statistics on BLUPs (Appendix 3. 2) show that the average yield of the panel was 6.94 t/ha with a maximum of 8.13 t/ha achieved by KWS-Glacier, comparable to current yields on farm. Retriever confirms its high yield potential under second cereal conditions with the second highest yield of 8.02t/ha while Saffron ranks 11^{th}. Substantial phenotypic variation was observed for the traits analysed within this

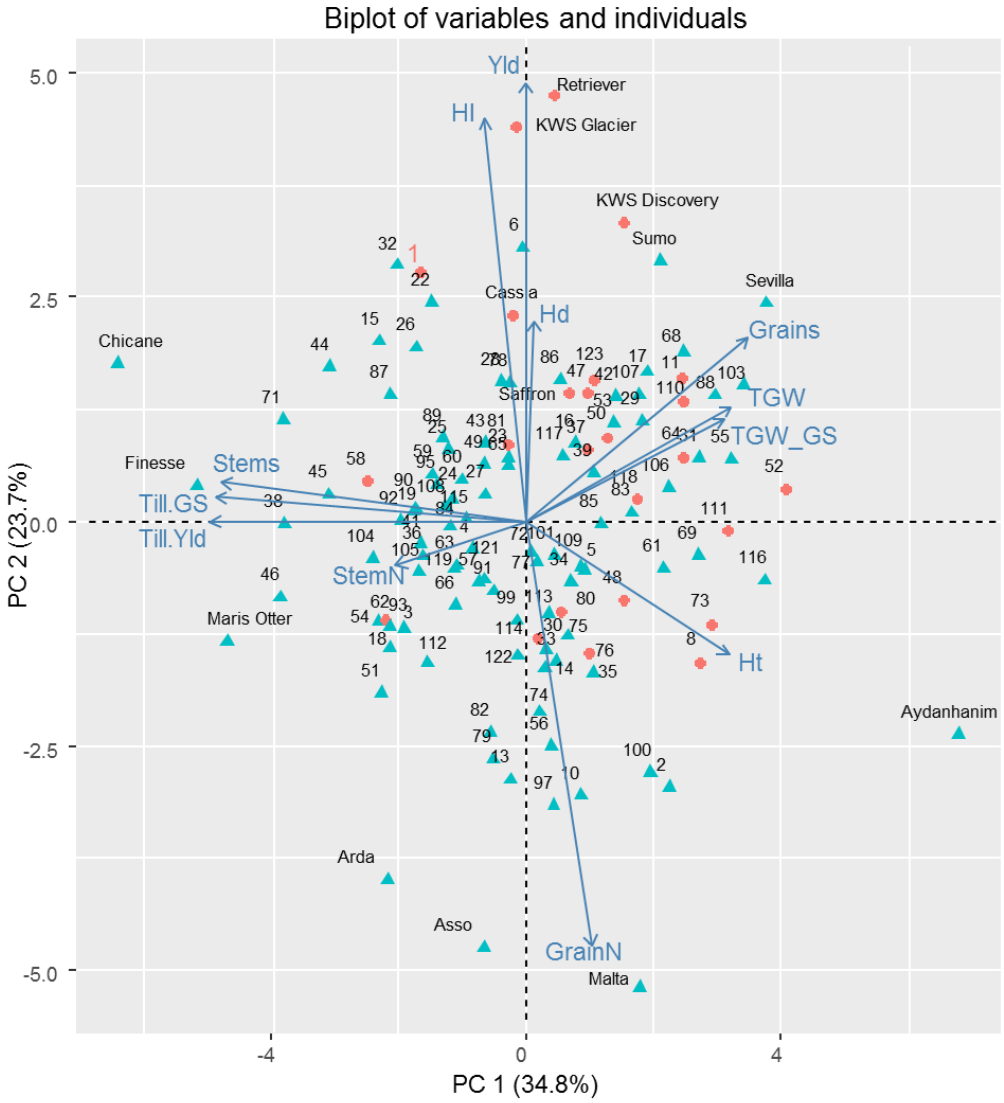
subset. A summarizing biplot including yield and yield components traits shows that 76.7 % of the panel variation could be captured by the first two first principal components (Figure 3.1a). It is noticeable that the two yield components of tillering and grains per ear have loading in opposite directions (Appendix 3. 3a), mirroring the negative correlation between the traits. The biplot shows that increased yield performance seems to benefit from higher TGW and is barely affected by either Grains per ear or tillering. Varieties at the centre of the biplot would therefore tend to be average for all those traits, including yield whereas in extreme phenotypes the balances between yields components can be pulled out. Saffron and Retriever are located along the TGW axis, with Retriever showing greater TGW and yield than Saffron, typical from the variety response to second cereal growing conditions (see Chapter 2). Chicane and Finesse have high tillering abilities whereas Marinka and Aydanhanim have higher grains number per ear but low yields. The recently released varieties KWS Discovery and KWS Glacier have average tillering and Grains per ear but are capable of achieving higher TGW and therefore higher yields. The spread of varieties over the yield and yield component axis underpin the large phenotypic variation of the panel to be targeted in association mapping. The increased complexity of the biplot caused by adding traits reduces the variance accounted for by the two principal components to 59% (Figure 3.1b) and increase the spread of varieties (e.g. *deficiens* varieties scattered across plot). The clear separation between yield components on the first two PC in Figure 3.1a) is now only captured through the first three PC in Figure 3.1b) (see also Appendix 3. 3b). The principal components loadings suggest that PC1 is attributable to the yielding potential ability of the varieties with Retriever, KWS Glacier and KWS Discovery being the highest yielding varieties of the panel associated with a better NUE (Appendix 3. 3b). Tillering is negatively correlated to TGW and Grains on PC2, although those three traits are separated out by the third PC.

The majority of nitrogen related traits are mainly positively correlated with yield with the exception of Grain Nitrogen which shows a strong negative correlation. The lower grain nitrogen associated with higher yields suggests a better utilisation of nitrogen resources for carbohydrates accumulation. Chicane and Finesse remain on the outside of the scatter with high tillering and short straw. Malta has very high grain nitrogen content and proved lower yielding in the field trials at all sites. The scatter of varieties overlaid over the traits in the biplot show that variation is substantial and well represented in the panel that is suited for further association mapping study.

a)

Figure 3.1 Biplot of BLUPs for NUE-CROPS traits
Principal component biplot presenting the distribution of 125 two-row winter barley varieties from NUE-CROPS for a) yield and yield components and b) (next page) an extended set of variables from the NUE-CROPS data set: (Grain Nitrogen, GrainN; Grains per ear, Grains; Harvest Index, HI; Heading date, Hd; Height, Ht; Stem Nitrogen: StemN; Stems per plant; Stems; TGW; TGW-Grab Samples, TGW-GS; tillering-Grab sample, Till-GS; tillering from yield, Till-Yld; Yield, Yld). The list of varieties is presented in b). The presence and absence of sterile spiklets is indicated as presence (▲) and absence (deficiens, ●).

Figure 3.1 b)



Varieties

1 Accrue	22 Clara	43 Frolic	64 Celebrity	85 Panda	106 Sevilla 3
2 Alpha	23 Clarine	44 Gaelic	65 Madrigal	86 Parasol	107 Tiffany
3 Amillis	24 Cobalt	45 Gleam	66 Magie	87 Pastoral	108 Torrent
4 Antigua	25 Coriolis	46 Halcyon	67 Malta	88 Pearl	109 Vanessa
5 Antonia	26 Diadem	47 Halifax	68 Malwinta	89 Pedigree	110 Vesuvius
6 Aquarelle	27 Diamond	48 Heligan	69 Marinka	90 Portrait	111 Vilna
7 Arda	28 Dolmen	49 Hermia	70 Maris Otter	91 Posaune	112 Vixen
8 Artist	29 Dolphin	50 Hurricane	71 Maris Trojan	92 Prelude	113 Willow
9 Asso	30 Duchess	51 Igri	72 Maritem	93 Puffin	114 Winner
10 Astrid	31 Duet	52 Intro	73 Masai	94 Retriever	115 Wintmalt
11 Avenue	32 Elmstead	53 Jessica	74 Medoc	95 Rifle	116 Babylone
12 Aydanhanim	33 Emeraude	54 Jewel	75 Melanie	96 Saffron	117 Cassata
13 Baraka	34 Emilia	55 Karisma	76 Molly	97 Sarah	118 Haka
14 Bronze	35 Epic	56 Kaskade	77 Murcie	98 Sevilla	119 KH Malko
15 Calcutta	36 Ethno	57 Kelibia	78 Musette	99 Sombrero	120 Cassia
16 Camion	37 Fahrenheit	58 Kestrel	79 Mystique	100 Sonja	121 Orchidea
17 Cannock	38 Fanfare	59 Kingston	80 Nectaria	101 Sprite	122 Rejane
18 Carat	39 Fighter	60 Kira	81 Saffron 2	102 Sumo	123 Tallica
19 Chestnut	40 Finesse	61 Labea	82 Nure	103 Sevilla 2	124 KWS Glacier
20 Chicane	41 Firefly	62 Leonie	83 Mortimer	104 SW Alison	125 KWS Discovery
21 Cinnamon	42 Flagon	63 Louise	84 Opal	105 Target	

AGOUEB

The BLUPs were sourced from the Germinate database available to breeders for mapping purposes. The range and summary statistics of the phenotypes recorded for the subset of the 179 two-row winter barley varieties is summarized in Table 3.1. In the official testing system, the only component of yield recorded is TGW. The measures of disease infection are extensively assessed. Substantial variation was observed for most traits, especially for disease resistance, which offer the possibility to enlarge the traits investigated in this study which were not present in the NUE CROPs experiment.

Table 3.1 Summary statistics on BLUPS of AGOUEB panel.
Summary statistics for the BLUPs of 20 traits of the subset of 179 varieties from the AGOUEB two-row winter barley panel. Traits were measured on treated conditions unless mentioned otherwise (U).The traits included were: heading date (Hd); Straw length (SL); straw strength (SS); Hectolitre weight (HLW); Thousand grain weight (TGW); Yield (Yld); Grain Nitrogen content (GrainN); Hot water extract (HWE); Winter Hardiness (Wint_Hard); Brown rust (BR); Midlew (Mil), Net blotch (NB); Rynchosporium (Ryncho); colour of aleurone layer (Aleu); Anthocyanin pigmentation (Antho); Ventral furrow hairs (VFH); lower leaf hairy leaf sheath (LLHLS); Sterile spikelets (StS).

	Hd	SL	SL (U)	SS (U)	HLW	TGW	Yld	GrainN	HWE	Wint Hard	BR (U)	Mil (U)	NB (U)	Ryncho (U)	Aleu	Antho	Ear_G	VFH	LLHLS	StS
Missing values	14	28	1	1	1	14	1	44	40	1	1	1	1	1	12	21	13	6	9	26
Mean	5	104	82	17	65	41.5	7.7	1.79	298	10	11.9	6.8	0.2	4.7	1.6	1.4	4.3	1.8	1.9	4.3
Minimum	1	88	62	4	61	32.1	7.0	1.68	286	9.4	7.5	0.8	-3.6	-0.7	1	1	1	1	1	1
Maximum	9	115	96	39	69	50.9	8.5	1.92	306	10.3	25.1	25.7	9.6	20.4	2	2	8	2	2	7
Range	8	27	33	36	8	18.9	1.5	0.23	20	0.9	17.6	24.9	13.1	21.1	1	1	7	1	1	6
Lower quartile	4	102	80	14	65	39.1	7.5	1.76	295	9.9	10.2	5.3	-0.7	2.3	1	1	3	2	2	4
Upper quartile	6	107	85	19	66	43.7	7.9	1.82	301	10.1	12.5	7.6	0.8	6.3	2	2	5	2	2	5
Standard deviation	1.8	4	5	5	1	3.4	0.3	0.04	4	0.2	3.0	3.1	2.0	3.5	0.5	0.5	1.7	0.4	0.3	1.6
Variance	3.4	16	23	26	2	11.5	0.1	0.00	19	0.02	8.8	9.5	4.2	11.9	0.2	0.2	2.8	0.2	0.1	2.5

Saffron was the highest yielding variety of the AGOUEB panel with a mean yield of 8.5 t/ha and was also one of the most recently released variety in that panel that was composed before NUE-CROPS. The distribution of disease resistance scores confirms the potential of the dataset for investigating disease resistance QTL. Because the official malting testing excluded feed barley varieties, traits related to malting such as HWE and GrainN have a higher number of missing values. This is potentially important when analysing those traits as it may impact on the allele frequency at some markers for the association tests. Similar to the NUE-CROPS results, grain nitrogen appears to be strongly negatively correlated to yield, supporting the effect of dilution mechanisms. The variety Willow had the highest GrainN and a yield of 7.28 t/ha ranked in the lower quartile. Additionally, a total of six DUS traits for two-row barleys was available in the

AGOUEB dataset. Those were investigated in GWAS using a smaller genotyping platform of 1536 SNP on the entire AGOUEB variety panel (Cockram et al., 2010) but the 9K chip may help to achieve higher mapping resolution on traits such as VFH and LLHLS.

3.3.2 Population structure in association mapping panels.

NUE-CROPS two-row winter barley

The genetic diversity of the NUE CROPS panel was captured with the correlation matrix based on 1639 SNP. The informative SNP of M_{NUE} were visualised by a dendrogram (Figure 3.2). The major clusters of genotypes reflect some underlying structure in the genotype data and this should be accounted for in the association mapping. As expected, the recently released varieties B100 (KWS-Glacier) and B99 (KWS-Discovery) are clustered close to their parental lines Saffron, Retriever and KWS-Cassia, which confirms the suitability of the marker information for capturing pedigree relationships. In Figure 3.2, B100 is genetically closer to its parent Retriever than its second KWS-Cassia. The varieties Sarah and Winner, both released in Europe in 1994, have a distinct allelic composition to the rest of the panel. Melanie and Orchidea have very similar genotypes suggesting that one of the varieties has been wrongly labelled in the project.

The distribution of the *deficiens* type into clear groupings suggests that the trait has been brought in the germplasm through a limited number of lines and is slowly being integrated into different genetic backgrounds. Although two main groups of *deficiens* varieties are visible, the trait is also found in more distant varieties (Leonie, Nectaria).

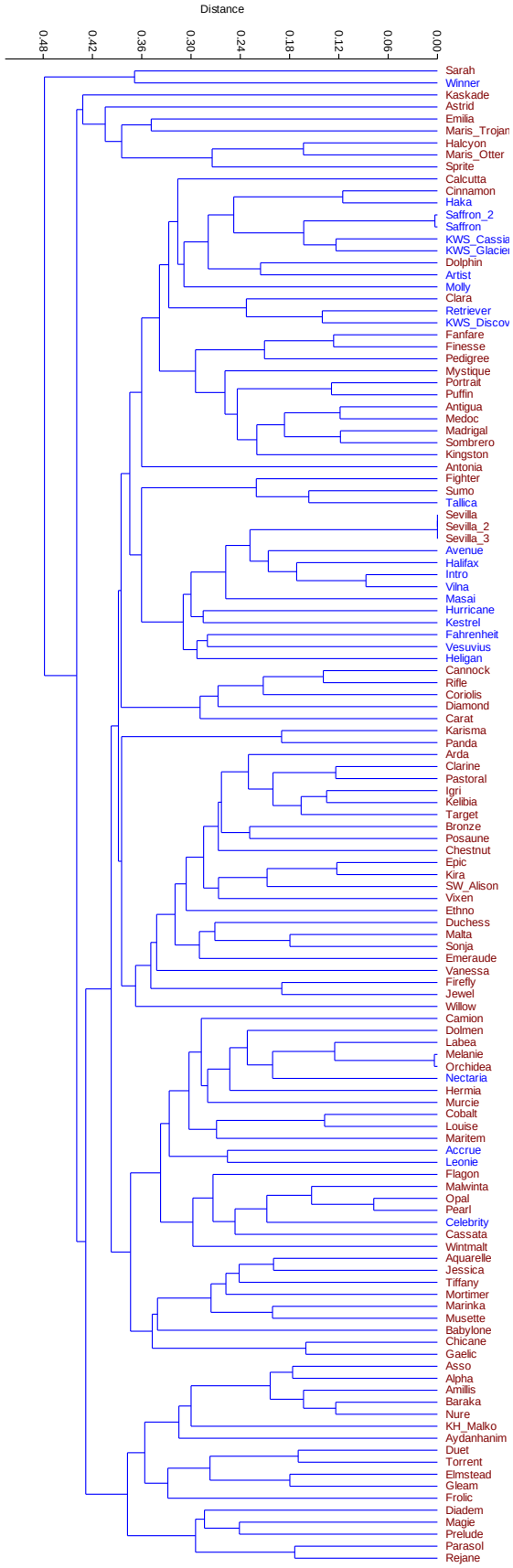


Figure 3.2 Dendrogram of diversity in NUE-CROPs two-row winter barley.
The dendrogram is based on a correlation matrix of the lines computed from M_{NUE}. Varieties indicated in brown have ears with sterile spikelets. The varieties in blue are *deficiens* type (i.e. displaying an absence of sterile spikelets).

AGOUEB two-row winter barley

A similar study of the population structure was carried out for the AGOUEB panel of varieties. After marker data processing, 1599 informative SNP were available for M_{AG}. The PCA presented in Figure 3.3 does not distinguish any major structural grouping in either NUE CROPS or AGOUEB panels unlike observed in other mapping panels of barley for GWAS where strong stratification was present (Cockram et al., 2010; Pasam et al., 2012). In addition, the 75 varieties in common between the NUE-CROPS and AGOUEB project cover most of the genetic diversity investigated. The strategy of analysing exclusively the two-row winter barley varieties shows that the strong levels of stratification associated with the two/six-row and winter/spring pools can be avoided. It should facilitate the comparison between mapping studies of the same crop type and the transfers of results to breeding. Although it was not apparent in the dendrogram (Figure 3.2), some of the oldest varieties (e.g. Malta and Alpha (1970's)) are on the periphery of the PCA cluster. Similarly to the dendrogram, the PCA confirm the pedigree relationship between varieties and highlights genotypes containing alleles of lower frequency in the winter barley germplasm or of more distant geographic origin. The lowest value on PC1 illustrate the genetic diversity of the variety Puffin ((Athos×Maris-Otter)×Igri) which has a quarter of its parentage from the French spring barley variety Athos. Puffin has thereafter been used as a parent in the pedigree of Opal and Pearl hence their close genotypic proximity (Figure 3.2, Figure 3.3). Nevertheless none of the lines appears to create a cluster significantly distant from the rest of the varieties.

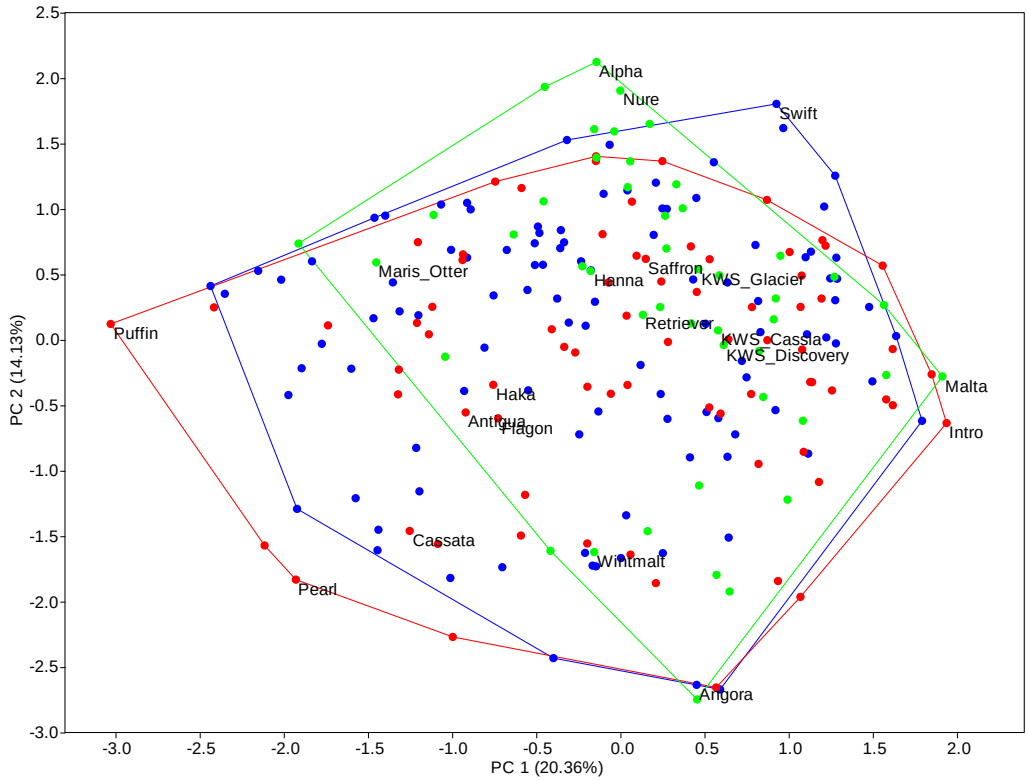


Figure 3.3 PCA of genetic relationships between 226 winter barley two-row varieties

The complete set of varieties used in the association panel is presented. For clarity, only a few varieties have been presented. The correlation based PCA uses a correlation matrix made from the 1284 common SNP markers between M_{AG} and M_{NUE}. The varieties represented by a blue dot are exclusive to AGOUEB winter barley two-row panel, in green to the NUE-CROPS panel and the red dots are varieties common to both panels. Axis are in Eigenvalue scale.

3.3.3 Association analysis

Models comparison

Different models were tested primarily to establish the need for population structure correction and secondly to find the optimal correction approach. Figure 3.4 presents a Q-Q plot example based on the TGW-GS for the NUE-CROPS data. The Q-Q plot shows the variable reduction in the number of significant associations from using the range of population structure corrections models in R and TASSEL in GWAS scans (see 3.2.3). In both mapping software, the Kinship structure correction using K_{AG} and K_{NUE} (Tassel-MLM_K and R-MLM_K models) was associated with a better fit than the correction by Q matrices only (Tassel-GLM-Q and R-GLM-Q models) (Figure 3.4). Alternatively, the Tassel mixed model using both K and Q (Tassel-MLM_K+Q) provided the best fit to the expected normal distribution of *p*-values. Similar observations were made from the analysis of the other traits in both mapping panels. For all comparable models, TASSEL resulted in a better correction than R with *p*-values

closer to the normal distribution (expected –log10(p)). The ranking of the marker associations between models were comparable and hence correlations between models for the *p*-values exceeded 0.97. This result suggests that the higher *p*-values obtained with the R models may inflate the number of QTL being falsely accepted at the detection threshold retained for this study. The Tassel-MLM_K+Q model was only slightly better than Tassel-MLM_K and both accounted adequately for population structure. The MLM_K method was also shown to better control for false positive and statistical power over structure correction models in barley GWAS (Pasam et al., 2012; Wang et al., 2012). Similar analysis done using the AGOUEB dataset alone confirmed that the Tassel-MLM_K model was the most appropriate to account for population structure. TASSEL software also provides additional information on the marker and statistical tests carried out (Marker effects and marker r^2) (Bradbury et al., 2007). Therefore, the GWAS scans for AGOUEB and NUE-CROPS presented in this study concentrate on the Tassel-MLM_K model.

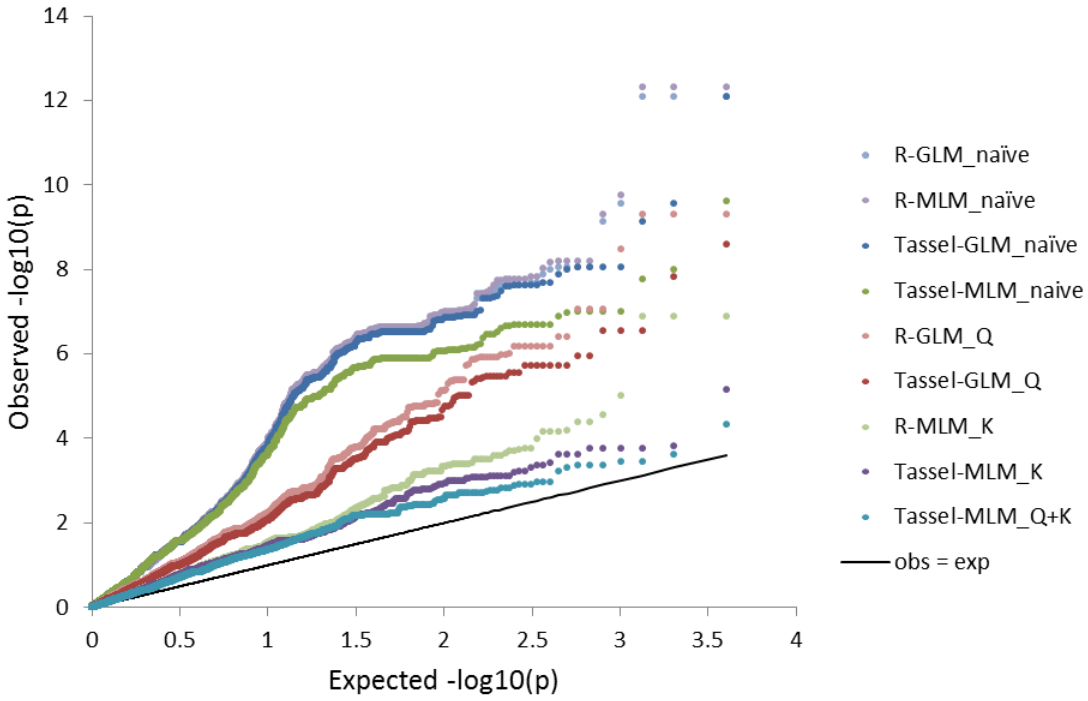


Figure 3.4 Quantile-quantile plots of *p*-values for the NUE-CROPS TGW-GS GWAS analysis using different population structure corrections.
Expected vs. observed P values are plotted for the generalised linear model (GLM) and mixed linear models (MLM) model in R and TASSEL including correction for population structure based on Kinship (K) and Principal components (Q). The x = y line (solid) expected distribution is indicated. The QQplot of p-values was computed from the TGW-GS GWAS results for 4041 SNP (GNUE).

NUE-CROPS: genome wide scans and QTL results

Several significant marker trait associations were found on the genome for the traits analysed (Appendix 3. 5). Manhattan plots for yield components illustrate which regions of the genome were involved in the genetic control of those traits (Figure 3.5). Generally, the QTL peaks were identified with a large number of markers above the detection threshold which included some markers with identical allelic distribution within the panel, hence with the same –log10(*p*). These markers often had identical map position and therefore could not provide additional resolution at the QTL peaks suggesting that multiple best markers for a QTL can be considered. The markers within a 10 cM window and with similar level of significance to the peak QTL (i.e. identical allele frequency and distribution in the panel) were used to define the QTL support interval.

The complete set of QTL identified in NUE-CROPS panel under standard fertilisation regime is summarised in Table 3.2 a) b) and c) and the genome wide results of tests of association are available as Supplementary data 2. The QTL detection threshold retained in that study (–log10(0.003)) enabled the reporting of up to 13 QTL depending on the trait analysed. NUE and GNYld have respectively 13 and 11 significant associations whereas StemN and Grains have only two distant significant associations. The range of higher significant markers (-log10(*p*)) was variable in the set of traits analysed and seemed to relate to their associated genetic complexity. For example, the DUS traits for presence-absence of pigmentation and sterile spikelets expected to be under the control of single genes *ant-2* and *vrs-1* respectively, were both mapped with a value for –log10(*p*)> 15. Pigmentation was mapped at marker I195164 at 96.8cM and sterile spikelet at A10287 at 85.9cM both on 2H in their expected positions. As the causative polymorphism is known and maps close to the peak SNP markers, both traits have additional significant MTA mapped at other locations in the genome. These results suggest that residual associations are present in the dataset and can be mapped with the QTL detection threshold. It may be that the traits are not completely controlled by the known major genes or that inter-chromosome LD hasn’t been entirely accounted for by the Kinship structure correction.

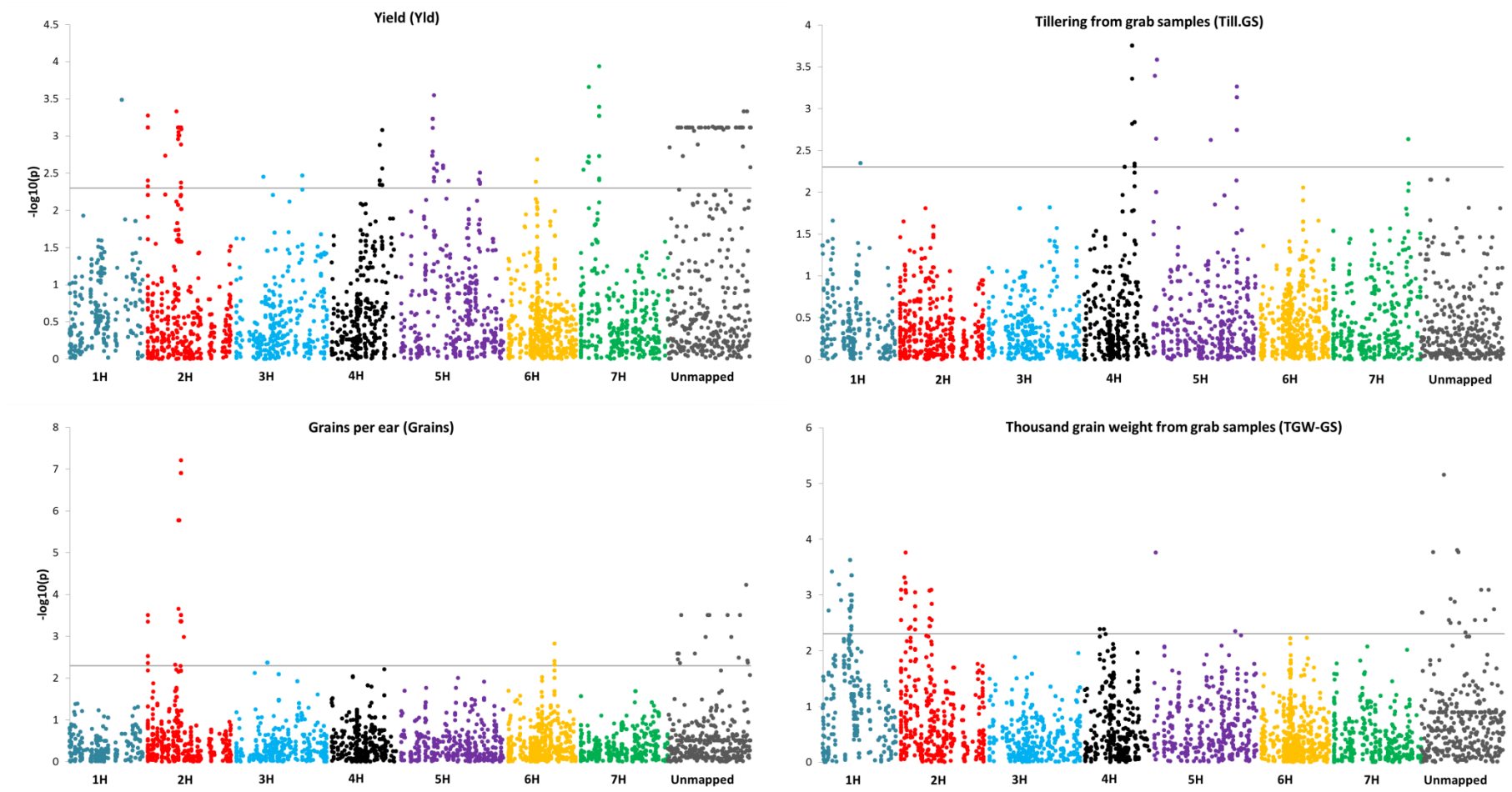


Figure 3.5 Manhattan plots for yield and yield components

Genome wide association scan for yield and yield components of the winter barley NUE-CROPS two-row varieties. 4041 SNP markers (G_{NUE}) were analysed for marker-trait associations using the Tassel-MLM_K model. Markers above a detection threshold of –log10(0.003) (horizontal line) were reported as QTL.

Traits with increased complexity tended to have more QTL with lower level of significance and rare associations with –log10(*p*)>5 (Table 3.2 a, b, c). Yield QTL were mapped at nine locations on six of the seven chromosomes. Six QTL are with LOD scores above three and with marker effects of similar magnitude. The strongest effects of Yld_1 and Yld_3 are observed for markers with the lowest MAF amongst the QTL for that trait. Yield components also show different levels of genetic complexity, with TGW having more QTL and Grains only showing significant associations at two loci but with Grains_1 having the highest –log10(*p*) for all yield components QTL with a LOD of 7.2 (Chromosome 2H, 63.5cM). The QTL of TGW obtained from the grab samples were all associated with QTL for TGW obtained from the combine sample except one, TGW-GS_5. The TGW obtained from the combine sample also had three additional QTL on 2H, 4H and 5H all with lower significance. The different methods of measuring tillering also produced a variable number of QTL. Till.GS_1 and Till.GS_2 however seem to be co localised with tillering from yield (Till.Yld_2; Till.Yld_3) and plant stems (Stems_2; Stems_3) suggesting that reliable markers could be identified to focus on the genetic control of the trait. However no outstanding large effects were observed for those QTL.

Five QTL for heading date were mapped on 1H 2H and 5H, four of which with a significant peak marker above –log10(*p*)=3. The strongest effects and significance level identified Hd_3 and Hd_4 on 2H in the regions of genes known to be involved with flowering time. Heading date (Hd_4) and height (Ht_2) were associated with strong effects and peak markers mapped at the same position on 2H at 63.5 cM. Although the common location of Hd_4 and Ht_2 suggests the presence of a gene with pleiotropic effect for the traits, the strong Hd_3 effect on 2H was not associated with a height QTL neither was Ht_7 on 5H with an effect on heading date. This result indicates that both traits are also genetically controlled by factors independent from each other.

Interestingly, a notable QTL UPOV_7 was found on 6H at 53 cM with a LOD of 4.21 and an effect of nearly 8 years and co-locating with yield QTL Yld_7, suggesting that the selected changes of alleles at that locus have also supported the genetic progress in yield over years.

All the traits corresponding to the plant nitrogen economy were found with significant associations. Thirteen QTL were found for NUE. The strongest –log10(*p*) score of 5.04 corresponded to the nitrogen uptake efficiency QTL NUpE_2. The expected associations of QTL of derived traits with their components were observed as the majority of yield QTL were also found at the positions significant for NUE, calculated

as the product of yield and available nitrogen. However, NUE had more QTL overall suggesting that the trait may highlight other important genetic regions for the control of yield in relation to available nitrogen. The more robust QTL for N utilization efficiency in the grain (NUtEg_2) was mapped at the same position as a Grain Nitrogen QTL (GrainN_2) on 4H 92.4cM, the latter being associated with QTL for NUE, harvest index, tillering and yield. NUpE_2 on 5H co-mapped with QTL for Grain nitrogen yield, TGW and tillering. This co-mapping of QTL clearly indicates the genetic relationship between traits and shows that the understanding of plant nitrogen economy is intimately associated with yield and its components. The biological relationship behind such co-mapping should be taken into account in the future comparison of mapping studies to avoid the misinterpretation of multiple gene effects and functions.

Table 3.2 Significant marker traits associations (QTL) of NUE-CROPs GWAS.
The table presents the significant QTL from the Tassel-MLM_K GWAS model on 20 phenotypes for the NUE-CROPS two-row winter barley study. Each QTL is presented with its peak marker (most significant SNP) with the associated marker map position, –log10(*p*) resulting from the test for association, the proportion of the genetic variation (R^2) of the association, the SNP alleles and their effects, identity of the minor allele (MA) and minor allele frequency (MAF) in the NUE-CROPS panel. A list of co-mapping SNP reported with identical association level at a particular QTL are presented in Appendix 3. 6.

a)

Trait	QTL	Chrom	SNP name	Dist	-log10(p)	R^2	Alleles	allele effects	MA	MAF
Anthocyanin color	Antho_1	2 H	A10326	6.45	**2.64**	0.09	G/A	-0.34/0	A	0.18
	Antho_2	2 H	I195164	96.80	**16.27**	0.98	C/T	-1/0	T	0.48
Grain Nitrogen Yield	GNYld_1	1 H	I154646	100.70	**2.59**	0.08	A/C	3.88/0	C	0.11
	GNYld_2	2 H	I151535	52.47	**3.39**	0.11	G/A	4.39/0	A	0.14
	GNYld_3	3 H	I115045	39.45	**2.56**	0.08	G/A	2.85/0	A	0.30
	GNYld_4	3 H	I164290	120.59	**3.52**	0.11	T/C	4.37/0	C	0.14
	GNYld_5	4 H	A20482	59.37	**3.50**	0.11	G/A	4.63/0	A	0.15
	GNYld_6	5 H	A20553	2.81	**4.06**	0.13	A/G	4.15/0	G	0.20
	GNYld_7	5 H	A21508	60.74	**4.88**	0.17	A/G	5.46/0	G	0.11
	GNYld_8	5 H	A20236	80.61	**2.61**	0.08	C/A	2.9/0	A	0.26
	GNYld_9	6 H	I123065	1.34	**2.69**	0.08	C/T	3.66/0	T	0.13
	GNYld_10	7 H	I186187	14.96	**2.65**	0.08	T/G	3.07/0	G	0.22
	GNYld_11	7 H	I138457	34.82	**3.04**	0.09	C/A	3.18/0	A	0.22
Gain Nitrogen	GrainN_1	2 H	A20862	63.50	**3.53**	0.11	T/A	0.08/0	A	0.42
	GrainN_2	4 H	I168399	92.40	**4.09**	0.13	C/A	-0.08/0	C	0.39
	GrainN_3	5 H	A21121	68.35	**3.46**	0.11	G/A	-0.08/0	A	0.35
	GrainN_4	6 H	B30120	52.75	**2.75**	0.08	C/A	-0.08/0	C	0.44
	GrainN_5	7 H	I138457	34.82	**2.60**	0.08	C/A	-0.07/0	A	0.22
Grains	Grains_1	2 H	A20862	63.50	**7.20**	0.27	T/A	-3.01/0	A	0.42
	Grains_2	6 H	I138716	88.90	**2.83**	0.09	G/A	-1.74/0	A	0.43
Heading date	Hd_1	1 H	B30241	20.82	**3.80**	0.12	C/A	2.67/0	C	0.29
	Hd_2	1 H	A21384	135.56	**3.02**	0.09	A/G	-2.19/0	G	0.27
	Hd_3	2 H	B30871	26.57	**4.96**	0.17	A/G	3.02/0	A	0.39
	Hd_4	2 H	A10191	63.53	**4.56**	0.15	C/A	-4.09/0	A	0.14
	Hd_5	5 H	B30867	136.43	**2.86**	0.09	C/A	-2.3/0	A	0.23
Harvest Index	HI_1	2 H	I10398	54.95	**3.56**	0.11	C/T	-2.79/0	T	0.17
	HI_2	3 H	I204057	51.70	**2.68**	0.08	C/T	2.46/0	T	0.10
	HI_3	3 H	I103215	126.27	**2.68**	0.08	A/G	2.07/0	G	0.14
	HI_4	4 H	I129218	92.40	**3.50**	0.11	C/A	1.71/0	C	0.42
	HI_5	5 H	A21121	68.35	**3.02**	0.09	G/A	1.64/0	A	0.35
	HI_6	5 H	A10183	80.02	**2.63**	0.08	G/A	-1.53/0	G	0.46
Height	Ht_1	1 H	A10338	117.80	**2.66**	0.08	C/A	-4.32/0	A	0.21
	Ht_2	2 H	B30265	63.53	**5.48**	0.19	A/G	-8.9/0	G	0.14
	Ht_3	3 H	A11016	58.64	**4.08**	0.13	G/C	-6/0	C	0.17
	Ht_4	3 H	A21163	80.89	**2.54**	0.07	A/G	-4.03/0	G	0.30
	Ht_5	4 H	I190401	48.72	**3.41**	0.11	A/G	-4.81/0	G	0.32
	Ht_6	5 H	I4717	34.25	**2.90**	0.09	G/A	-5.18/0	A	0.19
	Ht_7	5 H	B31257	48.11	**5.40**	0.19	A/T	-8.56/0	T	0.14
	Ht_8	6 H	I129756	80.52	**3.07**	0.09	T/G	-5.98/0	G	0.11

Table 3.2 cont.

b)

Trait	QTL	Chrom	SNP name	dist	-log10(p)	R^2	alleles	allele effects	MA	MAF
Nitrogen Use	NUE_1	1 H	I154646	100.70	**3.51**	0.11	A/C	2.54/0	C	0.11
Efficiency	NUE_2	2 H	A21304	33.74	**2.89**	0.09	A/G	-1.83/0	G	0.30
	NUE_3	2 H	A10358	59.21	**2.87**	0.09	C/A	-1.65/0	A	0.41
	NUE_4	3 H	I204057	51.70	**2.60**	0.08	C/T	2.37/0	T	0.10
	NUE_5	3 H	B31242	69.60	**2.67**	0.08	A/C	1.79/0	A	0.28
	NUE_6	3 H	I103215	126.27	**2.72**	0.08	A/G	2.05/0	G	0.14
	NUE_7	4 H	I129218	92.40	**3.40**	0.11	C/A	1.65/0	C	0.42
	NUE_8	5 H	I231238	63.31	**3.62**	0.11	T/C	2.19/0	C	0.21
	NUE_9	5 H	B31427	90.84	**2.58**	0.08	G/C	1.51/0	C	0.30
	NUE_10	5 H	A10080	151.36	**2.55**	0.07	G/A	1.77/0	A	0.24
	NUE_11	6 H	I118381	54.60	**2.58**	0.08	C/T	1.62/0	C	0.47
	NUE_12	7 H	I186187	14.96	**3.12**	0.10	T/G	1.91/0	G	0.22
	NUE_13	7 H	I138457	34.82	**4.14**	0.14	C/A	2.1/0	A	0.22
Nitrogen	NUpE_1	2 H	A10733	54.95	**2.58**	0.08	G/C	0.02/0	C	0.14
Uptake Efficiency	NUpE_2	5 H	A20553	2.81	**3.18**	0.10	A/G	0.02/0	G	0.20
	NUpE_3	5 H	A21508	60.74	**2.61**	0.08	A/G	0.02/0	G	0.11
	NUpE_4	5 H	I160288	129.41	**3.13**	0.09	G/A	0.02/0	A	0.11
Nitrogen	NUtEg_1	2 H	A20862	63.50	**4.19**	0.14	T/A	-2.53/0	A	0.42
Utilisation	NUtEg_2	4 H	I129218	92.40	**5.04**	0.17	C/A	2.51/0	C	0.42
Efficiency in Grain	NUtEg_3	5 H	I49958	68.35	**3.44**	0.12	A/G	2.21/0	G	0.38
	NUtEg_4	6 H	I124850	52.70	**3.25**	0.10	T/C	2.43/0	T	0.43
	NUtEg_5	7 H	I138457	34.82	**3.33**	0.10	C/A	2.22/0	A	0.22
	NUtEg_6	7 H	I14119	161.40	**2.54**	0.07	A/G	2.02/0	A	0.33
Nitrogen	NUtEt_1	3 H	I165444	99.89	**2.68**	0.08	A/G	-1.74/0	G	0.44
Utilisation	NUtEt_2	3 H	I154449	155.90	**3.26**	0.10	A/C	2.15/0	C	0.29
Efficiency total	NUtEt_3	5 H	I160288	129.41	**3.50**	0.11	G/A	3.18/0	A	0.11
	NUtEt_4	5 H	I156273	176.62	**3.15**	0.12	A/G	2.62/0	G	0.19
Stem Nitrogen	StemN_1	2 H	I177375	63.50	**3.88**	0.13	C/T	0.01/0	T	0.24
	StemN_2	5 H	A21318	53.18	**2.64**	0.08	G/A	-0.01/0	A	0.15
Stems	Stems_1	1 H	I182656	11.40	**3.19**	0.10	A/G	-0.28/0	G	0.13
	Stems_2	4 H	I129218	92.40	**3.71**	0.12	C/A	0.22/0	C	0.42
	Stems_3	5 H	B30975	6.40	**2.55**	0.08	A/C	-0.22/0	C	0.17
	Stems_4	5 H	I148402	135.72	**2.57**	0.08	G/A	-0.2/0	A	0.25
Sterile	StS_1	2H	A10823	46.98	**3.69**	0.11	A/G	0.28/0	A	0.37
Spikelets	StS_2	2H	A10287	85.92	**16.07**	0.67	A/G	0.92/0	A	0.20
	StS_3	5H	A21480	89.38	**3.89**	0.11	G/A	-0.37/0	A	0.26
	StS_4	5H	A10161	159.79	**2.57**	0.07	A/G	0.29/0	T	0.20
	StS_5	6H	I131992	81.88	**3.78**	0.11	A/G	0.31/0	A	0.42

Table 3.2 cont.

c)

Trait	QTL	Chrom	SNP name	dist	-log10(p)	R^2	alleles	allele effects	MA	MAF
Thousand Grain Weight	TGW_1	1 H	I232660	18.05	**3.24**	0.12	C/T	-2.75/0	C	0.36
	TGW_2	1 H	I128285	31.15	**3.45**	0.11	T/C	-2.64/0	T	0.36
	TGW_3	1 H	A20810	52.46	**3.56**	0.11	A/G	-2.76/0	A	0.46
	TGW_4	2 H	I146936	6.40	**3.33**	0.10	T/C	-3.53/0	C	0.18
	TGW_5	2 H	A10733	54.95	**4.33**	0.14	G/C	4.03/0	C	0.14
	TGW_6	2 H	I195051	156.72	**2.76**	0.10	T/C	2.71/0	C	0.40
	TGW_7	4 H	B30427	53.50	**2.61**	0.08	T/A	2.45/0	A	0.25
	TGW_8	5 H	A20553	2.81	**3.44**	0.11	A/G	3.04/0	G	0.20
	TGW_9	5 H	I194030	166.63	**2.83**	0.08	A/G	-2.62/0	A	0.38
Thousand Grain Weight from grab samples	TGW-GS_1	1 H	I232660	18.05	**3.42**	0.13	C/T	-3.25/0	C	0.36
	TGW-GS_2	1 H	I128285	31.15	**3.19**	0.10	T/C	-2.88/0	T	0.36
	TGW-GS_3	1 H	A20810	52.46	**3.63**	0.11	A/G	-3.2/0	A	0.46
	TGW-GS_4	2 H	I213799	8.57	**3.76**	0.12	A/C	-3.94/0	C	0.24
	TGW-GS_5	2 H	I143250	27.30	**3.05**	0.09	A/G	2.83/0	A	0.47
	TGW-GS_6	2 H	A10602	58.24	**3.09**	0.09	A/C	-3.51/0	C	0.26
	TGW-GS_7	5 H	A20553	2.81	**3.75**	0.12	A/G	3.68/0	G	0.20
Tillering from grab samples	Till.GS_1	4 H	A20732	92.38	**3.75**	0.12	G/A	55.67/0	G	0.42
	Till.GS_2	5 H	B30975	6.40	**3.59**	0.11	A/C	-66.86/0	C	0.17
	Till.GS_3	5 H	I147762	109.56	**2.62**	0.08	C/T	-48.35/0	T	0.26
	Till.GS_4	5 H	I720	159.80	**3.26**	0.10	A/G	55.52/0	A	0.37
	Till.GS_5	7 H	A10550	143.68	**2.63**	0.08	G/A	-51.15/0	A	0.25
Tillering from yield	Till.Yld_1	4 H	A21385	23.10	**2.73**	0.08	G/C	-58.15/0	C	0.26
	Till.Yld_2	4 H	A20732	92.38	**3.31**	0.10	G/A	57.53/0	G	0.42
	Till.Yld_3	5 H	A20553	2.81	**3.05**	0.09	A/G	-65.67/0	G	0.20
	Till.Yld_4	6 H	I4707	81.20	**2.81**	0.08	C/T	59.19/0	T	0.33
UPOV date of inscription	UPOV_1	2 H	I195164	96.80	**3.80**	0.12	C/T	-5.86/0	T	0.48
	UPOV_2	3 H	A10767	172.42	**2.93**	0.09	G/A	-4.93/0	A	0.32
	UPOV_3	4 H	I128723	54.98	**2.66**	0.08	A/G	5.98/0	G	0.18
	UPOV_4	5 H	I192396	19.40	**2.96**	0.09	T/A	5.93/0	T	0.46
	UPOV_5	5 H	I213753	64.00	**2.98**	0.09	C/A	5.45/0	A	0.25
	UPOV_6	6 H	I230959	4.90	**3.12**	0.10	G/T	5.89/0	T	0.23
	UPOV_7	6 H	I136897	53.29	**4.21**	0.14	A/G	7.92/0	G	0.23
	UPOV_8	7 H	A11222	4.90	**3.12**	0.10	G/C	5.89/0	C	0.23
Yield treated	Yld_1	1 H	I154646	100.70	**3.49**	0.11	A/C	0.53/0	C	0.11
	Yld_2	2 H	A21304	33.74	**2.73**	0.08	A/G	-0.37/0	G	0.30
	Yld_3	2 H	I10398	54.95	**3.33**	0.10	C/T	-0.54/0	T	0.17
	Yld_4	4 H	I182626	96.60	**3.08**	0.09	T/G	0.37/0	G	0.31
	Yld_5	5 H	I231238	63.31	**3.55**	0.11	T/C	0.45/0	C	0.21
	Yld_6	5 H	A20236	80.61	**2.60**	0.08	C/A	0.33/0	A	0.26
	Yld_7	6 H	I118381	54.60	**2.69**	0.08	C/T	0.34/0	C	0.47
	Yld_8	7 H	I186187	14.96	**3.66**	0.12	T/G	0.44/0	G	0.22
	Yld_9	7 H	I138457	34.82	**3.94**	0.13	C/A	0.43/0	A	0.22

AGOUEB panel QTL results

The AGOUEB data set proved to be informative and sufficiently variable to detect QTL for the traits available (Table 3.3). The additional DUS traits of ventral furrow hair, ear glaucosity, lower leaf hairy shealth (LLHS) and aleurone complemented the anthocyanin and sterile spikelet scored in the NUE. The strongest association was found for a marker A21087 on 4H indicating the position of the candidate gene controlling barley aleurone color (-log10(*p*) = 21.6). Both sterile spikelet and anthocyanin had similar highly significant QTL in the NUE-CROPS mapping. VFH had a noticeable QTL VFH_2 on 6H at 6.07 cM with –log10(*p*) = 12.5 suggesting proximity to the causal gene and the potential of the marker to be used by breeders to characterise germplasm for the DUS character. Despite having lower –log10(*p*) and effect of half the size of the LLHS variation, the QTL LLHS_1 and LLHS_2 had sufficiently strong significant markers to dissect the genetic control of that DUS trait.

The AGOUEB Yield and TGW traits had six and five QTL respectively, neither reported with strong effects on the traits. Amongst them, Yld_T_3, Yld_T_6, TGW_3 and TGW_5 (Table 3.3) appear to be the more significant QTL in that panel and may relate to QTL found in NUE-CROPS. Further investigation of those candidates will be carried out specifically in Chapter 4. The genetic control of heading date was localised in two positions on chromosome 2H with Hd_1 at 59cM and Hd_2 at 152 cM. Despite a high heritability and fewer genetic factors controlling this trait, the peak markers for the QTL were detected with –log10(*p*) around three.

Straw traits were measured in treated and untreated conditions in the AGOUEB project. The three QTL for straw length (corresponding to height) in treated conditions are also detected under untreated conditions with SL_T_1 and SL_U_1 having the highest significance. The strongest effect is found on 5H at QTL SL_T_3 and SL_U_5 where the allele of SNP A10236 associated to height reduction is only shared by 14 % of the varieties in the panel. Three additional QTL for straw length were found under untreated conditions suggesting the involvement of additional genes in the control of plant height under untreated conditions. The association of the straw length and straw strength traits was not evident as only SL_U_1 and SS_U_2 were potentially co-mapping in the centromere of 2H.

The mapping of disease resistance gave highly significant MTA which are promising for use in marker assisted breeding. A unique brown rust QTL was mapped on 2H at marker I146785 (70cM) with a –log10(*p*) value of 4.2. This location also has a

significant net blotch QTL NB(U)_1 potentially suggesting a common genetic factor involved in disease resistance. Net blotch had other two QTL, one with a strong effect on 6H NB(U)_3. The AGOUEB dataset also provided encouraging results with QTL associated with the resistance to Rynchosporium and mildew (Table 3.3). A key QTL ryncho(U)_1 was mapped on 2HL close to I129821 and two additional relevant QTL were mapped on 3H and 5H. QTL for mildew resistance were mapped at 6 locations. The main QTL mild-(U)_2 on 5H was mapped with a –log10(*p*) value of 5.21 and mild-(U)_6 was also highly significant.

These QTL from the AGOUEB GWAS may provide supporting evidence for the identification of candidate genes for disease resistance and yield related traits.

Table 3.3 Significant marker traits association for AGOUEB panel.

The table presents the significant QTL from the Tassel-MLM_K GWAS model on 20 phenotypes for the AGOUEB two-row winter barley study. Each QTL is presented with its peak marker (most significant SNP) with the associated marker map position, –log10(*p*) resulting from the test for association, the proportion of the genetic variation (R^2) of the association, the SNP alleles and their effects, identity of the minor allele (MA) and minor allele frequency (MAF) in the AGOUEB panel. A list of co-mapping SNP reported with identical association level at a particular QTL are presented in Appendix 3. 7.

a)

Trait	QTL	Chrom	SNP name	Dist	-log10(p)	R^2	Alleles	Allele effects	MA	MAF
Aleurone	Aleu_1	2 H	I171032	83.82	**2.87**	0.07	G/A	-0.28/0	A	0.28
	Aleu_2	4 H	A21087	62.10	**21.61**	0.79	G/A	-0.93/0	G	0.41
Anthocyanin	Antho_1	2 H	A10823	46.98	**3.15**	0.08	G/A	-0.34/0	A	0.29
Color	Antho_2	2 H	A10138	96.82	**20.83**	0.80	G/A	-0.93/0	A	0.45
Brown Rust	BR_1	2 H	I146785	70.50	**4.20**	0.09	G/C	3.21/0	G	0.13
Ear Glaucosity	EAR-G_1	1 H	I120059	0.75	**3.01**	0.07	T/C	1.05/0	T	0.30
	EAR-G_2	5 H	A10524	93.70	**2.54**	0.06	C/A	1.33/0	A	0.11
	EAR-G_3	6 H	B11455	42.36	**2.69**	0.06	G/A	1.08/0	A	0.40
	EAR-G_4	7 H	I1347	116.33	**3.44**	0.08	G/T	1.31/0	T	0.21
Grain Nitrogen	GrainN_1	2 H	A11384	60.68	**4.23**	0.13	C/G	0.04/0	G	0.42
	GrainN_2	4 H	I149873	0.74	**3.11**	0.09	G/A	-0.03/0	A	0.29
	GrainN_3	4 H	B31362	73.57	**2.67**	0.07	C/A	0.03/0	A	0.29
	GrainN_4	5 H	B30975	6.40	**4.86**	0.15	A/C	0.05/0	C	0.11
	GrainN_5	5 H	B30400	149.10	**2.57**	0.07	A/C	0.03/0	A	0.37
Heading date	Hd_1	2 H	B30042	59.21	**2.70**	0.06	A/G	-1.48/0	G	0.20
	Hd_2	2 H	B10937	152.79	**3.13**	0.07	G/C	-1.69/0	G	0.15
Hectoliter weight	HLW_1	6 H	B30025	117.68	**2.60**	0.05	A/G	-0.72/0	G	0.40
Hot Water	HWE_1	1 H	A10985	52.46	**4.75**	0.14	A/C	3.89/0	A	0.49
Extract	HWE_2	1 H	I165338	131.15	**2.53**	0.06	G/A	3.41/0	A	0.11
	HWE_3	2 H	I118168	9.28	**2.67**	0.07	A/T	2.84/0	T	0.50
LLHS	LLHLS_1	4 H	A10611	114.66	**8.71**	0.25	C/A	0.52/0	A	0.12
	LLHLS_2	5 H	A21355	153.50	**6.53**	0.17	A/G	0.44/0	G	0.11
	LLHLS_3	7 H	A20365	166.56	**3.20**	0.08	C/G	0.26/0	G	0.18
Mildew	Mild-(U)_1	4 H	I128147	86.27	**2.56**	0.05	T/C	-1.7/0	C	0.32
Untreated	Mild-(U)_2	5 H	I108541	19.40	**5.21**	0.12	C/G	-3.07/0	G	0.18
	Mild-(U)_3	5 H	I204494	51.30	**2.74**	0.06	C/A	-2.36/0	A	0.13
	Mild-(U)_4	6 H	I147090	33.74	**2.72**	0.06	T/C	-1.89/0	C	0.23
	Mild-(U)_5	6 H	I164156	90.15	**2.82**	0.06	T/C	-1.8/0	C	0.39
	Mild-(U)_6	7 H	I163976	29.82	**3.72**	0.08	T/C	-2.87/0	C	0.13
Net Blotch	NB(U)_1	2 H	I16024	71.12	**2.95**	0.06	C/T	1.19/0	C	0.48
Untreated	NB(U)_2	4 H	B10063	40.36	**3.05**	0.06	G/A	-1.67/0	A	0.16
	NB(U)_3	6 H	I128460	45.40	**5.01**	0.12	T/C	1.79/0	C	0.30
Rynchosporium	Ryncho(U)_1	2 H	I129821	158.15	**4.99**	0.12	T/G	-3.57/0	G	0.19
Untreated	Ryncho(U)_2	3 H	A20252	6.03	**3.52**	0.08	G/A	-2.45/0	A	0.31
	Ryncho(U)_3	5 H	B30456	113.11	**3.21**	0.07	A/G	2.42/0	G	0.26

Table 3.3 cont.

b)

Trait	QTL	Chrom	SNP name	Dist	-log10(p)	R^2	Alleles	Allele effects	MA	MAF
Straw Length	SL_T_1	2 H	I127347	63.53	**4.07**	0.11	T/C	-3.74/0	C	0.28
Treated	SL_T_2	5 H	I136777	123.08	**2.69**	0.07	C/T	2.42/0	T	0.32
	SL_T_3	5 H	A10236	181.43	**3.44**	0.09	A/G	-4.21/0	A	0.14
Straw Length	SL_U_1	2 H	I177375	63.50	**4.57**	0.11	C/T	-4.87/0	T	0.28
Untreated	SL_U_2	4 H	I129218	92.40	**3.95**	0.09	A/C	3.45/0	C	0.45
	SL_U_3	5 H	B30975	6.40	**3.45**	0.08	A/C	4.63/0	C	0.11
	SL_U_4	5 H	I136777	123.08	**2.58**	0.05	C/T	2.7/0	T	0.32
	SL_U_5	5 H	A10236	181.43	**3.31**	0.07	A/G	-4.9/0	A	0.14
	SL_U_6	6 H	A20745	28.39	**2.62**	0.05	C/A	-2.86/0	C	0.39
Straw Strenght	SS_U_1	1 H	I3336	15.74	**2.82**	0.06	A/C	-3.26/0	C	0.23
Untreated	SS_U_2	2 H	A10358	59.21	**4.14**	0.09	C/A	-4.06/0	A	0.42
	SS_U_3	3 H	I155763	83.23	**3.34**	0.07	A/G	4.25/0	G	0.18
	SS_U_4	4 H	A21035	113.92	**2.81**	0.06	A/G	-3.4/0	A	0.21
	SS_U_5	7 H	I138111	58.57	**2.55**	0.05	C/G	-2.94/0	G	0.28
	SS_U_6	7 H	I150049	104.78	**3.37**	0.07	T/C	-3.5/0	C	0.37
Sterile	StS_1	1 H	A21333	59.71	**3.13**	0.08	G/C	-0.93/0	C	0.46
Spikelets	StS_2	2 H	A10823	46.98	**3.60**	0.09	G/A	1.25/0	A	0.29
	StS_3	2 H	A10287	85.92	**13.98**	0.48	G/A	3.23/0	A	0.17
	StS_4	4 H	A10319	8.25	**2.68**	0.06	A/G	0.87/0	G	0.24
	StS_5	5 H	A10236	181.43	**2.59**	0.06	A/G	-1.25/0	A	0.14
	StS_6	6 H	I207933	4.41	**2.76**	0.07	C/G	0.83/0	C	0.32
	StS_7	6 H	I204148	60.23	**2.72**	0.07	T/C	-0.97/0	C	0.28
Thousand	TGW_1	1 H	I184784	40.99	**2.64**	0.06	T/C	-1.86/0	T	0.44
Grain Weight	TGW_2	2 H	I110647	31.00	**2.58**	0.06	T/C	2.23/0	C	0.18
	TGW_3	2 H	B30042	59.21	**3.32**	0.08	A/G	2.54/0	G	0.20
	TGW_4	4 H	I160461	103.10	**2.68**	0.06	T/C	1.77/0	C	0.46
	TGW_5	5 H	A20553	2.81	**3.90**	0.09	A/G	2.53/0	G	0.22
Ventral Furrow	VFH_1	2 H	I152485	101.78	**2.61**	0.05	G/A	0.23/0	A	0.27
Hairs	VFH_2	6 H	I194036	6.07	**12.50**	0.35	A/C	0.55/0	C	0.35
Winter	WintH_1	2 H	A21261	28.44	**3.01**	0.06	G/A	0.12/0	A	0.19
Hardiness	WintH_2	4 H	I110333	65.80	**2.80**	0.06	T/C	0.1/0	T	0.25
	WintH_3	5 H	I214760	18.72	**2.93**	0.06	G/A	0.08/0	A	0.42
	WintH_4	6 H	I114351	56.48	**2.84**	0.06	C/T	-0.1/0	T	0.32
Yield	Yld_T_1	4 H	I150603	48.72	**2.73**	0.06	G/A	0.23/0	A	0.14
Treated	Yld_T_2	4 H	I129218	92.40	**2.89**	0.06	A/C	-0.17/0	C	0.44
	Yld_T_3	5 H	I192396	19.40	**4.17**	0.09	A/T	-0.23/0	A	0.29
	Yld_T_4	5 H	I205853	86.63	**2.58**	0.05	T/G	-0.2/0	T	0.16
	Yld_T_5	6 H	I115369	55.90	**2.73**	0.06	T/C	-0.2/0	C	0.45
	Yld_T_6	7 H	B30380	138.17	**3.28**	0.07	G/A	0.19/0	G	0.42

3.4 Discussion

3.4.1 Population structure of two-row barley

The presence of population stratification in association mapping panels results from divergent selection and drift observed in material originating from different geographic regions which can lead to incorrect associations between markers and phenotype (Price et al., 2010). Most panels reported in barley association mapping studies include varieties with differential growth habit and number of row of grains (Comadran et al., 2009; Pasam et al., 2012; Wang et al., 2012). These traits are controlled by known major genes responsible for plant adaptation and evolution. Two main genes *vrs1* and *int-c* are involved in the partitioning between two- and six-row barley types (Komatsuda et al., 2007; Ramsay et al., 2011). Despite stringent population structure correction, the residual effects of those genes are sufficient to be mapped and associated with significant marker effects. In a GWAS on two and six-row spring barley varieties, the most significant MAT for grains per ear and tillering were found on SNP in close vicinity to *vrs1* and *int-c* (Comadran et al., 2011b). Further, in a study on spring barley including different row types, QTL for row type overlapped QTL TGW and grain proteins (Pasam et al., 2012). In winter barley, the analysis of the NUE-CROPS panel of two- and six-row barley varieties showed that diagnostic markers responsible for strong morphological division are also associated with effects on other traits such as yield and yield components (Thomas et al., 2013). The vernalization genes that divide spring and winter sown barleys have also been precisely mapped and showed associations to yield and yield component variation (Cockram et al., 2008; Wang et al., 2010).These observations suggest that despite population structure correction, residual variation due to structure remains effecting multiple phenotypes. Similar observations were made in a panel of barley varieties originating from very diverse geographical origins (Comadran et al., 2008). Although these earlier mapping experiments were successful in identifying the importance of major genes, they show that the residual structure needs to be accounted for when interpreting QTL results. In addition, the allele effects at partitioning genes (e.g. *vrs-1*, *int-c*) have limited interest when the objectives of crop improvement and evaluation of allelic variation of a trait are constrained to a specific crop type (i.e. two-row winter barley). The voluntary restriction of the panels to two-row winter barley in this study was associated with a reduction in apparent population structure whilst the genetic relationship could still be captured using the genetic marker data. Phenotypes from the high fertilisation regimes also had the advantage of

quantifying the effects of allelic variation directly available to breeders and expressed under conventional farming practices. As expected there was little evidence of gross population structure in either mapping panel with lines forming a single cluster (Figure 3.3). Moreover, the GWAS results showed that neither *vrs1* nor *int-c* loci were associated with significant effects on yield components. *Vrs1* was only associated with the sterile spikelet trait (StS_2 in NUE-CROPS mapping results). Therefore, the reduction of population structure in the GWAS study presented should allow for more precision in the investigation of the allelic variation responsible for breeding progress and facilitate the transfer of results to the two-row winter barley breeders.

The initial AGOUEB and NUE-CROPS variety panels were established to capture most of the allelic diversity in European two-row winter barley of the last 50 years (Thomas et al., 2013; Waugh et al., 2009). Although our study included only two-row winter barley types from those panels, the population structure mainly due to pedigree relationships had to be accounted for in GWAS. In order to best account for population structure in the panels, the correction using principal components and Kinship was investigated. Price et al, (2006) showed that principal components (Q) summarising a correlation matrix between lines obtained from genotypes could account for structure in GWAS models. Alternatively, Kinship matrix (K) of genetic correlation between lines can be included in mixed models for GWAS (Zhang et al., 2010). Here, the Q-Q plots investigating different GWAS models suggested that both K and K+Q correction were the most appropriate to account for structure in the AGOUEB and NUE-CROPS two-row winter barley panels. In addition, the simple K correction causes less reduction in the statistical power of the test than a more complex K+Q (Wang et al., 2012) and was shown to efficiently account for population structure when used in previous barley GWAS (Cockram et al., 2010; Pasam et al., 2012). In this study, the Kinship (K) correction was therefore considered adequate to account for pedigree relationships in both AGOUEB and NUE-CROPS panels.

3.4.2 Association mapping reveals known genes

The second objective of the study was the mapping of quantitative traits in the NUE CROP and AGOUEB panel in order to put emphasis on the genetic regions that contain interesting gene candidates and functional polymorphisms.

The simple DUS traits showing Mendelian segregation were used to validate the methodology. They were associated with highly significant MTA. The anthocyanin pigmentation gene *ant-2* was precisely located as the most significant marker

associations in both AGOUEB and NUE-CROPS (A10138) was located 10 gene models away from the functional polymorphism (Cockram et al., 2010). Highly significant MTA for sterile spikelets (A10287, 2H, 13 gene models from *vrs1*), aleurone colour (A21087, 4H) and LLHS (A10611 4H) at chromosome locations previously described in the GWAS of a larger panel (Cockram et al., 2010) confirm that the markers can be used to characterise those DUS traits in winter barley varieties. In this study, the level of significance of associations was found to reduce with the increased number of significant loci for the traits as more genes with smaller effects come into play to control the variation of a phenotype. The five and two QTL for heading date in NUE-CROPS and AGOUEB respectively indicate that the trait is only moderately complex given the strong selection for appropriate flowering time within the winter crop. The genetic control of heading in barley plays a crucial role in plant adaptation to geographical regions and its effects on yield responses (Cuesta-Marcos et al., 2009; Griffiths et al., 2003; Laurie, 1997). The NUE-CROPS panel detected a strong QTL effect for heading date at markers positioned in the gene sequence of the *Ppd-H1* (B30871) involved in the determination of flowering time in spring barley (Laurie, 1997; Stracke et al., 2009). The GWAS in winter barley demonstrates that the allelic variation at *Ppd-H1* is responsible to a moderate level for differences in heading date.

The GWAS results confirm some other QTL positions that have been detected in larger barley variety collections (Pasam et al., 2012). The most significant, both in terms of effect size and significance of MTA is a QTL on the centromere of 2H that targets a locus previously identified as *eam6*. The locus carries HvCEN, an homolog of *Antirrhinum CENTRORADIALIS* gene. That gene induces variation in heading date in spring barley and is responsible for divergent selection between spring and winter types (Comadran et al., 2012). Here, up to four days delay in heading date were detected in the two-row winter barley germplasm and associated with the allele A at marker A10191 and a stronger effect than SNP in the *Ppd-H1* gene. In addition, significant effects for Grain number, Yld, HI, TGW, Ht, Stem Nitrogen, Grain N, GrainNYld, NUE, NUtEg, NUpE were mapped in the same region with strong overlap of support intervals. With such a range of traits mapped at that cluster, the locus is a major contributor in the genetic variation of most traits in winter barley which can suggest a role in plant adaptation and fitness to geographical regions. Unlike *vrs1* and *int-c* which cause major ear morphology changes and fundamental subpopulation division (two-row vs six-row barley), HvCEN is involved in the control of plant phenology and the barley SNP B30265 was found to be segregating with the early and late alleles (Comadran et

al., 2012) When mapped on the genetic diversity of the winter barley panels of NUE-CROPS and AGOUEB, the alleles of B30265 reveal a cluster of varieties carrying the early allele within the overall winter barley diversity (Appendix 3. 8). It can be suspected that the two alleles are maintained in the winter barley germplasm as a result of selection over variable years and provide optimal adaptation for the range of seasons and environments. Nevertheless, the inclusion of a co-factor representing the SNP alleles in a GWAS can help to account for structure linked to phenology provide additional insight on the detection of smaller effects (Alqudah et al., 2014). Other NUE-CROPS heading date QTL were associated with known vernalization and photoperiod genes. Hd_2 SNP A21384 on 1H at 135.5cM (homologous to Os05g50800) mapped close to the *HvFT3* (*Ppd-H2*) locus proposed to be syntenic to Os05g44180 (Faure et al., 2007). Hd_5 SNP B30867 at 136.4cM (Os03g54084) is eight gene models away from *vrn-H1* gene (Os03g54160) on 5HL (Szucs et al., 2007). However, other QTL such as the NUE-CROPS QTL Hd_1 on 1H have the potential to provide novel sources of variation in the control heading date. While many associations remain to be validated, the study clearly identified a number of marker effects associated with genes known to be involved in the genetic control of the highly heritable traits.

3.4.3 GWAS of yield and yield components

Because yield is the product of yield components, its genetic architecture is expected to be characterised by higher genetic complexity resulting from a large number genes with small effects. The AGOUEB and NUE-CROPS GWAS detected five and nine QTL for yield respectively with rather low levels of significance and r^2 values suggesting that a number of lower effects were not detected at the threshold of the GWAS. The statistical power of the investigation can relate to different parameters such as the panel composition and size, the amount of recombination, the phenotyping protocols and heritabilities of the traits (Ingvarsson and Street, 2011). Pasam et al. (2012) reported that the sensitivity and strength of the model used for structure corrected GWAS can also affect the level of detection. Also the effects of major gene involved partitioning the population can affect the power of detection of QTL with smaller effects (Alqudah et al., 2014). In addition, in both NUE-CROPS and AGOUEB studies, a large proportion of the trait's variance was accounted for by the model for GxE interactions. In NUE-CROPS, the BLUPS were computed from 5 site and year environments (Thomas et al., 2013). Therefore the genetic variation captured in the BLUPs and used for the GWAS only summarises consistent main effects and the statistical approaches

used did not investigate epistatic interactions and GxE QTL effects. This may also explain the absence of significant effects at the *Ppd-H1* locus in AGOUEB. Therefore, the genetic clustering of the effects for correlated phenotypes independently measured (e.g yield components) can bring additional confirmation and help in the interpretation for these types of less prominent associations. Other alternatives such as genomic selection can be envisaged to exploit the genetic resources for improving complex traits by seeking to include the small effects to be able to make useful predictions of the phenotype from the genotype (Bernardo and Yu, 2007; Jannink et al., 2010).

Plant adaptation to the environment is critical in order to maximise yield. The interactions of the yield responses with environment have been partly described by QTL studies which identified polymorphic genes responsible for plant adaptation that maximise yield in different environments (Comadran et al., 2012; Snape et al., 2007). For example, the vernalization genes responsible for phenology differences have been identified as candidate genes for yield QTL in barley (Cuesta-Marcos et al., 2009). In the multi-site BLUP of the NUE-CROPS study, the effects of such genes were averaged out across the GxE components leaving only a small overall genetic effect to be mapped. Therefore the QTL from the NUE-CROPS capture genetic effects responsible for the genetic variation across the whole range of environments encountered in the five sites of the study. A highly significant yield QTL was mapped in AGOUEB on chromosome 5HS (I192396) that appears to co-localise with a winter hardiness and a UPOV QTL in NUE-CROPS. Interestingly, other QTL for yield Yld_5 on 5H and Yld_7 on 6H also overlapped with UPOV QTL. This trait captures alleles at markers under differential selection over time, and whilst it can relate to the mapping of the breeding progress, UPOV QTL may help in interpreting the traits and effects retained by selection. This initial observation on the co-mapping of QTL gives an insight on the complexity to consider for the biological interpretation of yield QTL so that they can be used effectively in yield improvement.

Although the trait of grains per ear was only mapped in the NUE-CROPS panel, a noticeable MTA in the region of *eam6* makes that candidate gene a primary target for the yield component. Polymorphism in the *Cen* gene family has been shown to affect the fate of terminal floral meristem in a range of species by regulating inflorescence development and the flowering time (Cremer et al., 2001; Foucher et al., 2003; Zhang et al., 2005). The barley homologue HvCEN was shown to be associated with heading date differences and the three major haplotypes are segregating in a winter barley

collection (Comadran et al., 2012). The NUE-CROPS findings confirm that both heading data and grains are controlled by HvCEN. In the NUE-CROPS panel, three main haplotypes at HvCEN (*eam6*) locus (2H centromere, 63 cM) were observed while segregating with a quasi-equal frequency within the only two haplotypes *Ppd-H1* gene. Such observations suggest that both mechanisms of earliness control are being maintained in the winter barley population while only the late alleles at the genetic factor at *eam6* may give an increase the number of grains per ear. On similar grounds, the second QTL for grains on 6H is located in a region homologue to rice chromosome 2 where the most significant barley SNP (I138716) is close to a barley gene homologue to rice gene OsGRF1 (LOC_Os02g53690) a plant growth regulating factor involved in regulating vegetative growth in rice (Choi et al., 2004). The barley homologue may be a candidate gene to investigate for a potential role in the control of ear elongation and increase in the grain number per ear through the control of inflorescence architecture as indicated by the co-localisation of the sterile spikelets QTL StS_5.

Increasing the number of sink organs per meter square by achieving a higher number of fertile tillers per plant is also a breeding strategy for increased yield. The ability of the plant to tiller is set in the early stages of development, when axillary meristems are being developed. Later, around GS31, developmental processes determine which of these tillers are converted into fertile ears (Sreenivasulu and Schnurbusch, 2012). The correlation between genetic pathways involved in the control of branching and tillering between species has been generally established (Kebrom et al., 2013). In barley, tillering mutant phenotypes have been used to describe the genes *als* and *lnt1* on 3HL (Dabbert et al., 2009; Dabbert et al., 2010). The barley *cul2* gene is involved in the control of inflorescence development and the transformation of axillary meristem into tillers (Babb and Muehlbauer, 2003). It has now been located in the centromeric region of 6H where unfortunately the level of recombination was insufficient to enable the gene to be cloned (Okagaki et al., 2013). Despite the low significance levels of the MAT for tillering found in the NUE-CROPS panel, the three independent methods of measuring tillering (see 3.2.1) identified a couple of convincing QTL on 4HL (Stems_2, Till.GS_1, Till.Yld_2) and 5HS (Stems_3, Till.GS_2, Till.Yld_3). The independent phenotype measurements give additional support to the validity of the genetic effects found. The tillering QTL locus on 4HL corresponds to a QTL for final tillering found on 4HL at 61.5cM in a bi-parental study by Borràs-Gelonch (2011). A significant MTA for tiller number was found 12cM away (A11500) from the Stems_2 best marker

I129218 that was also associated with a tillering QTL in a spring barley GWAS (Long et al., 2013). The corresponding rice chromosome region to this QTL cluster on 4HL is located on chromosome 3 in a region dense in genes and will require further investigation to identify candidate genes. The tillering QTL cluster on 5HS was also detected previously, associated with effects on harvest index and tillering (Comadran et al., 2011b). However, no previous QTL mapping results were found to support the tillering QTL on 7H. The candidate on 4H (Table 3.2) is the most robust tillering candidate identified in this two-row barley panel and offers a potential target for MAS and gene cloning.

TGW is the resultant of many factors occurring during grain fill. It is the last yield component to develop, following tillering and the construction of the number of grains per ear (Sreenivasulu and Schnurbusch, 2012). These predetermine the potential sink size by setting the available grains per meter square (Bingham et al., 2007b). QTL for TGW have previously been mapped in bi-parental populations of two-row barley (Bezant et al., 1997; Tinker et al., 1996). Advanced backcross studies have shown that QTL for TGW often co-map with yield and other yield component QTL (von Korff et al., 2006) and were affected by environment and nitrogen interactions (Saal et al., 2011). Research on the effects of introgressions from the wild *H. v. spontaneum* showed that the majority of exotic alleles reduced TGW (Li et al., 2005; Schmalenbach et al., 2009; von Korff et al., 2006) suggesting that positive alleles and epistatic interactions have been selected and maintained in elite germplasm. The changes in spike morphology form another mechanism affecting the potential sink size of plant with strong effects on TGW (Pasam et al., 2012) although there relevance for crop improvement may be limited due to their association to different crop types. Comadran et al., (2011b) detected three QTL for TGW, two of which associated with genes controlling two/six-row spike morphology (v*rs1* and *int-c*). The third QTL was confirmed by this study with a major effect at 58 cM on chromosome 2H (TGW_5 and TGW-GS_6). Although in the centromeric region of 2H, the close proximity to heading date QTL indicates that HvCEN is a plausible candidate gene for this effect. This can be associated with changes in the length of the construction phase and thus grains number in the ear affect the grain size (see Alqudah et al., 2014). At that stage, the variation cannot be attributed to a change of the sink size (ear) or plant adaptation (phenology). The independent measure of TGW gave higher confidence in the QTL detected as six QTL were confirmed by the two phenotypic scorings and co-mapped with other

agronomic traits. For example, the MTA on 5HS (TGW_8) coincided with a tillering QTL previously described (Comadran et al., 2011b) supporting the negative correlation between the two traits and the hypothesis that increased sink size via tillering at that locus penalises TGW. This biological interpretation should be borne in consideration when hypothesising the syntenic correspondence of candidate gene for that QTL with genes on rice chromosome 12. Here, the biological complexity and plasticity of TGW in two-row barley is revealed at the genetic level. Although interaction effects cannot be resolved, it becomes apparent that to fully understand the genetic architecture of TGW, it is required not only to have prior knowledge on the traits that contribute to its variation but also on the genetic architecture of those same traits.

The QTL results from GWAS in AGOUEB and NUE-CROPS panel offer an opportunity to delve into the genetic architecture of agronomic traits of interest. Optimising the population structure correction has enabled the detection of important QTL for agronomic traits. Whilst some traits suggest simple genetic control such as grains per ear, the widespread genetic factors associated to TGW and other traits like yield means they remain highly complex and partially understood. Therefore, in order to use these QTL results in germplasm improvement using MAS, a biological interpretation of each QTL is necessary to distinguish favourable from unfavourable alleles and understand the function of potential candidate genes. An in-depth interpretation of selected candidate regions is the subject of Chapter 4 in which the comparison of the previous mapping studies will be presented.

Chapter 4

Integration of QTL studies to target candidate genes

4.1 Introduction

QTL discovery has become a routine procedure to understand the genetic architecture of quantitative traits in barley (Cuesta-Marcos et al., 2009; Xue et al., 2010). These type of studies represent a necessary step to promote the efficient use of genomic information into breeding of improved varieties by the deployment of MAS (Rae et al., 2007). A range of QTL mapping methods have been proposed and developed along with the continuously developing technology (Bernardo, 2008). In Chapter 2, a bi-parental DH population of a cross between two elite lines Saffron and Retriever was used to map QTL for relevant agronomic traits. QTL for yield and yield components were found on most of the chromosomes and their location revealed that 23 regions were carrying genetic factors having effects on at least one trait. Particular attention was drawn to a region of chromosome 2HS at which highly significant QTL for the main effects of the yield components: tillering and TGW were detected. In that region, the *Ppd-H1* gene involved in the control of heading date in barley (Laurie, 1997) was proposed as a candidate gene. However it was pointed out that further investigation and validation was required owing to the absence of effects on heading at that locus in the DH population.

In Chapter 3, GWAS were conducted on the two-row winter barley varieties panels from a subset of NUE-CROPS (Thomas et al., 2013) and AGOUEB projects (Waugh et al., 2009). The GWAS approach offers improved resolution compared to bi-parental QTL mapping by exploiting the linkage disequilibrium between genetic markers segregating in panels of varieties (Gupta et al., 2005; Waugh et al., 2009). QTL for agronomic traits were found on all chromosomes in both panels, some of them showing consistent effects across studies. Amongst the strongest associations, the DUS traits of anthocyanin and aleurone [3] colour captured with a remarkable precision the chromosomes regions known to carry the functional polymorphisms (Cockram et al., 2010). A large number of traits had a significant association with a marker close to the centromere of 2H, in a region known to be associated with variation in heading date (Comadran et al., 2012). Generally, the complex traits such as yield and TGW also had large number of QTL detected with lower significance also associated to lower effects

[3] The gene controlling aleurone colour has not yet been cloned (Luke Ramsay pers. comm.)

and proportion of the genetic variation accounted for (r^2). This illustrates the complexity of the traits, resulting from multiple gene effects and interactions.

Despite an appreciable literature on barley QTL for yield and yield components, rather limited applications have been developed by the breeding industry (Bernardo, 2008; Rae et al., 2007). This reluctance to deploy MAS originates from the need for confidence in the presence and consistency of QTL effects, often obtained by QTL validation stages (Collard and Mackill, 2008). Therefore comparison of several independent studies may be a way to ascertain the true effect of the QTL. However this step may encounter difficulties because of differences in genotyping technology (i.e. different types of genetic markers or genotyping arrays) as well as varied experimental design and objectives in the published mapping studies. The Meta-QTL approach has been proposed to identify QTL of main effect from multiple mapping experiments and was applied to grass species using 15 mapping populations (Swamy et al., 2011). Although only three mapping experiments have been carried out in this study, similar principle can be applied to enhance the identification of consensus QTL. Indeed, bi-parental population mapping gives high confidence that an effect is segregating but has limited resolution whereas GWAS can deliver high resolution at the expense of risks for false positive. The integration of QTL results from these different mapping approaches by alignment based on common genetic markers should enable a primary validation of the effects of genetic factors and their position.

Popularised in the early 1990's, comparative genomics applied to grass species has become a powerful tool to understand genetic resources in species with large genomes (Gale and Devos, 1998). The approach uses the genomic similarities between different species based on gene-based marker and sequencing data to define co-linearity or synteny between the genomes (Feuillet and Keller, 2002). Major chromosome rearrangements are frequent between grass species (Gale and Devos, 1998) and are the results of plant and genome evolution (Bennetzen and Chen, 2008). In the case of barley (*H.vulgare*), the many comparative studies involve rice (*Oryza sativa*) for which the genome has been fully sequenced and many genes characterised. Five of the *barley* chromosomes show large scale homology with more than one rice chromosome and two: 3H and 6H show good co-linearity with single rice chromosomes (rice 1 and rice 2 respectively) (Mayer et al., 2011). The good genome homology of barley with rice and Brachipodium (*Brachypodium distachyon*) was also made from the flanking DNA

sequence of the SNPs contained in the Iselect 9K platform (Close et al., 2009). The resulting micro-co-linearity observed in small genome segments can subsequently be converted into more local analysis of the gene functions and evolution (Bennetzen and Chen, 2008) and used to enhance the connection of research results across several species in order to develop a candidate gene hypothesis (Swamy et al., 2011). Therefore the greater the resolution achieved by QTL mapping studies, the easier it is to focus on a small and manageable range of candidate genes homologous with other species and target the functional polymorphism.

This chapter follows on the QTL mapping experiments from Chapter 2 and Chapter 3 using both mapping experiments results to aim at the validation of QTL. The project will test the hypothesis that consistency and interpretation of agronomic QTL is achievable from a limited number of mapping studies which include a large number of traits. The possibility to obtain relevant candidate genes from mapping comparison and comparative genetics will be explored. The framework of the chapter is therefore defined by the following objectives: 1) to integrate the QTL results on a consensus map and identify consistent QTL effects across studies, 2) to characterise the complexity associated with identifying alleles associated with the control of agronomic traits as potential target in a MAS approach 3) to use comparative genomics with the rice genome to propose relevant candidate genes for loci of interest.

4.2 Material and Methods

4.2.1 QTL resources from association mapping and bi-parental studies

NUE-CROPs

The QTL mapping results of the NUE-CROPS two-row winter barley panel were sourced from Table 3.2 (page 93). For each of the QTL, the standard support intervals used the 5cM and 10cM distances from the most significant markers position. Any marker with a different map position but with identical allelic pattern within the panel (identical significance level in GWAS) was also interpreted as the most significant marker for the QTL, which can in some cases generate larger support intervals. Details on QTL associated with multiple SNP of similar allelic pattern but different map positions can be found in Appendix 3. 6.

In order to investigate the marker haplotype signature at the QTL locations, the size and direction of marker effects and their level of significance across the range of different traits were analysed. For the NUE-CROPS phenotypes, the results correspond to the tests of association from Tassel-MLM_K GWAS scans presented in Chapter 3 and available as Supplementary data 2.

AGOUEB

The QTL results for the two-row winter barley panel of AGOUEB were sourced from Table 3.3 (page 98). Similar to the NUE-CROPS dataset, standard support interval for each QTL were set at 5cM and 10cM from the most significant markers position. The list of the QTL mapped to several peak SNPs of similar allelic pattern but different map positions can be found in Appendix 3. 7. To compare marker effect across traits and haplotype signature, the size and direction of marker effects and level of significance were obtained from the AGOUEB Tassel-MLM_K GWAS scans.

The bi-parental Saffron × Retriever DH population

The QTL mapping results from the DH population Saffron × Retriever presented in Chapter 2 provide an additional dataset where significant genetic factors involved in the control of yield and yield components have been described. For the purposes of this chapter, each trait was considered once (e.g. only the TGW QTL from the QTL × Environment analysis using three sites Fowl09, Eld09 and Fowl12 are presented). The QTL information for traits measured exclusively in 2009 and 2012 were sourced from Table 2.6 and Table 2.7 respectively and QTL information for the traits measured on both years was obtained from Table 2.8 (see 2.3.3 page 59). Those specific QTL results were also presented in Figure 2.2.

4.2.2 Integration of QTL on multiple SNP map

The SNP map distances of the 9K iselect SNP chip used for GWAS scans in Chapter 2 were used as reference to integrate the QTL from the three mapping studies on a common genetic map. The 9K SNP distances are based on LD mapping of each individual SNP and have been kindly provided by the JHI in the context of NUE-CROPS project. The markers comprised in genotyping platforms BOPA1, BOPA2 and specific to 9K Illumina chip were identified by the prefix letter A, B and I respectively. The QTL from GWAS of NUE-CROPS and AGOUEB GWAS were positioned along the chromosomes based on peak marker positions and a standard support interval as described in 4.2.1. For each QTL of the bi-parental population, common SNP markers

between 9K Illumina array and BOPA1 were used as anchorage positions for the QTL and their support interval to give a realistic representation of the mapping resolution obtained in Chapter 2. Therefore, all 9K SNP (A, B and I) comprised in the bins of a BOPA1 marker that were mapped in a S×R QTL support interval were included in the adjusted support interval. This protocol was carried out to identify the 9K SNP markers corresponding to the decrease in magnitude of 1 LOD and 2 LOD scores from the QTL peak in the S×R dataset.

4.2.3 Synteny relationships

The genome co-linearity between the rice (*O. sativa)* and barley (*H.vulgare*) species was exploited to investigate putative candidate genes at the barley QTL position (Mayer et al., 2009). For each of the barley SNPs, the corresponding rice locus was obtained from the genome zipper available at http://mips.helmholtz-muenchen.de/plant/barley/gz/index.jsp. The detailed genome zipper for the SNP markers used in GWAS of NUE-CROPs and AGOUEB is presented in Supplementary data 1.

First, the pattern of significance of GWAS scans was analysed at each QTL cluster of interest by plotting the –log10(*p*) for barley SNP ordered by their homologous rice locus position. The patterns of significance enabled to define short chromosome segment bracketed by the most significant SNP. The homologous rice loci were then to identify and bracket a segment of the rice chromosome with putative rice gene models in physical order. The list of putative rice genes was screened for a preliminary identification of putative homologous candidate genes at the most relevant barley QTL clusters (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/).

4.3 Results

4.3.1 Integrated QTL locations for three mapping experiments

To visualise simultaneously QTL results from the three mapping experiments, a graphical display reporting their support interval along the chromosomes is presented in Figure 4.1. This offers the possibility to inspect on a chromosome basis the genetic regions associated with QTL for several traits rather than focusing on QTL for individual traits. The approach facilitates the biological interpretation of putative genetic factors underlying the clusters. The correspondence of the different

abbreviations used for the phenotypes analysed in previous chapter are mentioned in footnotes page 75 and 77.

The distal region of chromosome 1HS (Figure 4.1a) shows the co-localisation of TGW and tillering and heading date QTL in the NUE-CROPS. This region contains also significant effects for QTL of S×R on straw traits and mildew susceptibility which do not appear to be correspond to any of the traits mapped in both AM panels. However the straw strength QTL in AGOUEB could potentially relate to stem number (Stems_1) and straw collapse (SC.1_1). Ear glaucosity (EAR-G_1) also associates with the narrow QTL cluster of S×R containing effect for straw health and the very highly significant S×R QTL for ear glaucosity: E_Glau.1. These results confirmed across mapping studies strongly suggest the presence of a gene controlling leaf wax deposition and an associated effect on disease susceptibility. For the rest of chromosome 1H, the co-localisation between QTL from the different studies is rather limited; AGOUEB TGW_1 overlaps with a TGW effect of NUE (TGW_2); the quality trait HWE associates with TGW_3 and Hd_2 of NUE-CROPS. The yield derived traits NUE and GNYld also map at the same position in NUE-CROPS but do not seem to be associated with heading date or height in any of the mapping experiments.

Chromosome 2H contains numerous QTL from NUE-CROPS and AGOUEB (Table 3.2 Table 3.3) as well as S×R QTL with large effects (Chapter 2). The locations of QTL from NUE-CROPS suggest that at least five distinct chromosome regions are involved in the genetic control of the traits analysed and are mirrored by AGOUEB QTL positions. As expected, both QTL Antho_2 and StS_2 perfectly align with their causative genes *ant-2* and *vrs-1* which provides some validation of the analysis protocols used in both association panels. On 2HS, significant effect for TGW (TGW_4) and HWE (HWE_3) are found together in a chromosome segment monomorphic between Saffron and Retriever (precluding verification from this study as well). The main effect for the S×R TGW QTL on 2HS (Chapter 2) is reinforced by significant effects for TGW in both AGOUEB and NUE-CROPS but unlike the bi-parental mapping results no tillering QTL were found at that QTL cluster in the GWAS. This region located at around 27cM has been shown to be carrying the photoperiod gene *Ppd-H1*, at the exact same position where the highly significant NUE-CROPS Hd_3 QTL was reported. Although the QTL for heading date and TGW overlap, the comparison of experiments does not provide sufficient evidence in itself to confirm that

Ppd-H1 is the gene responsible for the TGW effects mapped but encourage additional investigation on the markers within that region. A considerable numbers of associations of QTL were found around 60cM both association mapping panels (Figure 4.1b) and in a region that was previously related to the flowering time locus *eam6* (Chapter 2). This cluster contains QTL for yield and yield components, especially the strongest effects for grain number (Grains_1), as well as phenology and nitrogen metabolism related traits. In the NUE-CROPS experiment, two slightly distinct QTL clusters seem to be present despite overlapping support intervals. Those can be divided into a cluster containing NUE-CROPS Yld_3 from a cluster containing Grains_1. It suggests that correlations between traits and haplotypes could have generated these different clustering positions. Indeed, the direct relationship and correlation between traits like Yld and GNYld or GrainN and NUtEg (Appendix 3. 4) suggest the effect of a unique underlying major gene. Marker haplotypes analysis may give additional information in that region. More strikingly, the clusters of QTL associated to genetic factors 5 and 6 on 2HL in the S×R experiment (Table 2.9) did not correspond to any significant marker effects found in the GWAS. These S×R clusters were the basis of considerable interest in Chapter 2 where GP.3_2 and SC.1_2 QTL were precisely mapped with high significance and close to independent tillering effects. In the NUE-CROPS variety panel, the Retriever haplotype for the SNPs mapped at 150.7 cM which include the best marker for GP.3_2 (A10791) is only shared by 16 varieties. In a wider chromosome segment, the Retriever haplotype described by the 9K chip SNP mapped between 100.4 to 150.7 cM (Supplementary data 1) and corresponding to the S×R genetic factors 5 and 6 (Table 2.9) is unique within the NUE-CROPs variety panel. Therefore the absence of comparable effects between the biparental mapping and association mapping is likely to be due to novel effects found exclusively for the alleles carried by Retriever which in very low frequency in association mapping panel. This integration and comparison of QTL results shows benefits in enabling the detection of QTL effects of different magnitude and associated with haplotype of variable frequencies. These can provide additional insight in the genetic control of traits.

Figure 4.1 Integrated QTL results from NUE-CROPS, AGOUEB and S×R studies
Diagram of representation of QTL mapped in bi-parental (Chapter 2) and association mapping studies (Chapter 2) on the seven chromosomes of barley (figs a-g). The QTL from the NUE-CROPS and AGOUEB two-row winter barley panels are presented in green and brown respectively. QTL of the bi-parental mapping experiment have been represented and correspond to results Chapter 2. Increasing allele effects from Saffron or Retriever are indicated in blue or red respectively while hatched QTL indicate significant QTL x Environment interaction. Only the polymorphic markers in the population S×R and known genes are presented at their 9K map position. Additional horizontal lines on chromosomes indicate markers used in association mapping.

a)

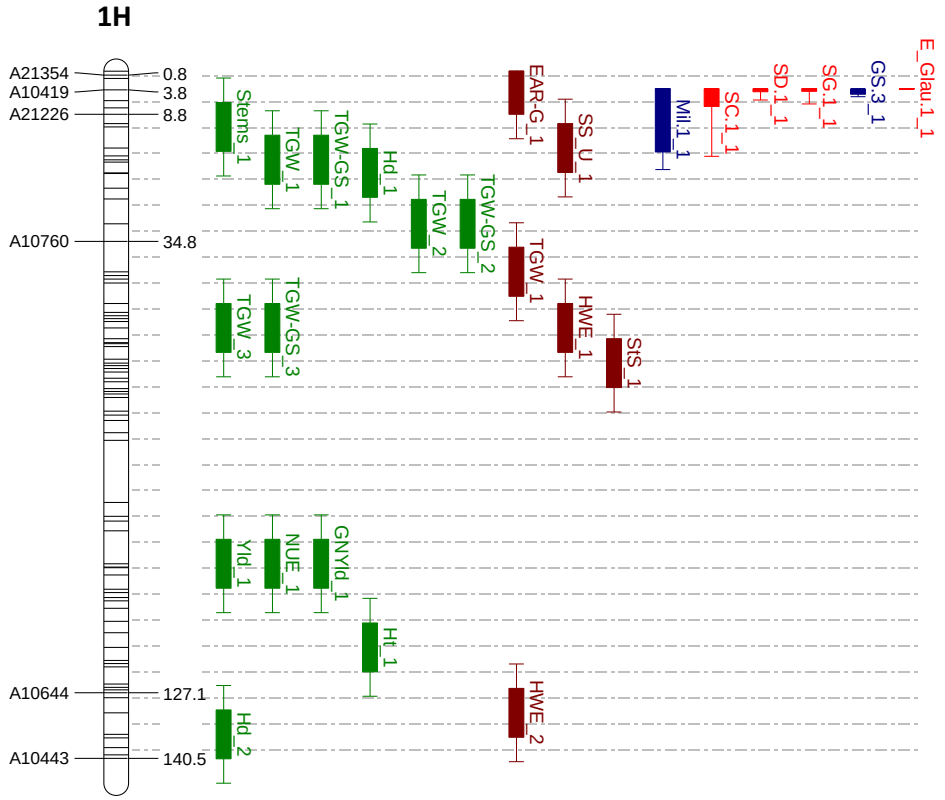


Figure 4.1 cont

b)

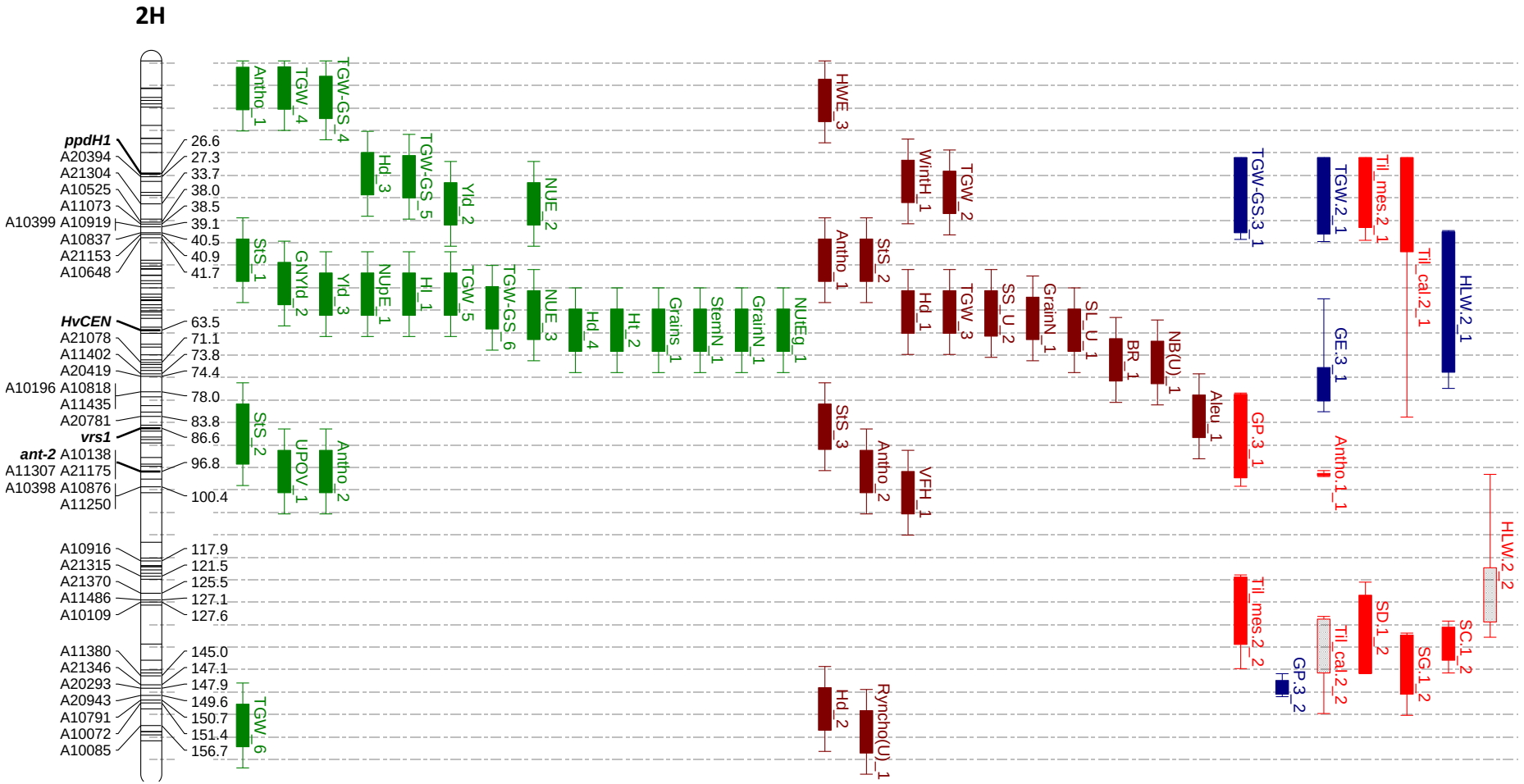
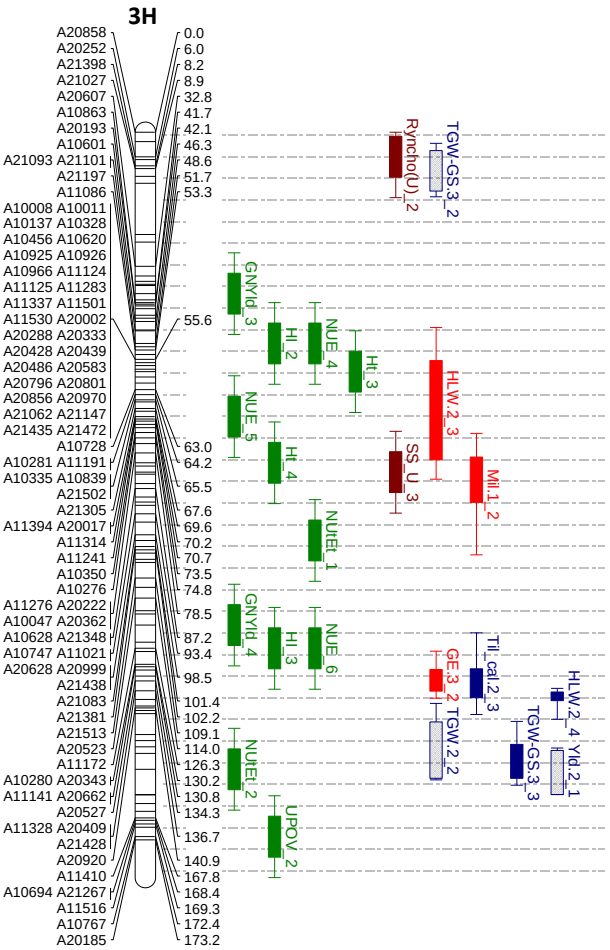
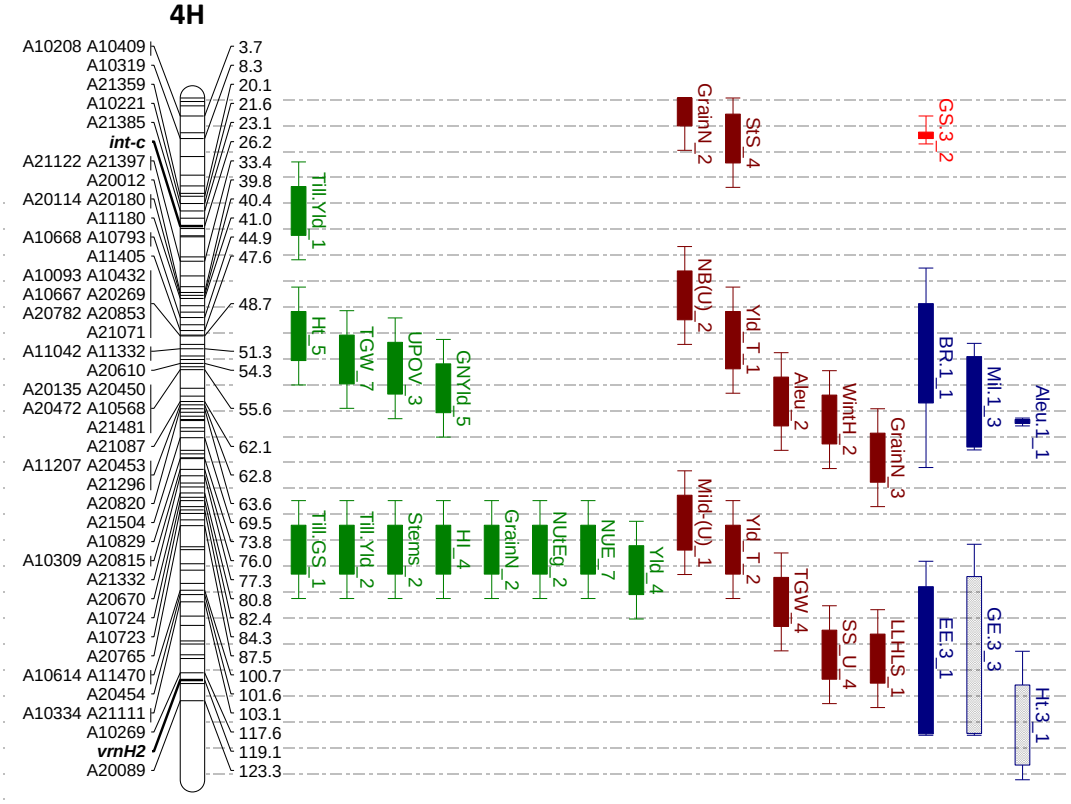


Figure 4.1 cont

c)



d)



Unlike genetic factors of 2H, the significant effects mapped on 3H are sparse and involved in fewer numbers of traits. Harvest index is associated with two of the QTL for NUE and shows strong overlap with GNYld_4. These traits are intimately related as HI is strongly correlated to yield and the derived traits NUE and GNYld. Interestingly, the straw strength QTL SS_U_3 in AGOUEB associates with a QTL for height (Ht_4) in NUE. Although no height QTL was directly associated with SS at that position in AGOUEB, the two effects could relate to a common genetic factor. The S×R QTL have on that chromosome low correspondence with AM results. The two genetic factors previously identified on 3HL in S×R map bins 10 and 11 (Table 2.9) overlap slightly with QTL from NUE-CROPS but none of the traits correspond. The NUtEt_2 QTL overlaps with significant effects found at Yld.2_1 and TGW-GS-3_3 in S×R suggesting that the Saffron haplotype is able to better utilise available nitrogen for the production of bigger grains and more yield. However neither yield nor TGW was significantly mapped together with the NUtEt_2 QTL in both AM panels.

On 4H, the distribution of QTL along the genome differs between studies. Three regions are clearly identifiable in the NUE-CROPS study whereas the support intervals of significant effects mapped in AGOUEB encompass nearly the entire chromosome 4H. The region containing the major gene *int-c* associates with a tillering QTL in NUE-CROPS only and not with effects on the grain characteristics. The gene *int-c* is functional in the determination of the two-row v. six-row characterisation (Ramsay et al., 2011), though the specifically design two-row barley panel studied here suggest that another candidate should be responsible for tillering effect. Around 54 cM, a cluster of QTL containing TGW_7, Ht_5, GNYld_5 and UPOV_3 is co-locating with Yld_T_1 of AGOUEB and disease QTL from S×R population. Although a direct relationship between the traits cannot be established at this position with the current data, the treated yield effect (Yld_T_1) may well be linked to the presence of disease resistance/susceptibility alleles. However only the AGOUEB Net-blotch disease QTL NB(U)_2 was co-located with this cluster and none of the brown rust or mildew QTL found in S×R were confirmed. Changes in allele frequency throughout the years of release can also be expected given the UPOV_3 QTL, although the pericentromeric region means that multiple genes may be driving this change. Additionally, variation in yield and TGW can also be linked to the differences in height of the plant (Ht_5). A notable second cluster of QTL at 96cM in NUE-CROPS and AGOUEB associates the yield QTL (Yld_4; Yld_T_2) with tillering, TGW and nitrogen related traits. Here the

three independent measures of tillering used in the NUE-CROPS mapping have substantial effects and are captured in a common location. Their association to QTL for yield and NUE gives this locus a high interest for understanding the genetic control of tillering and its impact on yield in two-row barley. Therefore further investigation is necessary to describe the genetic polymorphisms underpinning the effects on several traits at that locus. Despite a slight overlap of the S×R QTL in this position, there is little evidence for correspondence between the mapping results.

Figure 4.1 cont.

e)

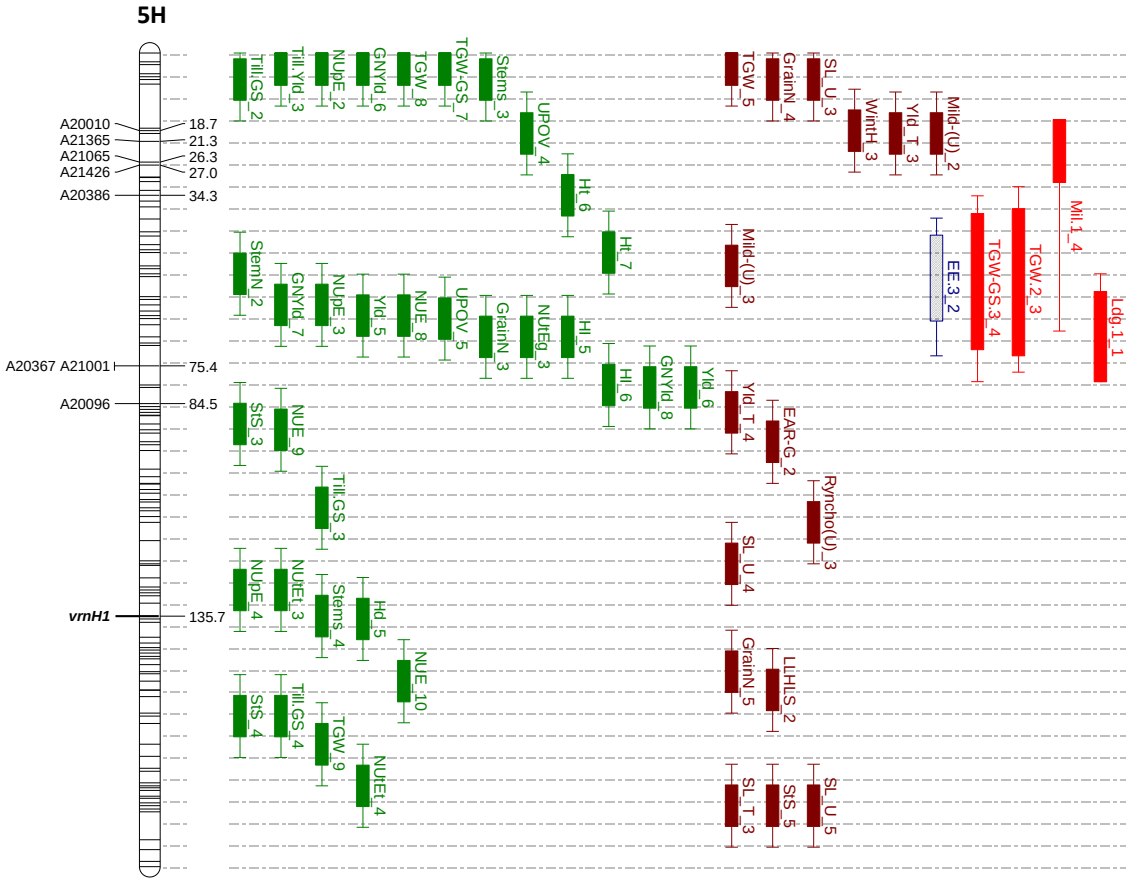
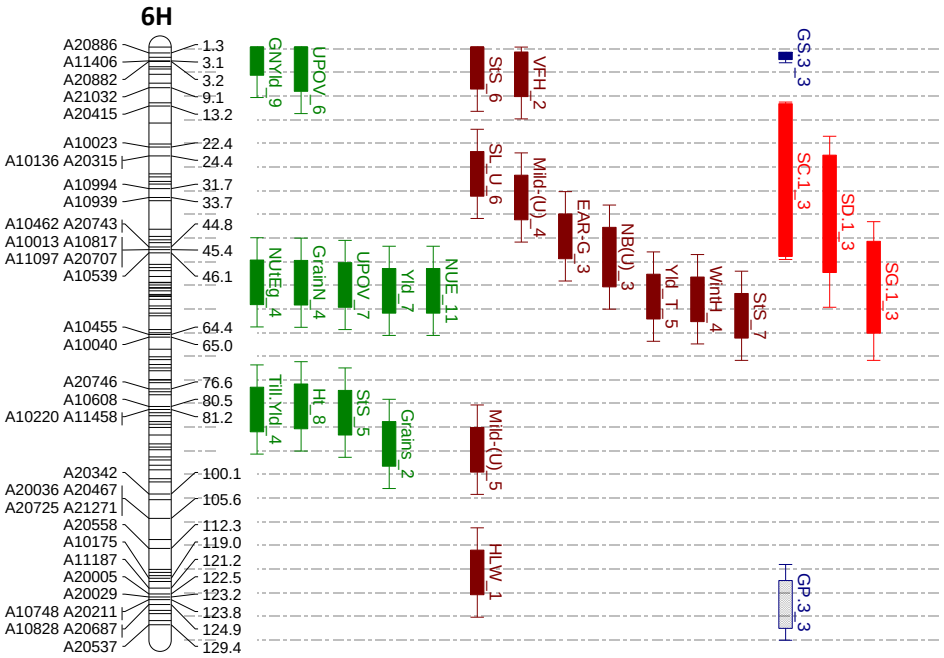
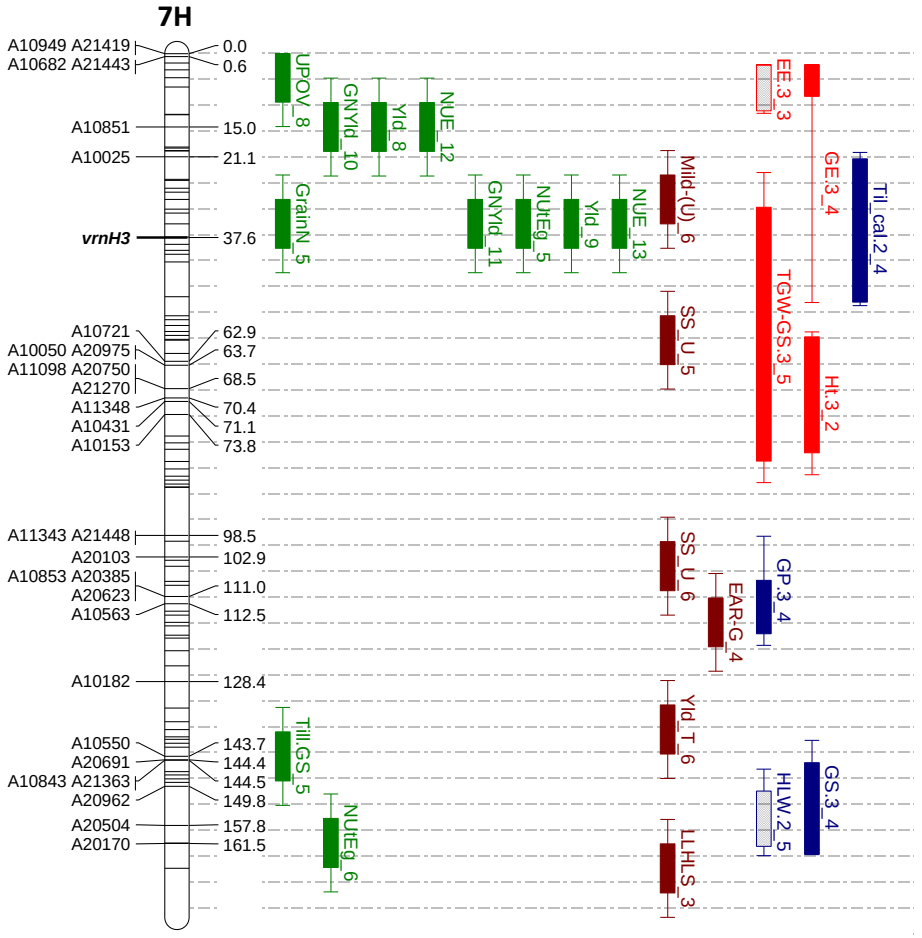


Figure 4.1 cont

f)



g)



On chromosome 5H, highly monomorphic in S×R, two genetic factors located on the short arm co-locate with more than five traits measured in the NUE-CROPS. In the first 10cM of the chromosome, significant effects for tillering and TGW were mapped in the AM panels. Variation in these traits may affect nitrogen related traits at NUpE_2, GNYld_6 and GrainN_4. The S×R Mil.1_4 QTL on 5HS was validated by the most significant mildew resistance QTL mild-(U)_2 detected in AGOUEB with a LOD score above 5. It suggests a strong resistance locus segregating in two-row winter barley. This QTL could relate to the mildew resistance effect recently reported in an Australian bi-parental cross that maps closely to the leaf rust *Rph20* locus (Hickey et al., 2012). Another well-defined cluster in the centromeric region shows significant effects for yield and nitrogen related traits. Mostly derived from the yield (Yld_5) and the nitrogen content figures, the strong correlation of the traits mapped at this cluster (Appendix 3. 4) can explain part of the co-location and strongly suggests the presence of an attractive genetic factor for yield improvement. On the long arm, Yld_T_4 co-locates with Yld_6 and NUE_9 of NUE-CROPS in an additional region of interest. In the remainder of 5HL the QTL are spread out with QTL co-location between mapping studies only found between different traits. For example, the vernalization gene *vrn-H1* (Szucs et al., 2007) associated with effects on heading date (Hd_5) and tillering (Stems_4) in NUE-CROPS was not detected in the AGOUEB panel. This may be due to differences in the phenotype data between the two panels suggesting that concurrently assessed phenotype is more accurate than the BLUP fitted data put together across seasons. In a more distal position on 5HL, AGOUEB straw length effects correspond to NUtEt_4 whereas the height QTL of NUE-CROPS (Ht_6; Ht_7) were mapped on 5HS. From the S×R results, only the mildew QTL was mirrored by AM results. No supporting evidence was found for the QTL cluster containing EE.3_2 which was considered as spurious in Chapter 2.

Three distinct regions are associated with QTL clusters on 6H in NUE-CROPS. Significant effects in the centromere are found in common with AGOUEB, namely for yield (Yld_7; Yld_T_5). This cluster underlines once again the correlation between yield and its derived traits. Here, the winter hardiness effect (WintH_4) may be directly linked with the nitrogen economy of the plant and determinant for yield in the early stages of plant establishment. It is also notable that UPOV_7 was mapped in that region and can relate to a locus capturing yield improvement over years. As observed in chromosomes 1H 2H, 4H and 5H, QTL for StS of low significance were also found on 6H where no ear-row number genes have been described which could potentially

indicate that residual variation associated with the newer varieties with *deficiens* character is not accounted for by the structure correction. Interestingly, the straw collapse QTL (SC.1_3) is overlapping a QTL for straw length while SD.1_3 and SG.1_3 are covering support intervals of markers significant for disease susceptibility, in particular net blotch (Table 3.3b). As observed on 2H, QTL co-location for grain number and height mapped on 6H at around 81cM suggests the existence of common genetic control for the two traits, possibly related to the control of meristem development (3.4.3). Finally on 6HS, the polymorphism for VFH trait is present in the panel due to recently released varieties (e.g. Saffron, B99, B100) which could explain its co-location with UPOV_6 although the trait in itself is of very little interest for varietal improvement.

On 7HS, the two main associations of QTL concern yield and nitrogen related traits (Figure 4.1) which have been shown to be strongly correlated and frequently co-located on other chromosomes. The vernalization locus *vrnH3* is co-located with a group of NUE-CROPS QTL and a tillering effect in S×R (Til.cal.2_2). This locus however was found to be monomorphic between Saffron and Retriever (Figure 2.2). On the remainder of 7H, QTL from the three studies do not match each other's location. Despite a little overlap for some support intervals, there is no specific segment of the chromosome that can be confidently described as a strong candidate region for selection.

The several associations of QTL within and between mapping experiments illustrated by Figure 4.1 suggest that convincing genetic factors can be targeted for further validation to get a better handle on the genetics of the traits. A striking example of multiple associations can be found on 2H at the *eam6* locus where both NUE-CROPS and AGOUEB have strong effects mapped (See 3.4.3). Association between significantly correlated traits were also frequently observed such as yield and NUE traits, sometimes in clusters with QTL for the year of release as well. The two grain number QTL were found associated with height QTL in NUE-CROPS GWAS. Remarkably, the QTL from the bi-parental mapping were rarely found in the regions of major clusters but often in association with QTL of minor effects.

Specific genetic factors are of interest for breeding either for of their significance level or for the number of traits involved in the cluster. A tillering effect was clearly identified on 4HL and the S×R TGW effect on 2H was partly confirmed in AM.

Multiple traits mapped over the *eam6* locus underline its major influence in plant development. However, the conservative approach taken by defining a standard support interval for displaying GWAS QTL emphasises the difficulty in resolving the precise genetic origin of the effect. These QTL regions of interest could as well encompass several putative causal genes or a single functional polymorphism with pleiotropic effects. To investigate further these genetic factors, it is necessary to look at the direction and size of the effects as well as the pattern of association with the different alleles of the markers.

4.3.2 Marker effects at QTL cluster position.

Within a cluster of QTL, the marker effects defined by magnitude and allelic direction can help to give a more detailed interpretation of the overall effect of genetic factors and to potentially characterise underlying candidate genes. No specific line of the AM panel was identified as a reference genotype so that the direction of marker effects were based on allele frequency in the panel considered. If a particular polymorphism has pleiotropic effects or simply associates with correlated traits, the direction of effects at QTL of those traits should be reflected by the same significant alleles. In other words, the similar pattern of significance should be visible across a set of SNP for traits under a similar genetic control. A summary of the effects of best marker for yield and yield components QTL in NUE-CROPS is presented in Table 4.1 and similar information for all traits mapped NUE-CROPS and AGOUEB is reported in Appendix 4. 1.

On 4HL, a convincing genetic factor was detected with significant QTL for yield, tillering and nitrogen related traits (Figure 4.1d). The best marker A20732 for Till.GS_1 had an increasing tillering effect given by the allele G associated with significant effects on NUE and NUtEg and reduced GrainN (Table 4.1). Yld_4 which is also part of the same cluster has identical effect direction to Till.GS_1 across the rest of the traits but has a different best marker I182626. However alleles at this marker are not significantly associated with tillering under the QTL detection threshold. Interestingly most of the QTL at that location have a similar allele for their positive effects which is encouraging for selection in the case of positive correlation between them. Additional information was obtained in exploiting the synteny of the 9K markers based on unigenes to the rice gen genome (Figure 4.2b).

The largest group of QTL was found in the centromere region of 2H and included the strongest effect for grain number (Grains_1) as well as strong heading date, yield and yield component effects. Allele T of A20862 at Grains_1 is associated with a reduction

of three grains per ear, significantly reduces Hd, Ht, NutEg and increases GrainN (Table 4.1). Additional non-significant positive effects are seen for that allele on TGW and tillering. As they are closely linked and within overlapping support intervals, the significant marker effects of Yld_3 (I10398), TGW_5 (A10733) and TGW-GS_6 (A10602) may be due to the effects of the same gene. The major alleles at these peak markers are all associated with a reduction in Hd and Grains but the direction of effect varies across the other traits (Appendix 4. 1). A10733 and A10602 have effects in the opposite direction for TGW and TGW-GS. The major alleles of A20862 and A10398 correspond respectively to a reduction and increase in plant height. In addition to the relatively large genetic distance between those markers in the centromeric region of 2H (54.95-63.5cM), the direction of the effects across traits suggests that the polymorphism associated with grain number and heading date may be distinct to that that causes variation in Ht or TGW. Both Grains and Hd can be the main traits to consider in the region located around HvCEN candidate that is homologue of LOC_Os04g33570 (Comadran et al., 2012), while effects on Ht and TGW need to be confirmed. Indeed, the analysis is complicated by the number of SNP describing the haplotypes in this chromosome region and the limitation of bi-allelic markers to resolve more contrasting haplotype. The investigation of significance patterns described by MTA for a set of traits over the QTL cluster could be a mean to identify the origin of the effects and the similarities between genetic controls.

The more distal QTL cluster on 5H is involved in the control of tillering and TGW. The most significant marker for TGW_8, TGW-GS_7, Till.Yld_3 (A20553) was mapped 4cM away from the best marker of Till.GS_2 (B30975). A20553 major allele (A) is associated with a significant reduction in tiller number (67 tillers/m^2) and an increase in TGW, GNYld and NupE as well as a non-significant but sizeable increase in grain number. The direction of effects of B30975 alleles across traits mirrors A20553 suggesting that the two markers capture the same effect and can be used to search for candidate genes in the rice genome.

As expected from Figure 4.1, Table 4.1 and paragraph 3.3.3, none of the markers significant for the yield QTL were found significant for yield components but were for nitrogen related traits which are often yield derived traits. In addition, the direction of marker effects tended to reflect the correlations of yield with those traits with slight differences in magnitude or significance (Appendix 3. 4). Some yield QTL had non-significant but sizeable effects of the same direction on yield components (e.g. Yld_2 and TGW; Yld_3 and Grains and TGW; Yld_4 and tillering). Others like Yld_5 did not

match any sizeable effect on yield components suggesting that the effect identified are targeting polymorphism which have a main impact on yield or biomass rather than through differential effect on the yield components.

This analysis of marker effects and the pattern of significance at nearby markers reveals the complexity of QTL comparisons and the necessity to account for a range of trais for interpretation. The better insight of QTL analysis using marker effects underlines the fact that although QTL are present in clusters, they may not relate to the same underlying functional polymorphism. The presence of different significantly associated markers at a QTL cluster and the variation in their allele effects in other traits highlights to the difficulty to identify of the choice for what can be considered as a positive allele and an optimal marker for MAS.

Table 4.1 Marker effects across traits at QTL locations for yield and yield components

Effect size and direction on NUE-CROPS agronomic traits for the most significant SNP associated with QTL of yield and yield components. The negative effect of allele 1 for a yield or yield component QTL on another trait are highlighted in grey. Effects associated with a significant marker association (defined as –log10(p) of the SNP > –log10(0.003)) are in bold. A complete table reporting all NUE CROPS QTL is presented in Appendix 4. 1.

Trait	QTL	Chorm	Marker	Dist	-log10(p)	alleles 1	alleles 2	MA	MAF	Effect of marker allele 1: Antho	GNYld	GrainN	Grains	Hd	HI	Ht	NUE	NupE	NutEg	NutEt	StemN	Stems	StS	TGW	TGW-GS	till.GS	till.Yld	UPOV	Yld
Grains	**Grains_1**	2 H	A20862	63.5	7.20	T	A	A	0.42	0.03	0.37	**0.08**	**-3.01**	**-2.43**	-1.35	**-3.93**	-1.52	0.00	**-2.53**	-1.68	0.01	0.07	0.13	0.66	0.60	33.72	46.37	0.87	-0.31
	Grains_2	6 H	I138716	88.9	2.83	G	A	A	0.43	0.00	-1.27	0.04	**-1.74**	0.02	-0.85	-0.31	-0.90	0.00	-1.56	-0.41	0.01	0.06	-0.09	0.82	1.12	24.65	26.03	-1.48	-0.21
Thousand	**TGW_1**	1 H	I232660	18.05	3.24	C	T	C	0.36	0.09	-1.43	-0.02	0.30	0.48	-0.32	0.87	-0.19	-0.01	0.15	0.13	0.00	0.15	0.06	**-2.75**	**-3.25**	30.18	24.80	1.35	-0.03
Grain Weight	**TGW_2**	1 H	I128285	31.15	3.45	T	C	T	0.36	0.06	-1.26	-0.02	0.32	0.49	-0.24	0.38	-0.21	-0.01	0.17	0.08	0.00	0.10	0.06	**-2.64**	**-2.88**	22.25	20.90	0.73	-0.03
	TGW_3	1 H	A20810	52.46	3.56	A	G	A	0.46	-0.02	-1.62	0.00	0.47	0.11	-0.29	1.00	-0.49	-0.01	-0.18	-0.54	0.00	0.09	0.08	**-2.76**	**-3.20**	17.85	8.64	-0.07	-0.11
	TGW_4	2 H	I146936	6.4	3.33	T	C	C	0.18	0.11	-1.34	0.01	0.10	-0.09	0.02	-0.92	-0.71	0.00	-0.48	-1.07	0.00	0.13	-0.18	**-3.53**	**-4.01**	2.08	13.14	-0.69	-0.15
	TGW_5	2 H	A10733	54.95	4.33	G	C	C	0.14	0.05	**3.82**	0.01	-1.60	-1.08	0.42	-3.72	0.86	**0.02**	-0.70	-0.02	0.01	-0.05	0.22	**4.03**	**3.76**	0.77	-1.81	-0.95	0.20
	TGW_6	2 H	I195051	156.72	2.76	T	C	C	0.40	0.10	2.33	0.00	0.28	0.41	0.58	1.85	0.86	0.01	0.63	1.01	-0.01	-0.06	0.05	**2.71**	2.29	-11.73	-25.63	-0.96	0.16
	TGW_7	4 H	B30427	53.5	2.61	T	A	A	0.25	0.15	0.86	0.00	-0.57	0.03	0.16	-1.08	0.46	0.01	-0.06	0.66	0.00	-0.05	0.14	**2.45**	2.46	-2.91	-19.02	1.59	0.08
	TGW_8	5 H	A20553	2.81	3.44	A	G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
	TGW_9	5 H	I194030	166.63	2.83	A	G	A	0.38	-0.17	-1.36	0.03	-0.55	0.44	-0.92	0.02	-1.20	0.00	-1.15	-0.69	0.00	0.13	0.03	**-2.62**	-2.61	26.24	27.98	-3.63	-0.24
Thousand	**TGW-GS_1**	1 H	I232660	18.05	3.42	C	T	C	0.36	0.09	-1.43	-0.02	0.30	0.48	-0.32	0.87	-0.19	-0.01	0.15	0.13	0.00	0.15	0.06	**-2.75**	**-3.25**	30.18	24.80	1.35	-0.03
Grain Weight	**TGW-GS_2**	1 H	I128285	31.15	3.19	T	C	T	0.36	0.06	-1.26	-0.02	0.32	0.49	-0.24	0.38	-0.21	-0.01	0.17	0.08	0.00	0.10	0.06	**-2.64**	**-2.88**	22.25	20.90	0.73	-0.03
from	**TGW-GS_3**	1 H	A20810	52.46	3.63	A	G	A	0.46	-0.02	-1.62	0.00	0.47	0.11	-0.29	1.00	-0.49	-0.01	-0.18	-0.54	0.00	0.09	0.08	**-2.76**	**-3.20**	17.85	8.64	-0.07	-0.11
grab samples	**TGW-GS_4**	2 H	I213799	8.57	3.76	A	C	C	0.24	0.10	-0.75	0.01	-0.77	0.62	-0.69	-2.05	-0.30	0.00	-0.54	-0.01	0.00	0.05	-0.06	**-3.09**	**-3.94**	33.59	48.61	-0.11	-0.12
	TGW-GS_5	2 H	I143250	27.3	3.05	A	G	A	0.47	0.14	1.31	-0.01	0.46	0.03	0.19	1.77	0.61	0.00	0.52	0.56	-0.01	-0.10	0.03	2.16	**2.83**	-19.36	-23.52	3.30	0.12
	TGW-GS_6	2 H	A10602	58.24	3.09	A	C	C	0.26	-0.29	-1.12	0.05	-0.14	-1.15	-1.02	-0.96	-0.79	0.00	-1.34	-0.72	0.00	0.10	-0.16	**-2.92**	**-3.51**	21.60	25.85	-2.06	-0.20
	TGW-GS_7	5 H	A20553	2.81	3.75	A	G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
Tillering from	**till.GS_1**	4 H	A20732	92.38	3.75	G	A	G	0.42	-0.03	0.31	**-0.08**	-0.51	0.65	1.45	-2.03	**1.56**	0.00	**2.47**	0.95	0.00	**0.22**	0.05	-0.86	-1.28	**55.67**	**57.53**	3.53	0.29
Grab samples	**till.GS_2**	5 H	B30975	6.4	3.59	A	C	C	0.17	-0.03	**3.56**	0.04	1.41	-0.05	-0.69	3.59	0.53	**0.02**	-1.04	1.24	-0.01	**-0.22**	0.06	1.37	1.62	**-66.86**	**-63.32**	0.82	0.13
	till.GS_3	5 H	I147762	109.56	2.62	C	T	T	0.26	-0.11	1.29	-0.03	1.36	1.13	0.43	2.00	0.95	0.01	1.16	1.29	-0.01	-0.11	0.06	1.68	1.87	**-48.35**	-45.97	0.78	0.18
	till.GS_4	5 H	I720	159.8	3.26	A	G	A	0.37	-0.04	-0.58	0.01	-1.32	0.03	-0.88	-1.66	-0.33	0.00	-0.79	0.07	0.00	0.13	0.03	-1.30	-1.28	**55.52**	49.56	-1.91	-0.08
	till.GS_5	7 H	A10550	143.68	2.63	G	A	A	0.25	-0.18	-0.03	0.02	0.70	0.66	-0.38	1.61	-0.61	0.00	-0.68	-0.24	0.00	-0.13	-0.05	0.05	0.24	**-51.15**	-40.98	-1.60	-0.12
Tillering from	**till.Yld_1**	4 H	A21385	23.1	2.73	G	C	C	0.26	0.06	-0.46	-0.01	0.68	0.96	-0.37	1.99	-0.66	-0.01	-0.14	-0.71	-0.01	-0.09	0.00	0.69	0.52	-36.20	**-58.15**	-2.71	-0.12
yield	**till.Yld_2**	4 H	A20732	92.38	3.31	G	A	G	0.42	-0.03	0.31	**-0.08**	-0.51	0.65	1.45	-2.03	**1.56**	0.00	**2.47**	0.95	0.00	**0.22**	0.05	-0.86	-1.28	**55.67**	**57.53**	3.53	0.29
	till.Yld_3	5 H	A20553	2.81	3.05	A	G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
	till.Yld_4	6 H	I4707	81.2	2.81	C	T	T	0.33	-0.19	-0.65	-0.02	-1.10	0.83	0.39	-2.00	0.10	0.00	0.52	-0.62	0.00	0.12	-0.07	-1.53	-1.48	41.69	**59.19**	-1.19	-0.03
Yield	**Yld_1**	1 H	I154646	100.7	3.49	A	C	C	0.11	-0.30	**3.88**	-0.07	0.87	1.53	1.88	-0.27	**2.54**	0.01	2.23	2.14	0.00	0.08	-0.04	1.12	0.10	20.37	16.07	1.67	**0.53**
treated	**Yld_2**	2 H	A21304	33.74	2.73	A	G	G	0.30	-0.08	-2.53	0.06	0.14	-0.22	-1.16	2.11	**-1.83**	-0.01	-1.84	-1.27	0.00	0.03	-0.11	-1.81	-2.01	2.65	-14.92	-3.94	**-0.37**
	Yld_3	2 H	I10398	54.95	3.33	C	T	T	0.17	-0.27	-2.20	**0.11**	-0.99	-0.10	**-2.79**	2.79	**-2.35**	-0.01	**-3.06**	-0.98	-0.01	-0.06	-0.09	-1.90	-1.70	1.54	7.85	-1.73	**-0.54**
	Yld_4	4 H	I182626	96.6	3.08	T	G	G	0.31	0.22	1.05	**-0.08**	-0.27	0.93	0.95	-3.04	**1.88**	0.01	1.65	1.65	0.00	0.16	0.16	0.02	-0.10	43.44	45.13	3.17	**0.37**
	Yld_5	5 H	I231238	63.31	3.55	T	C	C	0.21	0.12	**3.69**	-0.05	0.30	-0.47	**1.81**	-2.80	**2.19**	0.01	1.71	1.53	0.00	0.01	0.08	1.40	0.72	8.69	1.25	5.10	**0.45**
	Yld_6	5 H	A20236	80.61	2.60	C	A	A	0.26	-0.08	**2.90**	-0.04	0.51	-0.54	1.16	-0.59	1.45	0.01	1.24	0.57	0.00	-0.07	0.07	0.44	0.16	-7.50	-5.86	2.73	**0.33**
	Yld_7	6 H	I118381	54.6	2.69	C	T	C	0.47	-0.01	1.20	-0.07	-0.37	0.58	1.00	-1.96	**1.62**	0.00	1.92	1.04	0.00	0.10	0.04	1.68	1.69	28.22	26.87	4.68	**0.34**
	Yld_8	7 H	I186187	14.96	3.66	T	G	G	0.22	0.34	**3.07**	-0.07	0.73	0.27	1.57	-1.43	**1.91**	0.01	1.82	1.29	0.00	0.05	0.16	0.61	0.94	7.50	14.85	4.52	**0.44**
	Yld_9	7 H	I138457	34.82	3.94	C	A	A	0.22	0.09	**3.18**	**-0.07**	-0.14	1.22	1.32	-2.82	**2.10**	0.01	**2.22**	1.83	0.00	0.04	0.00	0.83	1.00	6.77	26.61	2.58	**0.43**

4.3.3 Synteny and candidate genes

The research concentrated on a subset of QTL clusters in order to search for candidate genes. The synteny with rice at each cluster was made by aligning the homologous rice loci to the significant GWAS SNP markers (Chapter 3) and their association test results (Figure 4.2). The putative rice genes encompassed by those homologous segments can be surveyed and the barley homologue considered as candidates with a putative polymorphism inducing variation in a trait of interest.

The focus was on the 4HL QTL cluster with a clear and large effect on tillering. Here, the support interval included markers that were derived from genes homologous with various rice chromosomes but a majority of SNP homologous to rice chromosome 3 (Figure 4.2b). A group of three SNPs correspond to QTL Till.GS_1, Till.Yld_2, Stems_2, HI_4, GrainN_2, NUtEg_2, NUE_7 and to a lesser extent to Yld_4 while the UPOV showed also a similar pattern of association, albeit non-significant. The homologous rice chromosome segment isolated from the best SNPs is comprised between LOC_Os03g09020 (I141214) and LOC_Os03g05430 (B30584) (Figure 4.2b) and all the homologous candidates genes in that interval reported in the rice genome database (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/) can be examined as potential candidate genes. Similar association levels for the traits were found at the SNP I168399 in a gene homologous rice locus LOC_Os03g09150 suggesting that co-linearity in gene order between barley and rice is not complete. On closer examination it is clear that for the region of interest on 4HL although largely syntenic with the short arm of rice chromosome 3 the co-linearity as presented is interrupted by a number of inversions. Using the information derived from the barley PopSeq map derived from next generation sequencing of lines from the Morex × Barke and Oregon × Wolfe mapping populations (Mascher et al., 2013b), there is a break in the co-linearity with synteny jumping from LOC_O3g09150 to LOC_O3g04960 in this region of 4H (Supplementary data 1). Indeed the break is confirmed by the delineation of markers BAC contigs with barley homologues of LOC_O3g04960 and LOC_03g_09070 – LOC_03g09150 on the same BAC contig (contig 44100) (IBGSC, 2012; http://pgsb.helmholtz-muenchen.de/cgi-bin/gb2/gbrowse/Barley_PhysMap/). This order of rice genes means that the three peaks evident in figure 4.2b relating to I129218, I168399 and I172072 collapse to a single peak. The most significant marker in that interval (I129218) across the traits corresponds to rice LOC_Os03g05280 identified as encoding a ‘ras-related protein’ (Figure 4.2b, Table 4.2b) a member of a GTPase family

(McCormick, 1995). The cluster interval considered also includes transcription factors as well as proteins with various functional domains; all putative candidates (Table 4.2b). However it is clear that distal to the peak marker there is a gap in the genetic map (from 92.4 to 96.6 cM) meaning potential candidate genes could be present within this unresolved region.

The same approach was carried out on the region of high interest for TGW and Hd QTL on 2HS which shows synteny with rice chromosome 7 (Figure 4.2a). The QTL mapped at that cluster were detected with low levels of significance except the Hd effects at six distinct gene based SNPs all homologous to LOC_Os07g49460 (Figure 4.2a, Table 4.2a). These SNPs correspond to gene based SNPs diagnostic for early and later alleles of the barley *Ppd-H1* gene (Faure et al., 2007). In Chapter 2 and Chapter 3, *Ppd-H1* was cautiously suggested as a candidate gene underpinning effects on tillering and TGW in this cluster. The patterns of significance for the MTA for both heading date and TGW enable us to differentiate between the SNPs associated to heading date which were within the *Ppd-H1* gene sequence and those associated to TGW. Indeed, the six most significant SNPs for heading date are not significantly associated to TGW effects suggesting different genetic controls for these two traits. The TGW effects can therefore be due to either polymorphism in a nearby gene or to additional alleles at *ppd-H1* that have no effects heading and have not been captured by genic SNP (Figure 4.2a and Table 4.1a). Further sequencing of the 3'UTR of *Ppd-H1* gene sequence in a range of varieties identified four different haplotypes within a set of winter barley varieties (Appendix 5. 4). These preliminary results indicate that additional haplotypes of the *Ppd-H1* gene exist in winter barley and may share identical allele distribution with markers significantly associated to TGW QTL (e.g. SNPs I143250 of QTL TGW-GS_5 and A20394 of TGW_GS_3.1 in S×R).

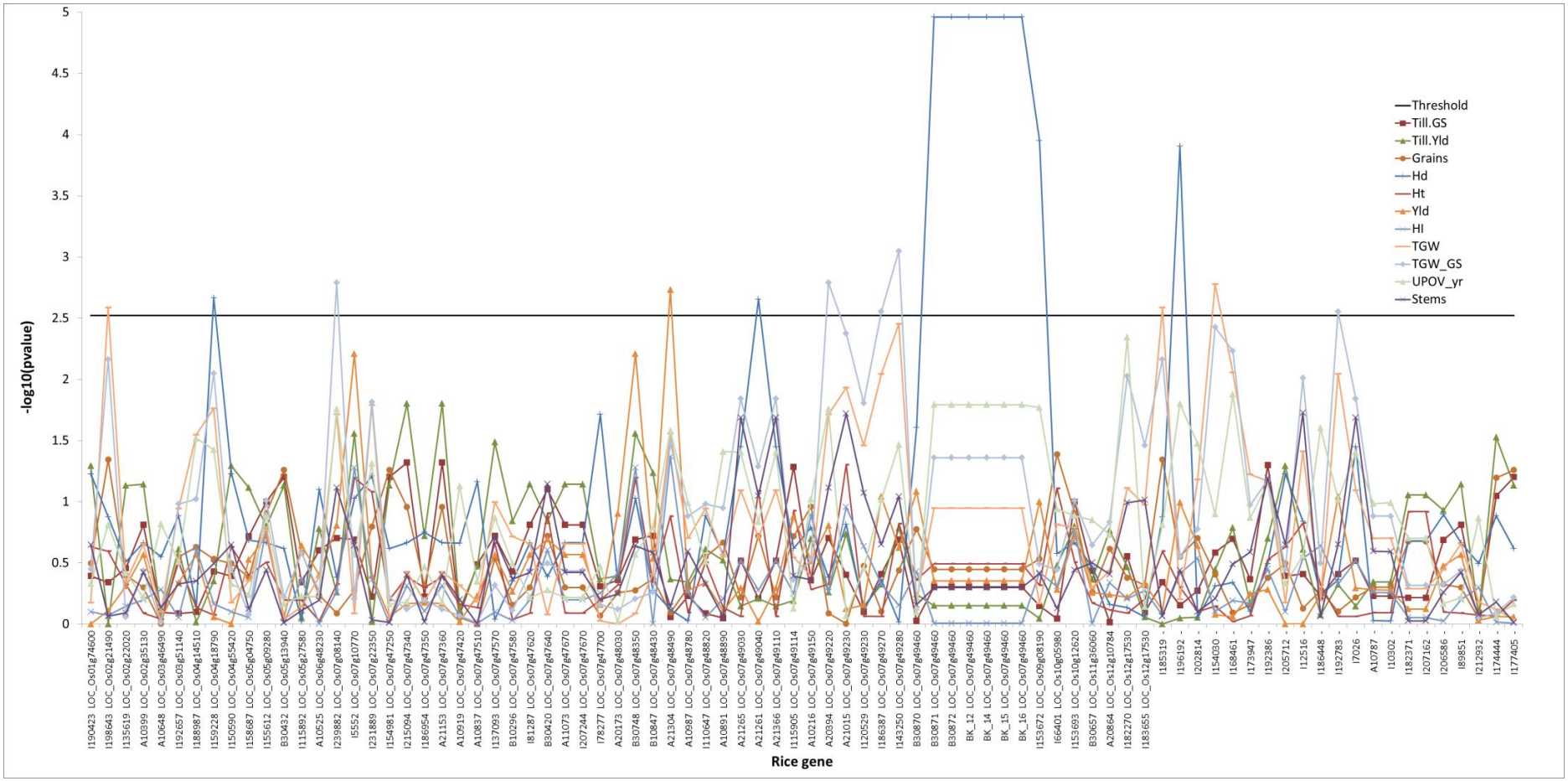


Figure 4.2 Patterns of marker trait associations at two QTL clusters identified from NUE-CROPS GWAS.

a) Pattern of marker trait associations for all 20 traits mapped in the NUE-CROPs GWAS at the QTL cluster showing significant TGW and heading date effects on 2HS at the Ppd-H1 gene locus (26-29cM). The X-axis plots barley SNP ordered according to the physical position of their homologous rice locus (e.g. SNP I190423 is homologous of rice locus LOC_Os01g74600). The Y-axis indicates the magnitude of the association as –log10(*p*). SNP markers BK_12, BK14, BK_15, BK15, B30872 and B30871 are diagnostic markers to the barley gene *Ppd-H1* homologous to rice gene 49460 on rice chromosome 7.

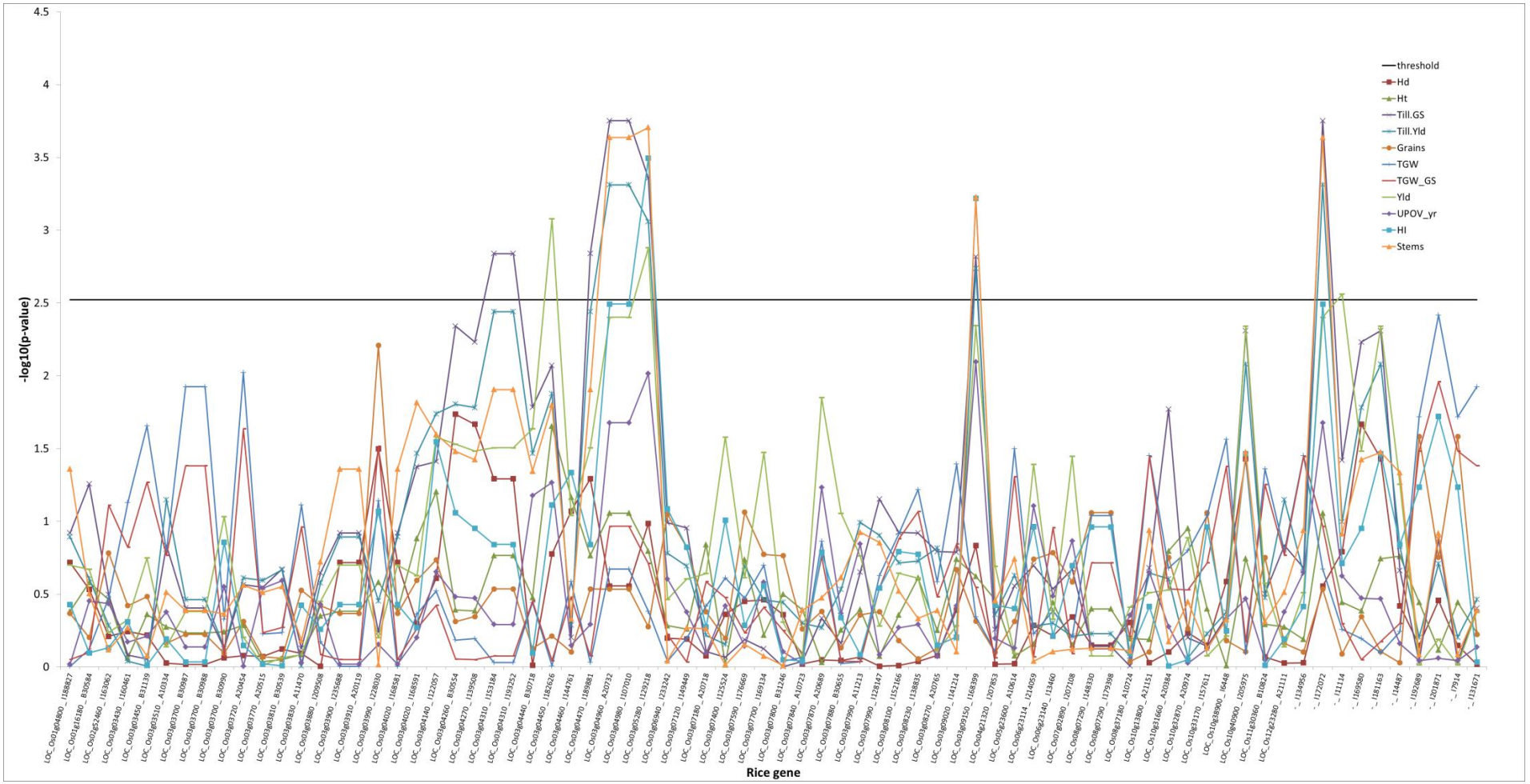


Figure 4.2 cont.

b) Pattern of marker trait association for all 20 traits mapped in the NUE-CROPs GWAS at the QTL cluster showing significant tillering effects on 4HL (92-96cM). The X-axis plots barley SNP ordered according to the physical position of their homologous rice locus (e.g. SNP I188827 is homologous of rice locus LOC_Os01g04800 found at the rice gene 04800 on Os01). The Y-axis indicates the magnitude of the association.

Table 4.2 List of rice gene models homologous to the barley QTL clusters of interest on 2HS and 4HL.

List of putative rice gene models homologous to the chromosome barley chromosome interval delimited by the best SNP markers (Figure 4.2) at **a)** the TGW and heading date QTL cluster on 2HS (26.5-28.4 cM) and **b)** the QTL cluster of significant tillering effects on 4HL (91.8-92.4 cM). The order of the list is based on the rice physical sequence ordering. The homologous barley SNP present on the 9K Illumina Iselect genotyping platform and their chromosome position are presented.

a)

Rice Locus	Gene product name	Barley SNP	Iselect cM
LOC_Os07g49000	DNAJ heat shock N-terminal domain-containing protein, putative, expressed		
LOC_Os07g49010	TOPBP1B - Similar to DNA replication protein TOPBP1 from, expressed		
LOC_Os07g49020	expressed protein		
LOC_Os07g49030	PHD-finger family protein, expressed	A21265	28.44
LOC_Os07g49040	protein phosphotase protein, putative, expressed	A21261	28.44
LOC_Os07g49050	expressed protein		
LOC_Os07g49070	expressed protein		
LOC_Os07g49080	COBRA-like protein 7 precursor, putative, expressed		
LOC_Os07g49090	WD-40 repeat family protein, putative, expressed		
LOC_Os07g49100	pectinesterase, putative, expressed		
LOC_Os07g49110	D-alanine--D-alanine ligase family, putative, expressed	A21366	28.44
LOC_Os07g49114	wound-induced protein WI12, putative, expressed	I115905	27.3
LOC_Os07g49120	sex determination protein tasselseed-2, putative, expressed		
LOC_Os07g49140	expressed protein		
LOC_Os07g49150	26S protease regulatory subunit 4, putative, expressed	A10216	26.53
LOC_Os07g49200	membrane associated DUF588 domain containing protein, putative, expressed		
LOC_Os07g49220	expressed protein	A20394	27.29
LOC_Os07g49230	ubiquitin-activating enzyme, putative, expressed	A21015	27.29
LOC_Os07g49240	MRH1, putative, expressed		
LOC_Os07g49250	thiamine pyrophosphate enzyme, C-terminal TPP binding domain containing protein, expressed		
LOC_Os07g49260	importin subunit beta, putative, expressed		
LOC_Os07g49270	AMP deaminase, putative, expressed	I186387	27.3
LOC_Os07g49280	PMR5, putative, expressed	I143250	27.3
LOC_Os07g49290	PHD finger family protein, putative, expressed		
LOC_Os07g49300	expressed protein		
LOC_Os07g49310	omega-3 fatty acid desaturase, chloroplast precursor, putative, expressed		
LOC_Os07g49320	HEAT repeat family protein, putative, expressed		
LOC_Os07g49330	phospholipase C, putative, expressed		
LOC_Os07g49350	expressed protein		
LOC_Os07g49360	peroxidase precursor, putative, expressed		
LOC_Os07g49370	glycosyltransferase family 43 protein, putative, expressed		
LOC_Os07g49380	PWWP domain containing protein, expressed		
LOC_Os07g49390	P-protein, putative, expressed		
LOC_Os07g49400	OsAPx2 - Cytosolic Ascorbate Peroxidase encoding gene 4,5,6,8, expressed		
LOC_Os07g49410	uncharacterized ACR, YagE family COG1723 containing protein, expressed		
LOC_Os07g49460	response regulator receiver domain containing protein, expressed	B30870 (ppdH1)	26.57
LOC_Os07g49470	protein kinase APK1B, chloroplast precursor, putative, expressed		
LOC_Os07g49480	KIP1, putative, expressed		
LOC_Os07g49510	expressed protein		
LOC_Os07g49520	2-oxoglutarate dehydrogenase E1 component, mitochondrial precursor, putative, expressed		
LOC_Os07g49530	MYB family transcription factor, putative, expressed		
LOC_Os07g49540	expressed protein		

Table 4.2 cont.

b)

Barley homologues of LOC_Os03g09150 and LOC_Os03g04960 present on the same BAC contig (contig 44100) (IBGSC, 2012; http://pgsb.helmholtz-muenchen.de/cgi-bin/gb2/gbrowse/Barley_PhysMap/) and highlighted in grey.

Rice locus	Gene product name	Barley SNP	Iselect cM
LOC_Os03g09020	dehydrogenase, putative, expressed	I141214	91.78
LOC_Os03g09030	expressed protein		
LOC_Os03g09040	hypothetical protein		
LOC_Os03g09060	prenyltransferase, putative, expressed		
LOC_Os03g09070	leucine rich repeat domain containing protein, putative, expressed		
LOC_Os03g09080	ubiquitin carboxyl-terminal hydrolase domain containing protein, expressed		
LOC_Os03g09090	expressed protein		
LOC_Os03g09100	calmodulin-binding transcription activator, putative, expressed		
LOC_Os03g09110	mitochondrial carrier protein, putative, expressed		
LOC_Os03g09120	expressed protein		
LOC_Os03g09130	expressed protein		
LOC_Os03g09140	ras-related protein, putative, expressed		
LOC_Os03g09150	pumilio-family RNA binding repeat domain containing protein, expressed	I168399	92.4
LOC_Os03g04960	cysteinyl-tRNA synthetase, putative, expressed	A20732	92.38
LOC_Os03g04970	T-complex protein, putative, expressed		
LOC_Os03g04980	PHD-finger domain containing protein, putative, expressed	I107010	92.38
LOC_Os03g04990	expressed protein		
LOC_Os03g05020	PIR, putative, expressed		
LOC_Os03g05030	dirigent, putative, expressed		
LOC_Os03g05040	expressed protein		
LOC_Os03g05049	expressed protein		
LOC_Os03g05060	exostosin family domain containing protein, expressed		
LOC_Os03g05070	exostosin family domain containing protein, expressed		
LOC_Os03g05080	expressed protein		
LOC_Os03g05100	expressed protein		
LOC_Os03g05110	xyloglucan galactosyltransferase KATAMARI1, putative, expressed		
LOC_Os03g05120	expressed protein		
LOC_Os03g05130	expressed protein		
LOC_Os03g05140	receptor-like protein kinase 2 precursor, putative, expressed		
LOC_Os03g05150	hypothetical protein		
LOC_Os03g05160	GATA zinc finger domain containing protein, expressed		
LOC_Os03g05170	expressed protein		
LOC_Os03g05180	expressed protein		
LOC_Os03g05200	DENN domain containing protein, expressed		
LOC_Os03g05210	WD domain, G-beta repeat domain containing protein, expressed		
LOC_Os03g05220	expressed protein		
LOC_Os03g05225	expressed protein		
LOC_Os03g05250	expressed protein		
LOC_Os03g05260	ankyrin repeat domain containing protein, expressed		
LOC_Os03g05270	RING finger and CHY zinc finger domain-containing protein 1, putative, expressed		
LOC_Os03g05280	ras-related protein, putative, expressed	I129218	92.4
LOC_Os03g05290	aquaporin protein, putative, expressed		
LOC_Os03g05300	cyclin-dependent kinases regulatory subunit 1, putative, expressed		
LOC_Os03g05310	pheophorbide a oxygenase, chloroplast precursor, putative, expressed		
LOC_Os03g05320	expressed protein		
LOC_Os03g05330	HEAT repeat family protein, putative, expressed		
LOC_Os03g05334	expressed protein		
LOC_Os03g05370	expressed protein		
LOC_Os03g05380	expressed protein		
LOC_Os03g05390	Citrate transporter protein, putative, expressed		
LOC_Os03g05420	MT-A70 domain containing protein, expressed		
LOC_Os03g05430	peptidase, putative, expressed	B30584	96.59

The homologous segment on the rice chromosome shows that the two most significant markers for TGW are actually located around 20 gene models away from *Ppd-H1* at genes homologous to LOC_Os07g49220 (A20394) and LOC_Os07g49280 (I143250) respectively and these SNPs are not associated with heading date effect (Figure 4.2a). The allele distribution for the two SNPs is nearly identical within the panel except for three varieties and the minor alleles frequencies of A20394 and I143250 within groups defined by early and late alleles at *ppd-H1* are 0.29 and 0.43 respectively. This distribution suggests that if an allele of *Ppd-H1* is involved in the control of TGW, it is not associated with the major heading date difference and is captured by haplotypes made from adjacent markers. Alternatively, the TGW variation could originate from a polymorphism in a gene nearby *Ppd-H1*. Several candidate genes can be observed in the syntenic region of rice chromosome 7 including a sex determination protein Tassel Seed-2 at LOC_Os07g49120 and a Glycosyltransferase protein at LOC_Os07g49370 (Table 4.2b).

As shown with *Ppd-H1*, the genes of known function can be precisely located using the rice synteny. The unique significant SNP for Hd_5 QTL on 5HL (B30867-LOC_Os03g54084) is seven rice gene models away from B30883, identified as the candidate gene for *vrn-H1* (OSU_VRN_H1_BM5A_intron1_vc_80 and LOC_Os03g50416) (Appendix 4. 3). This particular SNP diagnostic for differences between winter and spring types of barley was indeed monomorphic in the winter barley panel of NUE-CROPS. Although only a single marker was significant given the threshold retained, it strongly suggests the presence of an additional allele for the *vrn-H1* gene that would affect heading date in winter barley. Interestingly the pattern of the MTA results for stems (Stems_4) highlighted two SNPs with unknown rice correspondence in the same region but no significant effect for B30867 (Appendix 4. 3), suggesting that the co-location of tillering and heading date at that locus is not due to a pleiotropic effect on the *vrn-H1* gene. Other clusters of QTL such as the 2H centromeric cluster, highlight the need for investigating a wider chromosome segment to identify candidate gene location. In this example, the HvCEN gene homologous to LOC_Os03g33570 (Comadran et al., 2012) appears to be located in-between the best SNP for Grains (LOC_Os03g33220) and heading date (LOC_Os03g34080) and the overall support interval of the QTL cluster overlaps segments of rice chromosomes 3 and 7 (Appendix 4. 4). While HvCEN belongs to rice chromosome 3, the TGW QTL cluster (including TGW_5) pinpoint a narrow segment of rice chromosome 7 that contains LOC_Os07g42970 identified as a UDP-glucoronosyl and UDP-glucosyl

transferase domain containing protein, involved in the starch synthase activity (Baroja-Fernández et al., 2003; Singh et al., 1981).

The QTL cluster of yield components on 5HS (Containing Till.GS_2, TGW-GS_7) targets a rice segment containing 10 gene models comprised between LOC_Os12g44240 (N-acetylglucosaminyltransferase) and LOC_Os12g44310 (carotenoid cleavage dioxygenase). While on the centromeric region of 5H, the cluster containing the yield QTL Yld_5 and nitrogen related traits QTL (e.g. NUE_8) encompasses over 200 rice gene models based on the rice homologues of significant barley SNP suggesting that additional resolution is needed in order to suggest candidate genes. On 7H, the matching association patterns highlight a segment located between rice loci LOC_Os06g05880 and LOC_Os06g06090 (I138457) affecting yield and nitrogen related traits. This particular segment is located 21 rice gene models from the *vrnH3* barley gene (HvFT1) for which the rice homologue is LOC_Os06g06300 (Appendix 4. 5). However, none of the three barley SNPs that capture alleles of *vrnH3* (B30893, B30894, B30895) were significantly associated with any traits, including heading date despite being polymorphic in the panels. Because all varieties are winter barley, they are not expected to be polymorphic at the vernalization genes but the gene could still be considered as candidate for the QTL cluster. There may be another polymorphism in *vrnH3* not described by the 9K SNP array that associates with I138457.

The candidate genes for the corresponding effects on mildew resistance found on 5HS in S×R and AGOUEB were also investigated. The delimited segment from the GWAS scan is homologous to rice chromosome 12 and contain approximately 40 gene models located between LOC_Os12g43130 and LOC_Os12g43560 (Appendix 4. 6). Amongst them, different genes could be suggested to participate in the host response to mildew such as zinc finger protein, multiple copies of genes coding for thaumatin (LOC_Os12g43490) and an actin depolymerising factor (LOC_Os12g43340) (Appendix 4. 7).

The improved resolution obtained from the GWAS can successfully capture effects at documented genes involved in known phenotypic variation (*Ppd-H1*, *vrn-H1*) and can be confirmed by synteny. Therefore, when a relevant chromosome segment has been bracketed from highly significant SNP at a QTL cluster and confirmed by the pattern of MTA significance, the survey of homologous genes in rice offers the possibility to identify homologous gene candidates. Here putative genes involved in the control of tillering (4H), grain number (2H), TGW (2H) and mildew resistance (5H) have been

proposed. Nevertheless this approach highlights some limits of the QTL comparison to resolve the exact origin of polymorphism and to distinguish at the haplotype level between either unrepresented alleles or genuine multiple close polymorphic genes.

4.4 Discussion

4.4.1 Comparison of mapping studies

In breeding programmes, a substantial genetic diversity is maintained in the germplasm to enable selection for optimal allelic combinations for a range of environments. A large number of barley QTL studies published during the last few years have identified QTL with favourable alleles that can be selected in MAS (Chapter 1). In the case of bi-parental mapping (Chapter 2), the restricted diversity and mapping resolution can be a limitation to the rapid implementation of mapping results in breeding (Rae et al., 2007). Therefore GWAS (Chapter 3) was proposed as an alternative QTL mapping approach to investigate more allelic diversity with greater resolution than bi-parental mapping (Rostoks et al., 2006). In barley, the GWAS have included diverse panel origins such as American germplasm (Berger et al., 2012), Mediterranean barley panels (Comadran et al., 2011b; Visioni et al., 2013) and well investigated European germplasm (Wang et al., 2012). In this project both bi-parental and GWAS approaches were used on winter barley varieties from European germplasm (NUE-CROPS and AGOUEB). The genetic diversity in these panels is likely to represent the majority of that that can be encountered in UK two-row feed winter barley breeding programmes. This should enhance the transfer of the study results into breeding for improved varieties for the UK environment.

Meta-analysis of QTL aims at capturing consensus QTL from independent mapping studies so that QTL are confirmed and identified with better resolution. In wheat, 30 mapping studies were used to find meta-QTL for fusarium head blight (Loffler et al., 2009). Meta-QTL analysis together with comparative genomics was also used to enhance the mapping resolution in rice and identify the most likely candidate gene models involved in the control of yield (Swamy et al., 2011). Because the only three mapping experiments carried out in this project used different mapping approaches and germplasm, a conventional QTL meta-analysis was not appropriate to efficiently integrate the datasets. Nevertheless, the projection of QTL along the chromosomes using common genetic markers for anchorage provided a mean to identify genetic

factors and QTL clusters involved in the control of agronomic traits. Most QTL were found to form clusters along chromosomes both within and between mapping studies, with better correspondence between the GWA studies. As expected, the traits within QTL cluster mirrored to some extent the correlation between traits and derived traits such as grain yield, GNYld, HI and NUE but also yield components TGW and tillering (Appendix 3. 4). Although the alleles of the parents from the bi-parental population were represented in both GWAS panels only few correspondences mapping studies were found (e.g. TGW on 2HS and mildew on 5HS). It shows that bi-parental mapping is very useful in revealing specific effects of smaller magnitude that may be associated with marker alleles at a lower frequency in a GWAS panel.

One considerable QTL cluster was identified from the GWAS near the 2H centromere and also overlaps with the wide support interval of S×R QTL for grains per ear (GE.3_1) where Saffron and Retriever are monomorphic for the SNP implying that it was a phantom effect. This chromosome region was previously highlighted as containing the candidate gene HvCEN associated with adaptation to environmental conditions in the spring and winter barley germplasm (Comadran et al., 2012) and to affect the fate of floral meristems (Cremer et al., 2001) (see 3.4.2). With numerous traits included in the cluster, it is highly likely that alleles relevant for breeding are present at that locus and control variation in winter barley agronomic performance, including nitrogen traits and yield components. However it remains unclear to what extent correlations between traits and pleiotropic effects are responsible for the co-mapping of the QTL at that locus. A better understanding of winter alleles at that locus is necessary if one wants to use MAS is to be applied, in the selection for grain number, heading date and height. Other correspondences of QTL cluster between AM panels were found on chromosomes 4H, 5H and 6H highlighting these genomic segments to be considered for selection. It is indicative of partial QTL validation and shows that using both mapping methods could narrow down the QTL position and number of underlying candidate genes.

When investigating candidate genes, it needs to be born in mind that the interpretation from QTL comparison remains partly subjective. In most cases, the QTL clusters have been defined based on the overlap of large QTL support intervals. This reflects the variation in the phenotypic data between traits and also the statistical error associated to the phenotypes. The comparison between GWAS is also affected by the allele frequencies of the SNP in each of the panels. These differences can be manifested by different peak markers within the support interval of a cluster of QTL and set of SNP to

capture the underlying haplotype patterns. In addition, the NUE-CROPS study confirms that complete QTL interpretation within a cluster can only be drawn when the phenotyping of multiple traits is included to allow independently assessed traits with common genetic control to be co-mapped. It is common that when QTL loci are suggested for MAS very little is mentioned on their effects on other apparently related or unrelated traits. The results presented here show that for most of situations of co-mapping of tillering and TGW, the decreasing effect for tillering is associated with an increasing effect for TGW (e.g. NUE CROPs QTL TGW_8 and till.Yld_3 on 5HS). This underlines at the genetic level the well-established negative genetic correlation of these two traits. Only rare loci where no significant association of these critical yield components were found suggest an independent genetic control between those traits (e.g. NUE CROPs QTL TGW_5 on 2H). These hidden associations that are not always captured by studies may explain the reasons behind a delay in MAS implementation and the complexity of conventional selection as well.

This project shows that the integration of QTL from few mapping studies is a valid alternative approach to meta-QTL analysis and enables the identification of conserved genetic factors for traits of interests. The association of both GWAS and bi-parental approaches provides confidence in the validity of an effect and a greater resolution on the QTL. The knowledge of several contrasting genotypes agronomic performance and morphological traits as well as an understanding of interconnected phenotypes can help in the biological interpretation for the putative function of candidate genes. These complementary approaches can reduce the number and size of chromosome segments to consider for functional polymorphism of candidate genes.

4.4.2 From QTL clusters to candidate genes

The QTL clusters were used to bracket specific chromosome segments worthy of further study for candidate genes. The different peak markers found for the different QTL in the same cluster reflect the phenotypic variation across genotypes and suggest that it is necessary to consider simultaneously multiple SNPs and multiple traits for interpretation. Haplotypes can be defined by the specific combination of alleles on a chromosome segment, co-segregating in sub-sets of genotypes owing to common descent. Haplotypes formed by the polymorphisms across several tightly linked loci are necessary to capture multi-allelic polymorphic gene loci (Stram et al., 2012). Several bi-allelic SNP markers must be used in order to resolve all haplotypes (Stram et al., 2012). In barley haplotypes have been shown to associate with geographical adaptation and

can describe alleles that confer adaptation to environment (Comadran et al., 2012; Jones et al., 2008). SNP based haplotypes have also been used to estimate the genetic diversity for genes involved in nitrogen metabolism (Matthies et al., 2013). In this study, the SNPs contained in the 9K chip have been designed from clustered sequence fragments (contigs) corresponding to different barley genes (Close et al., 2009; Rostoks et al., 2005). This approach maximises the number of genes covered by the assay, but most SNP therefore represent only two alleles of a single gene. In this study, the similarities between patterns of associations of different traits observed over groups of SNPs were interpreted as the haplotype signatures on the effects. Indeed, the traits sharing the same genetic control at one locus would tend to show a similar pattern of association for the alleles represented by their haplotypes. Although this approach can detect multi-allelic effects it has limited resolution to disentangle the effect of two closely linked genes. For example the 2HS *Ppd-H1* locus had significant heading date effects at the SNPs from the gene whereas the TGW effects were found associated with alleles from few closely linked SNPs. Although the data were insufficient to indicate the functional polymorphism causing the TGW difference, the pattern of associations did discriminate between markers linked to the functional heading date alleles and those linked to a TGW related locus. Technically the functional polymorphism for TGW could be from an additional allele of the *Ppd-H1* or from a closely linked putative gene. In this case, supporting evidence from the bi-parental mapping confirms that the variation in TGW is not a consequence of the heading date polymorphism. Recent sequencing of the gene revealed that 40 polymorphisms were present in the sequence of *Ppd-H1* in European landraces but very few were in complete association to heading date (Jones et al., 2008). Hence other sequence polymorphism of *Ppd-H1* could be investigated across the panel for association to TGW before that gene can be excluded. Alternatively, an association analysis using haplotypes may offer additional power to the mapping experiment (Lorenz et al., 2010; Stram et al., 2012). This underlines that the ability to point at candidate genes depends on the alleles represented by the SNP tested in the array.

The comparative genomics approach was used to give an insight on candidate genes. It uses the collinearity of the genomes to anchor genetic markers from different species onto a reference genome (Feuillet and Keller, 2002). Both rice and Brachypodium offer a well conserved synteny with barley with some variations depending on particular regions of the genome (Mayer et al., 2011). Homologous regions identified in grasses can also help to broaden the genetic studies available that describe the control of a trait and help the candidate genes identification (Swamy et al., 2011). Therefore, the results

of GWAS using genic SNP makers can be ordered based on the gene order of the reference species to target putative candidate genes. Comadran et al., (2012) used rice to target homologous genes of HvCEN. In NUE-CROPS and AGOUEB GWAS, the mapping results correctly located the genes involved in major phenotype differences. *Ant-2* and *vrs1* were precisely mapped by SNP to location within 30 gene models from the actual gene (Cockram et al., 2010). Diagnostic SNPs of *Ppd-H1* identified by Turner et al, (2005) were also significant in the mapping of heading date. These results confirmed the working hypothesis that comparison of mapping experiments can better indicate the location of candidate genes. Within QTL clusters the majority of the barley genes that contained the SNPs had homology to a specific rice chromosome segment. The interval delimited by the best markers could be used to isolate in most cases less than a hundred gene models. Some homologies of barley SNP to several rice chromosomes were also noticeable suggesting micro rearrangement of the gene sequences between species in the form of duplications, deletions and inversions of gene sequences (Bennetzen and Chen, 2008; Feuillet and Keller, 2002), while others may be spurious. These types of rearrangements and breaks in colinearty were observed on 4HL which illustrated that the resolution achieved by the GWAS may be hampered by imprecision in the mapping of the barley SNPs. The synteny of rice with barley is very good but it is likely that there will be some micro-synteny rearrangements. This means some of the spreading visible on the significant SNP associations can be a reflection of imprecision in the SNP ordering compared to the barley physical gene ordering.

On 5HL, a heading date effect was detected at a peak SNP located seven gene models away from LOC_Os03g54160, a MADS-box family gene candidate for barley vernalization gene *vrn-H1* (Szucs et al., 2007). Effect of vernalization on heading date and plant adaptation are well documented in barley (Cockram et al., 2008) and make *vrn-H1* a convincing candidate gene for that cluster. It is therefore reasonable to propose that additional alleles of the *vrn-H1* gene are present in the winter barley germplasm that do not correspond to the alleles partitioning major winter/spring difference but that are captured in this study in the form of haplotypes from closely mapped SNP of the 9K array. The patterns of associations also suggest that tillering effects found in that cluster correspond to a different haplotype to the one significant for heading date effects, a similar situation to the cluster with *Ppd-H1* and TGW on 2HS. As in the latter, an independent control of both traits is plausible. Therefore, excluding genic *Ppd-H1* SNPs, other barley gene models homologous to the syntenic rice loci could be proposed as candidates for the control of TGW. These include a sex determination protein tassel

seed-2 (LOC_Os07g49120) which modifies inflorescence structure in maize (Irish, 1997) is located five genes models from A20394. The involvement of homologous genes controlling the same trait in maize and barley was shown in a study on two-/six-row barleys, where polymorphisms in the gene *int-c*, an ortholog of the plant architecture altering gene Teosinte Branched 1, affected barley lateral floret fertility and seed size (Ramsay et al., 2011). In addition to genes involved in the control of plant structure, genes controlling the metabolic pathways of starch and sugar should have primary consideration because of their potential influence on starch accumulation in grains and TGW (Tang et al., 2009). Therefore an alternative potential candidate gene could be the glycosyltransferase family 43 protein homologous to the linked rice gene (LOC_Os07g49370). Similar proteins influenced the rate of starch synthesis in Sorghum (Singh et al., 1981) while the proteins involved in the sugar metabolism can affect variation in TGW (Ishimaru et al., 2013).

The barley and rice synteny at the 4HL tillering QTL cluster revealed that the three most significant barley SNPs bracket a homologous segment on rice chromosome 3 showing a break of collinearity with barley genome sequence. The regions of 54 rice gene models between LOC_Os03g09020 and LOC_Os03g05430 could be analysed for putative candidates. A20732 was found to be the SNP most significant across traits and is homologous to LOC_Os03g04960 coding for a putative cysteinyl-tRNA synthetase. Comadran et al., (2011b) identified a QTL for grains per spike at A20732 and a significant effect for grain Nitrogen was found nearby (Pasam et al., 2012) while neither reported effects on tillering. Nearby, I129218 is homologous to a putative expressed ras-related protein (LOC_Os03g05280). When transformed in tobacco plants, a ras-related gene *rpg1* isolated from rice produced notable phenotype changes of reduction in apical dominance and increased tillering (Kamada et al., 1992). This protein was suggested to favour the signal transmission pathways and cell growth (Kamada et al., 1992; McCormick, 1995) and is a realistic candidate gene that could explain changes in meristem growth leading to a variation of tiller number. However there are potentially other genes comprised in the homologous segment that need to be investigated as putative gene, some of them transcription factors containing functional domains like PHD-finger (LOC_Os03g04980), DENN (LOC_Os03g05200), RING finger (LOC_Os03g05270). In addition the micro-rearrangements of gene order between species should be considered as it may suggest additional candidates by taking into account the association at I168399 (LOC_Os03g09150) (Feuillet and Keller, 2002).

On 5HS, the tight correspondence of a mildew susceptibility QTL S×R Mil.1_4 QTL (Chapter 2) with the strong effects in AGOUEB GWAS suggested that high resolution could be achieved for candidate genes. This QTL locus corresponded to QTL observed in other barley mapping studies (Aghnoum et al., 2009; Comadran et al., 2009; Hickey et al., 2012). HvWIR1 coding for a glycine- and proline-rich protein of unknown function was proposed as a candidate gene responsible for the resistance (Douchkov et al., 2010), in a similar region of *Rph20* (Hickey et al., 2012). The results of the present study pointed to an homologous rice segment with 40 gene model on chromosome 12 (Appendix 4. 7), amongst which there is a complex of 7 genes coding for thaumatin proteins (LOC_Os12g43410) and an actin-depolymerizing factor (LOC_Os12g43340). The thaumatin-like genes have been described as 'pathogenesis related proteins' in multiple species. In barley, a thaumatin-like protein Hv1 was characterised after inoculation of an incompatible race of mildew (Bryngelsson and Green, 1989). The transformation of a wheat plant with a thaumatin gene increased significantly the resistance to mildew by delaying the development of the disease (Xing et al., 2008). Similar effects affecting mildew development were observed in grapes as the thaumatin gene VVTL1 was found to correlate with the inability of the mildew pathogen to initiate further infections (Tattersall et al., 1997). Given these associations between mildew resistance and thaumatin proteins, the putative complex of thaumatin genes on 5HS should be considered as convincing candidate genes for the resistance without excluding other candidates as evolution of resistance genes in populations is very dynamic (Meyers et al., 2005). Resistance to pathogens, especially mildew, can be partly dependent on cell cytoskeletal rearrangements that enable the plant to interfere with the fungal development (Miklis et al., 2007). Resistance pathways involving actin have been shown to be controlled by the MLO protein (Miklis et al., 2007). It is conceivable that a sequence polymorphism in the actin depolymerizing factor (LOC_Os12g43380) could modify the standard cytoskeleton rearrangements and affect plant response to mildew infection. It is worth noting however that the rapid evolution of specific disease resistance genes means that the comparison with rice may not be the optimal strategy for this trait.

Candidate genes could be proposed for other QTL. The Grains_2 QTL on 6H was located in a rice homologous segment containing a growth regulator factor protein (LOC_Os02g53690). The gene models between LOC_Os06g05860 and LOC_Os06g06130 (glutamate receptor) can also be considered for the yield and

nitrogen derived traits QTL due to the number of significant positive effects associated with allele C of I138457 (LOC_Os06g06040). The results of comparative genomics highlight the potential to actively select desired alleles at those loci using SNP markers. However, despite narrowing the number of candidate genes based on homology, further research will be necessary in order to fully characterise the functional polymorphism. Increases in marker density may help to capture additional alleles of the gene candidates and thus refining the potential candidates but the diversity of the panels is probably the most limiting factor to further identify of variation in gene of quantitative traits.

4.4.3 Challenges for yield component marker assisted breeding

The successful implementation of MAS and genomic resources toward increasing crop performance in yield and yield components is the main objective for competitive breeding programmes. It seems however that the wider use of MAS anticipated a few years ago still remains in its early stages (Collard and Mackill, 2008). In barley, QTL and genes directly involved in major morphological differences inducing considerable yield effects have been described (Marquez-Cedillo et al., 2001; Ramsay et al., 2011). However within a crop type the relevance of such loci for concrete yield improvement is limited because this type of genetic variation is continuously under selection. In a specific crop type, the application of MAS for trait such as yield and yield components will benefit from a better understanding of the targeted polymorphism and the possible effects associated to alleles linked to the targets. In this study, a wide range of traits was investigated to reinforce the confidence in the putative genetic factors and facilitate biological interpretation. From the GWAS studies, the TQL clustering across the genome identified 26 genetic factors involved with single or multiple yield related traits. A number of yield QTL did not cluster with QTL for yield components suggesting that additional genetic control of yield is present and may be attributable to other traits such as resource use efficiency and biomass produced.

The selection for yield also remains complex at the genetic marker level despite the high resolution of the GWAS. The patterns of marker-trait associations can help the identification of marker haplotypes and multiple alleles involved in the control of a set of traits and that can reflect independent and different genetic control of traits mapped in a QTL cluster. This was illustrated by the major QTL target for grains per ear (Grains_1) on 2H located at a junction between rice chromosome 4 and 7 (Mayer et al., 2011) which partitioned the set of candidate genes into two groups of traits. On one hand a group containing Grains, Hd, Ht, StemN QTLs seemed to be capturing the same

allele associated genes homologous with rice chromosome 4 while yield derived traits (Yld, HI, NUE, NutEg) matched another haplotype. On the other hand, TGW effects also significantly associated to the cluster suggested the presence of a functional polymorphism in a gene nearby LOC_Os07g42924 and LOC_Os07g43040 on the homologous segment of rice chromosome 7 (Appendix 4. 4). LOC_Os07g42970 is a UDP-glucoronosyl and UDP-glucosyl transferase domain containing protein. UDP-glucosyl transferase, also referred to as sucrose synthase, catalyses the conversion of ADP and sucrose into ADP-glucose necessary for starch production (Baroja-Fernández et al., 2003). In rice, the variation in activity of proteins involved in the sugar metabolism was shown to affect starch accumulation and the rate at which the grain is filled (Ishimaru et al., 2013; Tang et al., 2009). These results from studies on rice have not been confirmed in barley but it seems reasonable to assume that a gene with similar roles could cause differential grain filling in barley too. Based on the convergence of mapping results, the TGW increasing allele G of marker B10733 is located five rice gene models distant from LOC_Os07g42970 could be a potential target for MAS. Therefore it will be necessary to consider multiple SNP haplotypes to enable the characterisation of the allele or haplotype linked to the desired polymorphism until diagnostic markers can be identified. The haplotype analysis in association mapping can be used to distinguish relevant associations (Lorenz et al., 2010; Stram et al., 2012) although this approach was not investigated in this study.

The clustering of QTL of correlated traits was also illustrated by the direction of the effects for the significant markers. The significant cluster on 5HS showed that alleles G or A of A20553 associated with an increase of tillering or TGW (TGW_8) respectively. Therefore the choice of the adequate allele for MAS will depend on additional information on the variety characteristics, especially the identification by the breeder of the traits that needs to be improved so that the increasing effects on one trait will need to consider the potential pleiotropic effects on correlated traits. The direction and intensity of allele effects at a marker can vary depending on the trait considered and also on non-allelic interactions. In breeding programmes, these loci carrying opposite allele effects in interaction with the environment and genetic background experience a push/pull selection pressure that reflects the favourable polymorphisms in a given season and environment. Therefore the better characterisation of breeding material for the alleles at those loci may be useful for optimising selected lines toward targeted environments. The GWAS on individual sites in the NUE-CROPS panel could provide

further evidence of the variability in QTL effects and highlight the stability of particular QTL cluster.

The genetic complexity of yield components renders the quest for yield candidate genes and diagnostic markers even more challenging. The comparative genomic approach with rice is a primary attempt at investigating potential candidate genes underlying barley QTL clusters. Despite the study being able to identify rice segments with candidate homologous gene models, the broad diversity of gene models and families that could be involved in the genetic control of yield and yield components should be considered for candidate gene discovery. The most obvious are genes involved in critical adaptive traits influencing plant fitness to the environment and architecture such as major genes involved in the control of plant phenology and height as well as genes involved in the control of underpinning grain filling (e.g. starch metabolism) (Comadran et al., 2011b; Tang et al., 2009). However a large range of other physiological traits with individual genetic control participate in yield variation and therefore phenotyping methods for those traits have been used alongside QTL mapping to study and improve yield (Reynolds et al., 2011). For traditional breeding, the comparison of mapping studies and the SNP and haplotypes associated to the targets may be sufficient to acquire confidence on the presence of a genetic effect to select for by MAS. However the validation of the allele effects is essential to fully understand potential pleiotropic and antagonist effects associated to a polymorphism before enriching or maintaining favourable alleles in breeding programmes.

Chapter 5
Validation of QTL using a Near Isogenic Lines approach

5.1 Introduction

The large number QTL described in cereal species would suggest that breeders have a concrete opportunity to develop material with improved traits by using efficiently genomic resources. However the actual implementation in breeding programmes of the results of genetic research has been little reported suggesting that crop improvement has not yet benefited from the genomics revolution to the expectations described in Bernardo (2008). One reason could be the lack of QTL validation studies reported in the literature which are necessary to give more reliability and accuracy of the QTL in order to make a decisive contribution in crop improvement (Asíns, 2002). Often disregarded, the QTL validation step is at the interface between QTL discovery and the MAS implementation (Collard and Mackill, 2008; Romagosa et al., 1999). While QTL validation confers additional support for allele effects and their associated genetic markers, it is also describes and assigns the breeding value of the QTL and its associated allele, or haplotype, in the context of a larger germplasm. Few validation studies in barley QTL have been reported and these have had only moderate success in validating traits. Cockram et al., (2010) used a complementation of GWAS and DH population mapping to validate a QTL for a DUS trait controlled by a single gene. The validation of QTL involved in quantitative traits such as grain yield, height, disease resistances have been reported (Navara and Smith, 2013; Spaner et al., 1999; Yun et al., 2006). The lack of validation experiments for agronomic traits such as yield and yield components illustrates the missing link between the sustained description of QTL information and the real use of QTL and MAS in breeding.

Single gene locus that are involved in quantitative traits are generally of relatively small magnitude. They can be difficult to resolve when traits have low heritability. Any efforts aiming at achieving accurate phenotyping through the careful design and control of environmental variation is therefore essential to better isolate target the genetic origin of the effect. Similarly, the control of background genetic variation segregating alongside a target QTL effect help at minimising undesired phenotypic variation in comparisons of lines. Therefore the validation of a genetic effect at a QTL or gene is

optimized by the reduction of all sources of additional variation whether these are genetic or environmental. Near Isogenic Lines (NILs) are individual genotypes that differ at a specific target locus and otherwise share an identical genetic background. When compared in a common environment, NILs are a powerful research tool to validate both the position and the magnitude of the QTL effect. QTL validation studies using comparisons between NILs can consist of the introgression of specific alleles or haplotypes into a recurrent genetic background using a backcross scheme (Kandemir et al., 2000; Kongprakhon et al., 2009) or by using advanced backcross strategy on material directly linked to the QTL mapping study (Yun et al., 2006). These validation studies often utilise material related to that used in the mapping studies which limits the estimation of the consistency of the effect over a larger range of genetic background often encountered in breeding programmes. Heterogeneous Inbred Family (HIF) has been presented as an alternative approach for developing NIL validation material (Tuinstra et al., 1997). With HIF, the NILs are developed from a founder line heterozygous at a locus of interest (e.g. at a QTL) that is used to fix alleles in two opposite phases in sister lines produced in subsequent generations of self-fertilisation (Tuinstra et al., 1997) (Figure 5.1). HIF from multiple crosses can be generated to test allele effects associated with QTL in mapping studies. Hence it is possible to simultaneously utilise the material continuously generated in breeding programmes and test for allele effects in a relevant background (Pumphrey et al., 2007).

In Chapter 2, a bi-parental mapping population resulted from a cross of elite winter barley varieties Saffron and Retriever was used to map QTL for yield and yield components. Within that cross, a promising QTL on chromosome 2HS had sizeable effects on both TGW and tillering. The QTL for other traits and the commercial success of the parents reinforced the scientific value of the population as it segregates for alleles that have been maintained in superior varieties. Consequently, it was recognised that the validation of some of the agronomic QTL found in that population would benefit winter barley breeding progress. Furthermore, a QTL validation experiment offered the possibility to complement the results of GWA mapping conducted simultaneously (Chapter 3). This association of mapping and validation experiments can be used to increase the resolution of some of the QTL targets (Navara and Smith, 2013) while the winter barley varieties used for validation can directly contribute as a source of beneficial alleles and haplotypes for the commercial breeding of elite lines.

This chapter reports a QTL validation experiment for that aims at identifying significant effects for agronomic traits attractive for the breeding industry and to implement in a MAS strategy. The validation project exploits QTL mapping results obtained in Chapter 2 and late generation breeding material developed and genotyped at KWS UK. The principal objectives are to develop HIF of NILs using late generation breeding material identified for segregating at the relevant QTL target defined in Chapter 2 and to validate the effects of specific haplotypes and alleles of SNP markers on agronomic traits in a yield trial experiment.

5.2 Material and methods

5.2.1 Plant material

The NIL development exploits existing plant material and genotypic resources available at KWS-UK winter barley breeding programme. The procedure consisted of identifying HIF founder lines in segregating breeding material, heterozygous at relevant loci in order to develop NILs with opposite alleles and maximised homozygous genetic background (Figure 5.1). The several residual segregating regions found in HIF founders can also be exploited to create additional NIL pair comparisons and test for effects at more than a single locus. Two batches of plant material with corresponding genotypic data were screened for HIF founder lines.

NILs batch-A

Three validation populations of F_4 Recombinant Inbred Lines (RILs) from crosses 06-03×B88 (147 lines), Saffron×Retriever (49 lines) and (B78×Retriever)×B88 (181 lines) were made available with genotypic information from the custom subset of 384 SNPs using the Illumina BeadXpress platform (Bx384). The choice of target genetic factors and source genotypes for HIF founder were confined to those identified in Chapter 2. In autumn 2009, a total of nine F_4 HIF founder lines were chosen for their residual heterozygosity at relevant loci while having minimal background heterozygosity. Between 12 and 24 $F_{4:5}$ plants (F_5 plants derived from a single F_4 plant) for each F_4 founder were vernalized for seven weeks before being transplanted in the field as spaced plants to provide seeds in summer 2010. Each F_5 was genotyped with Bx384 and bagged to prevent cross pollination. At that stage the aim was to select pairs of individual plants that were homozygous for opposite alleles at the target loci. However very poor correspondence of the F_4 founder line genotype with the expected subsequent

$F_{4:5}$ genotypes. This was later found to be caused by inaccuracies in the labelling of the HIF founder. Nevertheless, it was decided to continue with these $F_{4:5}$ progenies of unknown pedigree and identify the segregating loci at alternative positions within the HIF and their haplotypes. These alternative targets were associated with QTL of lower initial interest or corresponded to haplotypes of unknown pedigree. The selected $F_{4:5}$ plants were harvested and oven dried at 40C for 48 hours in August 2010. The $F_{4:6}$ NILs were sown as 1, 2, 6 or 12 ear rows in untreated field conditions for seed multiplication and harvested in August 2011. For most lines, residual background heterozygosity was present at more than one locus. Depending on the pertinence of the trait and effects associated with the novel candidate regions, additional $F_{4:6}$ plants from a subset of $F_{4:5}$ were genotyped and screened to select for reduced background heterozygosity. The genotyped $F_{4:6}$ plants were bagged and harvested as individual plants or ears while the $F_{4:6}$ with no additional genotyping were harvested as row bulk. The $F_{4:7}$ plants were grown in a replicated mini-plot experiment in 2011-12 (see 5.3.2). Two ear-rows of each of the lines were harvested to provide a bulk of $F_{4:8}$ seeds for the 2013 replicated yield trial experiment (see 5.2.3).

NILs batch-B

In 2010, a set of F_5 RILs from two crosses (Saffron×B78)×Retriever (28 lines) and Saffron×Retriever (15 lines) in pedigree breeding scheme was genotyped using the Bx384 array. The F_5 HIF founder lines that contained heterozygous haplotype segments that co-localised with targeted genetic factors were selected for further self-pollination (Figure 5.1 stage C). From each F_5, three ears were bagged to prevent cross pollination and harvested in summer 2010. Only five F_5 lines (4187, 4042, 4188, 4190, 4045) contained heterozygous segments of chromosomes in the regions of the QTL of interest together with a reasonably homozygous genetic background. From the bagged ears of this subset, 12 or 24 $F_{5:6}$ seedlings were genotyped and NILs that carried the contrasting homozygous haplotypes within the HIF were kept. A preliminary seed multiplication of selected $F_{5:6}$ NIL plants was carried in pot sin the glasshouse during winter 2010-11. For each NIL, 22 $F_{5:7}$ seeds were sown, vernalized for seven weeks and transplanted as spaced plants into the field in spring 2011. A further marker assisted background selection was carried out on a subset of the $F_{5:7}$ plants issued from $F_{5:6}$ plants homozygous at the QTL target but highly heterozygous in their background. $F_{5:7}$ plants were harvested in autumn 2011 either individually if genotyped or as a bulk. The

resultant $F_{5:8}$ NILs were included in the replicated mini-plot experiment with batch-A in 2011-12 (see 5.2.3).

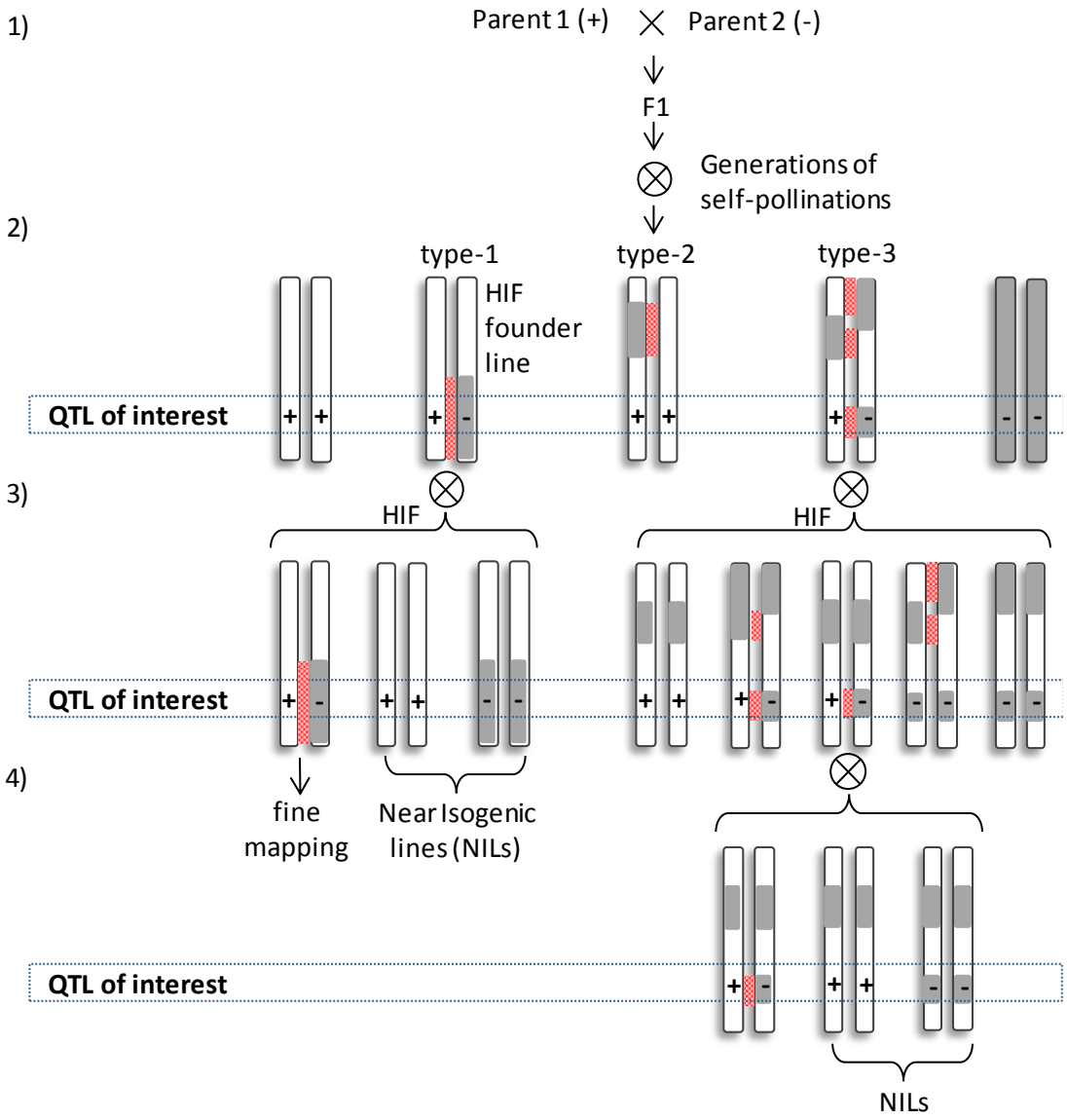


Figure 5.1 Development scheme of Heterogeneous Inbred Families (HIF) from breeding material recombinant inbred lines.
HIF and NIL development scheme is adapted from Tuinstra et al., (1997). 1) A cross segregating for alleles associated with a QTL undergoes early generations of self-pollinations until generations F_4 and F_5. 2) Heterogeneous lines or HIF founders are identified with heterozygous chromosome segments (in red) at the QTL-locus of interest in a panel F_4 and F_5 plants (type-1 and type-3 lines) while lines homozygous for these regions are discarded (type-2 lines). 3) Progenies of F_4 and F_5 derived plants are screened to identify homozygous lines for both alleles at the target to form a HIF. The type-3 lines with excess of residual background heterozygosity undergo an additional cycle of self-pollination. 4) Marker assisted screening in subsequent generations is used to identify novel NIL with increased background homogeneity.

5.2.2 NILs Genotyping

Plant DNA was extracted from leaf material sampled from seedling or adult green plants. The genotypes of the NILs and HIF founder lines were obtained using custom subset of 384 SNP using the Illumina BeadXpress platform (Bx384) available at KWS-UK. The SNP markers were ordered according to distances of the map used with the Illumina iSelect 9K chip (Comadran et al., 2012) used in Chapter 3 and Chapter 4 (Supplementary data 1). The 384 SNP array allowed characterisation of haplotypes at the loci of interest and of the genetic background of NILs. Some genotype verification was done on a subset of the material in 2012 using KASP markers (LGC Genomics, UK) designed to some of the SNPs on the Bx384 available at KWS-UK. In 2013, the genotypes of NILs tested in yield trial experiment were verified with the Bx384 chip.

5.2.3 Experimental trials and phenotyping

Season 2012

In season 2011-12, a preliminary collection of phenotypes was assessed using miniplots of NILs from 14 HIF (9 and 5 for batches A and B respectively). 4 to 13 NILs per HIF were grown in replicated miniplots blocked by families to minimize environmental variation between the NILs (Appendix 5. 1). The mini-plots were formed of 6 rows of 1 metre length and were grown under untreated conditions. Date of heading (GS59) (Hd) was recorded as number of days from sowing. At maturity, five plants were sampled from two inside rows of each mini-plot and these were further handled as a grab sample bulk for phenotyping. The number of fertile tillers per plant (Till_GS) and grains per ear (Grains) were counted. The thousand grain weight (TGW) was computed from the total grains of each grab sample dried at 40°C for 48h. A subsample of grains was milled using a 0.8 mm sieve and the flour analysed by NIR to measure grain proteins (GP) and grain sugars (GS) following the procedure and calibrations described in 2.2.2.

Season 2013

In 2012-13, the phenotype data collected in 2012 and genotypes were used to identify a subset of NILs in the seven most promising HIF to test further testing in a replicated yield trial experiment. The yield trial was drilled on the 28th October 2012. Plot size was 4x1.6 metres and managed following the KWS standard input program for fertilisation (180 kgN/ha in three applications), plant growth regulators and fungicide. The 96 plots yield trial was formed of two replicates of randomised block structures characterising

each HIF. To further reduce environmental error, each of the NILs tested was replicated and randomised within its HIF block replicate (Appendix 5. 2).

Date of heading (Hd) and plant height (Ht) were recorded during the growing season. At maturity, the grab samples were collected from entire plants were pulled from a 30 cm length of the 4th row at one meter within the plot to avoid edge effects. The number of ears and grains was recorded from each sample and grains per ear calculated (Grains). The samples of grains were dried at 40°C for 48 hours, weighed and TGW computed Each plot was combined and the yield figures recorded at 15% moisture. The phenotype score of tillering (Till_yld) was derived from yield data and yield components TGW and Grains. A subsample of grains was milled using a 0.8 mm sieve and the flour analysed by NIR to measure grain proteins(GP) and grain sugars (GS) (see 3.2.1 Phenotyping).

5.2.4 Statistical analysis

The phenotypes collected in 2012 and 2013 were analysed with the ANOVA procedure in Genstat 14th Edition (Payne *et al.*, 2009). In both years, a two-step approach was carried out to identify significant differences in each of the phenotype measured first between HIF and then the NILs within HIF using appropriate error variance components.

The first step consisted in obtaining a residual error for the whole trial while testing for significant differences between HIF and NILs within HIF. The phenotype means for each NIL could be obtained from the following model:

(year 2012) $y_{i1} = M_1 + F_{j1} + L_{ij1} + e_1$

(year 2013) $y_{i1} = M_1 + B_1 + F_{j1} + L_{ij1} + B_{bij1} + e_1$

Where y_{i1} is the mean of a NIL $_i$ resulting from the constant M_1, F_j the effect of family $_j$, L_{ij} the effect of line $_i$ in family $_j$, B_1 the block effect, B_{bij} the effect of block $_b$ on line $_i$ in family $_j$ and e_1 the residual error.

In order to identify subtle differences between NILs grown in close proximity, each HIF was analysed individually by ANOVA to estimate the NIL effect L_{ij}, For each experimental year, the mean y_{i2} of NIL $_i$ was obtained from a constant M_2, B_2 the block effect, L_i the NIL effect, L_{bi2} the effect of block $_b$ on NIL $_i$ and e_{2j} the residual error:

(year 2012) $y_{ij2} = M_2 + L_{i2} + e_2$

(year 2013) $y_{ij2} = M_2 + B_2 + L_{i2} + L_{bi2} + e_2$

To test the significance of differences between the NILs within their families (L_{i2}), the variance ratio was computed using the overall trial error e_1 and appropriate degrees of freedom and the F probability calculated. The significant differences between NILs

were considered with a lower stringency at a threshold of $p<0.1$. Although it may cause type 1 errors, this enabled retention of most of the possible real effects that would be tested further.

The HIF with significant differences between NILs in 2013 were then tested for significant association of phenotypes with the segregating haplotypes for the relevant trait. Within each HIF, the segregating chromosome segments for different haplotypes were identified. The SNPs with identical allele distribution within the HIF and mapped at a similar position were described as components to be used as a factor in further statistical analysis. For each HIF, a reference line was arbitrarily identified so that the alleles at each component could be associated to factor levels of -1, 0 and 1 corresponding to alleles similar to the reference line, heterozygote or homozygous for opposite allele respectively. ANOVA was used to test for significance of the within family components (haplotypes) with the traits variation using the error variance calculated from the whole trial under the null hypothesis of no association between the factor level (i.e. alleles of a haplotype) and the NILs' phenotypes. The optimal model and allelic effects were obtained with the REML procedure by fitting all possible haplotype factors of a HIF in a maximal model as fixed effects and dropping alternatively individual terms to exclude non-significant and redundant factors. This optimisation was limited by the experimental design which lacked orthogonality between some of the haplotype factors and was over parameterised (more contrasting factors than experimental units). Only the factors accounting for a significant proportion of the variance between lines were kept in the optimal model to validate the genetic effect of the loci.

5.3 Results

5.3.1 Development scheme for NIL pairs.

The NIL development followed the scheme presented in Figure 5.1. It was primarily intended to test the effects of known alleles found to be segregating in breeding material at positions of QTL and genetic factors identified in the S×R DH population (Table 5.1b). Therefore, the segregating breeding material was screened for potential HIF founders by genotyping and selecting for its relationship to the parents of DH population. For example, the cross (B78×Retriever)×B88 was considered for testing the Retriever allele effects in a genetic background different to Saffron. The study exploits

the SNP correspondence between genotyping platforms to associate BOPA1 markers at genetic factors in the S×R DH population with their equivalent on the Bx384 NILs genotypes. Despite the reduction in the number of SNP between the platforms, 184 SNPs were polymorphic across the breeding material screened.

The screening of genotypes (Figure 5.1 step 2) enabled the identification of nine and five HIF founder lines in batches A and B respectively with heterozygous chromosome segments at relevant locations of genetic factors or QTL (Table 5.1a). The genotypes observed for the lines involved in NIL creation (including the HIF founders and subsequent lines) can be found for chromosome 2H in Figure 5.2, and for other chromosomes the details are given in Appendix 5. 3.

The haplotypes segregating in founder lines of batch A matched their expected pedigree while the B-4 and B-5 HIF founders showed haplotypes different to their expected Saffron × Retriever pedigree. Each of the founder lines carried between one to four heterozygous segments scattered along the genome. HIF founders A-1, A-6 and A-9 had a heterozygous segment associated with a single known target and very little residual heterozygosity and this material followed the type-1 NIL scheme (Figure 5.1 step 2). The founder line A-1 was heterozygous for five co-segregating SNPs associated with genetic factor 2 (Figure 5.2) also associated with the S×R QTL for tillering and TGW on 2HS (Table 5.1) while the remaining chromosome length has Saffron alleles.

Founder lines A-4 and B-5 were heterozygous at multiple loci and chromosomes and the use of this material for NIL development was more complicated than founder A-1 and involved an additional breeding generation (Figure 5.1 step 2, type-3 lines). In 2010, the spaced plant experiment from seeds issued from self-pollination of the initial lines (Figure 5.1 step 3) showed discrepancies between the segregating haplotypes and the expectations from the haplotypes of initial founders in batch-A. This highlighted an error in seed identification for these families and genotyping discrepancies were also observed in some families such as A-3 for which the monomorphic segment on 2HS (SNPs A10525 to A11302) was found segregating subsequent generations (Figure 5.2). However, the unexpected A-3 HIF was found to segregate for markers associated to other genetic factors on chromosomes 1H, 3H 4H and 7H (Appendix 5. 3; Figure 5.3; Table 5.1a). Despite these unfortunate early results, the NILs development was continued. The unexpected segregating segments were analysed with regards to the positions of alternative genetic factors and the new targets identified for each new founder line planted in 2011 (Table 5.1a). The founder of HIF A2 was conserved to develop NILs targeting bin 13 while A-5 offered the potential insight into effects of bin

6 although substantial residual segregation was observed for other different genetic factors. The HIF from unexpected founder lines were kept despite unresolved pedigree and segregation at loci associated with for QTL targets of lower interest. Fortunately, the lines of batch B grown in 2012 (at stage 3 Figure 5.1) had a genotype matching the founder line (Figure 5.2).

In 2011, each HIF of batch-A consisted of five to nine NILs with different allelic status at the segregating regions. The additional cycle of self-fertilisation benefited family A-8 to fix a background segregating segment on 5H (stage 4 Figure 5.1) and enabled the creation of alternative segregating haplotypes at the interesting region of 2HS for family B-4. In B-4, the G and A alleles of SNP A10287 (*vrs1* locus) were associated respectively with the presence of sterile spikelets in lines 4187F3/9 and 4187H3/4 and their absence in lines 4187F3/8 and 4187C3/1 (Figure 5.3).

In 2012, all the HIF grown consisted of only the optimal NILs available that carried homozygous opposite alleles at the target loci and had minimal background segregation. The genotypic and phenotypic data collected on lines grown in 2012 was used to identify two to four lines for each of seven HIF that were then tested in replicated yield plots experiment in 2013 (Figure 5.2). The residual segregation in other families such as A-9 remained too large for further field testing (Appendix 5. 3). Four NILs of family A-2 and B-4 were selected for their haplotypes at the genetic factor 2 for TGW-tillering (2HS). Due to the additional loci that are segregating in the background of the NILs, the combination of multiple NILs within a HIF may help in resolving cases of situations of multiple testing. In addition to A-2, the NILs development experiment retained HIF A-3, A-5, A-8, B-2, B-3 and B-4 for testing haplotypes at different relevant genetic factors (Figure 5.1 and Figure 5.2).

Table 5.1 Details of the genetic factors and associated QTL co-located with the segregating haplotypes in the 14 founder lines of HIF tested in 2013.
The genetic factors or bin numbers correspond to results presented in Chapter 2 (Table 2.9)

a) Genetic factors co-located with the segregating haplotypes for each of the founder lines of 14 Homogenous Inbreed Families.

HIF	Bin or genetic factor		
	Founder line	New founder line	2013 yield trial
A-1	2	15 21	
A-2	18	2 13 18	2 8 13
A-3	18 10	2 10 13 18 20 21	13 21
A-4	1 5 12 20	18 19	
A-5	21 4 5	3 4 5 6 8 13 18 21	5 6 13
A-6	6	13	
A-7	10 21	13 18	
A-8	8 13	10 11	10 11
A-9	13	10 11 13 18 20 21	
B-1	2 18		
B-2	7 12		12
B-3	18 21		18
B-4	2 3		2 3
B-5	5 6 13 21		

b) Details on the QTL and traits corresponding to the genetic factor bins targeted in 2013 yield trial.

Genetic factor Bin	chr	Bin interval (cM)	Clustered SxR QTL						
2	2H	0.0 - 39.2	HLW.2_1	TGW.2_1	TGW_GS.2_1	TGW_GS.3_1	TGW-GS.1_1	Til_cal.2_1	Til_mes.2_1
3	2H	0.0 - 48.7	GE.3_1	GP.2_1	HLW.2_1	Til_cal.2_1	TGW-GS.1_1		
5	2H	52.3 - 124.1	HLW.2_2	SC.1_2	SD.1_2	SG.1_2	Til_cal.2_2	Til_mes.2_2	
6	2H	95.6 - 124.1	GP.2_2	GP.3_2	SG.1_2	Til_cal.2_2	Til_mes.2_2		
8	3H	61.1 - 90.4	HLW.2_3						
10	3H	120.7 - 141.8	GE.3_2	GE.2_1	HLW.2_4	Til_cal.2_3			
11	3H	120.7 - 172.1	HLW.2_4	TGW.2_2	TGW_GS.2_2	TGW_GS.3_3	TGW-GS.1_3	Til_cal.2_3	Yld.2_1
12	4H	0.0 - 18.8	GS.2_2	GS.3_2					
13	4H	41.2 - 64.4	BR.1_1	Mil.1_3	Aleu.1_1				
18	6H	18.9 - 63.7	SC.1_3	SD.1_3	SG.1_3				
21	7H	38.7 - 91.3	Ht.3_2	TGW_GS.3_5					

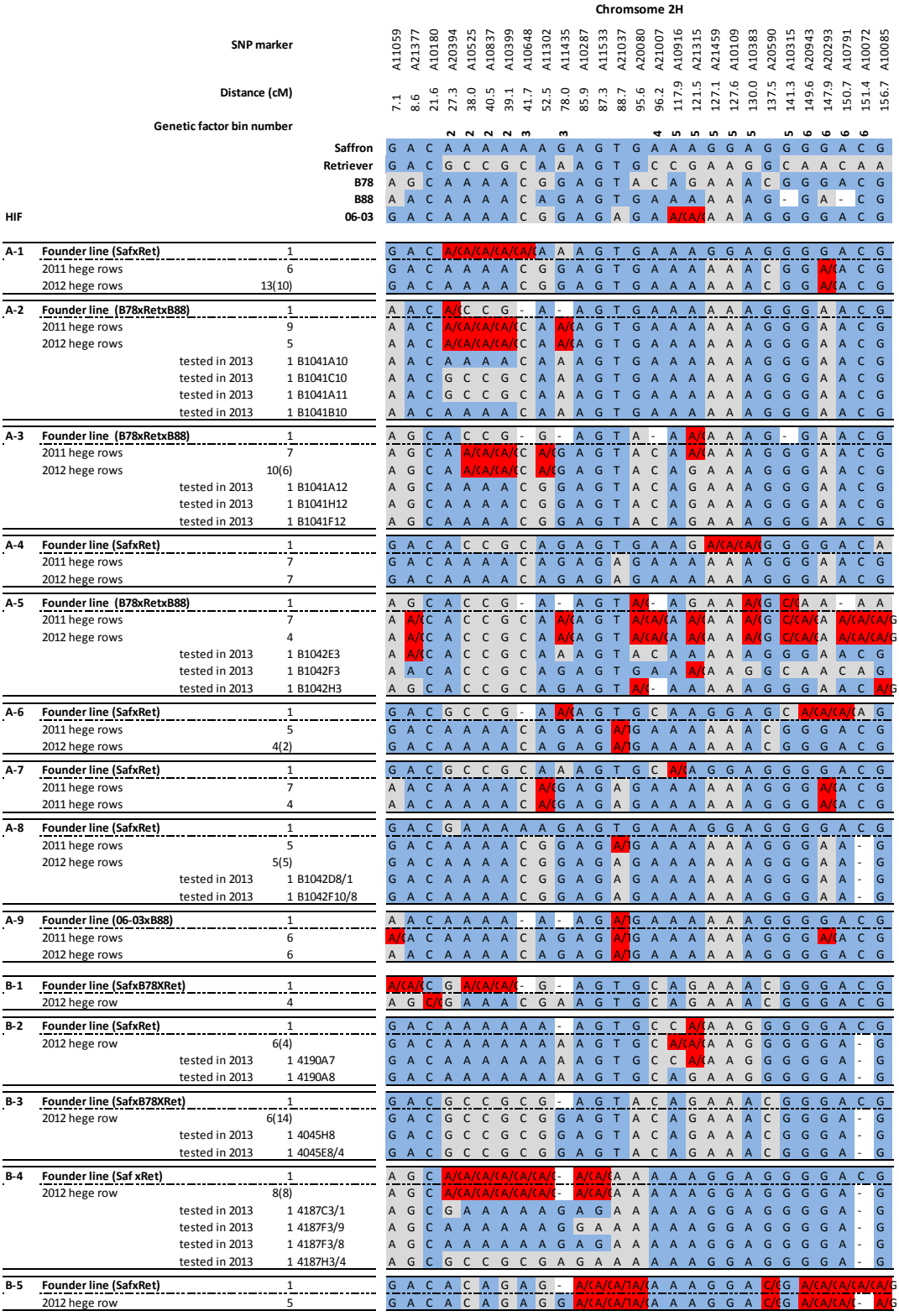


Figure 5.2 HIF founders and NILs genotypes for chromosome 2H.

SNP marker alleles on chromosome 2H for the 14 HIF including the founder lines, progenies grown in 2012 and NILs present in 2013 experiment. Only informative markers of the Bx384 chip are presented (polymorphic across the whole set of HIF). Markers have been ordered from left to right based on the OPA1 consensus genetic distance. Genotype of NILs multiplied in hege row in 2011 and 2012 are presented as a consensus haplotype (homozygous haplotypes within HIF could also be present at these same multiplication stages). Heterozygous markers are highlighted in red. The genetic factor number (bin numbers) correspond to results presented in Table 2.9 and were associated with the Bx384 SNP based on their colocation with the OPA1 SNP used for mapping in Chapter 2.

5.3.2 2012 Preliminary phenotype analysis

A preliminary analysis was done on the phenotypes for the 14 HIF multiplied in 2012. This experiment was not optimised for NIL testing as the 2012 phenotypes were obtained from mini-plots (field grown spaced plants) and therefore subject to a different macro-environment to the original full plot trials used to make the original S×R QTL map. The 2012 experiment was used as a first indication for possible differences between NILs to complement the NILs genotypes in selecting for a subset of NILs to test in 2013. These differences should in principle correspond to the traits associated to genetic factors segregating within the HIF (Table 5.1). The full trial ANOVA revealed that families and lines within families were significantly different for all traits but tillering (Table 5.2) (the genotypes corresponding to the line tested are labelled as '2012 hege rows' in Figure 5.2 and Appendix 5. 3). The within family test of the 14 HIF indicated that differences between NILs could be identified, suggesting significant effects in the segregating chromosome regions. The number of significant HIF varied for each traits and contrast with differences estimated with F pr. <0.2 are reported in Table 5.2. Significant differences for grain number per ear were found between NILs in 7 out of 14 HIF while fewer significant comparisons were found for the other yield components. The significant difference between the control varieties Saffron and Retriever was observed for all traits which can be explained by the high replication over the whole multiplication experiment. Some families such as B-4 had significant differences between NILs for multiple traits suggesting genetic factors of major effect. It needs to be born in mind that Therefore

The analysis of phenotypes was complemented with a genotype analysis to select the 2013 testing subset. The additional generations of self-fertilisation made a reasonable reduction in the background segregation for most of the families (Figure 5.2, and Appendix 5. 3). However, residual segregation at two or more loci was still present between NILs within HIF (Figure 5.3). The material screened for families A-2 and A-5 however did not generate sufficient optimal combinations of alleles to extract clear NILs comparisons and leaving instead lines within those families segregating at 10 or more loci. In the case of families A-4 and A-9 that were significantly different in the 2012 experiment (Table 5.2), their residual background segregation was too important for adequate testing in 2013. In addition, the size and position of the segregating segments of families A-6 and B-1 did not capture the targeted genetic factors (Table 5.1). The combination of phenotypic and genotypic data was used to select a total of

seven HIF to be tested in a replicated yield trial in 2013. The segregating regions of the seven HIF are presented in Figure 5.3. All HIF segregated for SNPs associated with target genetic factors (Table 5.1) or at loci where no effects were identified by the mapping experiment of Chapter 2. None of the families could be found with a unique locus segregating at a candidate position but a minimum of two (A-8, B-3), with a candidate genetic factor associated with a single locus only (Figure 5.3). The family B-2 was only significant for tillering effects in 2012 despite some expected effects in grain sugars associated with genetic factor 12 (Table 5.1). The component 1 of family B-3 contained SNPs associated with genetic factor 18 which was associated with the straw characteristics of stay green (SG), straw degradation (SD) and straw collapse (SC). Because NILs were obtained from a heterozygous chromosome segment of the founder line, different haplotypes occurred at the same segment in the subsequent generations due to recombinations. In B-4 the founder heterozygous segment between 27.3 and 87.3 cM led by self-pollination to three components regions on chromosome 2H (Figure 5.2, Figure 5.3) that can be tested for effects relating to the genetic factors 2 and 3 (Table 5.1) with expected TGW, tillering effects.

Examples of component co-located with alleles for DUS traits can be found in families A-5 and B-4 segregating for anthocyanin pigment at component 4 and sterile spikelets at component 3 respectively (Figure 5.3). The allelic effect associated with the presence of sterile spikelets associated with allele G and A at SNP A10287 and A11533 respectively (85.9-87.3cM 2H) can be estimated in an optimal NIL comparison involving the B-4 lines 4187F3/9 and 4187F3/8 which differ for component 3. This same comparison can also be used to investigate the effects of allelic differences in this region on other agronomic traits. Additionally, lines 4187C3/1 and 4187H3/4 from B-4 contain alternative alleles at other components that can better describe effects of component 2 on 2H. It was observed that in some cases, the segregating components between NILs were made of a single SNP indicating some heterozygosity in a short chromosome fragment of the founder line (e.g. A-2 component 9) or contained missing information (e.g. A-3 component 7 Figure 5.3). It suggests that the choice of genotyping platform and its genome coverage used for the NIL development is critical to describe and interpret the effects between NILs and of their genetic background. Although a single SNP cannot be informative about the actual size of the segregating region, it cannot be excluded from the NILs description. In turn, one can overcome the missing genotype information around that SNP by additional genotyping.

Table 5.2 Summary of ANOVA for the 2012 HIF.

The full trial ANOVA was carried out using the complete set of data to identify significant family.line effect and give an estimate of the trial residual error variance. The within family ANOVA utilises the residual error variance (residual m.s.) from full trial ANOVA to calculate variance ratio (v.r.) and F probability (F pr.). Only the HIF with an *F pr.*<0.2 are reported.

Traits	Full trial ANOVA					Within family ANOVA			
Heading date	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	family	14	496.27	35.45	58.45	**<.001**	A-3	9	5.13	**0.000**
	family.line	72	268.10	3.72	6.14	**<.001**	A-9	5	5.52	**0.000**
	Residual	115	69.74	0.61			B-1	3	9.78	**0.000**
	Total	201	834.10				B-4	7	11.41	**0.000**
							Cont	1	252.89	**0.000**
Height	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	family	14	3554.8	253.9	31.8	**<.001**	A-2	4	2.31	**0.062**
	family.line	72	1412.0	19.6	2.5	**<.001**	A-9	5	2.35	**0.045**
	Residual	115	918.1	8.0			B-4	7	2.99	**0.006**
	Total	201	5884.9				Cont	1	100.36	**0.000**
TGW	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	family	14	1464.87	104.63	14.34	**<.001**	A-2	4	2.12	**0.083**
	family.line	72	2139.01	29.71	4.07	**<.001**	A-4	6	2.39	**0.032**
	Residual	115	838.83	7.29			A-8	4	1.83	**0.127**
	Total	201	4442.71				A-9	5	3.31	**0.008**
							B-3	5	2.50	**0.035**
							B-4	7	1.90	**0.075**
							Cont	1	183.68	**0.000**
Grains	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	family	14	369.79	26.41	20.13	**<.001**	A-3	9	2.54	**0.011**
	family.line	72	281.47	3.91	2.98	**<.001**	A-5	3	6.40	**0.000**
	Residual	114	149.62	1.31			A-6	3	4.87	**0.003**
	Total	200	800.70				A-8	4	6.69	**0.000**
							B-3	5	3.11	**0.012**
							B-4	7	5.18	**0.000**
							B-5	4	16.60	**0.000**
							Cont	1	2.45	**0.120**
Tillering	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	family	14	250.80	17.91	2.72	**0.002**	A-4	6	1.62	**0.149**
	family.line	71	538.18	7.58	1.15	0.248	A-5	3	2.57	**0.058**
	Residual	114	749.81	6.58			A-8	4	2.17	**0.077**
	Total	199	1533.20				B-2	3	2.76	**0.046**
							B-3	5	1.51	**0.192**
							B-4	6	10.26	**0.000**
							Cont	1	6.49	**0.012**
Grain proteins	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	family	14	28.34	2.02	11.22	**<.001**	A-3	9	1.79	**0.077**
	family.line	72	18.22	0.25	1.40	**0.052**	A-4	6	2.08	**0.061**
	Residual	115	20.75	0.18			A-8	4	2.40	**0.054**
	Total	201	67.31				A-9	5	2.53	**0.032**
							Cont	1	33.46	**0.000**
Grain sugars	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	family	14	9.17	0.65	9.98	**<.001**	A-3	9	1.45	**0.176**
	family.line	72	7.14	0.10	1.51	**0.024**	A-4	6	1.83	**0.100**
	Residual	113	7.41	0.07			A-8	4	10.67	**0.000**
	Total	199	23.69				A-9	5	1.94	**0.094**
							B-1	3	2.33	**0.078**
							B-4	7	1.73	**0.110**
							Cont	1	10.89	**0.001**

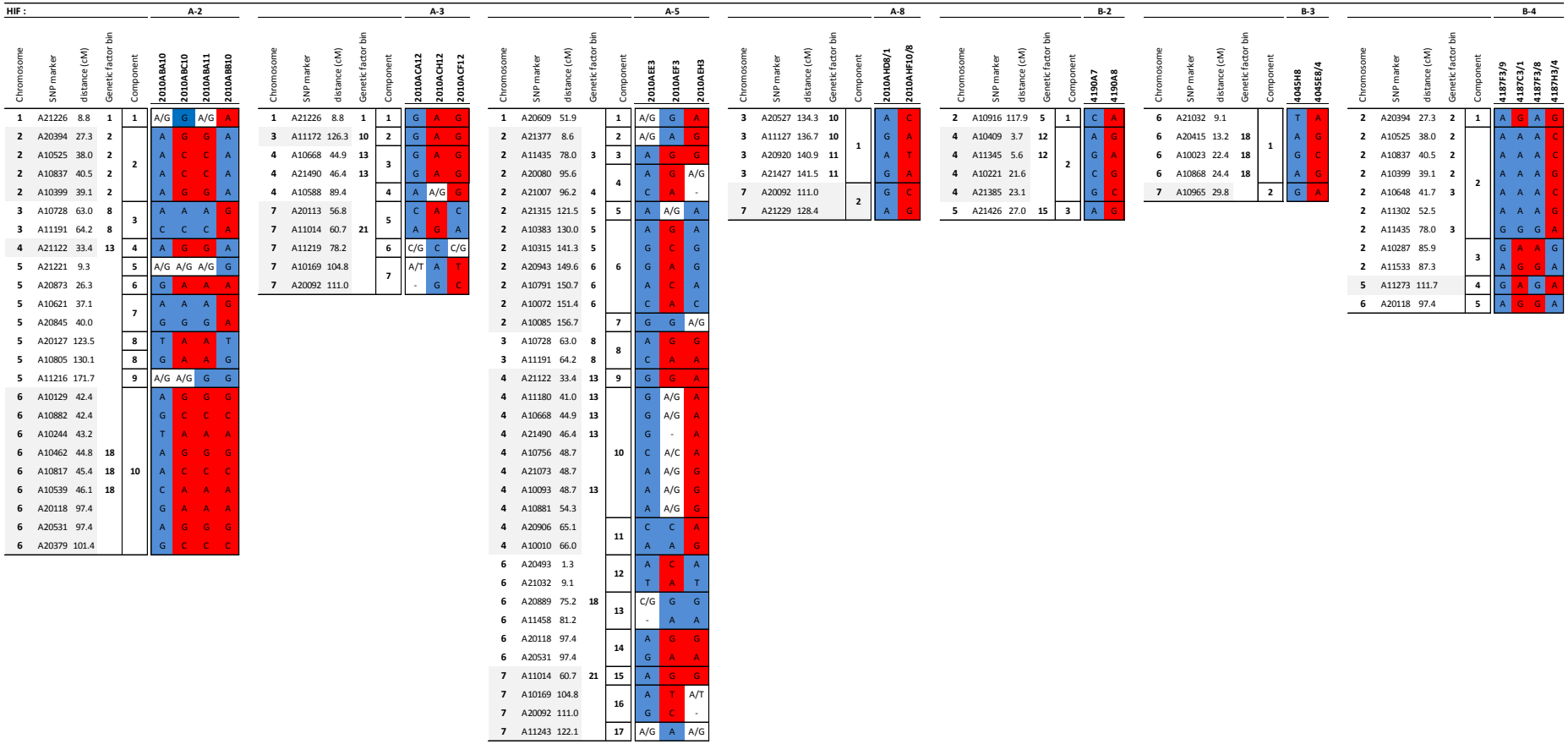


Figure 5.3 Segregating haplotypes at each component of the NILs and HIF grown in 2013 field experiment

For each of the HIF, only the polymorphic markers from the Bx384 platform are presented. The genetic factor bin refers to the targeted regions identified in Chapter 2 and segregating between NIL (Table 5.1). The components are defined by a set of nearby SNPs with a conserved allelic pattern across the NILs of a HIF. The component numbering is specific to each HIF. The colours describe each haplotype contained in the components based of SNP alleles. A colour coding is made in based on the alleles of SNP of a reference NIL in each family. The colour enable each components to be analysed as a multilevel factor: 1 (blue), -1 (red) and 0 (heterozygous or "–" missing value). The SNP alleles should be used for comparison of common markers between HIF.

5.3.3 2013 replicated yield trial

As for 2012, the initial analysis of the trial focused on estimating the variance component attributable to the residual error from a full trial ANOVA with factors set as block, family, lines within family and block x lines within family (Table 5.3). The HIF were significantly different for most traits with the exception of yield which only approached the significance level (P=0.088) (Table 5.3). The significant differences illustrate the different genetic potential of the families attributable to the different founder lines. The overall significance differences between lines within their families (family.line term) was found significant for heading date, height and TGW. The closeness to the significance level obtained for grains (P=0.057) and yield (P=0.081) suggests that specific NIL pairs could be potentially significantly different. This may require increased replication. There were no significant differences found between lines within families for tillering or grain proteins. This partly confirms earlier observations made in 2012 where the two traits only approached the significance threshold (Table 5.2). In order to identify the significant differences between lines within family, a one way ANOVA on individual families was carried out with a variance ratio using overall trial error (Table 5.3). None of the NILs were found significantly different for grain proteins or tillering, although the test did not include all alleles from Saffron and Retriever. The significant differences observed indicate that families B-4 and A-5 could potentially validate genetic factors for grains and yield respectively. The NILs in families A-2, A-3, A-5 and B-4 were found to have significant differences for heading date. This significant effect may benefit from high heritability in this trial. Remarkably, the NILs in family B-4 were found to be significantly different for a range of traits including Grains, TGW, Hd and yield. In B-4, lines 4187C3/1 and 4187H3/4 were found to have a recorded heading date 5 and 9.7 days later than 4187F3/8 or 4187F3/9, strongly supporting the presence of earliness alleles segregating with alleles at a component of B-4 (Figure 5.3). In addition to Hd, significant differences in yield were found between NILs of A-5 as well as for TGW and Ht between A-3 NILs. The NILs of B-3 confirmed 2012 differences for TGW with 4045E8/4 significantly higher than 4045H8. This family also suggest a segregation of alleles for grain sugars. The remaining NILs in families A-8 and B-2 did not capture significant differences in the traits measured despite significant differences in the 2012 mini-plot experiment Table 5.2.

Table 5.3 ANOVA table for phenotypes collected on the 2013 NILs trial experiment.

The full trial ANOVA was carried out for each of the phenotypes using the complete set of data to give an estimate of the trial residual error variance. The within family ANOVA utilises the residual error variance (residual m.s.) from full trial ANOVA to calculate variance ratio (v.r.) and F probability (F pr.).

Traits	Full trial ANOVA					Within family ANOVA			
Yield	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	block	1	1.83	1.83	10.43	**0.002**	A-2	3	0.86	0.467
	family	7	2.35	0.34	1.91	0.088	A-3	2	0.35	0.705
	family.line	16	4.76	0.30	1.69	0.081	A-5	2	4.44	**0.017**
	block.family.line	23	4.12	0.18	1.02	0.461	A-8	1	0.17	0.682
	Residual	48	8.43	0.18			B-2	1	0.12	0.727
	Total	95	21.50				B-3	1	2.00	0.164
							B-4	3	2.95	**0.042**
							cont	3	1.27	0.294
Heading date	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	block	1	197.80	197.80	281.5	**<.001**	A-2	3	7.27	**0.000**
	family	7	189.96	27.14	38.62	**<.001**	A-3	2	20.84	**0.000**
	family.line	16	360.58	22.54	32.07	**<.001**	A-5	2	9.41	**0.000**
	block.family.line	23	22.53	0.98	1.39	0.164	A-8	1	2.06	0.158
	Residual	48	33.73	0.70			B-2	1	0.00	1.000
	Total	95	804.60				B-3	1	0.09	0.769
							B-4	3	124.41	**0.000**
							cont	3	18.48	**0.000**
Height	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	block	1	693.38	693.38	118	**<.001**	A-2	3	0.16	0.925
	family	7	532.04	76.01	12.94	**<.001**	A-3	2	3.63	**0.034**
	family.line	16	605.92	37.87	6.45	**<.001**	A-5	2	0.00	1.000
	block.family.line	23	115.63	5.03	0.86	0.650	A-8	1	0.00	1.000
	Residual	48	282.00	5.88			B-2	1	0.09	0.772
	Total	95	2228.96				B-3	1	0.09	0.772
							B-4	3	12.41	**0.000**
							cont	3	19.22	**0.000**
TGW	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	block	1	220.59	220.59	45.81	**<.001**	A-2	3	1.87	0.147
	family	7	970.62	138.66	28.8	**<.001**	A-3	2	2.97	**0.061**
	family.line	16	401.02	25.06	5.2	**<.001**	A-5	2	0.64	0.534
	block.family.line	23	185.20	8.05	1.67	0.067	A-8	1	0.00	0.968
	Residual	48	231.14	4.82			B-2	1	0.00	0.989
	Total	95	2008.57				B-3	1	4.38	**0.042**
							B-4	3	3.30	**0.028**
							cont	3	18.74	**0.000**
Grains	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	block	1	0.77	0.77	0.34	0.564	A-2	3	0.37	0.775
	family	7	165.88	23.70	10.47	**<.001**	A-3	2	1.95	0.153
	family.line	16	65.77	4.11	1.82	0.057	A-5	2	0.50	0.611
	block.family.line	23	30.65	1.33	0.59	0.915	A-8	1	1.14	0.292
	Residual	48	108.66	2.26			B-2	1	0.59	0.447
	Total	95	371.74				B-3	1	0.06	0.810
							B-4	3	3.59	**0.020**
							cont	3	3.50	**0.022**

Table 5.3 cont.

Traits	Full trial ANOVA						Within family ANOVA		
Tillering	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
	block	1	50280	50280	4.75	**0.034**	A-2	3	0.24	0.865
	family	7	335557	47937	4.53	**<.001**	A-3	2	0.86	0.428
	family.line	16	86691	5418	0.51	0.928	A-5	2	0.28	0.760
	block.family.line	23	186275	8099	0.77	0.753	A-8	1	1.51	0.225
	Residual	48	507795	10579			B-2	1	0.18	0.672
	Total	95	1166598				B-3	1	0.32	0.574
							B-4	3	0.82	0.491
							cont	3	0.24	0.868
Grain	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
proteins	block	1	22.37	22.37	154.7	**<.001**	A-2	3	0.80	0.499
	family	7	7.02	1.00	6.94	**<.001**	A-3	2	0.07	0.928
	family.line	16	2.91	0.18	1.26	0.263	A-5	2	1.65	0.203
	block.family.line	23	4.79	0.21	1.44	0.142	A-8	1	2.17	0.147
	Residual	48	6.94	0.14			B-2	1	0.10	0.748
	Total	95	44.04				B-3	1	1.11	0.297
							B-4	3	1.90	0.142
							cont	3	1.73	0.174
Grain	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.	Family	ndf	v.r.	F pr.
sugars	block	1	0.00	0.00	0.01	0.906	A-2	3	2.04	0.121
	family	7	2.43	0.35	7.61	**<.001**	A-3	2	0.89	0.417
	family.line	16	1.77	0.11	2.41	**0.01**	A-5	2	1.02	0.368
	block.family.line	23	1.34	0.06	1.27	0.236	A-8	1	0.75	0.390
	Residual	48	2.19	0.05			B-2	1	0.30	0.589
	Total	95	7.73				B-3	1	4.22	**0.045**
							B-4	3	0.75	0.530
							cont	3	7.06	**0.001**

The project investigated further the significant effects associated with the segregating loci between NILs within HIF. The genetic components segregating between NILs (Figure 5.3) were tested in a one way ANOVA using the overall trial error. The additional degrees of freedom and hence more accurate estimate of the between plot error for the majority of the traits increased the significance of the component (Table 5.4). The resolution of the experiment and the number of NILs tested in each HIF limits the number of components that could be included in a maximal model at a time (i.e. the number of alleles at components was generally equivalent or lower to the number of lines tested). Although significant components were considered in a maximal model for each trait, the optimal model was often simplified to contain a single most significant component accounting for most of the genetic variation between NILs.

The main QTL targeted by the HIF family A-2 was the TGW and tillering genetic factor 2 co-located with component 2 on 2HS for which NILs C10-A11 and A10-B10 had the Retriever and Saffron haplotypes respectively (Figure 5.3). A-2 is also segregating for genetic factors 1, 8, 13 and 18 which were not associated with yield or yield components effects but straw and disease traits (Table 5.1b). The 2013 experiment ANOVA on A-2 HIF of did not capture significant differences for TGW (F pr. = 0.147) but only for heading date (Table 5.3). This heading date effects was found to be significantly associated with component 2 (made of four SNP) with an effect approximating two days between the late allele carried by lines A10 and B10 and the early allele carried by lines C10 and A11 (Table 5.4). Based on prior knowledge and especially the location of the photoperiod gene *Ppd-H1*, component 2 of A-2 is the more plausible factor responsible for Hd difference but the small confounding (co-segregating) fragments component 4 and component 8 on chromosomes 4H and 5H cannot be formally excluded. It is therefore possible that the haplotypes segregating in that family carry alleles with Hd that have been detected in the NUE-CROPS GWAS (Chapter 3) rather than effects found in the bi-parental population (Chapter 2).

NILs of family A-3 aimed to test possible allele effects on tillering at genetic factor 10 and TGW at genetic factor 21. The family was also segregating for SNP associated with genetic factors 1 and 13 (Figure 5.3). Significant differences between NILs were found for heading date, height and TGW (Table 5.3) despite neither heading date nor height being candidate QTLs expected to segregate in that family. The significant TGW effect was found to associate with component 5 (confounder with component 6) in a region on

7H where the S×R QTL TGW_GS.3_5 and Ht.3_2 were mapped. Here, the polymorphic markers of component 5 are located 10.4 cM away from the S×R peak QTL at SNP A10431 (Table 2.8, Supplementary data 1). Although these results agree with the location of S×R TGW effects, the haplotypes segregating between the NILs of A-3 do not correspond to Saffron or Retriever. The secondary hits and significant differences between NILs for Hd and Ht were found to associate with component 4 also confounded with component 7 (Table 5.4) despite none of them being associated to corresponding genetic factors form the bi-parental mapping. Although these components are represented by three alleles, the allele of line 2010ACF12 appear to be recessive and associated to a delay in Hd of 3 days and height increase of 4 cm compared to the heterozygotes state and homozygote opposite allele (Table 5.4). These effects may correspond to QTL identified in Chapter 3 where the mapping included a greater allelic diversity than the bi-parental mapping study.

The HIF A-5 was retained to investigate effects at genetic factors 3 where grains per ear, tillering and TGW QTL were mapped in the bi-parental population (Table 2.9) and for genetic factors 5 and 6 also associated to tillering QTL (Til_cal.2_2 and Til_mes.2_2). A-5 NILs were also kept based on results from the 2012 experiment which showed strongly significant grains per ear (Table 5.2) which could have related to the segregating component 3 associated to genetic factor 3 and the associated S×R QTL GE_3.1. The complexity of the NILs of Family A-5 resides in the large segregating components associated to additional genetic factors 4, 8, 13, 18 and 21. In 2013, the NILs of HIF A- 5 validated significant differences for yield and heading date while the expected effects of grains per ear and tillering could not be confirmed (Table 5.3). The yield difference of 0.7 t/ha was found to be significant for alleles segregating in component 5 (confounded with 6, 12 and 17) (Table 5.4) associated to the tillering genetic factor 5 and 6. Retriever was found to have the increasing tillering alleles on 2HL (Table 2.6). However in the A-5 NILs, the Retriever haplotype carried by 2010AEF3 at component 5 and 6 (Figure 5.2) was associated with a reduction in tiller number (Table 5.4). Therefore the direct relationship between yield and tillering in association with the Retriever alleles at that locus cannot be established from that NIL contrast.

The significantly delayed heading date of line 2010AEE3 of 2.2 days compared to 2010AEF3 and 2010AEH3 was significantly associated with the allelic pattern of component 3 (confounding with components 8,9,13,14,15). None of these components

were associated with a genetic factor affecting Hd in Table 2.9 but could relate to QTL found in the association mapping of Chapter 3.

The HIF A-8 was conserved in 2013 for its segregation at genetic factors 10 and 11 associated to tillering and TGW QTL (Table 5.1b). In 2012, the NILs of A-8 were significantly different for grains per ear and grain sugars and close to being significant for tillering and TGW (Table 5.2). However, no significant differences between the NILs of this HIF for any of the phenotypes analysed was found in 2013.
Similar, HIF B-2 offered an opportunity to investigate segregating alleles at the genetic factors 5 and 12 associated to QTL for tillering and grain sugars respectively (Table 5.1b). Although a significant tillering effect was observed in 2012, the HIF grown in 2013 did not enable the discrimination of significant effects for the phenotypes measured.

HIF B-3 was developed with only two segregating haplotypes on chromosome 6H and 7H. The 6H haplotype co-localise with genetic factor 18 describing the which straw strength and disease QTL were mapped in the original bi-parental analysis (Table 5.1b). The significant differences in TGW and grains per ear between the B-3 NILs in 2012 suggested that the haplotypes segregating in that HIF could potentially validate effects for yield components (Table 5.2). In 2013, the testing revealed significant effects for TGW and grain sugars (Table 5.3). It should be noted that both Saffron and Retriever are monomorphic for the single SNP segregating at component 2 in B-3 (A10965). The NIL 4045E8/5 which carries the Saffron haplotype at component 1, was found to be associated with an increasing effect on TGW (+1.62g) confirming 2012 results. This same haplotype also associates with higher concentration of grain sugars. Although no yield components effects were detected in this region in the bi-parental study, the consistent TGW effects in both 2012 and 2013 experiments strongly support the presence of a genetic control of TGW on 6HS (9-24cM) or 7H (around SNP A10965 at 29.8 cM). These components could potentially enable the selection of increased TGW independently from yield.

The HIF B-4 targeted the genetic factors 2 and 3 in association with the strong effect TGW (TGW_GS.3_1) and tillering (Til_mes.2_1) QTL observed in the bi-parental mapping (Table 2.9)). The family was also segregating at single SNP loci on 5H and 6H but with no particular QTL effects expected. In 2012, B-4 NILs were significantly

different for Hd, Ht, Grains per ear and approximated significance for TGW (Table 5.2). The 2013 yield trial experiment confirmed significant effects for Hd, Ht and Grains per ear. In addition, the NILs showed differences in yield and TGW (Table 5.3). The model optimisation enabled to associate the NIL segregating allelic components to the trait variation. The component 1 at 27.3cM (confounded with component 4) explained most of the variation in TGW. The A allele of SNP A20394, also shared with Saffron, is associated to a TGW increase of 1.66 gr (s.e. 0.89). A20394 was found highly significant in Chapter 2 for yield components TGW and tillering included in genetic factor 2. It was also shown to be mapped proximal to the photoperiod gene *Ppd-H1* influencing heading date in barley. In this family, heading date appears to be significantly controlled by alleles at component 1 and component 2 simultaneously, both with allelic effects of similar size estimated around 2.5 days (s.e. 0.5) (Table 5.4). Such large effects on delayed heading were visible in the field conditions (Appendix 5. 2). Although these two components did not correspond to the Hd QTL found in the bi-parental DH mapping, the chromosome location suggests correspondence with significant loci identified in Chapter 3. Component 2 is particularly interesting as it was found with significant effects for Ht, Hd and grains per ear in a 2H region where similar QTL effect were observed in GWAS. In B-4, the delayed heading date was associated with taller plants, more grains per ear but lower yield. The yield differences between NILs was mapped to component 2 .The haplotypes described by component 3 corresponded with the presence (NILs 4187H3/4 and 4187F3/9) and absence (4187C3/1 and 4187F3/8) of the sterile spikelet (Sts) on barley ears. This simple DUS trait confirms the quality of the material developed to isolate and validate simple allelic effects. Component 3 however was not found to be significantly associated with any of the quantitative traits measured suggesting that the DUS trait of presence and absence of sterile spikelets is not relevant for improving yield and yield components in two-row winter barley.

Table 5.4 Modelling of best significant component and associated allelic effects for NILs differing significantly in agronomic traits of interest.

Each component (Comp) of HIF found significant in Table 5.3 are tested by ANOVA for association with the trait variation. The confounded components are indicated in brackets. The REML optimal model retains the significant components that best explain explaining the variation between NILs within HIF. The effects of alleles of the significant components retained in the optimal model are reported. For each component, the SNP alleles corresponding to component level -1, 0 and 1 can be found in Figure 5.3.

Trait	HIF	Source of variation (Component)	df	1 way ANOVA (within family error)			With overall trial error		Optimal model	Optimal model effects (REML)							
				ms	v.r.	F pr.	v.r.	F pr.		Constant mean (s.e.)	rep 1	rep 2	s.e. rep	Component level -1	0	1	allele s.e.
Heading date	**B-4**	Comp1	1	218	53.21	<.001	309.6	1.44E-22	**Const + rep + Comp1 + Comp2**	42.84 (0.47)	1.44	-1.44	0.42	2.51	-	-2.51	0.52
		Comp2	1	195	33.61	**<.001**	277.80	**1E-21**						2.36	-	-2.36	0.60
		Comp3	1	24	1.27	**0.281**	34.17	**4.3E-07**									
	A-3	Comp1(2,3)	1	**9.75**	3.69	0.087	13.88	**0.001**	**Const + rep + Comp4**	36.43 (0.42)	1.78	-1.78	0.42	1.74	-1.56	-1.74	0.51
		Comp4 (7)	2	14.6	27.62	**<.001**	20.84	**3.1E-07**									
		Comp5 (6)	1	5.13	1.63	0.234	7.31	**0.009**									
	A-2	Comp1	2	3.36	2.4	0.133	4.78	**0.013**	**Const+rep+Comp2**	37.18 (0.35)	1.98	-1.98	0.4	0.98	-	-0.98	0.40
		Comp2 (4,8)	1	15.2	23.79	**<.001**	21.65	**0.000**									
		Comp3 (5,7)	1	5.07	3.57	0.081	7.22	**0.010**									
		Comp 6 (10)	1	5.07	3.57	0.081	7.22	**0.010**									
		Comp 9	1	0.06	0.03	0.855	0.09	0.768									
	A-5	Comp1 (2,4,10,16)	2	6.61	5.74	**0.028**	9.41	**0.000**	**Const + rep + Comp3**	33.1 (0.47)	1.30	-1.30	0.59	-1.10	-	1.10	0.63
		Comp3 (8,9,13,14,1	1	12.9	12.18	**0.007**	18.37	**0.000**									
		Comp 5 (6,12,17)	1	1.71	0.74	0.412	2.43	0.126									
		Comp 7 (9,11)	1	5.23	2.73	0.133	7.44	**0.009**									
Height	**B-4**	Comp1 (4)	1	90.3	8.19	**0.013**	15.36	**0.000**	**Const + rep + Comp2**	107.9 (0.63)	1.88	-1.88	0.57	4.25	-	-4.25	0.66
		Comp2	1	217	168.2	**<.001**	36.89	**0.000**									
		Comp3 (5)	1	56.3	4.13	0.063	9.57	**0.003**									
	A-3	Comp1 (2,3)	1	10.7	0.71	0.423	1.82	0.184	**Constant + rep + Comp4**	99.33 (1.04)	4.00	-4.00	2.08	2.00	-2.00	-2.00	2.55
		Comp4 (7)	2	21.3	1.64	0.253	3.63	**0.034**									
		Comp5 (6)	1	10.7	0.71	0.423	1.82	0.184									
Grains	**B-4**	Comp1 (4)	1	13.3	10.61	**0.006**	5.87	**0.019**	**Constant + rep + Comp2**	25.93 (0.56)	0.01	-0.01	0.5	1.17	-	-1.17	0.58
		Comp2	1	16.4	16.14	**0.001**	7.24	**0.010**									
		Comp3 (5)	1	0	0.001	0	0.00	0.983									
Thousand Grain Weight	**B-4**	Comp1 (4)	1	44.1	14	**0.002**	9.16	**0.004**	**Const + rep + Comp1**	44.57 (0.44)	-1.61	1.61	0.89	-1.66	-	1.66	0.89
		Comp2	1	28.3	6.5	**0.024**	5.89	**0.019**									
		Comp3 (5)	1	1	0.15	0.7	0.21	0.650									
	B-3	Comp1 (2)	1	21.1	2.51	0.174	4.38	**0.042**	**Const + rep + Comp1**	45.1 (1.02)	-0.65	0.65	2.05	1.62	-	-1.62	2.05
	A-3	Comp1 (2,3)	1	3	0.46	0.514	0.62	0.434	**Const + rep + Comp5**	51.61 (0.59)	-0.94	0.94	1.12	-1.61	-	1.61	1.19
		Comp4 (7)	2	14.3	3.48	0.082	2.97	0.061									
		Comp5 (6)	1	27.5	7.3	**0.024**	5.71	**0.021**									
Yield	**B-4**	Comp1 (4)	1	0.57	2.77	0.12	3.27	0.077	**Const + rep + Comp2**	10.29 (0.21)	-0.12	0.12	0.16	-0.34	-	0.34	0.22
		Comp2	1	1.39	9.6	**0.008**	7.89	**0.007**									
		Comp3 (5)	1	0.85	4.57	0.052	4.83	**0.033**									
	A-5	Comp1 (2,4,10,16)	2	0.78	4.87	**0.041**	4.44	**0.017**	**Const + rep + Comp5**	11.25 (0.24)	-0.08	0.08	0.25	0.35	-	-0.35	0.26
		Comp3 (8,9,13,14,1	1	1.05	5.3	**0.047**	5.99	**0.018**									
		Comp5 (6,12,17)	1	1.28	7.33	**0.024**	7.26	**0.010**									
		Comp7 (9,11)	1	0.01	0.03	0.858	0.06	0.807									
Grain Sugars	**B-3**	Comp1 (2)	1	0.19	0.193	0.071	4.22	**0.045**	**Constant + rep + Comp1**	2.03 (0.12)	-0.18	0.18	0.14	0.16	-	-0.16	0.14

5.4 Discussion

5.4.1 Breeding resources for QTL validation

NILs development

Currently the application of MAS for QTL is only cautiously implemented in plant breeding programmes because further steps for validating the QTL effects are required. Numerous QTL from mapping studies provide extensive information on genetic regions apparently with associated useful effects to breed in improved crop varieties (Bernardo, 2008; Collins et al., 2008). Breeders therefore require effective and efficient tools to help with the validation of these QTL and transfer research results into progress. In this study NILs have been developed to test and validate the effect of alleles at specific chromosome locations. Optimal NILs should only differ for a single chromosome segment in order to exclude any other background genetic variation. However, methods for NIL development often lead to residual segregation which when minimised should still provide valuable contrasts. The continuously renewed breeding material can provide a efficient genetic resource for the validation of QTL and alleles effects relevant to elite breeding. This study shows that heterogeneous inbred families (HIF) made of near isogenic individuals is perhaps best suited to a direct application to a breeding programme (Pumphrey et al., 2007; Tuinstra et al., 1997). Indeed, barley lines in F4 and F5 generations following either a Pedigree or Single seed descent breeding schemes contain sufficient heterozygote regions that can be matched with QTL locations and targeted in what would become a HIF founder line.

Here, the barley QTL identified in Chapter 2 and characterised as genetic factors constituted the set of target QTL used for screening founder lines for HIF. Populations of breeding material expected to segregate for alleles present in the DH mapping population were screened in order to identify sufficient HIF founders. In order to investigate the possibility of more than two alleles at a genetic factor, breeding material from crosses unrelated to Saffron or Retriever was also screened using the same genotyping array. This approach was successful as it was indeed possible to find potential HIF founder lines with different segregating haplotypes at the targeted genetic factors. This very speculative approach may identify effects of multiple alleles and haplotypes, especially if it is supported by the prior knowledge on the presence of effective factors. However in this case, the production of NILs was complicated by errors of seed tracking that may have happened between the plant genotyping at F4 and

the F5 seeds leading to an unfortunate discrepancy between the originally planned QTL targets for validation and subsequent genotypes. Nevertheless, the new HIF founder lines could be used for validating different targets and haplotypes by the introduction of different parental haplotypes to those mapped in the Chapter 2. These unknown haplotypes and genetic backgrounds do indeed present a limit to the interpretation of the potential effects as the phasing of the alleles cannot be referred to the mapped effects of haplotypes segregating in Chapter 2. The experiment generated a valuable resource of 14 HIF containing each 8 to 24 NILs that could be used to validate different haplotypes at multiple genetic factors (Table 5.1).

Number and size of haplotypes

The concept of validating QTL effects by NILs resides in the comparison of haplotypes expected to carry the different alleles of a functional gene controlling the trait of interest. Haplotypes can be defined using a set of markers associated with the QTL interval but will vary in size depending on the original mapping accuracy, the number of markers available to define the region and the recombination events found and selected. In a NILs development approach, Oh et al., (2010) identified a set of 47 descendant lines with informative recombination breakpoints using six SSR markers and grouped them based on similar haplotypes. The haplotypes were then used in a substitution mapping approach identifying which recombination breakpoint was associated to the difference in grain weights. To validate the DON QTL in wheat, Navara and Smith (2013) defined two haplotypes with a limited number of markers in each of the NILs families so that each haplotype was represented one to seven times in a family. Both these approaches allow the haplotypes to be replicated within families in order to overcome the potential residual background effects. In this study, the small number of lines developed within each HIF restricted the choices for optimal comparison. Although lines for several of the candidate loci shared the same haplotype and could have formed larger polymorphic groups for testing, the genetic background information revealed that residual segregation was confounded with the targeted alleles. Therefore the confounding residual segregation in the background cannot be excluded to interact and modify an effect on the phenotypes that are expected to be controlled by the targeted regions. In cases of numerous segregating regions, an additional cycle of self-pollination can be used to reduce the background polymorphism. Pumphrey et al. (2007) used a similar strategy to improve the validity of the NILs contrasts. Overall, the

NILs contrasts with homogenous background were favoured and a subset of seven HIF containing two to four NILs was retained for 2013 field testing.

NIL development using HIF is not without practical issues as residual segregation needs to be accounted for to generate sufficient progenies in order to identify optimal NILs comparisons. In this study, the large residual segregation led a very limited number of acceptable NILs comparisons and did not entirely resolved the confounding background segregation. Such a situation generates over parameterized contrasts due to more haplotype combinations across the genome than available NILs in an HIF. Each NIL comparison is unique as segregating background could be associated with confounded effect. In order to improve the contrasts, it is suggested to consider a larger number of NILs within HIF to identify optimal contrasts (perfect NILs) or a group of NILs that contain replicated haplotypes at the target region in a randomised genetic background. This could be achieved by generating a larger number of inbred progenies from the HIF founder lines. The optimal approach would be to generate new NILs by crossing or backcrossing fairly homozygous NILs tested in 2013 and developing segregating populations. This extra cycle of crossing can make a more homozygous background andalso smaller haplotypes with recombination breakpoints around the target that can be in substitution mapping (Oh et al., 2010).

5.4.2 Haplotype effects and QTL validation

In order to test the hypothesis of an effect of the components segregating between NILs, a two-step approach was implemented. The comparisons of haplotypes segregating within families was favoured following the approach used by Navara and Smith (2013). However, this conservative approach in estimating effects reduces the statistical power that can be obtained if all HIF are included in the analysis but excludes potentially strong confounding background effect linked to the different pedigrees of the HIF. The test of SNP and haplotype effects using the whole set of NILs from different HIF would require a standardisation of the haplotypes based on a reference genotype. However, this approach would often produce comparisons of single SNP haplotypes from unrelated and highly structured lines which could generate false associations. Instead, the study focused on within HIF analysis to maximise the estimations of effects between contrasts. The HIFs with significant differences between NILs were first identified so that the segregating haplotypes within HIF between NILs could be tested and their effect measured. The subdivision of haplotypes into components enabled for some

contrasts to add replication while diluting the background effects and also fine map the effect to the component itself. Such approach is only effective for the HIF constituted of three NILs or more.

The major QTL effect of TGW and tillering found on 2HS in Chapter 2 (genetic factor 2) was one of the principal targets for validation in the NILs. HIF A-2 was developed to focus on validating this genetic factor by having NILs segregating for contrasting haplotypes at component 2 associated with the TGW/tillering genetic factor. In 2012, the A-2 HIF showed only limited significance for TGW differences. In 2013, the A-2 NILs experiment did not capture significant differences for either TGW or tillering initially targeted by component 2 (confounder with 4 and 8) and therefore did not validate the genetic factor 2 QTL from the S×R mapping. However, the A-2 NILs were significantly different for heading date at component 2 suggesting that the NILs were validating effects that had not been mapped in the bi-parental population. This may be due to the different pedigree of the founder line and therefore the comparison of different alleles at the component 2 rather than the Saffron and Retriever alleles. Because the component 2 in A-2 was only described by four SNPs, the haplotypes of the NILs were a priori identical to either Saffron or Retriever haplotypes although the NIL pedigree suggests different origins. The heading date effects located at component 2 agreed with the location of the photoperiod response gene *Ppd-H1* (Turner et al., 2005). Although the four SNPs in A-2 HIF component 2 were not significantly associated with heading date in the GWAS (Chapter 3), the allele effect direction was similar across the two experiments (Table 5.4 and Supplementary data 2) suggesting that *Ppd-H1* is a strong candidate to explain the variation in heading in A-2 but the alleles do not have an effect on TGW variation in that particular experiment. However, it needs to be born in mind that the contrasts in HIF A-2 were insufficient to completely resolve the origin of the heading date effect as the confounder component 8 on chromosome 5HL was also located in a region where NUE-CROPS QTL Hd_5 was identified. The marker A10805 of component 8 (130.1 cM) is located 6cM from the peak SNP of Hd_5 close to the candidate gene *vrn-H1*. An increase in the number of NIL within that A-2 HIF as well as additional genotyping of NILs with the significant markers identified in GWAs (e.g *Ppd-H1* SNP) would help to characterise further the haplotypes at those loci and to achieve greater resolution of the effects location.

The HIF A-3 validated in 2013 a significant TGW effects in the centromere region of chromosome 7H (component 5) which was in alignment with the location of TGW-GS.3_5 QTL associated to genetic factor 21. In both S×R mapping and HIF A-3, the Retriever (G) allele for A11014 SNP is associated with an increasing effect on TGW. However the haplotypes segregating in that HIF do not correspond to the Saffron or Retriever. Although it is possible that alleles in the centromere of 7H are involved in controlling the variation of TGW in winter barley, these TGW effects were not confirmed by the GWAS of NUE-CROPS and AGOUEB. This may be due to the prevalence of the relevant alleles in these larger germplasm sets. Alternatively this NIL contrast appears to identify an effect expressed exclusively in this specific genetic background which may be due to non-allelic interactions (Bocianowski, 2014). Also, the height effect expected to map at that same genetic factor 21 on 7H was not confirmed by the NILs. Instead, significant differences between NILs for height and also heading date mapped with confounding components on 4HL (89.4cM) and 7HL (104.8cM). On 7HL, no QTL for relevant traits were mapped that would support the presence of a genetic factor affecting height. On 4HL, the relation of the effect to the height QTL Ht3_1 is unclear due to the significant distance between the QTL and component 4. In addition, the single SNP in that component A10588 is monomorphic between Saffron and Retriever. Although no height effect was found in GWAS on 4HL, component 4 is mapped in the region of a multiple QTL for tillering described in Chapter 3 and for which putative candidate genes have been proposed (see 4.3.3 and 4.4.2). It may be worth investigating further the polymorphism on 4HL in A-3 by increasing the marker density and developing recombinants.

The complexity of QTL validation using NILs and HIF was exemplified by HIF A-5. First the numerous co-segregating loci (components) were found over multiple loci targets despite a cycle of self-fertilisation. This situation provided limited and not sub-optimal NIL contrasts within that HIF. Second, the inconsistency of significantly different phenotypes between the NILs illustrates the seasonal Genotype x Environment interaction. Indeed, the HIF A-5 was initially kept for the NILs with contrasting effects on grains per ear found in 2012 but these were found significant only for yield and heading date in 2013. Because of the confounding loci, the correspondence of the segregating factors with the QTL mapped in that study cannot be established with certainty. The components associated with both significant yield and heading date differences did not correspond to any effects observed in the mapping studies presented

in Chapter 2 and Chapter 3. These results suggest that the NILs of A-5 are identifying significant genetic factors for yield and heading date but the resolution of the HIF is insufficient to capture the exact location of the effects. The A-5 HIF could be developed further to optimise the contrasts (see 5.4.1).

Genetically, the B-4 family extends the polymorphic region on 2HS beyond that seen in A-2 to include pericentric parts of the chromosome, up to the SNP A10287 associated with *vrs1*. The haplotypes at that chromosome section were dissected into three components (component 1, 2 and 3). Alleles at SNP A20394 of B-4 component 1 matched direction of the significant heading effects found for A-2 component 2 with a stronger effect but did not validate the TGW and tillering effects seen in Chapter 2. Aided by the GWAS results (Chapter 3), the study suggests that allele A of A20394 associates to alleles of earliness at *Ppd-H1* in the B-4 NILs haplotype. However, the component 2 in B-4 was also significantly involved in the control of heading date in addition to height, grains per ear, TGW and yield. This particular component 2 location corresponds to a chromosome segment in 2H that has been found to carry a larger QTL cluster in the GWAS although not the S×R population. In winter barley, the locus was associated with three days difference in heading but also had an effect on height, grains per ear, TGW, yield and nitrogen related traits (Chapter 3). In spring barley, heading date effects in the centromeric region of 2H have been described and associated with the gene candidate HvCEN (Comadran et al., 2012). In a GWAS on spring barley Alqudah et al. (2014) showed that the alleles at HvCEN were associated with delaying effects on heading date within population groups defined by their alleles at *ppd-H1*. Here, the optimal modelling of components strongly suggests that the heading date is controlled by two additive genetic factors of equivalent size in this winter barley B-4 family. The modelling of component 2 untangles the genetic architecture of heading date in this HIF and is consistent with the segregation of HvCEN in this pericentric region of 2H and the observations of Alqudah et al. (2014). The combination of late alleles at both the 2H loci led to an impressive delay in heading date greater than 10 days (NIL 4187H3/4) compared to the combination of early alleles. Such material could be used to investigate the sequence polymorphisms in winter barley for HvCEN especially at the winter-spring SNP at intron-exon splicing site and P135A (Comadran et al., 2012).

The B-4 component 2 effects reinforces the presence of strong pleiotropic effects associated with the genetic factor on 2H centromere and herein the difficulties in identifying an optimal allele to select for in breeding programmes. The lateness is

associated with taller and lower yielding plant type in 2013 but increased the number of grains per ear, a very valuable yield component. The effect on Grains was not expected in this area from the S×R data however the SNP and alleles effects at component 2 of B-4 are in alignment with the position and direction of the GWA major QTL for Grains in two-row winter barley panel (Grains_1, Table 3.2) and with the co-located height QTL Ht_2. This increases the value of those NILs for QTL validation of the potential gene candidate affecting meristem elongation (Cremer et al., 2001) but also the multiple traits found associated with that locus in Chapter 3. Additional material development may be required to obtain full combination set of alleles at component 1 and component 2 to test hypotheses of an ideal combination of alleles favouring both yield components. This would aim to maintain the large number of grains at component 2 (late allele) with a reasonable TGW and earliness at component 1 (early allele). A cross could be carried out between 4187F3/9 and 4187H3/4 in order to find a recombination event on 2HS between component 1 and component 2 producing sufficient material to allow for both the fixation of the genetic background and a reduction in size of component 2.

Although significant effects were validated with the segregating components of the NIL contrasts, the study was unsuccessful in validating QTL corresponding to the S×R population despite specifically targeting the relevant genetic factors. Inconsistency in the significance of effects has been frequently observed in QTL mapping experiments indicating the presence of non-additive genetic effects due to pleiotropy, QTL × genetic background and QTL × Environment interactions (Bernardo, 2008; Cooper et al., 2009; Romagosa et al., 1999). Epistatic effects can differentiate allele effects in different populations and lead to contrasting selection (Asíns, 2002). In this study, the low success of validation can partly be due to the material itself that did not only concentrate on the Saffron or Retriever alleles inherited in a Saffron or Retriever genetic background. Also, the Saffron and Retriever haplotypes found in a different genetic background of a HIF may not have expressed the same effect and phenotype due to non-additive interactions. In most cases, the haplotypes at QTL targets were from an unknown pedigree reducing the potential to detect a corresponding effect at co-located S×R genetic factor. However the material enabled the mapping of unexpected effects for quantitative traits in chromosome regions that have been found associated with relevant QTL clusters and known genes.

Moreover, the variability of the effects observed across years highlights the strong effects of seasonal G×E interactions to account for in validation experiments. This was

exemplified by the contrasting results of HIF in 2012 and 2013. HIF A-8 and B-2 did not confirm the significant NIL differences observed for yield components in the 2012 experiment. QTL×E interactions have been observed in several QTL mapping studies (Emebiri, 2013; Hayes et al., 1993; Saal et al., 2011) and need to be considered in the investigation of genetic architecture of quantitative traits (Cooper et al., 2009). Here, the NILs results show that yield components are strongly affected by the extent of G×E associated with variable growing seasons and sites. The experiment describes the environment specific allele effects which may be used by breeders in MAS to optimise crop performance under specific environments which may be a limiting factor for commercial breeding. However the inconsistency in the effects imply that breeders aiming at targeting those alleles will have to first validate the effect in their targeted environment and second validate the allele effect in an elite genetic background. This strategy may be unrealistic in commercial breeding programmes.

Chapter 6
General discussion

Breeding for improved varieties relies first and foremost on the ability to select for favourable allele combinations associated with traits of interest, resulting in better agronomic and yield performance. From a commercial breeding perspective, a thorough understanding of the genetic control of traits results in increased breeding efficiency and reliability. The results obtained in this project illustrate that considerable genetic knowledge on quantitative traits such as yield and yield components can be collected using a range of complementary disciplines and offers real opportunities for achieving genetic progress.

6.1 QTL mapping studies

The foundations of this project consisted of large scale QTL mapping studies aimed at understanding the genetic architecture of yield and yield components, both quantitatively inherited traits. The success of any mapping approach relies largely on the quality of the phenotypic data of which the proportion explained by genetic variation is maximised by accounting appropriately for environmental variation or error variance (Bernardo, 2008; Piepho et al., 2008). To reach that goal, the strategy targeted optimal growing conditions to allow the full expression of the crop yield potential by minimising environmental stress using the best possible field conditions and agronomic practices. Furthermore, this experimental protocol designed for high input regimes relates directly to the majority of conventional breeding and farming practices in the UK (DEFRA, 2010), making the results more interpretable and comparable to commercial crop production. The choice for experimenting in optimal growth conditions also maximised trait heritability which helped the accuracy of phenotyping and may compensate for the reduced replication available (Chapter 2). The bi-parental mapping QTL analysis of multiple environment data, confirmed consistent QTL effects despite significant GxE and highlighted the presence of 23 genetic factors associated with one or more QTL across the genome. A similar reasoning was behind the NUE-CROPs GWA study (Chapter 3) to aim at maximising the transferability and application of the results into applied breeding. The reliability of the phenotypic data was also enhanced by the use of statistical mixed modelling. In GWAS, the phenotypes recorded as BLUPs

were mapped as they represent the best genetic value of the varieties (Piepho et al., 2008). Therefore this phenotypic data obtained under optimal conditions should maximise the quality QTL of mapping, and ensure the relevance for conventional breeding and farming practices.

The other aspect addressed in this study is related to the complementarity of different QTL mapping approaches. On one hand, the bi-parental population enables the mapping of QTL in a well understood and unstructured genetic environment. Because there is only a restricted set of alleles segregating in bi-parental QTL mapping populations, the choice of parental lines is critical to determine the relevance of alleles in a breeding germplasm and impact of the results. In barley, crosses between extreme phenotypes have been used to identify major genes involved in ear morphology and the control of phenology (Decousset et al., 2000; Turner et al., 2005; Xue et al., 2010). From a breeding perspective, the pleiotropic effects on agronomic traits attributed to these major genes (e.g *vrs1*) are hardly exploitable because the traits affected are essentially defining the crop type itself. Therefore studies that aim at avoiding the mapping of pleiotropic effects at major genes should be more attractive for a MAS approach in breeding (Rae et al., 2007). This project highlights the potential of DH populations created from elite parents (Saffron and Retriever) of the same crop type: two-row winter barley. It confirms the benefits of using breeding related material to carry out QTL mapping of direct relevance to the breeding activities (Würschum, 2012). The value of the Saffron×Retriever population was further increased by the UK registration and release of the elite variety KWS-Tower that came from the same cross, while its progeny KWS Orwell (pedigree: KWS-Tower×KWS-Salsa) was UK listed in 2014. This confirms that the beneficial alleles brought in from Saffron and Retriever have been important in subsequent breeding cycles. However certain limitations of the population were evident such as the monomorphic chromosome segments on 1H, 5H and 7H suggesting common ancestry and impeding the investigation of genes and alleles present in those regions. In addition, some QTL had wide confidence intervals indicating that markers associated with these QTL may be less precise for targeting the causal polymorphism or of lower interest for MAS.

On the other hand, GWA mapping allows for screening marker-trait associations within a larger and more diverse set of alleles than bi-parental mapping and with increased mapping precision and including of rare alleles in the panel (Gupta et al., 2005). These alleles are likely dispersed throughout the breeding germplasm, segregating in mapping

populations and available for introgression into UK varieties. A major concern in association mapping is population structure which can lead to incorrect marker-trait associations (Comadran et al., 2009; Zhang et al., 2009). Although structure is accounted for by statistical models, major partitioning loci (e.g. *vrs1*) continue to have pleiotropic effects on a range of other agronomic traits and are detected in the QTL analysis (Cockram et al., 2008; Pasam et al., 2012). The approach followed in this project aimed at reducing the main structural components of the mapping panel by focusing exclusively on the two-row winter barley crop type (Chapter 3). The mixed model approach using a Kinship matrix based on genetic marker information theoretically accounted for the population structure present in the two-row winter barley panels investigated. This Kinship approach implemented in the study reduced significantly and appropriately cases of spurious associations as it was reported previously in barley GWAS (Pasam et al., 2012; Wang et al., 2012). It needs to be borne in mind that other correction methods should not be set aside and that even if they produce very comparable results, additional information can be gained by using multiple analyses (Cockram et al., 2010).

This project addressed QTL mapping from a plant breeding perspective. First, the varieties selected for the mapping experiments were to be genetically close relevant to the germplasm used for breeding commercially two-row winter barley. Secondly the phenotypic data collected was similar to the data routinely generated in a breeding programme. This strategy intends to optimise the transferability of research results into commercial breeding with the aim of utilising the MTA discoveries in the selection process. Furthermore, the complementarities between mapping methodologies can be used to identify targets for further investigation.

6.2 Detection of constitutive QTL

The assets of this project were the extensive data collected from diverse genetic resources used in three main QTL mapping experiments. In order to identify relevant QTL and define the position of genetic targets with associated markers, the mapping results were gathered using the advantages of having common SNP markers across the genotyping platforms (Close et al., 2009). Unlike meta-QTL analysis which requires the statistical integration of multiple genetic maps (Swamy et al., 2011), the approach was

simplified by using an existing the reference map based on bi-parental and LD mapping (Comadran et al., 2012). The SNP markers were considered sufficient to anchor the QTL and their support intervals along chromosomes and allowed the clear identification of QTL clusters across studies.

In general, the positions of QTL clusters showed more correspondence between AGOUEB and NUE-CROPS GWAS than with the bi-parental QTL. This can partly be explained by a considerable proportion of varieties overlapping between the GWA mapping panels implying a large proportion of shared alleles. Less cases of correspondence between GWAS and bi-parental mapping in this study were seen suggesting that these mapping approaches capture different genetic factors, even though Saffron and Retriever were present in the NUE-CROPs variety panels. In addition, both AGOUEB and NUE-CROPs panels were historic which could be seen by the mapping of QTL for date of variety date of inscription (UPOV). Therefore it is likely that some alleles in especially Retriever may be represented in minority in the GWAS analysis. The S×R results are particularly relevant to current varieties, especially those from KWS-UK breeding programme for which a considerable current elite germplasm originates from those two key varieties. Consistent effects across studies were however identified, especially for single traits such as the precise mapping of the anthocyanin trait for which the gene *ant-2* was identified and cloned by Cockram et al, (2010). Also, the positon of a main mildew resistance QTL on 5HS in the S×R bi-parental population was also observed in the AGOUEB GWAS (Figure 4.1e). This evidence of correspondence between QTL of highly heritable traits across the three mapping studies supports the strategy followed in comparing mapping studies despite QTL correspondences for complex traits between studies more difficult to establish.

Although the QTL alignment does not provide QTL validation as such, the confirmation between studies is a strong argument toward the presence of a genuine genetic factor. This is reinforced if common significant markers and the phase of these for specific genotypes are found. Some overlap between QTL of the mapping studies were found on all chromosomes (e.g TGW QTL on 2HS and 5HS). The most important QTL cluster between studies was found close to the centromere of chromosome 2H at the HvCEN locus. This flowering time locus was shown to have effects on multiple agronomic traits described in a panel of spring and winter barley varieties (Comadran et al., 2012). The NUE-CROPS and AGOUEB GWAS show that alleles at this candidate gene segregate

within the two-row winter barley germplasm. In addition, the range of traits represented in the QTL cluster emphasise the amplitude of the pleiotropic effects of that locus on major agronomic traits. The control of phenology is well known to play a large role in plant adaptation and agronomic trait variation in barley (Kandemir et al., 2000; Li et al., 2005; Wang et al., 2010). In the present case of HvCEN, the alleles associated with early or late types may have been maintained in the elite genepool by breeders in the two-row winter barley germplasm to adapt to a range of seasonal variation. In addition, there was no evidence of major population structure in the European two-row winter barley panels used in this study to suggest that this locus was responsible for stratification. This was especially important as it could have affected further the GWAS QTL results.

This project shows that the comparison of QTL studies can be an alternative method to meta-QTL analysis to identify constitutive QTL from various studies using complementing mapping methods sharing genetic markers. Further exploitation of the variety panels can also be envisaged to investigate specific alleles by developing material segregating for specific allelic combinations at the loci of interest. One possibility could be the development of a DH population designed to exclude segregation of major alleles such as the Saffron × Retriever population which was monomorphic at the HvCEN locus. Also, specific crosses could be made to generate recombination around the tillering locus on 4HL. Recently, the potential of Multiparent Advanced Generation Inter-Cross (MAGIC) populations has been demonstrated for quantitative trait research in cereals (Huang et al., 2012) though it includes segregation of major genes. This approach introduces segregation of multiple alleles while reducing the genetic population structure. A set of relevant parents from NUE-CROPs or AGOUEB panels could be identified for the creation of a two-row winter barley MAGIC population aimed at reducing the effects of major genes. This resource could also complement the present study in validating further the GWAS results and constitute a valuable breeding tool. However, the time and resources needed for the elaboration of such population restrain the more frequent development of this type of validation approach.

6.3 Yield and yield components

QTL for yield and yield components were found on all chromosomes in the three mapping experiments reported in Chapter 2 and Chapter 3. In the two GWAS, six and nine yield QTL were identified with relatively small effects and low significance scores. Such results highlight the genetic complexity of this trait and the challenge to identify genetic markers associated with significant effects leading to improvement in yield. As expected, the locations of most yield QTL coincide often with those of correlated traits, either component of yield or derived traits (e.g. nitrogen related traits). This correspondence is similar to the findings in other studies investigating multiple agronomic traits (Comadran et al., 2011b; Pasam et al., 2012; Tong et al., 2010). Interestingly, the yield QTL also co-localised with QTL associated with the date of release of varieties which suggests that these alleles are among the genetic factors breeders have selected over the past 40 or 50 years. At the centromere of 6H, the composition of the cluster including UPOV_7 and Yld_7 QTL may correspond to a polymorphism relating to nitrogen related traits and winter hardiness. Moreover, the NUE-CROPs QTL Yld_2, Yld_3 and Yld_4 all coincided with yield component QTL. Yin et al, (2002) demonstrated that the dissection of yield into simpler yield components provides additional insights into the genetic architecture of the trait. In this study, the results show that the dissection of traits is required in order to correctly interpret the role of effective loci before embarking on MAS. Indeed, one of the main challenges in understanding the genetic control of a quantitative trait is to consider the influence of pleiotropic, environmental and epistasic effects on the trait variation (Mackay et al., 2009). Phenology is one of the most important traits conferring adaptability to the environment with substantial pleiotropic effects on yield. In wheat, the alteration of the photoperiod response by *ppd-D1* gene was shown to affect the ratio between source and sink organs in the plant and yield components (Foulkes et al., 2004). Unfortunately the large effects of such genes can hide a range of genes with smaller effects which could be of interest for breeding and impact directly on the significance of allelic differences at other genes working in related pathways (Alqudah et al., 2014). Indeed, Cuesta-Marcos et al. (2009) characterised novel yield QTL after identifying that different maturity groups were present in a mapping population. It could be therefore possible that the strong pleiotropic effects at the HvCEN locus affect indirectly the detection of secondary genetic factors for traits other than heading date. This may be investigated by further mapping by including HvCEN alleles as a co factor

in the mapping study or by enlarging and dividing the mapping panels so that the locus are kept monomorphic in each panel. Yield is not the only trait expected to have a complex genetic control. Components of yield are also prone to interactions and pleiotropic effects of adaptive loci that need to be considered in the interpretation of results.

The number of fertile tillers per plant is a major contributor to yield. It is initiated during the foundation phase of plant development before any other yield component and is therefore less affected by compensatory mechanisms. However during the later growth stages, tillering may influence the balance of source and sink organs and impact on significant compensation mechanisms which can cause the correlation between traits (Sreenivasulu and Schnurbusch, 2012). For example, high tillering may increase the sink size and reduce the ability of the plant to fill the grains resulting in overall decline of grain quality and perhaps of yield. Consideration of this interdependence between traits is important in helping the interpretation of clustered QTL effects as the mapping showed that significantly correlated traits often had clustered QTL. This emphasises the need to carry out and include a complete range of phenotypes in mapping studies in order to sensitively interpret results from QTL studies. An example is provided by the initially appealing tillering QTL cluster on 5HS that suggested that SNP A20553 could be used to select for increased tillering. However, further analysis highlighted the significant negative effect of this QTL on TGW, indicating that the positive effect is balanced out by the decreasing effect of the other yield components (Table 4.1). This same locus and markers were identified with strong associations for tillering and height in a more diverse panel, supporting the evidence of a functional polymorphism present in two-row barley types (Comadran et al., 2011b). A similar observation of an allele with opposing effects on different yield components was made on chromosome 2HS for the tillering and TGW QTL described in Chapter 2. This highlights the recurrent dilemma faced by plant breeders when selecting appropriate QTL targets to be implemented in MAS. If perhaps observed solely as a TGW QTL, active selection could lead to disappointing results as low tillering selections continue to fail to improve in yield. However the project also revealed some promising candidate regions with QTL having unidirectional positive effect across traits. A tillering QTL cluster with no apparent negative effects on other yield components was identified on 4HL and should highly be considered as target for MAS. Associated with SNP A20732, it appeared independent to heading date and sufficiently distant from the phenology controlling

gene *vrn-H2* (Dubcovsky et al., 2005). QTL for this trait has indeed previously been identified in the region but without any candidate genes being proposed (Borràs-Gelonch et al., 2011; Long et al., 2013). These results confirm the importance on 4HL in the control of tillering,

The major QTL found for number of grains per ear in two-row winter barley was a surprising outcome of the GWAS (Chapter 4). Because the panels were composed exclusively of two-row types, it was expected that the genetic control of this yield component would involve multiple loci however the results identified only one main QTL for grains per ear located at the 2H centromeric region and comapping with the HvCEN locus (3.4.2 and 3.4.3). The direction of allelic effects showed that delayed flowering was associated with an increase in the number of grains per ear. This contrasts with the observations made by Comadran et al. (2011b) who concluded the opposite effect in a two-row spring barley germplasm. These different interpretations may reflect differences in gene by environment interactions between the spring and winter germplasm in relation to plant adaptation and further emphasises the evidence for strong pleiotropic effects at that locus. The involvement of this gene in inflorescence architecture has been reported in other studies; the *CENTRORADIALIS* gene family was shown to be involved in the fate of floral meristems in snapdragon (Cremer et al., 2001) and, the overexpression of the gene in tobacco delayed the shift to flowering (Amaya et al., 1999). The CEN like gene MdCENa in apple trees was shown to be involved in the proliferation of tissues and able to complement the flowering time pathway of FT1 gene when transformed in Arabidopsis (Mimida et al., 2009). The observation of a QTL cluster at the HvCEN locus supports the hypothesis of a similar function for the gene in barley in controlling meristem behaviour during the ear development. Recently, Boden et al. (2015) showed that the wheat photoperiod gene *Ppd-1* was involved in controlling floral architecture and the development of paired spikelets which reinforce the strong association between phenology and floral architecture found in the NUE-CROPS GWAS and the pleiotropic effects of such genes. Ultimately the gene affects the length of development phases and timing of heading date, which if manipulated carefully could result in increased grains per ear and a potential increase in overall yield (Alqudah et al., 2014).

It should be noted that another four QTL were found for grains per ear in the mapping experiment of Chapter 2 although none of them coincide with the GWAS results. These differences demonstrate the power of biparental mapping populations monomorphic for

major alleles, in characterising loci with smaller effects. Although the GWAS accounted for population structure, the adaptive role of HvCEN for the two experimental seasons may be reflected across the whole range of phenotypes, especially yield and yield components. Further QTL analysis accounting for the variation at the HvCEN locus could help in identifying loci independent from this major gene effects.

TGW is the last yield component to be determined in the plant cycle after tillering and grains per ear. In this study, the genetic complexity of TGW was highlighted by the large number of QTL found for the trait which may relate to the high heritability of the trait. In fact, most TGW QTL coincided with QTL for other traits such as heading date, tillering, and grains per ear. These types of QTL associations between traits have also been observed in many other mapping studies analysing several yield components (Pasam et al., 2012; von Korff et al., 2006). On one hand, QTL clusters can be caused by the direct pleiotropic effects of the polymorphic candidate genes at those loci while on the other hand, the indirect effect of plant compensatory mechanisms influencing the balance between source and sink organs may partly explain the QTL clustering. The importance of phenology and plant adaptation in the final performance of sink organs was confirmed by the convincing associations between TGW and heading date QTL (e.g HvCEN cluster on 2H). The 5HS QTL cluster (Figure 4.1e) clearly suggests a strong relationship between tiller number and grain weight and an equivalent interpretation can be made at the HvCEN locus. This complex network of pleiotropic effects and interactions impede the process of isolating independent TGW effects and ideal QTL targets for a MAS approach in two-row winter barley. Nevertheless, results from rice studies have suggested that specific attention should be directed to genes involved in plant sugar metabolism (Ishimaru et al., 2013; Tang et al., 2009). The investigation of candidate genes in this project suggests that genetic control of this pathway may be involved at the small effect QTL found in NUE-CROPs on chromosome 1H.

This study demonstrates that the strong relationship between yield components is also observed at the genetic level. The mechanisms of compensation between yield components and the GxE interactions for yield have been associated to QTL for yield suggesting that the identification of key traits are necessary to understand the physiological mechanisms in play and the genes involved (Slafer, 2003). Physiological models that attempt to understand yield analyse the yield architecture based on the

generation of yield components through the stages of plant development. The number of grain m^{-2} is initiated in the pre-anthesis phase of development and made of tiller m^{-2} and grains per ear (Slafer, 2003).The positive correlation of grain m^{-2} to yield and its role in establishing the yield potential is well recognized. During pre-anthesis period, the crop phenology impact on the control of stem elongation and plant growth at 20-30 days before anthesis and is critical to determine the yield potential set by grain m^{-2} (Borràs-Gelonch et al., 2011). Indeed, phenology genes have a major influence on yield performance through plant adaptation to its environment in both barley and wheat, (Comadran et al., 2011b; Cuesta-Marcos et al., 2009; Foulkes et al., 2004). It was also confirmed by the key role of phenology controlling genes like HvCEN and *Ppd-H1* in this study and their pleiotropic effects at QTL clusters. The alteration of phenology during the phases of yield potential establishment can modify the survival rate of tillers and florets that define grain m^{-2} and the timing of growth stages leading to seasonal adaptation as observed in the NIL experiments. The trait of grain m^{-2} was not investigated in this study and it may be used to complement and accommodate the negative correlations and give additional insight on QTL clusters and dissecting the genetic interactions responsible for the architecture of yield in barley. In physiological models, TGW is determined during post-anthesis and appears to be determined with little relationship to the components determining the number of grains m^{-2}. This suggest that the negative relationship it has with grain number may not be due to feedback processes (Slafer, 2003). Other key traits such as the ratios of source and sink, the resource economy within the plant will affect the competition of grains for assimilate and individual grain weight (Bingham et al., 2007a; Bingham et al., 2007b; Reynolds et al., 2011). Therefore, the analysis of grains m^{-2} as an additional yield component would give a better understanding of the underlying effects at a QTL cluster and their role in phenotypic plasticity observed through relationships and pleiotropic effects on other traits such as TGW.

The final interpretation of the QTL cluster relies mainly on a validation of candidate genes and candidate gene effects and therefore any simplification of the process involved in yield establishment should be considered. This project is a substantial step towards understanding the yield architecture of two-row winter barley and provides marker trait associations for most of the phenotypes that can be used to optimise allelic combinations in improved crops. However it brings into question the potential application of MAS for a complex trait independently of other traits such as phenology. Theoretically, a constitutive QTL effect independent from the environmental conditions

could offer consistent genetic progress however the reality implies that a more in-depth analysis of the plant physiology, epistatic interactions and gene × environment interactions may be necessary before implementation of MAS for complex traits (Romagosa et al., 1999). Genomic selection may also offer alternative approach to take into consideration minor and pleiotropic effect as well as gene × environment interactions in a commercial breeding context (Jannink et al., 2010).

6.4 Insight on novel candidate genes

Gene discovery and gene cloning in barley has mainly focused on major genes of highly heritable traits involved in the control of ear row number, response to photoperiod and monogenetic traits (Cockram et al., 2010; Ramsay et al., 2011). By definition, a QTL maps on a chromosome position where a sequence polymorphism causes the variation in a trait. In this project, the QTL results have confirmed the critical role of phenology genes in environmental adaptation and subsequent variation of yield and yield components. To take a step further to mapping, an attempt was made to identify potential candidate genes underpinning interesting QTL regions on 2HS, 4HL, 5HS and 5HL (4.3.3) with putative genetic control of yield and yield components. These could give a real advantage to breeders in their goal to achieve consistent progress, especially if candidate genes could be diagnosed with a sequence polymorphism responsible for the trait variation. This approach was aided by the genomic collinearity between grass species. Comparative genomics has enabled the identification and cloning of candidate genes in barley and gives an insight into the gene function and the possible alteration effects of a sequence polymorphism (Comadran et al., 2012; Koppolu et al., 2013; Ramsay et al., 2011). The concept of the ‘genome zipper’ between barley and rice (Mayer et al., 2011) was used in an attempt to anchor the significant barley SNPs on rice chromosomes and bracket rice genes as potential homologous gene for candidates genes in barley. The number of homologous rice genes bracketed by the SNPs varied depending on the QTL and chromosome position. One reason could be the different gene density observed along chromosomes of grass species (Feuillet and Keller, 2002). Also the position of the barley QTL on the chromosome will affect the ratio between physical and genetic maps. In general, the gene models contained in the rice chromosome segments could not be associated with certainty to a trait variation and a greater understanding of gene function is necessary to further refine and use these

results. However, some more plausible genes were proposed as candidates for a subset of QTL. A ras-related protein known for its implication in the control of meristem growth (Kamada et al., 1992; McCormick, 1995) was proposed as the most convincing gene candidate for the tillering locus on 4HL. The analysis of the 5HS mildew QTL homed in on a rice gene cluster of seven thaumatin protein genes (LOC_Os12g43410) which have been shown to be involved in the control of development and resistance to mildew infections (Tattersall et al., 1997; Xing et al., 2008) although comparison of disease resistance genes between species is complicated by the speed and complexity of their evolution (Meyers et al., 2005). Keeping in mind that differences in copy number can be a source of polymorphism, the results indicate that the 5HS QTL could be an ideal candidate gene to investigate and may confirm similar observations of a resistance locus in barley (Aghnoum et al., 2009; Comadran et al., 2009)

It is worth mentioning that the candidate genes proposed at the QTL targets are part of a non-exhaustive list of homologous rice genes. First, the delimitation of the homologous rice chromosome segment used only the best SNP of GWAS for each of the QTL rather than the whole QTL support interval. This approach therefore applies a high level of stringency to reduce the size of the homologous chromosome region investigated and may miss the causal gene. Secondly, for each QTL investigated, the list of homologous genes contains other potentially interesting candidates including transcription factors and gene coding for proteins with functional domains. Finally, although there is a strong homology between the barley and rice genomes (Mayer et al., 2011), one should bear in mind the probable cases of micro rearrangements of gene order, insertions and deletions but also differences in gene functions between the two species. In addition the study did not investigate the genome of Brachypodium distachyon which also a well conserved synteny which cereal species (Mayer et al., 2012; Mayer et al., 2011).Taken together, additional research into gene function and a wider screening including other cereal species should provide more opportunities to find and validate candidates genes.

Interestingly, the barley-rice genome zipper and the pattern of marker traits associations gave additional perspective to the GWAS results and the understanding of QTL clusters. In the case of the clusters located over the *Ppd-H1* and HvCEN candidate genes, the marker trait associations suggested different haplotype signatures for different traits. At *Ppd-H1*, the results of associations showed that approximately 20 rice gene models were separating the highly significant TGW SNP to the diagnostic SNP of *Ppd-H1*

(homologous to LOC_Os07g49460). Therefore either *Ppd-H1* may contain additional alleles that do not affect heading date but only TGW or closely linked genes are controlling TGW (e.g glycosyltransferase protein, LOC_Os07g49370). Similar dissociation of TGW and heading date was found in a cluster overlapping the HvCEN gene. Although the QTL overlapped at the centromere, the haplotype signatures suggested that different genetic factors (or genes) rather than different alleles at the same gene are associated with the QTL. These results illustrate that despite being mapped in the same QTL cluster, an increased genetic resolution can help to separate out the different genetic controls. This resolution depends essentially on the composition of the SNP array used in genotyping the material and the alleles segregating in the population genotyped. For important QTL targets, both the rice gene models and barley sequence information (Mayer et al., 2012) can be used to develop additional markers and increase marker density in the region to screen the panel of varieties and increase the resolution of GWAS.

6.5 QTL validation

Validation of identified QTL is an essential step to confirm their potential for MAS, especially for complex quantitative traits such as yield and yield components (Collard and Mackill, 2008). Tuinstra et al., (1997) demonstrated that inbred lines of advanced generations could be used as founder lines to develop HIF made of NIL segregating for a chromosome segment matching a QTL position. This method was implemented in barley to validate a deoxynivalenol QTL previously mapped in a barley GWAS (Navara and Smith, 2013). Here we implemented the HIF approach in winter barley using breeding material at F5 generation and later which ideally shared alleles with parents used in a QTL mapping experiment in which target loci were identified (Chapter 2, Table 2.9). This approach has the advantage to validate the effects of alleles that are also undergoing selection in parallel breeding activities and can potentially become parental lines (Pumphrey et al., 2007). A good insight into the genetic background and the level of homozygosity was obtained by using a 384 SNP array. In addition, the genotyping platform revealed substantial residual heterozygous regions, even at the F5 generation, which would not have been identified if a smaller array of SNP was used. Despite additional cycles of self-fertilisation to generate a more homozygous genetic background, some NIL pairs had residual background segregation at two or more loci. It

is therefore possible that those residual segregating loci contain potential sources of confounding effects to the targeted QTL when being tested (Pumphrey et al., 2007). More accurate pairs of NILs can be achieved by additional cycles of self-fertilisation and by increasing the number of the progeny from the HIF founders (Pumphrey et al., 2007) while maintaining the use of good marker coverage of the genome. On the other hand, this extra cycle will increase the length of the validation process delaying the implementation of the results in the breeding programme. From a total of 14 initial HIF founder lines heterozygous for QTL targets that included the major 2HS QTL for TGW and tillering (Chapter 2, Table 2.6), a subset of seven HIF families were identified for a replicated field trial experiment in 2013. The results showed that a three year period was necessary to characterise HIF founders, multiply NILs and carry out replicated yield plot experiments. The process could be reduced to two years for replicated mini-plot experiments that would be sufficient for validating disease resistance QTL although it may compromise the homozygosity of the background. However, an additional year of testing should confirm the robustness of the effect across environments. Our results agree with Pumphrey et al, (2007) and show that NILs and HIF approach is attractive to breeders seeking to simultaneously utilise and validate the effects of alleles present in advanced breeding material.

The traditional QTL validation by NILs often uses a restricted gene pool and backcross scheme of alleles in a well-defined genetic background which is often the other parent in the QTL mapping experiment (Kandemir et al., 2000; Yun et al., 2006). One aspect of this study was the use of diverse parents of the HIF founders compared to the parents used for the biparental QTL mapping. Therefore, the haplotypes and alleles tested in the NILs could differ from the expected results of the bi-parental mapping population. On one hand this gives the opportunity to test for effects of different alleles and haplotypes within HIF which can correspond to the allelic diversity in a breeding germplasm. On the other hand the multiples HIF are highly structured genetically which makes the alleles of individual SNP difficult to compare across HIF and as it would reduce the statistical power of the NILs comparisons (Tuinstra et al., 1997). Indeed it is not appropriate to compare individual SNP alleles across HIF due to strong effect attributable to drastically different genetic backgrounds. Consequently, the experiment was designed to minimise environmental variation between NILs within HIF and the statistical analysis focused on comparisons between contrasting haplotypes of NILs within HIF in a three step process. First, the whole experiment was analysed to estimate

error variance; secondly the significance of NILs contrasts were tested within HIF families using the error variance from the whole trial and finally the segregation of genetic factors between NILs within HIF was modelled using the markers to define haplotype components.

Overall, the confirmation of the QTL effects in the replicated field trial was moderate with relatively low correspondence between the significant NILs differences in the two years of phenotyping. This may primarily be due to different growing conditions between untreated miniplots in 2012 and treated yield plots in 2013 which may account for a large proportion of GxE interactions in the phenotypic variation. The HIF A-5 exemplified both the potential of the NILs to validate QTL effects but also the limit of this experiment in resolution and power. The NIL pair contrasting for anthocyanin pigmentation (component 4) and the literature information on the gene position (Cockram et al., 2010), validated polymorphism at component 4. However, the reduced HIF size with a high number of remaining segregating regions (components) caused very limited resolution and led to over parameterised QTL effect modelling. Further resolution on the origin of the variation amongst the components can be achieved by reducing background heterozygosity and increasing the allele replication by the number of NILs in a HIF.

The positive results of HIF B4 confirmed the utility of the strategy undertaken by identifying the significant QTL effects expected from the GWAS of Chapter 3. Based on marker alleles at SNP A10287, optimal NIL contrasts could be made for the presence or absence of sterile spikelets at the *vrs1* gene (Figure 5.3). Other contrasts were significant for heading date at components matching the locations of both *Ppd-H1* and HvCEN. More importantly, the family confirmed the GWA results and the significant effect on HvCEN locus on height and number of grains per ear. These results strongly support the presence of pleiotropic effects of the polymorphism and the presence of a QTL hotspot with those traits at the centromere of 2H. It is also in agreement with previous QTL mapping results that identified a major effect on 2H centromere in spring germplasm (Comadran et al., 2011b; Pasam et al., 2012). The NILs comparison brings further evidence for an important polymorphism in 2H centromere responsible for significant variation in grain number, stem extension and heading date and yield. In fact, the haplotype associated with late heading at the 2H centromere significantly reduced yield despite increasing the number of grains. Most likely, the later heading resulted in a sub-optimal grain filling period in the 2013 season, which was reflected by

a lower TGW at harvest. Because of its association with heading date, HvCEN is a highly probable candidate genes for that locus (Comadran et al., 2012). Moreover the results obtained from the NILs suggest that the gene activity is initiated in the early stages of plant development and is directly or indirectly involved in pathways of meristem elongation (6.3). The delays observed in plant development have adirect effect on plant adaptation to its environment hence its yield performance.

6.6 Project outlook

The results produced in this project have strengthened our understanding of the genetic architecture of complex traits and suggested potential novel research.

The mapping material used and generated is available to breeders to exploit allelic diversity, and carry out mapping and validation experiments. This can complement the different sources of allelic diversity found in genetically distant barley germplasm have been identified and can be targeted for introgressions in elite germplasm (Ellis et al., 2000; Wang et al., 2010). However, the introgressions from un-adapted sources of diversity require subsequent cycles of selection to breed out negative linkage drag and retrieve the original plant fitness and adaptation. Therefore the material used in that study is more relevant to European breeders than distant diverse material. This study shows that QTL for quantitative traits and disease resistance are also present and segregating in recent and adapted European germplasm. These ought to be associated with near or fully diagnostic markers and used in efficient MAS routines. Further mapping studies using the NUE-CROPS winter barley panel in a non-fungicide treated experiment would add valuable information to the dataset created by the project. For example it could confirm the mildew QTL on 5HS as disease resistance QTL present a strong interest in breeding. In addition, an increase in both the size of the mapping panel marker coverage can be envisaged by adding in more recent varieties in order to reduce LD and increase mapping resolution. This project confirms that any future mapping of especially yield-related traits should collect a range of phenotypic data in order to capture trait correlations which will help in interpreting QTL clusters.

A direct follow up to this project's results should aim at fine mapping and validation of candidate genes. More priority should be given to targets of high potential such as HvCEN candidate gene on 2H centromere, the tillering QTL on 4HL and a mildew

resistance QTL on 5HS. In addition to the available NILs and mapping population, the design of novel fine mapping populations and validation material can be envisaged. The NILs generated can be used to test the combinations of earliness alleles at both *Ppd-H1* and HvCEN loci while additional crosses between these NILs would generate recombination at the targeted loci. The variety panels used for GWAS also presents a source for other relevant alleles and haplotypes to design fine mapping and validation tools such as MAGIC and NAM populations (Buckler et al., 2009; Huang et al., 2012). Furthermore, comparative genomics complemented by the barley genome sequence information can be used to identify and increase marker density around gene candidates, while bulk segregant analysis in F2 populations could be an alternative route to identify diagnostic markers (Mayer et al., 2012; Ramirez-Gonzalez et al., 2014).

As QTL hotspots suggest strong pleiotropic effects, the prospect of increased MAS use in the breeding pipeline will certainly depend on the understanding of the roles of candidate genes and their impact of alleles in the multiple physiological pathways they are involved in. Based on other species, this project proposes a putative role of the candidate gene HvCEN in meristem development affecting development time of the ear and plant and thus increasing organ size and delayed heading date.

Although the GWAS provided greater insight on the effect across traits at the *Ppd-H1* locus and the 2HS TGW and tillering QTL identified in the S×R mapping population, the results obtained were insufficient to draw a definite conclusion on the origin of the effects. Indeed, the *Ppd-H1* SNP alleles associated with earliness in GWAS did not correspond to the significant SNP for TGW and the contrasting NILs pairs did not confirm the effect either. These results suggested the presence of a phenotypic variation under a different genetic control. Preliminary research was initiated to identify winter barley haplotypes by sequencing a segment of the *Ppd-H1* gene. A thorough analysis of the diversity of *Ppd-H1* haplotypes in winter barley panels of NUE CROPS and AGOUEB and NILs of HIF B4 and A-2 can also be envisaged to discriminate between significant alleles for both traits at that locus. Also, a simple marker assisted breeding strategy using Saffron, Retriever or related DH lines can be envisaged in a range of elite germplasm of varying genetic backgrounds.

Last but not least, this barley project established a valuable resource to be used in wheat (*Triticum aestivum* L.) research. While significant progress has been made in

sequencing the wheat genome (IWGSC, 2014), barley has remained the European crop species of choice for genetic research in small grain cereals. The diploid genome of barley can be exploited for genome collinearity with the hexaploid genome of wheat. For example, Distelfeld et al., (2008) showed that polymorphisms at orthologous genes relates to similar phenotypic variation in grain proteins in both species. In addition, both wheat and barley follow similar breeding strategies for yield adaptability to the environment, better resource use efficiency, resistance to biotic and abiotic stresses, as well as quality traits. This makes the barley QTL and candidate genes presented in this project very promising targets to be followed up in wheat. SNP can be used as anchors on the wheat genome to zoom in putative regions where the effects are expected with higher probabilities.

6.7 Project main findings

This project aimed at increasing the understanding of genetic architecture of yield and yield components in barley for potential implementation in commercial breeding. It was designed to contribute directly to breeding by using relevant QTL mapping material and *in situ* validation approaches as outlined in the research objectives in Chapter 1 paragraph 1.5. The major research findings for this project are:

- The QTL mapping experiment presented in Chapter 2 produced an in-depth understanding of the genetic architecture of yield, yield components and other agronomic traits in a cross between elite commercial varieties.
  - QTL for yield and yield components were mapped on five chromosomes. QTL clusterings were in agreement with the phenotypic correlations.
  - A major QTL with opposite effects for TGW and tillering was found on chromosome 2HS. At this locus, the gene *Ppd-H1* involved in the control of photoperiod was advanced as a candidate gene for its effects phenology and plant adaptation.
  - The QTL results from the 2012 untreated experiment conditions confirmed a number of significant loci for yield components found in the 2009 experiment, especially the TGW effects on 2HS. The genotype by environment CIM identified constitutive QTL for most evaluated traits.
  - The disease resistance QTL for brown rust and mildew found on chromosomes 3H, 4H and 5H present an opportunity for rapid implementation of MAS approach in related elite material.
  - A total of 23 genetic factors composed from either QTL clusters or single QTL were identified. They constitute potential targets for QTL validation and MAS approaches as well as chromosome segments of interest in comparative QTL studies.

- The GWAS studies of NUE-CROPS and AGOUEB illustrated the extent of the genetic variation for complex agronomic traits in European two-row winter barley and available to breeders in adapted and readily usable germplasm (Chapter 3).
  - Based on genetic markers information, the population of European two-row winter barley showed no obvious stratification and could be used effectively in GWAS using a kinship matrix.

- The mapping results supported the presence of known genes involved in the control of phenology (*Ppd-H1*, *eam6*) and responsible for the variation of DUS traits (*Ant-2*, *vrs1*).
- In the NUE-CROPS GWAS, a total of nine QTL were mapped for yield. TGW appeared to be the yield component with a complex genetic architecture attributable to nine significant QTL. A few QTL were associated with increasing effects of 7% or more over the mean. Tillering was the second most complex yield component based on QTL number (up to five QTL depending on the method of phenotyping). The grains per ear appeared to be mainly controlled by two loci, on chromosome 2H and 6H.
- The major QTL for grains per ear in two-row winter barley was found in the centromeric region of 2H with an increasing allele effect of three grains and a LOD score of 7.2. This QTL is located at the HvCEN candidate gene locus, a gene involved in the control of phenology in barley.
- A consistent QTL found for all tillering phenotypes was mapped on chromosome 4HL in the vicinity of the vernalization gene *vrn-H2*.
- In AGOUEB, six QTL were found for yield and five for TGW. The strongest QTL for TGW was found on the short arm of chromosome 5H. A remarkable QTL for mildew resistance also identified in the same vicinity on 5HS. Other QTL for brown rust and net blotch resistance were mapped and are of potential interest for breeding.

▪ The integration of QTL results from diverse mapping studies gives additional insight into the genetic architecture of quantitative traits, especially the range of effects of the genetic factors. This approach enables breeders to characterise and identify chromosome regions as breeding targets for future MAS and fine mapping strategies (Chapter 4).

- The QTL clusters within the individual mapping studies were more frequent than across studies and better correspondence was found between NUE-CROPS and AGOUEB than with the DH population mapping. Nevertheless, the co-location of TGW and mildew resistance between GWAS and the bi-parental mapping study confirmed the benefit of comparing mapping results from different studies to identify consistent effects.
- The QTL clusters were composed of traits known to be correlated phenotypically. Most of the QTL for yield were found in clusters with derived traits measuring the

plant nitrogen economy such as NUE and GrainN. For clusters containing at least two yield components, the alleles associated with increasing effects were generally in opposition.

- The genes involved in the control of phenology were found in association with QTL clusters. The HvCEN gene was identified at the main QTL cluster on 2H with potential pleiotropic effects on 13 other traits, including yield and yield components. Vernalization genes *vrn-H1* on 5H and *vrn-H3* on 7H were also associated with clusters of yield and yield components.
- A detailed analysis of significance patterns of marker-trait association at the QTL clusters confirmed the presence of different genetic control between some traits. The TGW variation at the *Ppd-H1* locus could not be directly related to the alleles responsible for variation in heading date. Similarly, a differential genetic control of TGW and grain per ear was suggested for at the HvCEN locus.
- The comparative genomics with rice at the best barley SNP of QTL clusters could identify relevant candidate genes for investigation. On 5HS, a cluster of thaumatin-like genes was proposed for candidate genes associated with the mildew resistance QTL. The implication of HvCEN and Ras-related genes in the control of meristem development was advanced for the grain per ear QTL on 2H and the tillering QTL on 4HL respectively.

▪ The validation of QTL for agronomic traits can be carried out alongside breeding cycles to validate alleles under selection in advanced material (Chapter 5)

- The residual heterozygosity of advanced barley breeding lines at the F5 generation can be exploited to develop HIF of NILs and validate agronomic traits such as yield in field conditions.
- The NIL development protocol requires a genotyping platform offering good marker coverage to characterise the haplotypes and all residual background segregation. Alternatively, sufficient replication of the NILs within each HIF will increase the statistical power of the validation. After two cycles of self-fertilisation, the residual heterozygous loci between NILs detected by the whole genome genotyping ranged from two to 17 creating confounding segregating factors with the targeted QTL.
- Significant heading date effects were detected for alleles at the *Ppd-H1* and HvCEN loci. This confirmed the QTL results from GWAS. The additive effects of early

alleles at both loci reduced ear emergence by five days on average. These alleles are have a major influence in controlling winter barley seasonal adaptation.

- The major effect on grains per ear was validated with the HIF B-4. The increasing grain number allele produced on average one more grain per ear. The increasing allele was also significantly associated with a delay in heading date and an increase in plant height.
- Three HIF identified a segregating segment in association with TGW variation. Family B-4 confirmed that haplotypes at the early allele at the *Ppd-H1* locus was associated with an increase of 1.7 g on TGW in the 2013 growing season.

Conclusion

This project has contributed to the development of novel scientific resources and knowledge that have increased our understanding of the genetic architecture of agronomic traits in UK winter barley. The particular emphasis given to the yield and yield components QTL has enabled the characterisation of loci of interest for breeders who will be able to select for favourable alleles and haplotypes. Already, the genetic SNP markers at some of the QTL loci are used to characterise the genetic identity of elite germplasm in order to design appropriate combinations in novel crosses. In addition, the QTL obtained by screening for diversity of alleles in European germplasm in NUE-CROPS and AGOUEB GWAS confirmed the substantial prospects for improving agronomic traits using winter barley closely related to current UK elite varieties and thus more desirable for the breeders to cross with.

However, this study also highlights the challenges faced by breeders using genetic markers to select for yield and yield components. While most attempts to explain complex traits are based on additive statistical models, it needs to be borne in mind that epistatic and environment interactions also contribute to the plasticity of the traits. When analysed individually, the mapping results showed that a large number of loci can be considered for yield improvement but this approach may lead to misinterpretation of the loci effects. Indeed, the complexity of the traits, relationship between yield components and pleiotropic effects observed at QTL clusters illustrate the real challenge of characterising and defining what would be a beneficial allele to select for. Amongst the different trait associations found in the clusters, this project identified key role of loci controlling phenology, in particular the HvCEN candidate gene at a major QTL cluster located at the 2H centromere. At that locus, the allele for earliness was associated with effects of opposite direction for yield and yield components, especially a major reduction effect on grains per ear. Thus selecting for increased grains per ear would also affect phenology which may modify the adaptation to a given environment and impact negatively on yield. In cases of QTL clusters independent from phenology, the direction of the effects for the yield components QTL also showed opposite allele effect. This was illustrated by co-localised TGW and tillering QTL on 2H and on 5HS suggesting that any attempt to improve one trait via any of these QTL would be done at the expense of the other trait. Such opposite effects on important selection traits may be

the main reason behind the delays in implementation of MAS approaches for quantitative traits other than disease resistance major phenology controlling genes in breeding programs. The resolution of the origin of pleiotropic effects at QTL clusters may also be achieved by a better description of the haplotypes and alleles associated with each QTL effects. Indeed, the GWAS results have shown that the patterns of significant SNP associations revealed different alleles involved in the genetic control of traits mapped at the same QTL clusters. The fine mapping strategies using increased marker density from exploiting genome sequence information, genome co-linearity and described functions of the gene models of sequenced grass species will enable breeders to identify the diagnostic alleles for MAS.

Because they have a rather strong influence on yield, the genetic factors involved in fine tuning plant adaptation have been indirectly maintained by breeders as a consequence of environmental variation. From a MAS perspective, these factors represent limited interest for the long term improvement of crops, though they need to be maintained and highlight the need to be differentiated from the genetic factors controlling agronomic traits of interest independently of phenology. This study showed that these genetic factors are less frequent and consist of effects of lower magnitude therefore harder to detect compared to effects of major phenology controlling factors. However they are ideal targets for MAS and should generate consistent progress as their effect is perceived independently of environmental variation with no interference of genes for adaptation. The disease resistance QTL typically enter into this category of genetic factors. The brown rust QTL on 2H and the mildew resistance QTL on 5H found in AGOUEB could be accessible MAS targets. Similarly, the genetic factors for yield and yield components should only be suggested for MAS if they are dissociated from phenology factors. Once the genetic architecture of the traits in a crop is understood, the QTL mapping approaches can be designed to investigate the effects of specific loci. For example, the QTL mapping with a DH population offers the possibility to exclude the segregation of undesired alleles at major genes and to focus on genetic factors with lower effects for any traits which interact with an adaptive locus. In GWAS, comprehensive trait phenotyping can also identify relevant target loci, even with lower effects. The tillering and yield increasing locus identified on 4HL presents a concrete opportunity to improve both agronomic traits since no association with phenology at this locus was detected. The project results provide breeders with a range of other genetic targets associated with significant markers that can also be used to screen for

specific haplotypes in elite European germplasm. Naturally, these optimistic prospects should not dismiss the validation of the effects in diverse, yet relevant, genetic background which can be tackled by the development of NILs and HIF. This attractive approach for breeders to validate QTL effects in a timely manner can be further refined in order to evaluate allele comparisons with more precision and with sufficient statistical power. Eventually, the validated positive effects will be routinely selected to sustain breeding progress in future barley varieties.

References

Abeledo, L.G., D.F. Calderini, and G.A. Slafer. 2003. Genetic improvement of yield responsiveness to nitrogen fertilization and its physiological determinants in barley. Euphytica 133:291-298.

Abeledo, L.G., D.F. Calderini, and G.A. Slafer. 2008. Nitrogen economy in old and modern malting barleys. Field Crops Research 106:171-178.

Aghnoum, R., T.C. Marcel, A. Johrde, N. Pecchioni, P. Schweizer, and R.E. Niks. 2009. Basal Host Resistance of Barley to Powdery Mildew: Connecting Quantitative Trait Loci and Candidate Genes. Molecular Plant-Microbe Interactions 23:91-102.

Alqudah, A.M., R. Sharma, R.K. Pasam, A. Graner, B. Kilian, and T. Schnurbusch. 2014. Genetic Dissection of Photoperiod Response Based on GWAS of Pre-Anthesis Phase Duration in Spring Barley. PLoS ONE 9:e113120.

Amaya, I., O.J. Ratcliffe, and D.J. Bradley. 1999. Expression of CENTRORADIALIS (CEN) and CEN-like genes in tobacco reveals a conserved mechanism controlling phase change in diverse species. The Plant Cell 11:1405-1418.

Asíns, M.J. 2002. Present and future of quantitative trait locus analysis in plant breeding. Plant Breeding 121:281-291.

Atkinson, N.J., and P.E. Urwin. 2012. The interaction of plant biotic and abiotic stresses: from genes to the field. Journal of Experimental Botany 63:3523-3543.

Ayoub, M., S. Symons, M. Edney, and D. Mather. 2002. QTLs affecting kernel size and shape in a two-rowed by six-rowed barley cross. Theoretical and Applied Genetics 105:237-247.

Babb, and Muehlbauer. 2003. Genetic and morphological characterization of the barley *uniculm2* (*cul2*) mutant. Theoretical and Applied Genetics 106:846-857.

Backes, G., A. Graner, B. Foroughi Wehr, G. Fischbeck, G. Wenzel, and A. Jahoor. 1995. Localization of quantitative trait loci (QTL) for agronomic important characters by the use of a RFLP map in barley (Hordeum vulgare L.). Theoretical and Applied Genetics 90:294-302.

Badr, A., K. M, R. Sch, H.E. Rabey, S. Effgen, H.H. Ibrahim, C. Pozzi, W. Rohde, and F. Salamini. 2000. On the Origin and Domestication History of Barley (Hordeum vulgare). Molecular Biology and Evolution 17:499-510.

Baroja-Fernández, E., F.J. Muñoz, T. Saikusa, M. Rodríguez-López, T. Akazawa, and J. Pozueta-Romero. 2003. Sucrose Synthase Catalyzes the de novo Production of ADPglucose Linked to Starch Biosynthesis in Heterotrophic Tissues of Plants. Plant and cell physiology 44:500-509.

Bastergue, P., A. Cauderon, L. Jestin, J. Legouis, and M. Madre. 2006. L'orge, p. 495-515, *In* V. F. Doré C, ed. Histoire et amélioration de cinquante plantes cultivées. INRA, Paris.

Beales, J., A. Turner, S. Griffiths, J. Snape, and D. Laurie. 2007. A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics 115:721-733.

Beatty, P.H., Y. Anbessa, P. Juskiw, R.T. Carroll, J. Wang, and A.G. Good. 2010. Nitrogen use efficiencies of spring barley grown under varying nitrogen conditions in the field and growth chamber. Annals of Botany 105:1171-1182.

Bennetzen, J., and M. Chen. 2008. Grass Genomic Synteny Illuminates Plant Genome Function and Evolution. Rice 1:109-118.

Berger, G., S. Liu, M. Hall, W. Brooks, S. Chao, G. Muehlbauer, B.K. Baik, B. Steffenson, and C. Griffey. 2012. Marker-trait associations in Virginia Tech

winter barley identified using genome-wide mapping. Theoretical and Applied Genetics 126:693-710.
Bernardo, R. 2008. Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. Crop Science 48:1649-1664.
Bernardo, R., and J. Yu. 2007. Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Science 47:1082-1090.
Bertholdsson, N.O., and A. Kolodinska Brantestam. 2009. A century of Nordic barley breeding - Effects on early vigour root and shoot growth, straw length, harvest index and grain weight. European Journal of Agronomy 30:266-274.
Bezant, J., D. Laurie, N. Pratchett, J. Chojecki, and M. Kearsey. 1997. Mapping QTL controlling yield and yield components in a spring barley (Hordeum vulgare L.) cross using marker regression. Molecular Breeding 3:29-38.
Bingham, I.J., J. Blake, M.J. Foulkes, and J. Spink. 2007a. Is barley yield in the UK sink limited?: I. Post-anthesis radiation interception, radiation-use efficiency and source-sink balance. Field Crops Research 101:198-211.
Bingham, I.J., J. Blake, M.J. Foulkes, and J. Spink. 2007b. Is barley yield in the UK sink limited?: II. Factors affecting potential grain size. Field Crops Research 101:212-220.
Bocianowski, J. 2014. Estimation of epistasis in double haploid barley poulations considering interactions between all possible marker pairs. Euphytica 196:105-115.
Boden, S.A., C. Cavanagh, B.R. Cullis, K. Ramm, J. Greenwood, E. Jean Finnegan, B. Trevaskis, and S.M. Swain. 2015. *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. Nature Plants 1:4016.
Borràs-Gelonch, G., G.A. Slafer, A.M. Casas, F. van Eeuwijk, and I. Romagosa. 2011. Genetic control of pre-heading phases and other traits related to development in a double-haploid barley (*Hordeum vulgare* L.) population. Field Crops Research 119:36-47.
Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.
Brenchley, R., M. Spannagl, M. Pfeifer, G.L.A. Barker, R. D/'Amore, A.M. Allen, N. McKenzie, M. Kramer, A. Kerhornou, D. Bolser, S. Kay, D. Waite, M. Trick, I. Bancroft, Y. Gu, N. Huo, M.-C. Luo, S. Sehgal, B. Gill, S. Kianian, O. Anderson, P. Kersey, J. Dvorak, W.R. McCombie, A. Hall, K.F.X. Mayer, K.J. Edwards, M.W. Bevan, and N. Hall. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705-710.
Breseghello, F., and M.E. Sorrells. 2006. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars 172:1165-1177.
Bryngelsson, T., and B. Green. 1989. Characterization of a pathogenesis-related, thaumatin-like protein isolated from barley challenged with an incompatible race of mildew. Physiological and molecular plant pathology 35:45-52.
Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J.C. Glaubitz, M.M. Goodman, C. Harjes, K. Guill, D.E. Kroon, S. Larsson, N.K. Lepak, H. Li, S.E. Mitchell, G. Pressoir, J.A. Peiffer, M.O. Rosas, T.R. Rocheford, M.C. Romay, S. Romero, S. Salvo, H.S. Villeda, H. Sofia da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M.D. McMullen. 2009. The Genetic Architecture of Maize Flowering Time. Science 325:714-718.
Casao, M.C., E. Igartua, I. Karsai, J.M. Lasa, M.P. Gracia, and A.M. Casas. 2011. Expression analysis of vernalization and day-length response genes in barley

(*Hordeum vulgare* L.) indicates that *VRNH2* is a repressor of *PPDH2* (*HvFT3*) under long days. Journal of Experimental Botany 62:1939-1949.

Cheverud, J.M., and E.J. Routman. 1995. Epistasis and its contribution to genetic variance components. Genetics 139:1455-61.

Chloupek, O., B. Forster, and W. Thomas. 2006. The effect of semi-dwarf genes on root system size in field-grown barley. Theoretical and Applied Genetics 112:779-786.

Choi, D., J.H. Kim, and H. Kende. 2004. Whole Genome Analysis of the *OsGRF* Gene Family Encoding Plant-Specific Putative Transcription Activators in Rice (*Oryza sativa* L.). Plant and Cell Physiology 45:897-904.

Close, T., P. Bhat, S. Lonardi, Y. Wu, N. Rostoks, L. Ramsay, A. Druka, N. Stein, J. Svensson, S. Wanamaker, S. Bozdag, M. Roose, M. Moscou, S. Chao, R. Varshney, P. Szucs, K. Sato, P. Hayes, D. Matthews, A. Kleinhofs, G. Muehlbauer, J. DeYoung, D. Marshall, K. Madishetty, R. Fenton, P. Condamine, A. Graner, and R. Waugh. 2009. Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582.

Cockram, J., H. Jones, F.J. Leigh, D. O'Sullivan, W. Powell, D.A. Laurie, and A.J. Greenland. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. Journal of Experimental Botany 58:1231-1244.

Cockram, J., J. White, F. Leigh, V. Lea, E. Chiapparino, D. Laurie, I. Mackay, W. Powell, and D. O'Sullivan. 2008. Association mapping of partitioning loci in barley. BioMed Central Genetics 9:16.

Cockram, J., J. White, D.L. Zuluaga, D. Smith, J. Comadran, M. Macaulay, Z. Luoc, M.J. Kearsey, P. Wernerd, D. Harrap, C. Tapsell, H. Liu, P.E. Hedley, N. Stein, D. Schulte, B. Steuernagel, D.F. Marshall, W.T.B. Thomas, L. Ramsay, I. Mackay, D.J. Balding, T.A. Consortium, R. Waugh, and D.M. O'Sullivan. 2010. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. PNAS 107:21611-6.

Collard, B., M. Jahufer, J. Brouwer, and E. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196.

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society of London, Serie B: Biological Sciences 363:557-572.

Collins, N.C., F. Tardieu, and R. Tuberosa. 2008. Quantitative Trait Loci and Crop Performance under Abiotic Stress: Where Do We Stand? Plant Physiology 147:469-486.

Comadran, J., L. Ramsay, K. MacKenzie, P. Hayes, T. Close, G. Muehlbauer, N. Stein, and R. Waugh. 2011a. Patterns of polymorphism and linkage disequilibrium in cultivated barley. Theoretical and Applied Genetics 122:523-531.

Comadran, J., W. Thomas, F. van Eeuwijk, S. Ceccarelli, S. Grando, A. Stanca, N. Pecchioni, T. Akar, A. Al-Yassin, A. Benbelkacem, H. Ouabbou, J. Bort, I. Romagosa, C. Hackett, and J. Russell. 2009. Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. Theoretical and Applied Genetics 119:175-187.

Comadran, J., J.R. Russell, A. Booth, A. Pswarayi, S. Ceccarelli, S. Grando, A.M. Stanca, N. Pecchioni, T. Akar, A. Al-Yassin, A. Benbelkacem, H. Ouabbou, J. Bort, F.A. Eeuwijk, W.T.B. Thomas, and I. Romagosa. 2011b. Mixed model association scans of multi-environmental trial data reveal major loci controlling

yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. Theoretical and Applied Genetics 122:1363-1373.

Comadran, J., B. Kilian, J. Russell, L. Ramsay, N. Stein, M. Ganal, P. Shaw, M. Bayer, W. Thomas, D. Marshall, P. Hedley, A. Tondelli, N. Pecchioni, E. Francia, V. Korzun, A. Walther, and R. Waugh. 2012. Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 44:1388-1392.

Comadran, J., J. Russell, F. van Eeuwijk, S. Ceccarelli, S. Grando, M. Baum, A. Stanca, N. Pecchioni, A. Mastrangelo, T. Akar, A. Al-Yassin, A. Benbelkacem, W. Choumane, H. Ouabbou, R. Dahan, J. Bort, J.L. Araus, A. Pswarayi, I. Romagosa, C. Hackett, and W. Thomas. 2008. Mapping adaptation of barley to droughted environments. Euphytica 161:35-45.

Cooper, M., F.A. van Eeuwijk, G.L. Hammer, D.W. Podlich, and C. Messina. 2009. Modeling QTL for complex traits: detection and context for plant breeding. Current Opinion in Plant Biology 12:231-240.

Cremer, F., W.-E. Lonnig, H. Saedler, and P. Huijser. 2001. The Delayed Terminal Flower Phenotype Is Caused by a Conditional Mutation in the *CENTRORADIALIS* Gene of Snapdragon. Plant Physiology 126:1031-1041.

Cuesta-Marcos, A., A.M. Casas, P.M. Hayes, M.P. Gracia, J.M. Lasa, F. Ciudad, P. Codesal, J.L. Molina-Cano, and E. Igartua. 2009. Yield QTL affected by heading date in Mediterranean grown barley. Plant Breeding 128:46-53.

Dabbert, T., R. Okagaki, S. Cho, J. Boddu, and G. Muehlbauer. 2009. The genetics of barley low-tillering mutants: *absent lower laterals* (*als*). Theoretical and Applied Genetics 118:1351-1360.

Dabbert, T., R. Okagaki, S. Cho, S. Heinen, J. Boddu, and G. Muehlbauer. 2010. The genetics of barley low-tillering mutants: *low number of tillers-1* (*lnt1*). Theoretical and Applied Genetics 121:705-715.

Decousset, L., S. Griffiths, R.P. Dunford, N. Pratchett, and D.A. Laurie. 2000. Development of STS markers closely linked to the *Ppd-H1* photoperiod response gene of barley (*Hordeum vulgare* L.). Theoretical Applied Genetics 101:1202-1206.

DEFRA. 2010. Fertiliser Manual (RB209), *In* F. R. A. Department for Environment, (ed.), 8 ed. DEFRA.

Devlin, B., and K. Roeder. 1999. Genomic control for association studies. Biometrics 55:997-1004.

Distelfeld, A., A. Korol, J. Dubcovsky, C. Uauy, T. Blake, and T. Fahima. 2008. Colinearity between the barley grain protein content (GPC) QTL on chromosome arm 6HS and the wheat *Gpc-B1* region. Molecular Breeding 22:25-38.

Douchkov, D., A. Johrde, D. Nowara, A. Himmelbach, S. Lueck, R. Niks, and P. Schweizer. 2010. Convergent evidence for a role of WIR1 proteins during the interaction of barley with the powdery mildew fungus *Blumeria graminis*. Journal of Plant Physiology 168:209.

Doust, A.N. 2007. Grass architecture: genetic and environmental control of branching. Current Opinion in Plant Biology 10:21-25.

Dreiseitl, A., and K. Krianová. 2012. Powdery mildew resistance genes in spring barley cultivars registered in Slovakia from 2000 to 2010. Cereal Research Communications 40:494-501.

Dreiseitl, A., and G. Platz. 2012. Powdery mildew resistance genes in barley varieties grown in Australia. Crop and Pasture Science 63:997-1006.

Druka, A., J. Franckowiak, U. Lundqvist, N. Bonar, J. Alexander, J. Guzy-Wrobelska, L. Ramsay, I. Druka, I. Grant, M. Macaulay, V. Vendramin, F. Shahinnia, S.

Radovic, K. Houston, D. Harrap, L. Cardle, D. Marshall, M. Morgante, N. Stein, and R. Waugh. 2010. Exploiting induced variation to dissect quantitative traits in barley. Biochemical Society Transactions 38:683-688.

Dubcovsky, J., C. Chen, and L. Yan. 2005. Molecular characterization of the allelic variation at the *VRN-H2* vernalization locus in barley. Molecular Breeding 15:395-407.

Dunford, R.P., M. Yano, N. Kurata, T. Sasaki, G. Huestis, T. Rocheford, and D.A. Laurie. 2002. Comparative mapping of the barley *Ppd-H1* photoperiod response gene region, which lies close to a junction between two rice linkage segments. Genetics 161:825-834.

Ellis, R.P., B.P. Forster, D. Robinson, L.L. Handley, D.C. Gordon, J.R. Russell, and W. Powell. 2000. Wild barley: a source of genes for crop improvement in the 21st century? Journal of Experimental Botany 51:9-17.

Emebiri, L. 2013. QTL dissection of the loss of green colour during post-anthesis grain maturation in two-rowed barley. Theoretical and Applied Genetics 126:1873-1884.

Fan, C., Y. Xing, H. Mao, T. Lu, B. Han, C. Xu, X. Li, and Q. Zhang. 2006. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and Applied Genetics 112:1164-1171.

Faure, S., J. Higgins, A. Turner, and D.A. Laurie. 2007. The *FLOWERING LOCUS T*-Like Gene Family in Barley (*Hordeum vulgare*). Genetics 176:599-609.

Feuillet, C., and B. Keller. 2002. Comparative Genomics in the Grass Family: Molecular Characterization of Grass Genome Structure and Evolution. Annals of Botany 89:3-10.

Foucher, F., J. Morin, J. Courtiade, S. Cadioux, N. Ellis, M.J. Banfield, and C. Rameau. 2003. *DETERMINATE* and LATE FLOWERING are two *TERMINAL FLOWER1/CENTRORADIALIS* homologs that control two distinct phases of flowering initiation and development in pea. Plant Cell 15:2742-54.

Foulkes, M.J., R. Sylvester-Bradley, A.J. Worland, and J.W. Snape. 2004. Effects of a photoperiod-response gene *Ppd-D1* on yield potential and drought resistance in UK winter wheat. Euphytica 135:63-73.

Fufa, H., P.S. Baenziger, B.S. Beecher, R.A. Graybosch, K.M. Eskridge, and L.A. Nelson. 2005. Genetic improvement trends in agronomic performances and end-use quality characteristics among hard red winter wheat cultivars in Nebraska. Euphytica 144:187-198.

Furbank, R.T., and M. Tester. 2011. Phenomics--technologies to relieve the phenotyping bottleneck. Trends in Plant Sciences 16:635-44.

Gale, M.D., and K.M. Devos. 1998. Comparative genetics in the grasses. PNAS 95:1971-1974.

Gallais, A., and B. Hirel. 2004. An approach to the genetics of nitrogen use efficiency in maize. Journal of Experiment Botany 55:295-306.

Good, A.G., A.K. Shrawat, and D.G. Muench. 2004. Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? Trends in Plant Science 9:597-605.

Gregersen, P.L., P.B. Holm, and K. Krupinska. 2008. Leaf senescence and nutrient remobilisation in barley and wheat. Plant Biology 10:37-49.

Griffiths, S., R.P. Dunford, G. Coupland, and D.A. Laurie. 2003. The Evolution of CONSTANS-Like Gene Families in Barley, Rice, and Arabidopsis. Plant Physiology 131:1855-1867.

Gupta, P., S. Rustgi, and P. Kulwal. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Molecular Biology 57:461-485.

Hammer, Ø., D.A.T. Harper, and P.D. Ryan. 2001. Past: paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4:9 pp.

Hargreaves, C., P. Gregory, and A. Bengough. 2009. Measuring root traits in barley (*Hordeum vulgare* ssp. *vulgare* and ssp. *spontaneum*) seedlings using gel chambers, soil sacs and X-ray microtomography. Plant and Soil 316:285-297.

Haseneyer, G., S. Stracke, C. Paul, C. Einfeldt, A. Broda, H.P. Piepho, A. Graner, and H.H. Geiger. 2010. Population structure and phenotypic variation of a spring barley world collection set up for association studies. Plant Breeding 129:271-279.

Hayes, P.M., B.H. Liu, S.J. Knapp, F. Chen, B. Jones, T. Blake, J. Franckowiak, D. Rasmusson, M. Sorrells, S.E. Ullrich, D. Wesenberg, and A. Kleinhofs. 1993. Quantitative trait locus effects and environmental interaction in a sample of North American barley germ plasm. Theoretical and Applied Genetics 87:392-401.

Heisey, P.W., M.A. Lantican, and H.J. Dubin. 2002. Impacts of International Wheat Breeding Research in Developing Countries, 1966-97. Mexico D.F.:CIMMYT.

Hickey, L.T., W. Lawson, G.J. Platz, R.A. Fowler, V. Arief, M. Dieters, S. GermÃ¡n, S. Fletcher, R.F. Park, D. Singh, S. Pereyra, and J. Franckowiak. 2012. Mapping Quantitative Trait Loci for Partial Resistance to Powdery Mildew in an Australian Barley Population. Crop Science 52:1021-1032.

Hirel, B., J. Le Gouis, B. Ney, and A. Gallais. 2007. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. Journal of Experimental Botany 58:2369–2387.

Hirel, B., P. Bertin, I. Quilleré, W. Bourdoncle, C. Attagnant, C. Dellay, A. Gouy, S. Cadiou, C. Retailliau, M. Falque, and A. Gallais. 2001. Towards a Better Understanding of the Genetic and Physiological Basis for Nitrogen Use Efficiency in Maize. Plant Physiology 125:1258–1270.

Huang, B.E., A.W. George, K.L. Forrest, A. Kilian, M.J. Hayden, M.K. Morell, and C.R. Cavanagh. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnology Journal 10:826-839.

Hussien, A., E. Tavakol, D.S. Horner, M. Muñoz-Amatriaín, G.J. Muehlbauer, and L. Rossini. 2014. Genetics of Tillering in Rice and Barley, pp. -, Vol. 7.

IHGSC. 2004. The International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 431:931-945.

Ingvarsson, P.K., and N.R. Street. 2011. Association genetics of complex traits in plants. New Phytologist 189:909-922.

Inostroza, L., A. del Pozo, I. Matus, D. Castillo, P. Hayes, S. Machado, and A. Corey. 2009. Association mapping of plant height, yield, and yield stability in recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. Molecular Breeding 23:365-376.

Irish, E.E. 1997. Experimental Analysis of Tassel Development in the Maize Mutant *Tassel Seed* 6. Plant Physiology 114:817-825.

Ishimaru, K., N. Hirotsu, Y. Madoka, N. Murakami, N. Hara, H. Onodera, T. Kashiwagi, K. Ujiie, B.-i. Shimizu, A. Onishi, H. Miyagawa, and E. Katoh. 2013. Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. Nat Genetics 45:707-711.

IWGSC, T. 2014. The International Wheat Genome Sequencing Consortium : A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788.

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Briefing in functional genomics 9:166-177.

Jones, H., F.J. Leigh, I. Mackay, M.A. Bower, L.M.J. Smith, M.P. Charles, G. Jones, M.K. Jones, T.A. Brown, and W. Powell. 2008. Population-Based Resequencing Reveals That the Flowering Time Adaptation of Cultivated Barley Originated East of the Fertile Crescent. Molecular Biology and Evolution 25:2211-2219.

Jukanti, A.K., and A.M. Fischer. 2008. A high-grain protein content locus on barley (*Hordeum vulgare*) chromosome 6 is associated with increased flag leaf proteolysis and nitrogen remobilization. Physiologia Plantarum 132:426-439.

Kamada, I., S. Yamauchi, S. Youssefian, and H. Sano. 1992. Transgenic tobacco plants expressing *rgp1*, a gene encoding a *ras*-related GTP-binding protein from rice, show distinct morphological characteristics. The Plant Journal 2:799-807.

Kandemir, N., B.L. Jones, D.M. Wesenberg, S.E. Ullrich, and A. Kleinhofs. 2000. Marker-assisted analysis of three grain yield QTL in barley (*Hordeum vulgare* L.) using near isogenic lines. Molecular Breeding 6:157-167.

Kang, H.M., J.H. Sul, S.K. Service, N.A. Zaitlen, S.Y. Kong, N.B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. Nature Genetics 42:348-354.

Kant, S., Y.-M. Bi, and S.J. Rothstein. 2010. Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. Journal of Experiment Botany 6:1-11.

Kebrom, T.H., W. Spielmeyer, and E.J. Finnegan. 2013. Grasses provide new insights into regulation of shoot branching. Trends in plant science 18:41-48.

Kjaer, B., and J. Jensen. 1996. Quantitative trait loci for grain yield and yield components in a cross between a six-rowed and a two-rowed barley. Euphytica 90:39-48.

Komatsuda, and Mano. 2002. Molecular mapping of the intermedium spike-c (*int-c*) and non-brittle rachis 1 (*btr1*) loci in barley (*Hordeum vulgare* L.). TAG Theoretical and Applied Genetics 105:85-90.

Komatsuda, T., M. Pourkheirandish, C. He, P. Azhaguvel, H. Kanamori, D. Perovic, N. Stein, A. Graner, T. Wicker, A. Tagiri, U. Lundqvist, T. Fujimura, M. Matsuoka, T. Matsumoto, and M. Yano. 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. PNAS 104:1424–1429.

Kongprakhon, P., A. Cuesta-Marcos, P.M. Hayes, K.L. Richardson, P. Sirithunya, K. Sato, B. Steffenson, and T. Toojinda. 2009. Validation of rice blast resistance genes in barley using a QTL mapping population and near-isolines. Breeding Science 59:341-349.

Koppolu, R., N. Anwar, S. Sakuma, A. Tagiri, U. Lundqvist, M. Pourkheirandish, T. Rutten, C. Seiler, A. Himmelbach, R. Ariyadasa, H.M. Youssef, N. Stein, N. Sreenivasulu, T. Komatsuda, and T. Schnurbusch. 2013. *Six-rowed spike4 (Vrs4)* controls spikelet determinacy and row-type in barley. PNAS 110:13198-13203.

Kraakman, A.T.W., R.E. Niks, P.M.M.M. Van den Berg, P. Stam, and F.A. Van Eeuwijk. 2004. Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars. Genetics 168:435-446.

Kurth, J., R. Kolsch, V. Simons, and P. Schulze-Lefert. 2001. A high-resolution genetic map and a diagnostic RFLP marker for the *Mlg* resistance locus to powdery mildew in barley. Theoretical and Applied Genetics 102:53-60.

Lander, E.S., and D. Botstein. 1989. Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. Genetics 121:185-199.

Laurie, D.A. 1997. Comparative genetics of flowering time. Plant Molecular Biology 35:167-177.

Le Gouis, J., D. Béghin, E. Heumez, and P. Pluchard. 2000. Genetic differences for nitrogen uptake and nitrogen utilisation efficiencies in winter wheat. European Journal of Agronomy 12 163-173.

Le Gouis, J., O. Delebarre, D. Beghin, E. Heumez, and P. Pluchard. 1999. Nitrogen uptake and utilisation efficiency of two-row and six-row winter barley cultivars grown at two N levels. European Journal of Agronomy 10:73-79.

Li, J.Z., X.Q. Huang, F. Heinrichs, M.W. Ganal, and M.S. Röder. 2005. Analysis of QTLs for yield, yield components, and malting quality in a BC_3-DH population of spring barley. Theoretical and Applied Genetics 110:356-363.

Liu, T., Y. Zhang, W. Xue, C. Xu, X. Li, and Y. Xing. 2010. Comparison of quantitative trait loci for 1,000-grain weight and spikelets per panicle across three connected rice populations. Euphytica 175:383-394.

Loffler, M., C.-C. Schon, and T. Miedaner. 2009. Revealing the genetic architecture of FHB resistance in hexaploid wheat (*Triticum aestivum* L.) by QTL meta-analysis. Molecular Breeding 23:473-488.

Long, N., O. Dolstra, M. Malosetti, B. Kilian, A. Graner, R.F. Visser, and C.G. Linden. 2013. Association mapping of salt tolerance in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 126:2335-2351.

Lorenz, A.J., M.T. Hamblin, and J.-L. Jannink. 2010. Performance of Single Nucleotide Polymorphisms versus Haplotypes for Genome-Wide Association Analysis in Barley. PLoS ONE 5:e14079.

Lorieux, M. 2012. MapDisto: fast and efficient computation of genetic linkage maps. Molecular Breeding 30:1231-1235.

Mackay, I., A. Horwell, J. Garner, J. White, J. McKee, and H. Philpott. 2011. Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. Theoretical and Applied Genetics 122:225-38.

Mackay, T.F.C., E.A. Stone, and J.F. Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10:565-577.

Malosetti, M., J. Voltas, I. Romagosa, S.E. Ullrich, and F.A. van Eeuwijk. 2004. Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139-145.

Mariotti, F., T. D, and P.P. Mirand. 2008. Converting nitrogen into protein--beyond 6.25 and Jones' factors. Critical Reviews in Food Science and Nutrition 48:177-184.

Marquez-Cedillo, L.A., P.M. Hayes, A. Kleinhofs, W.G. Legge, B.G. Rossnagel, K. Sato, S.E. Ullrich, and D.M. Wesenberg. 2001. QTL analysis of agronomic traits in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. Theoretical and Applied Genetics 103:625-637.

Mascher, M., S. Wu, P. Amand, N. Stein, and J. Poland. 2013a. Application of Genotyping-by-Sequencing on Semiconductor Sequencing Platforms: A Comparison of Genetic and Reference-Based Marker Ordering in Barley. PLoS ONE 8:e76925.

Mascher, M., G.J. Muehlbauer, D.S. Rokhsar, J. Chapman, J. Schmutz, K. Barry, M. Muñoz-Amatriaín, T.J. Close, R.P. Wise, A.H. Schulman, A. Himmelbach, K.F.X. Mayer, U. Scholz, J.A. Poland, N. Stein, and R. Waugh. 2013b. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). The Plant Journal 76:718-727.

Matthies, I., S. Weise, J. Forster, V. Korzun, N. Stein, and M. Roder. 2013. Nitrogen-metabolism related genes in barley - haplotype diversity, linkage mapping and associations with malting and kernel quality parameters. BMC Genetics 14:77.

Mayer, K.F., R. Waugh, J.W. Brown, A. Schulman, P. Langridge, M. Platzer, G.B. Fincher, G.J. Muehlbauer, K. Sato, T.J. Close, R.P. Wise, and N. Stein. 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711-6.

Mayer, K.F.X., S. Taudien, M. Martis, H. Simková, P. Suchánková, H. Gundlach, T. Wicker, A. Petzold, M. Felder, B. Steuernagel, U. Scholz, A. Graner, M. Platzer, J. Dolezel, and N. Stein. 2009. Gene Content and Virtual Gene Order of Barley Chromosome 1H. Plant Physiology 151:496-505.

Mayer, K.F.X., M. Martis, P.E. Hedley, H. Šimková, H. Liu, J.A. Morris, B. Steuernagel, S. Taudien, S. Roessner, H. Gundlach, M. Kubaláková, P. Suchánková, F. Murat, M. Felder, T. Nussbaumer, A. Graner, J. Salse, T. Endo, H. Sakai, T. Tanaka, T. Itoh, K. Sato, M. Platzer, T. Matsumoto, U. Scholz, J. Doleželc, R. Waugh, and N. Stein. 2011. Unlocking the Barley Genome by Chromosomal and Comparative Genomics. The Plant Cell 23:1249-1263.

McCormick, F. 1995. Ras-related proteins in signal transduction and growth control. Molecular reproduction and development 42:500-6.

Meyers, B.C., S. Kaushik, and R.S. Nandety. 2005. Evolving disease resistance genes. Current Opinion in Plant Biology 8:129-134.

Mickelson, S., D. See, F.D. Meyer, J.P. Garner, C.R. Foster, T.K. Blake, and A.M. Fischer. 2003. Mapping of QTL associated with nitrogen storage and remobilization in barley (*Hordeum vulgare* L.) leaves 54:801-812.

Miklis, M., C. Consonni, R.A. Bhat, V. Lipka, P. Schulze-Lefert, and R. Panstruga. 2007. Barley MLO Modulates Actin-Dependent and Actin-Independent Antifungal Defense Pathways at the Cell Periphery. Plant Physiology 144:1132-43.

Milne, I., P. Shaw, G. Stephen, M. Bayer, L. Cardle, W. Thomas, A. Flavell, and D. Marshall. 2010. Flapjack - graphical genotype visualization. Bioinformatics 26:3133-3134.

Mimida, N., N. Kotoda, T. Ueda, M. Igarashi, Y. Hatsuyama, H. Iwanami, S. Moriya, and K. Abe. 2009. Four *TFL1/CEN*-like genes on distinct linkage groups show different expression patterns to regulate vegetative and reproductive development in apple (*Malus x domestica* Borkh.). Plant and cell physiology 50:394-412.

Moll, R.H., E.J. Kamprath, and W.A. Jackson. 1982. Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization. Agronomy Journal 74:562-564.

Muehlbauer, G.J., C. Feuillet, and J. Salse. 2009. Comparative Genomics in the Triticeae, p. 451-477 Genetics and Genomics of the Triticeae, Vol. 7. Springer US.

Mühleisen, J., H.P. Maurer, G. Stiewe, P. Bury, and J.C. Reif. 2013. Hybrid Breeding in Barley. Crop Science 53:819-824.

Muurinen, S., G.A. Slafer, and P. Peltonen-Sainio. 2006. Breeding effects on nitrogen use efficiency of spring cereals under northern conditions. Crop Science 46:561-568.

Navara, S., and K. Smith. 2013. Using near-isogenic barley lines to validate deoxynivalenol (DON) QTL previously identified through association analysis. Theoretical and Applied Genetics 127:633-645.

NIABtag. 2016. seedstats, pp. 50 Seedstats. NIABtag.

Oh, J.-M., D.-B. Yoon, and S.-N. Ahn. 2010. Fine mapping of grain weight QTLs using near isogenic lines from a Cross between *Oryza sativa* and *O. grandiglumis*. Journal of Crop Science and Biotechnology 13:7-12.

Okagaki, R., S. Cho, W. Kruger, W. Xu, S. Heinen, and G. Muehlbauer. 2013. The barley *UNICULM2* gene resides in a centromeric region and may be associated with signaling and stress responses. Functional & Integrative Genomics 13:33-41.

Parrott, D.L., J.M. Martin, and A.M. Fischer. 2010. Analysis of barley (Hordeum vulgare) leaf senescence and protease gene expression: a family C1A cysteine protease is specifically induced under conditions characterized by high carbohydrate, but low to moderate nitrogen levels. New Phytologist 187:313-331.

Pasam, R., R. Sharma, M. Malosetti, F. van Eeuwijk, G. Haseneyer, B. Kilian, and A. Graner. 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biology 12:16.

Passarella, V.S., R. Savin, and G.A. Slafer. 2005. Breeding effects on sensitivity of barley grain weight and quality to events of high temperature during grain filling. Euphytica 141:41-48.

Payne, R.W., D.A. Murray, S.A. Harding, D.B. Baird, and D.M. Soutar. 2009. GenStat for Windows (12th Edition) Introduction. VSN International, Hemel Hempstead.

Piepho, H.P., J. Möhring, A.E. Melchinger, and A. Büchse. 2008. BLUP for phenotypic selection in plant breeding and variety testing. . Euphytica 161:209-228.

Price, A.L., N.A. Zaitlen, D. Reich, and N. Patterson. 2010. New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics 11:459-463.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904-909.

Pritchard, J.K., M. Stephens, N.A. Rosenberg, and P. Donnelly. 2000. Association Mapping in Structured Populations. American Journal of Human Genetics 67 170–181.

Pumphrey, M.O., R. Bernardo, and J.A. Anderson. 2007. Validating the *Fhb1* QTL for Fusarium Head Blight Resistance in Near-Isogenic Wheat Lines Developed from Breeding Populations, pp. 200-206, Vol. 47.

Rae, S.J., M. Macaulay, L. Ramsay, F. Leigh, D. Matthews, D.M. O'Sullivan, P. Donini, P.C. Morris, W. Powell, D.F. Marshall, R. Waugh, and W.T.B. Thomas. 2007. Molecular barley breeding. Euphytica 158:295-303.

Ramirez-Gonzalez, R.H., V. Segovia, N. Bird, P. Fenwick, S. Holdgate, S. Berry, P. Jack, M. Caccamo, and C. Uauy. 2014. RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant Biotechnology Journal 13:613-24.

Ramsay, L., J. Comadran, A. Druka, D.F. Marshall, W.T.B. Thomas, M. MacAulay, K. MacKenzie, C. Simpson, J. Fuller, N. Bonar, P.M. Hayes, U. Lundqvist, J.D. Franckowiak, T.J. Close, G.J. Muehlbauer, and R. Waugh. 2011. *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. Nature Genetics 43:169-172.

Raun, W.R., and G.V. Johnson. 1999. Improving Nitrogen Use Efficiency for Cereal Production. Agronomy Journal 91:357-263.

Raun, W.R., J.B. Solie, G.V. Johnson, M.L. Stone, R.W. Mullen, K.W. Freeman, W.E. Thomason, and E.V. Lukina. 2002. Improving Nitrogen Use Efficiency in Cereal Grain Production with Optical Sensing and Variable Rate Application. Agronomy Journal 94:815–820.

Ren, X., C. Li, M. Cakir, W. Zhang, C. Grime, X.Q. Zhang, S. Broughton, D. Sun, and R. Lance. 2012. A quantitative trait locus for long photoperiod response mapped on chromosome 4H in barley. Molecular Breeding 30:1121-1130.

Reynolds, M., D. Bonnett, S.C. Chapman, R.T. Furbank, Y. Manès, D.E. Mather, and M.A.J. Parry. 2011. Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. Journal of Experiment Botany 62:439-452.

Romagosa, I., F. Han, S.E. Ullrich, P.M. Hayes, and D.M. Wesenberg. 1999. Verification of yield QTL through realized molecular marker- assisted selection responses in a barley cross, pp. 143-152, Vol. 5.

Rossini, L., A. Vecchietti, L. Nicoloso, N. Stein, S. Franzago, F. Salamini, and C. Pozzi. 2006. Candidate genes for barley mutants involved in plant architecture: an in silico approach. Theoretical and Applied Genetics 112:1073-1085.

Rostoks, N., L. Ramsay, K. MacKenzie, L. Cardle, P.R. Bhat, M.L. Roose, J.T. Svensson, N. Stein, R.K. Varshney, D.F. Marshall, A. Graner, T.J. Close, and R. Waugh. 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. PNAS 103:18656–18661.

Rostoks, N., S. Mudie, L. Cardle, J. Russell, L. Ramsay, A. Booth, J.T. Svensson, S.I. Wanamaker, H. Walia, E.M. Rodriguez, P.E. Hedley, H. Liu, J. Morris, T.J. Close, D.F. Marshall, and R. Waugh. 2005. Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Molecular Genetics and Genomics 274:515-527.

Saal, B., M. von Korff, J. Léon, and K. Pillen. 2011. Advanced-backcross QTL analysis in spring barley: IV. Localization of QTL × nitrogen interaction effects for yield-related traits. Euphytica 177:223-239.

Salse, J., and C. Feuillet. 2009. Comparative Genomics in the Triticeae, p. 451-477, *In* G. J. Muehlbauer and C. Feuillet, eds. Genetics and Genomics of the Triticeae, Vol. 7. Springer New York.

Sanchez-Garcia, M., C. Royo, N. Aparicio, J.A. Martín-Sánchez, and A. F. 2013. Genetic improvement of bread wheat yield and associated traits in Spain during the 20th century. The Journal of Agricultural Science 151:105-118.

Schmalenbach, I., J. Léon, and K. Pillen. 2009. Identification and verification of QTLs for agronomic traits using wild barley introgression lines. Theoretical and Applied Genetics 118:483-497.

Schnaithmann, F., and K. Pillen. 2013. Detection of exotic QTLs controlling nitrogen stress tolerance among wild barley introgression lines. Euphytica 189:67-88.

Serrano, M., F. Coluccia, M. Torres, F. L'Haridon, and J.-P. Métraux. 2014. The cuticle and plant defense to pathogens. Frontiers in Plant Science 5:274.

Singh, R., J. Kaur, A. K. Arya, and I. S. Bhatia. 1981. Accumulation of Starch, Protein and the Activities of Invertase and Sucrose ADP (UDP) Glycosyl Transferase in Developing Sorghum vulgare Grains. Biochemie und Physiologie der Pflanzen 176:606-613.

Slafer, G.A. 2003. Genetic basis of yield as viewed from a crop physiologist's perspective. Annals of Applied Biology 142:117-128.

Snape, J., M. Foulkes, J. Simmonds, M. Leverington, L. Fish, Y. Wang, and M. Ciavarrella. 2007. Dissecting gene × environmental effects on wheat yields via QTL and physiological analysis. Euphytica 154:401-408.

Spaner, D., B.G. Rossnagel, W.G. Legge, G.J. Scoles, P.E. Eckstein, G.A. Penner, N.A. Tinker, K.G. Briggs, D.E. Falk, J.C. Afele, P.M. Hayes, and D.E. Mather. 1999. Verifications of a Quantitative Trait Locus Affecting Agronomic Traits in Two-Row Barley. Crop Science 39:248-252.

Sreenivasulu, N., and T. Schnurbusch. 2012. A genetic playground for enhancing grain number in cereals. Trends in Plant Science 17:91-101.

Stein, N., D. Perovic, J. Kumlehn, B. Pellio, S. Stracke, S. Streng, F. Ordon, and A. Graner. 2005. The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in *Hordeum vulgare* (L.). Plant Journal 42:912-922.

Stich, B., J. Mohring, H.-P. Piepho, M. Heckenberger, E.S. Buckler, and A.E. Melchinger. 2008. Comparison of Mixed-Model Approaches for Association Mapping. Genetics 178:1745-1754.

Stracke, S., G. Haseneyer, J.-B. Veyrieras, H. Geiger, S. Sauer, A. Graner, and H.-P. Piepho. 2009. Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theoretical and Applied Genetics 118:259-273.

Stram, D., V. Seshan, R.C. Elston, J.M. Satagopan, and S. Sun. 2012. Multi-SNP Haplotype Analysis Methods for Association Analysis, p. 423-452 Statistical Human Genetics. Humana Press.

Su, Z., C. Hao, L. Wang, Y. Dong, and X. Zhang. 2011. Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics 122:211-223.

Svačina, P., T. Středa, and O. Chloupek. 2014. Uncommon selection by root system size increases barley yield. Agronomy for Sustainable Development 34:545-551.

Swamy, B.P.M., P. Vikram, S. Dixit, H.U. Ahmed, and A. Kumar. 2011. Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. BMC Genomics 12:319.

Sylvester-Bradley, R., and D. Kindred. 2009. Analysing nitrogen responses of cereals to prioritize routes to the improvement of nitrogen use efficiency. Journal of Experimental Botany 60:1939-1951.

Szucs, P., J. Skinner, I. Karsai, A. Cuesta-Marcos, K. Haggard, A. Corey, T. Chen, and P. Hayes. 2007. Validation of the *VRN-H2/VRN-H1* epistatic model in barley reveals that intron length variation in *VRN-H1* may account for a continuum of vernalization sensitivity. Molecular Genetics and Genomics 277:249-261.

Tanaka, W., M. Pautler, D. Jackson, and H.-Y. Hirano. 2014. Grass Meristems II: Inflorescence Architecture, Flower Development and Meristem Fate. Plant and cell physiology 54:313-324.

Tang, T., H. Xie, Y. Wang, B. Lü, and J. Liang. 2009. The effect of sucrose and abscisic acid interaction on sucrose synthase and its relationship to grain filling of rice (*Oryza sativa* L.). Journal of experimental botany 60:2641-52.

Tattersall, D.B., R. van Heeswijck, and P.B. Hoj. 1997. Identification and Characterization of a Fruit-Specific, Thaumatin-Like Protein That Accumulates at Very High Levels in Conjunction with the Onset of Sugar Accumulation and Berry Softening in Grapes. Plant Physiology 114:759-769.

R Foundation for Statistical Computing, Vienna, Austria. 2013. R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0,

URL http://www.R-project.org. R Foundation for Statistical Computing, Vienna, Austria.

Thomas, W., G. Barral-Baron, D. Harrap, A. Booth, R. Keith, C. Warden, J. Russell, R. Waugh, P. Werner, and L. Ramsay. 2013. NUE-CROPS report: Interaction of variation of nitrogen use efficiency with sites and seasons in winter barley. Newcatle University.

Tinker, N.A., D.E. Mather, B.G. Rossnagel, K.J. Kasha, A. Kleinhofs, P.M. Hayes, D.E. Falk, T. Ferguson, L.P. Shugar, W.G. Legge, R.B. Irvine, T.M. Choo, K.G. Briggs, S.E. Ullrich, J.D. Franckowiak, T.K. Blake, R.J. Graf, S.M. Dofing, M.A. Saghai Maroof, G.J. Scoles, D. Hoffman, L.S. Dahleen, A. Kilian, F. Chen, R.M. Biyashev, D.A. Kudrna, and B.J. Steffenson. 1996. Regions of the genome that affect agronomic performance in two-row barley. Crop Science 36:1053-1062.

Tong, H.-h., L. Chen, W.-p. Li, H.-w. Mei, Y.-z. Xing, X.-q. Yu, X.-y. Xu, S.-q. Zhang, and L.-j. Luo. 2010. Identification and characterization of quantitative trait loci for grain yield and its components under different nitrogen fertilization levels in rice (*Oryza sativa* L.). Molecular Breeding.

Touraev, A., B. Forster, S.M. Jain, P. Devaux, and K.J. Kasha. 2009. Overview of Barley Doubled Haploid Production, p. 47-63 Advances in Haploid Production in Higher Plants. Springer Netherlands.

Tuinstra, M.R., G. Ejeta, and P.B. Goldsbrough. 1997. Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theoretical and Applied Genetics 95:1005-1011.

Turner, A., J. Beales, S. Faure, R.P. Dunford, and D.A. Laurie. 2005. The Pseudo-Response Regulator *Ppd-H1* Provides Adaptation to Photoperiod in Barley. Science 310:1031-1034.

Uauy, C., A. Distelfeld, T. Fahima, A. Blechl, and J. Dubcovsky. 2006. A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. Science 314:1298-1301.

van Eeuwijk, F.A., M.C.A.M. Bink, K. Chenu, and S.C. Chapman. 2010. Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology 13:193-205.

Varshney, R.K., A. Graner, and M.E. Sorrells. 2005. Genomics-assisted breeding for crop improvement. Trends in Plant Science 10:621-630.

Venuprasad, R., S.M. Impa, R.P.V. Gowda, G.N. Atlin, and R. Serraj. 2011. Rice near-isogenic-lines (NILs) contrasting for grain yield under lowland drought stress. Field Crops Research 123:38-46.

Verma, V., M.J. Foulkes, A.J. Worland, R. Sylvester-Bradley, P.D.S. Caligari, and J.W. Snape. 2004. Mapping quantitative trait loci for flag leaf senescence as a yield determinant in winter wheat under optimal and drought-stressed environments. Euphytica 135:255-263.

Visioni, A., A. Tondelli, E. Francia, A. Pswarayi, M. Malosetti, J. Russell, W. Thomas, R. Waugh, N. Pecchioni, I. Romagosa, and J. Comadran. 2013. Genome-wide association mapping of frost tolerance in barley (*Hordeum vulgare* L.). BMC Genomics 14:424.

von Korff, M., J. Léon, and K. Pillen. 2010. Detection of epistatic interactions between exotic alleles introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). Theoretical and Applied Genetics 121:1455-1464.

von Korff, M., H. Wang, J. Léon, and K. Pillen. 2005. AB-QTL analysis in spring barley. I. Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley. Theoretical and Applied Genetics 111:583-590.

von Korff, M., H. Wang, J. Léon, and K. Pillen. 2006. AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (H. vulgare ssp. spontaneum). Theoretical and Applied Genetics 112:1221-1231.

Wang, G., I. Schmalenbach, M. von Korff, J. Léon, B. Kilian, J. Rode, and K. Pillen. 2010. Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC_2DH population and a set of wild barley introgression lines. Theoretical and Applied Genetics 120:1559-1574.

Wang, M., N. Jiang, T. Jia, L. Leach, J. Cockram, R. Waugh, L. Ramsay, B. Thomas, and Z. Luo. 2012. Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. Theoretical and Applied Genetics 124:233-246.

Waugh, R., J.-L. Jannink, G.J. Muehlbauer, and L. Ramsay. 2009. The emergence of whole genome association scans in barley. Current Opinion in Plant Biology 12:218-222.

Würschum, T. 2012. Mapping QTL for agronomic traits in breeding populations. Theoretical and Applied Genetics 125:201-210.

Xing, L.-P., H.-Z. Wang, Z.-N. Jiang, J.-L. Ni, A.-Z. Cao, L. Yu, and P.-D. Chen. 2008. Transformation of Wheat Thaumatin-Like Protein Gene and Analysis of Reactions to Powdery Mildew and Fusarium Head Blight in Transgenic Plants. Acta Agronomica Sinica 34:349-354.

Xue, D.-w., M.-x. Zhou, X.-q. Zhang, S. Chen, K. Wei, F.-r. Zeng, Y. Mao, F.-b. Wu, and G.-p. Zhang. 2010. Identification of QTLs for yield and yield components of barley under different growth conditions. Journal of Zhejiang University - Science B 11:169-176.

Yan, C.-J., S. Yan, Y.-C. Yang, X.-H. Zeng, Y.-W. Fang, S.-Y. Zeng, C.-Y. Tian, Y.-W. Sun, S.-Z. Tang, and M.-H. Gu. 2009. Development of gene-tagged markers for quantitative trait loci underlying rice yield components. Euphytica 169:215-226.

Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, and J. Dubcovsky. 2003. Positional cloning of the wheat vernalization gene *VRN1*. PNAS 100:6263-6268.

Yan, L., A. Loukoianov, A. Blechl, G. Tranquilli, W. Ramakrishna, P. SanMiguel, J.L. Bennetzen, V. Echenique, and J. Dubcovsky. 2004. The Wheat *VRN2* Gene Is a Flowering Repressor Down-Regulated by Vernalization. Science 303:1640-1644.

Yan, L., D. Fu, C. Li, A. Blechl, G. Tranquilli, M. Bonafede, A. Sanchez, M. Valarik, S. Yasuda, and J. Dubcovsky. 2006. The Wheat and Barley Vernalization Gene VRN3 Is an Orthologue of *FT*. PNAS 103:19581-19586.

Yin, X., S.D. Chasalow, P. Stam, M.J. Kropff, C.J. Dourleijn, I. Bos, and P.S. Bindraban. 2002. Use of component analysis in QTL mapping of complex crop traits: a case study on yield in barley. Plant Breeding 121:314-319.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels or relatedness. Nature Genetics 38:203-208.

Yu, J., S. Hu, J. Wang, G.K.-S. Wong, S. Li, B. Liu, Y. Deng, L. Dai, Y. Zhou, X. Zhang, M. Cao, J. Liu, J. Sun, J. Tang, Y. Chen, X. Huang, W. Lin, C. Ye, W. Tong, L. Cong, J. Geng, Y. Han, L. Li, W. Li, G. Hu, X. Huang, W. Li, J. Li, Z. Liu, L. Li, J. Liu, Q. Qi, J. Liu, L. Li, T. Li, X. Wang, H. Lu, T. Wu, M. Zhu, P. Ni, H. Han, W. Dong, X. Ren, X. Feng, P. Cui, X. Li, H. Wang, X. Xu, W. Zhai, Z. Xu, J. Zhang, S. He, J. Zhang, J. Xu, K. Zhang, X. Zheng, J. Dong, W.

Zeng, L. Tao, J. Ye, J. Tan, X. Ren, X. Chen, J. He, D. Liu, W. Tian, C. Tian, H. Xia, Q. Bao, G. Li, H. Gao, T. Cao, J. Wang, W. Zhao, P. Li, W. Chen, X. Wang, Y. Zhang, J. Hu, J. Wang, S. Liu, J. Yang, G. Zhang, Y. Xiong, Z. Li, L. Mao, C. Zhou, Z. Zhu, R. Chen, B. Hao, W. Zheng, S. Chen, W. Guo, G. Li, S. Liu, M. Tao, J. Wang, L. Zhu, L. Yuan, and H. Yang. 2002. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*). Science 296:79-92.

Yun, S.J., L. Gyenis, E. Bossolini, P.M. Hayes, I. Matus, K.P. Smith, B.J. Steffenson, R. Tuberosa, and G.J. Muehlbauer. 2006. Validation of Quantitative Trait Loci for Multiple Disease Resistance in Barley Using Advanced Backcross Lines Developed with a Wild Barley. Crop Science 46:1179-1186.

Zabka, V., M. Stangl, G. Bringmann, G. Vogg, M. Riederer, and U. Hildebrandt. 2008. Host surface properties affect prepenetration processes in the barley powdery mildew fungus. New Phytologist 177:251-263.

Zadoks, J.C., T.T. Chang, and C.F. Konzak. 1974. A Decimal Code for the Growth Stages of Cereals. Weed Research:415-421.

Zeng, Z.B. 1994. Precision Mapping of Quantitative Trait Loci. Genetics 136:1457-1468.

Zhang, G., C. Li, C.D. Li, M. Cakir, and R. Lance. Genetic Improvement of Malting Quality through Conventional Breeding and Marker-assisted Selection, p. 260-292 Genetics and Improvement of Barley Malt Quality. Springer Berlin Heidelberg.

Zhang, L., S. Marchand, N. Tinker, and F. Belzile. 2009. Population structure and linkage disequilibrium in barley assessed by DArT markers. Theoretical and Applied Genetics 119:43-52.

Zhang, S., W. Hu, L. Wang, C. Lin, B. Cong, C. Sun, and D. Luo. 2005. *TFL1/CEN*-like genes control intercalary meristem activity and phase transition in rice. Plant Science 168:1393-1408.

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42:355-360.

Zhao, H., and T.P. Speed. 1996. On Genetic Map Functions. Genetics 142:1369-1377.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and Prospects of Association Mapping in Plants. The plant genome 1:5-20.

Zitzewitz, J., P. Szucs, J. Dubcovsky, L. Yan, E. Francia, N. Pecchioni, A. Casas, T.H. Chen, P. Hayes, and J. Skinner. 2005. Molecular and Structural Characterization of Barley Vernalization Genes. Plant Molecular Biology 59:449-467.

Appendices

Appendix 2. 1 Table of fixed and random terms used for computing S×R phenotype means (BLUE).

Statistical REML models used to calculate the Best linear unbiased estimates of the phenotypes measured for the S×R DH population in the field experiments of 2009 and 2012 (see 2.2.2). The BLUE were used for QTL mapping.

Trait	Prediction model terms	
	fixed terms	random terms
Fowl09		
EE	Constant + Genotype	row + topography
Ht	Constant + Genotype	column + topography
Til-cal	Constant + Genotype	row + column + topography
Til-mes	Constant + Genotype	topography
TGW	Constant + Genotype	row + topography + lodging
TGW_GS	Constant + Genotype	row + column + topography
GE	Constant + Genotype	column
Yld	Constant + Genotype	topography + lodging
HLW	Constant + Genotype	row + lodging
GP	Constant + Genotype	-
GS	Constant + Genotype	-
Elm09		
EE	Constant + Genotype	row
Ht	Constant + Genotype	row
Til-cal	Constant + Genotype	row + column
Til-mes	Constant + Genotype	row
TGW	Constant + Genotype	row + column
TGW_GS	Constant + Genotype	row
GE	Constant + Genotype	row + column
Yld	Constant + Genotype	row + column
HLW	Constant + Genotype	row + column
GP	Constant + Genotype	row
GS	Constant + Genotype	row + column
Fowl12		
EE	Constant + Genotype	Block
Ht	Constant + Genotype	column
TGW_GS	Constant + Genotype	smallblock
GE	Constant + Genotype	row
Mil	Constant + Genotype	row + column
BR	Constant + Genotype	smallblock
GP	Constant + Genotype	row
GS	Constant + Genotype	row + column
Ldg	Constant + Genotype	row + column
SC	Constant + Genotype	smallblock
SD	Constant + Genotype	row
SG	Constant + Genotype	smallblock

Appendix 2. 2 Heatmap of field topograhy and lodging percentage at Fowl09.

Heatmap of the field topography of the Fowlmere site in 2009 (Fowl09, see 2.2.2) and the percentage lodging per plots of the same trial. These factors have been included in statistical models for the estimations of BLUEs.

field topography scores (1 field deep, 6 field top)

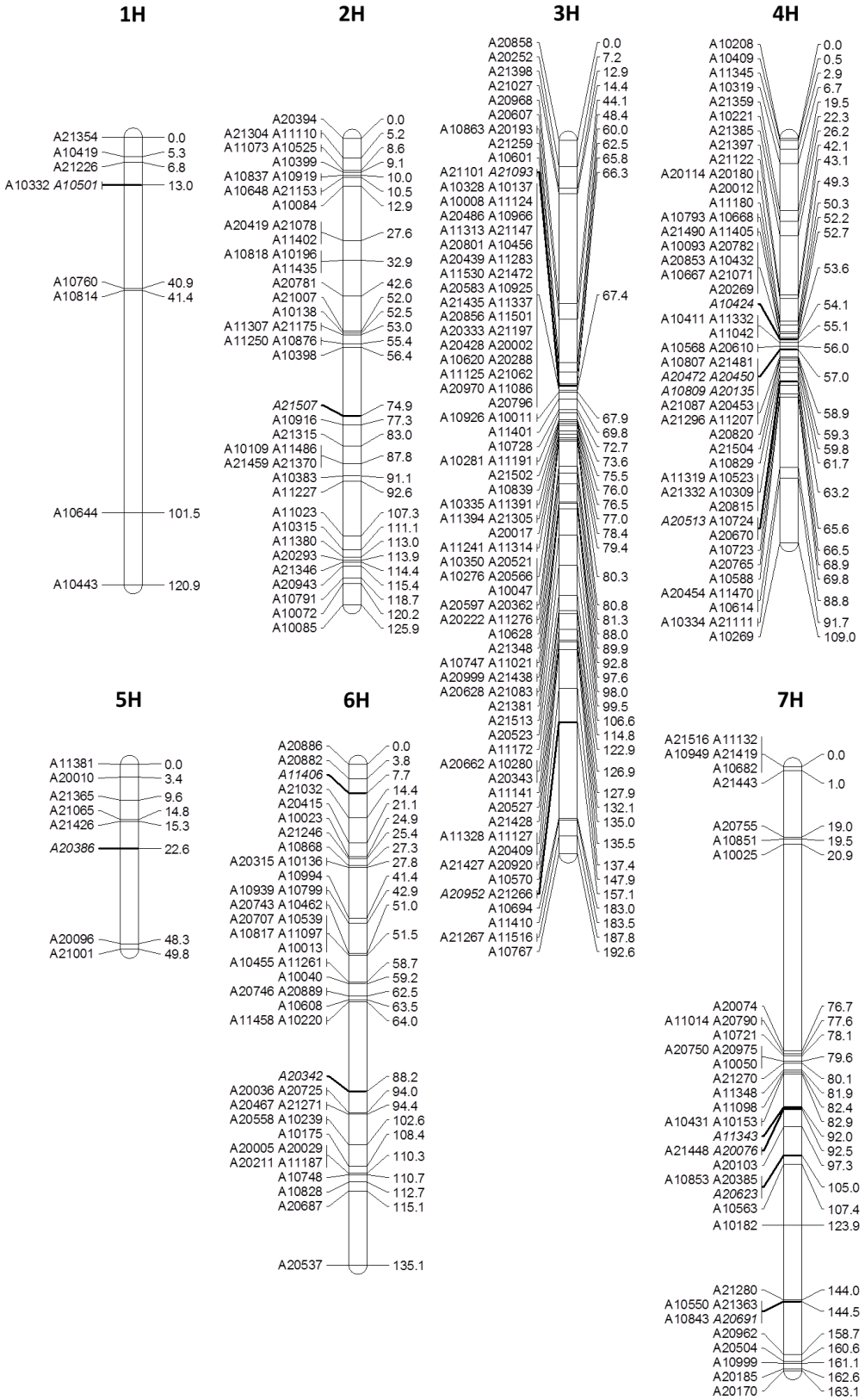
row\col	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
9	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5	5	5	5
8	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
7	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	4
6	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	5	5	5	5	5	5	5	4	4	4	4
5	2	2	2	2	2	2	2	2	2	2	3	3	3	3	4	4	4	4	4	5	5	5	4	4	4	4
4	1	1	1	1	1	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4
3	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	3	3	3	4	4	4	4	4	4	4	4
2	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3
1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3

Plot lodging scores (% of plot area)

row\col	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	45	75	90	90	85	85	20	75	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	90	95	85	95	90	95	90	80	97	95	8	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	5	10	25	0	20	30	60	80	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 2. 3 Genetic map of the Saffron × Retriever population.

The genetic map was obtained using Mapdisto software (2.2.3) and the 309 polymorphic SNP markers from the BOPA1 genotyping platform across 211 DH lines. SNP names in italic correspond to markers mapped in S×R but unmapped in the 2009 consensus OPA genetic map (Close et al., 2009).



Appendix 3. 1 List of varieties included in NUE-CROPS and AGOUEB Two-row winter barley GWAS panels.

Each GWAS panels was composed of varieties for which correct genetic marker information from the 9K Illumina chip was available. The NUE-CROPs panel contains 125 varieties and an AGOUEB panel contains 179 varieties.

	Mapping panel	
Variety name	NUE CROPs	Agoueb
Accrue	x	
Alpha	x	
Amillis	x	
Angora		x
Antelope		x
Antigua	x	x
Antonia	x	x
Anvil		x
Aquarelle	x	
Archimedes		x
Arda	x	
Artist	x	x
Asso	x	
Astrid	x	
Avenue	x	
Aydanhanim	x	
Babylone	x	
Baraka	x	
Barcelona		x
Baton		x
Becket		x
Bistro		x
Blythe		x
Boreale		x
Breeze		x
Bronze	x	x
Calcutta	x	x
Calliope		x
Camion	x	x
Campion		x
Candy		x
Cannock	x	x
Caption		x
Carat	x	
Cassata	x	x
Cassia	x	
Cathay		x
Cedar		x
Celebrity	x	x
Cellina		x
Celsius		x
Chamomile		x
Charleston		x
Chestnut	x	x
Chicane	x	x
Chintz		x
Chord		x
Cinnamon	x	x
Clara	x	x
Clarine	x	x

	Mapping panel	
Variety name	NUE CROPs	Agoueb
Cobalt	x	x
Concept		x
Connoisseur		x
Coriolis	x	x
Credo		x
Crescendo		x
Cynthia		x
Cypress		x
Diadem	x	
Diamond	x	x
Digby		x
Dolmen	x	
Dolphin	x	x
Druid		x
Duchess	x	
Duet	x	
Eagle		x
Electron		x
Elmstead	x	x
Emeraude	x	x
Emilia	x	x
Epic	x	x
Ethno	x	x
Fahrenheit	x	x
Fanfare	x	x
Faraday		x
Fighter	x	x
Finesse	x	
Firefly	x	x
Flagon	x	x
Flute		x
Frolic	x	x
Gaelic	x	x
Gazelle		x
Gleam	x	
Glint		x
Goldmine		x
Goldrush		x
Gypsy		x
Haka	x	x
Halcyon	x	
Halifax	x	x
Hanna		x
Harland		x
Heligan	x	x
Hermia	x	x
Honey		x
Houston		x
Hurricane	x	
Igri	x	

	Mapping panel	
Variety name	NUE CROPs	Agoueb
Imogen		x
Intro	x	x
Jessica	x	x
Jet		x
Jewel	x	x
Karisma	x	x
Kaskade	x	
Kelibia	x	
Kestrel	x	x
KH Malko	x	
Kingston	x	x
Kira	x	x
Kite		x
KWS Discovery	x	
KWS Glacier	x	
Labea	x	
Lambada		x
Lark		x
Laurel		x
Leonie	x	
Linnet		x
Louise	x	x
Madrigal	x	x
Magie	x	
Magnolia		x
Mahogany		x
Malta	x	
Malwinta	x	x
Mariner		x
Marinka	x	x
Maris Otter	x	
Maris Trojan	x	
Maritem	x	x
Masai	x	x
Masquerade		x
Mead		x
Medoc	x	x
Melanie	x	x
Melusine		x
Merode		x
Milena		x
Molly	x	x
Montage		x
Moonshine		x
Mortimer	x	x
Murcie	x	
Musette	x	x
Mystique	x	x
Nectaria	x	x
Nocturne		x

Appendix 3. 1 cont

	Mapping panel	
Variety name	NUE CROPs	Agoueb
Nure	x	
Opal	x	x
Orchidea	x	
Outlook		x
Panda	x	
Parasol	x	x
Pastoral	x	x
Pearl	x	x
Pedigree	x	x
Peridot		x
Perth		x
Pilot		x
Pipkin		x
Pippa		x
Portrait	x	x
Posaune	x	x
Prelude	x	x
Puffin	x	x
Punch		x
Rattle		x
Ravel		x
Regina		x
Rejane	x	
Retriever	x	
Rhythm		x
Rifle	x	x
Saffron	x	x
Saffron 2	x	
Sapphire		x
Sarah	x	
Scylla		x
Selection		x
Sevilla	x	x
Sevilla 2	x	
Sevilla 3	x	
Silverstone		x
Sombrero	x	x
Sonic		x
Sonja	x	
Spectrum		x
Spinner		x
Spirit		x
Sprite	x	
Steeple		x
Sumo	x	x
Sunrise		x
Surtees		x
SW Alison	x	
Swallow		x
Swift		x

	Mapping panel	
Variety name	NUE CROPs	Agoueb
Tabetha		x
Tallica	x	x
Target	x	x
Tempo		x
Thalia		x
Tiffany	x	
Tipster		x
Toffee		x
Tokyo		x
Torrent	x	x
Tosca		x
Tucker		x
Tudor		x
Turine		x
Vanessa	x	
Vanilla		x
Vesuvius	x	x
Vilna	x	x
Vivaldi		x
Vixen	x	
Volley		x
Weaver		x
Wigwam		x
Willow	x	x
Winner	x	x
Wintmalt	x	
Wizard		x
Wombat		x
Zulu		x

Appendix 3. 2 Summary statistics on BLUPs used for NUE-CROPs GWAS.

The table of summarises statistics for the phenotypic means (BLUPs) obtained for panel of varieties of NUE-CROPs. (see 3.3.1).

	Antho	GNYld	GrainN	Grains	Hd	HI	Ht	NUE	NupE	NupEg
Missing values	22	0	0	0	0	0	0	0	0	0
Mean	1.52	116	1.99	19.9	219	52.2	74	25.6	0.64	40.2
Standard deviation	0.50	4	0.10	2.3	3	2.2	6	2.2	0.02	2.6
Variance	0.25	17	0.01	5.2	8	5.0	31	4.8	0.0004	7.0
Minimum	1	101	1.72	15.9	210	45.8	65	20.1	0.58	32.7
Maximum	2	124	2.35	25.7	229	59.0	104	31.8	0.68	48.8
Range	1	22	0.63	9.8	19	13.3	39	11.7	0.09	16.2
Lower quartile	1	113	1.92	18.2	217	50.8	70	24.3	0.63	38.4
Upper quartile	2	119	2.05	21.3	221	53.5	76	27.1	0.65	41.8

	NutEt	StemN	Stems	StS	TGW	TGW_GS	Till.GS	Till.Yld	UPOV	Yld
Missing values	0	0	0	0	0	0	0	0	0	0
Mean	49.7	0.56	2.9	2.2	53.6	59.2	644	745	1994	6.94
Standard deviation	2.6	0.02	0.3	0.4	3.8	4.2	73	82	7	0.46
Variance	6.5	0.0003	0.1	0.2	14.3	17.4	5309	6735	54	0.21
Minimum	42.4	0.51	2.2	2	42.9	48.2	477	515	1966	5.80
Maximum	54.9	0.61	3.9	3	62.0	69.0	867	1029	2009	8.13
Range	12.5	0.09	1.7	1	19.1	20.7	390	514	43	2.33
Lower quartile	48.0	0.55	2.7	2	51.4	56.4	595	683	1990	6.67
Upper quartile	51.4	0.57	3.1	2	56.0	62.1	686	794	1999	7.22

Appendix 3. 3 Descriptive statistics of principal component analysis of the agronomic traits measured in NUE-CROPs.

a) Percentage of variation of accounted for by the main principal components and loadings of each trait for the principal component analysis presented in **Figure 3.1 a).**

Principal Component :		**1**	**2**	**3**	**4**	**5**	**6**	**7**
% Variance accouned for :		54.4%	22.3%	16.4%	3.2%	2.0%	1.1%	0.6%
Latent vectors	Yld	-0.07	0.34	-0.82	0.32	-0.16	0.26	0.12
(loadings)	TGW	-0.37	0.53	0.08	-0.18	0.12	0.08	-0.72
	Grains	-0.31	-0.42	-0.52	-0.35	0.32	-0.47	-0.12
	Stems	0.46	0.13	-0.17	-0.71	-0.49	-0.01	-0.06
	Till.Yld	0.47	0.25	-0.04	0.39	-0.01	-0.71	-0.24
	Till.GS	0.46	0.24	-0.08	-0.19	0.79	0.22	0.14
	TGW_GS	-0.36	0.54	0.14	-0.21	0.00	-0.38	0.61

b) Percentage of variation of accounted for by the main principal components and loadings of each trait for the principal component analysis presented in **Figure 3.1 b)**.

Principal Component :		**1**	**2**	**3**	**4**	**5**	**6**	**7**
% Variance accouned for :		34.8%	23.7%	17.7%	7.8%	5.3%	3.7%	2.2%
Latent vectors	Yld	0.00	0.54	-0.05	0.07	-0.08	0.24	-0.47
(loadings)	Hd	0.01	0.25	0.40	0.35	0.70	-0.30	0.23
	Ht	0.29	-0.16	0.37	0.10	0.05	0.74	0.12
	TGW	0.30	0.14	-0.46	0.31	0.10	0.16	0.13
	GrainN	0.09	-0.52	-0.13	0.02	-0.05	-0.10	0.27
	Grains	0.32	0.23	0.37	-0.28	-0.09	0.11	0.08
	StemN	-0.19	-0.05	-0.32	-0.60	0.64	0.29	-0.04
	HI	-0.06	0.50	-0.11	-0.32	-0.22	-0.11	0.55
	Stems	-0.44	0.05	0.06	0.08	-0.08	0.32	0.42
	Till.Yld	-0.45	0.00	-0.02	0.25	0.05	0.11	-0.26
	Till.GS	-0.45	0.03	-0.01	0.24	-0.11	0.18	0.18
	TGW_GS	0.28	0.13	-0.47	0.31	0.10	0.11	0.17

Appendix 3. 4 Table of correlations between agronomic and phenotypic trait means used in NUE-CROPs and AGOUEB GWAS.

a) Pearson correlation coefficient between traits mapped in the NUE CROPs GWAS. The correlations were calculated using the R and significant correlation factor at the p<0.01level indicated in bold. The abbreviations of traits are: Yld: Yield; Till_GS and Till_Yld: the fertile tillers measured from grab samples and yield respectively, Stems: number of tillers, TGW: thousand grain weight, TGW-GS: thousand grain weight from grab samples, Grains: number of grains per ear, NUE nitrogen use efficiency, NUpE: nitrogen uptake efficiency, NutEg: nitrogen utilisation efficiency for grain, NutEt: the Nitrogen utilisation efficiency total, GrainN: Grain Nitrogen, StemsN: Nitrogen content of stems, GrainNYld: grain nitrogen yield, Hd: heading date, Ht: height, HI: harvest index, UPOV: year of release, Sts: sterile spikelets.

	Yld	Till.GS	Till.Yld	Stems	TGW	TGW_GS	Grains	NUE	NupE	NutEt	NupEg	GrainN	StemN	GrainNYld	Hd	Ht	HI	UPOV	Antho
Yld	**1**																		
Till.GS	0.06	**1**																	
Till.Yld	0.06	**0.88**	**1**																
Stems	0.07	**0.84**	**0.81**	**1**															
TGW	**0.29**	**-0.43**	**-0.47**	**-0.53**	**1**														
TGW_GS	**0.23**	**-0.43**	**-0.44**	**-0.51**	**0.95**	**1**													
Grains	**0.3**	**-0.6**	**-0.69**	**-0.48**	0.06	0.02	**1**												
NUE	**0.98**	0.08	0.07	0.07	**0.26**	0.21	**0.32**	**1**											
NupE	**0.51**	-0.15	-0.03	-0.17	**0.38**	**0.3**	0.07	**0.48**	**1**										
NutEt	**0.64**	-0.04	0.04	-0.06	0.19	0.13	**0.3**	**0.66**	**0.64**	**1**									
NupEg	**0.79**	0.16	0.1	0.18	0.02	0.01	**0.35**	**0.82**	-0.05	**0.4**	**1**								
GrainN	**-0.75**	-0.21	-0.17	**-0.25**	0.04	0.04	**-0.3**	**-0.78**	0.03	**-0.45**	**-0.93**	**1**							
StemN	-0.08	0.21	**0.26**	**0.25**	-0.06	-0.05	**-0.39**	-0.13	0.08	**-0.35**	-0.22	0.05	**1**						
GrainNYld	**0.68**	-0.15	-0.1	-0.18	**0.47**	**0.39**	0.13	**0.63**	**0.81**	**0.46**	0.17	-0.05	-0.07	**1**					
Hd	**0.28**	0	0.04	0.04	-0.14	-0.16	**0.34**	**0.32**	-0.04	**0.43**	**0.43**	**-0.44**	**-0.27**	-0.06	**1**				
Ht	**-0.23**	**-0.48**	**-0.49**	**-0.41**	0.02	-0.01	**0.55**	**-0.23**	0.06	0.21	**-0.24**	**0.24**	**-0.39**	-0.11	0.17	**1**			
HI	**0.7**	0.11	0.03	0.16	0.14	0.13	0.23	**0.7**	0.03	0.01	**0.79**	**-0.67**	0.12	**0.32**	0.1	**-0.42**	**1**		
UPOV	**0.43**	0.02	-0.04	0.01	0.22	**0.24**	0.15	**0.44**	0	0.21	**0.51**	**-0.49**	-0.11	0.14	0.11	**-0.25**	**0.42**	**1**	
Antho	**0.26**	-0.14	-0.2	-0.03	0.17	0.12	0.24	0.24	0.07	0.15	0.22	**-0.26**	0.02	0.09	0.07	0.01	0.19	**0.3**	**1**
StS	**0.24**	-0.14	**-0.24**	-0.17	**0.41**	**0.43**	0.12	**0.24**	0.08	0.07	0.18	-0.1	-0.12	0.23	-0.03	-0.05	0.22	**0.35**	0.25

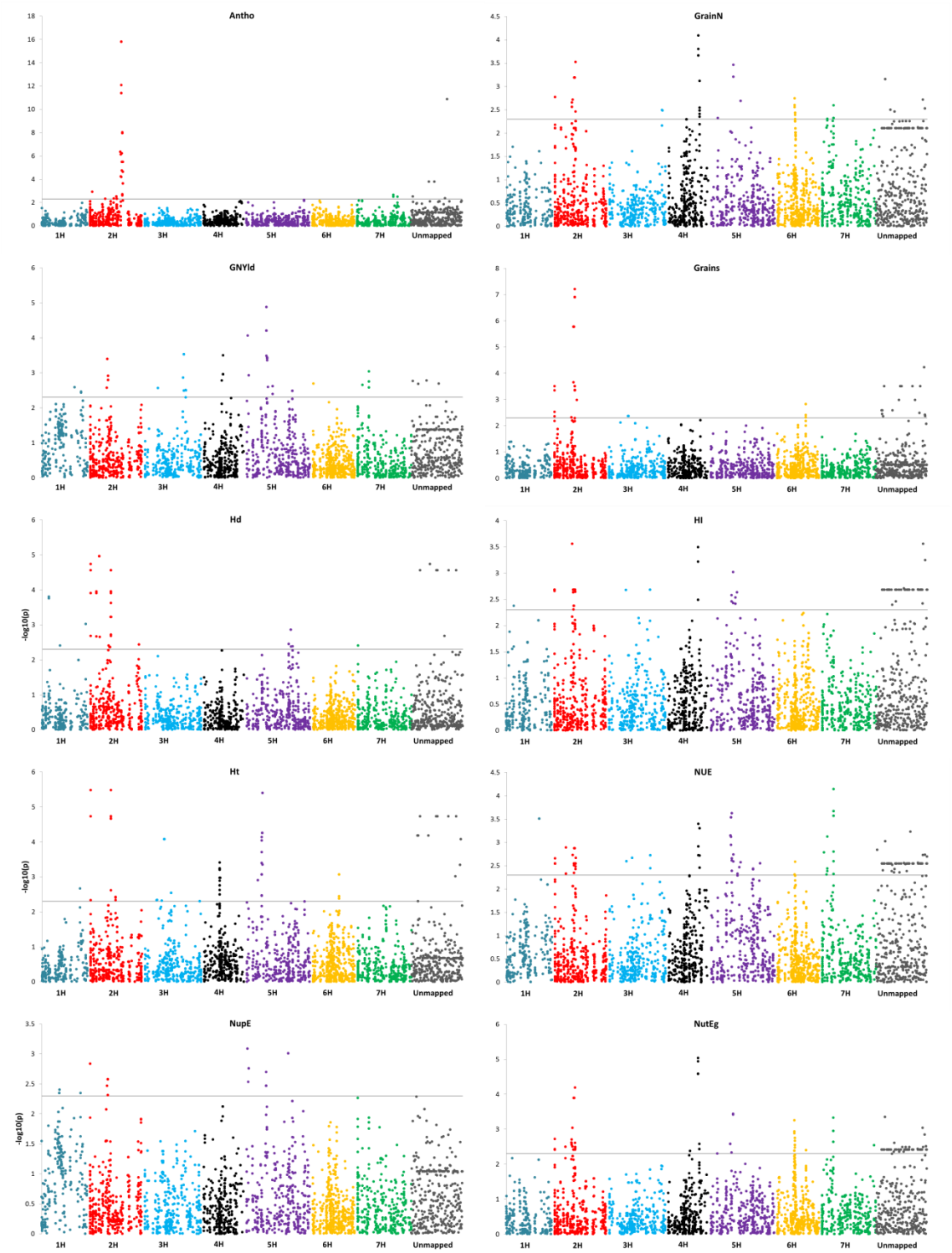
Appendix 3. 4 Cont.

b) Pearson correlation coefficient between traits mapped in the AGOUEB GWAS. The correlations were calculated using the GenStat 14th Edition directive FCORRELATION and significant correlation at the p<0.001 level are in bold. The abbreviations of traits are: BR: Brown rust, Ear_Gl: ear wax layer, Hd: heading date, HWE: hot water extract, LLHS: lower leaf hairy sheath, GrainN: grain nitrogen, NB: net blotch, Ht: straw length treated, Ht_U: straw length untreated, SS_U, straw strength untreated, HLW: Hectoliter weight, TGW, thousand grain weight, VFH: ventral furrow hair, Wthd: winter hardiness, Yld: yield treated, Aleu: aleurone layer, Sts: sterile spikelets, Mildew, Rynchо, rynchosporium.

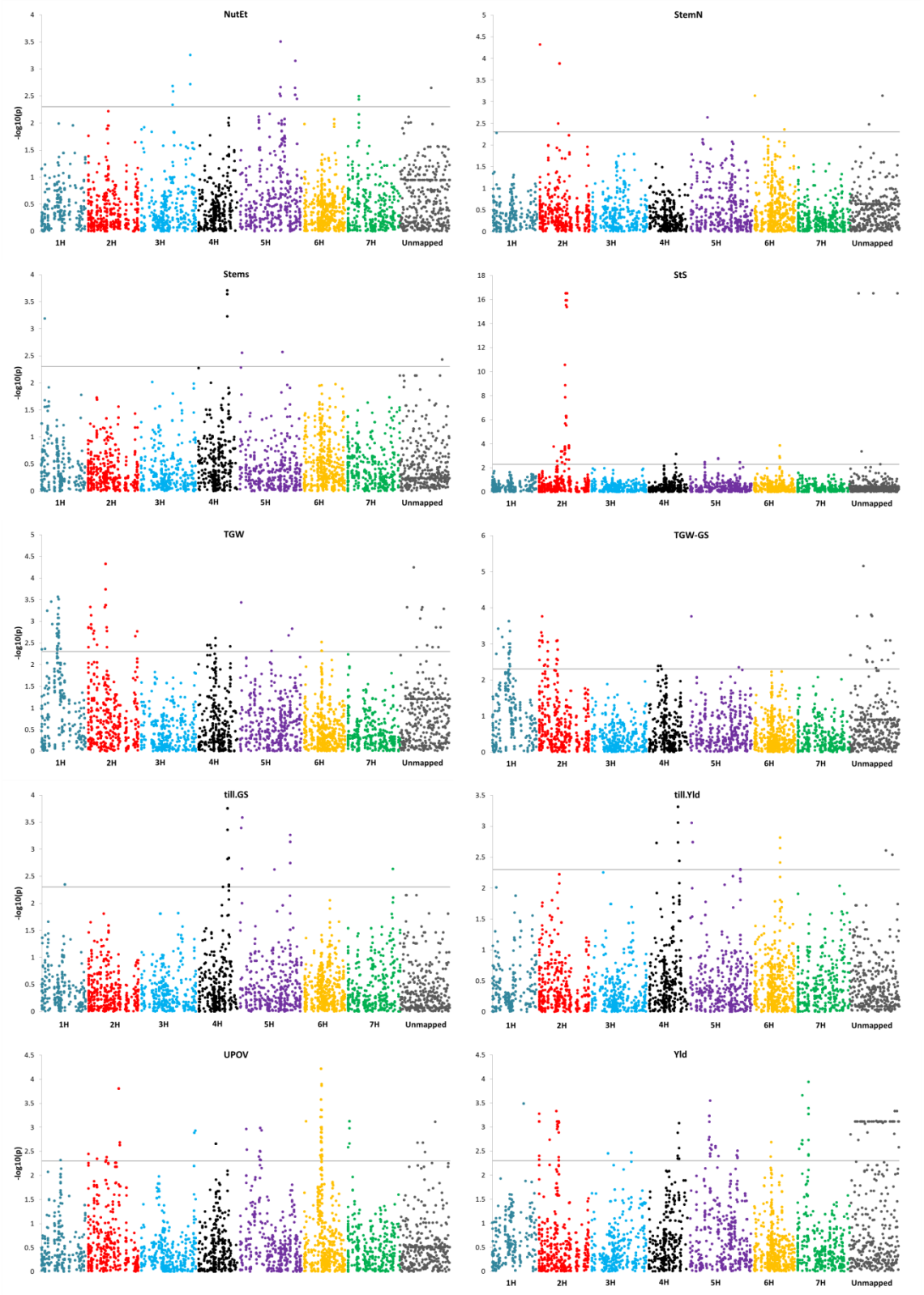
	Antho	BR	EAar_Gl	Hd	HWE	LLHLS	GrainN	NB	Ht	Ht_U	SS_U	HLW	TGW	VFH	WtHd	Yld	Aleu	Sts	Mildew
BR	0.06	-																	
Ear_Gl	-0.09	-0.06	-																
Hd	-0.08	**-0.40**	0.02	-															
HWE	-0.04	-0.02	0.27	0.15	-														
LLHLS	-0.02	0.02	-0.05	0.17	**0.35**	-													
GrainN	-0.04	-0.02	-0.11	-0.09	**-0.44**	-0.15	-												
NB	0.00	0.10	0.24	-0.14	0.21	0.11	-0.05	-											
Ht	0.07	-0.22	-0.11	0.28	-0.12	-0.04	0.07	-0.03	-										
Ht_U	0.07	-0.27	-0.03	0.30	-0.16	-0.07	0.18	0.03	**0.86**	-									
SS_U	-0.13	-0.09	0.04	0.29	0.14	0.15	-0.21	0.06	**0.34**	0.30	-								
HLW	0.06	-0.12	0.29	-0.06	0.08	-0.09	0.23	-0.02	-0.08	-0.03	-0.15	-							
TGW	0.09	0.03	-0.19	0.00	-0.21	0.10	**0.35**	-0.17	0.13	0.24	**-0.35**	0.03	-						
VFH	0.18	0.13	-0.10	-0.17	-0.22	-0.09	0.07	-0.14	0.06	0.00	-0.03	**-0.34**	0.09	-					
WtHd	0.17	0.07	-0.04	0.07	0.17	0.01	-0.02	0.15	-0.02	-0.09	0.11	0.06	-0.18	-0.17	-				
Yld	0.18	-0.05	-0.14	-0.09	0.19	0.10	-0.31	-0.19	-0.06	-0.04	-0.28	-0.07	0.27	-0.02	-0.03	-			
Aleurone	0.06	-0.17	0.02	-0.06	-0.02	-0.25	0.09	-0.06	0.03	0.07	0.10	0.15	-0.15	-0.14	-0.14	-0.01	-		
Sts	-0.16	0.04	0.30	-0.03	0.13	0.03	-0.13	0.11	-0.16	-0.17	0.23	-0.05	**-0.41**	0.04	-0.02	-0.26	-0.12	-	
Mildew	0.04	**0.42**	-0.08	-0.14	-0.08	-0.15	-0.06	-0.18	-0.11	-0.15	0.11	0.13	-0.07	0.11	-0.02	-0.08	-0.03	-0.01	-
Rynchо	0.01	-0.20	0.11	0.01	-0.22	-0.12	0.09	-0.10	0.09	0.03	0.08	0.04	0.00	0.13	-0.17	-0.25	0.14	-0.07	-0.03

Appendix 3. 5 Manhattan plots for the GWAS of NUE-CROPs for 20 traits.

The plots represent the results of the tests for marker-traits associations from the genome wide association scans (4041 SNP) of the NUE-CROPs two-row winter barley experiment (3.3.3). The X-axis plots SNP ordered on their chromosomal position based on the consensus map used in this study. The Y-axis represents the logarithm of the *p*-values from the tests of association carried out by the GWAS after correction for population structure (Tassel-MLM_K model) (Supplementary data 2). Markers above a detection threshold of –log10(0.003) (horizontal line) were retained as belonging to a QTL.



Appendix 3. 5 cont.



Appendix 3. 6 List of additional significant peak SNP for NUE-CROPs QTL.

The table lists the significant markers sharing identical map position and association test result across the variety panel with the peak marker reported at the NUE CROPs QTL in Table 3.2. The shared position are indicative of markers with identical allele distribution across the panel of varieties used for the mapping and can be illustrated by r^2 between marker of 1.

Trait	QTL	Chrom	Main reported marker	Other marker in complete R2
Anthocyanine color	Antho_2	2 H	I195164	A10138; A11307; A21175; B10649
Grain Nitrogen Yield	GNYld_1	1 H	I154646	I236160
	GNYld_4	3 H	I164290	I193132; I151711
	GNYld_8	5 H	A20236	A21133
Grains	Grains_2	6 H	I138716	I147599
Heading date	Hd_3	2 H	B30871	B30872; BK_12; BK_14; BK_15; BK_16
	Hd_4	2 H	A10191	A20438; A21399
Harvest Index	HI_2	3 H	I204057	A10380
	HI_3	3 H	I103215	I165334
Height	Ht_3	3 H	A11016	A10653; I114566
	Ht_4	3 H	A21163	I171062; I219894; I27417; I148020; I238157
	Ht_8	6 H	I129756	I144337
Nitrogen Use Efficiency	NUE_1	1 H	I154646	I236160
	NUE_3	2 H	A10358	A20251; A20631; A11384
	NUE_4	3 H	I204057	A10380
	NUE_5	3 H	B31242	I200508
	NUE_6	3 H	I103215	I165334
	NUE_10	5 H	A10080	B31109
Nitrogen	NUpE_1	2 H	A10733	B30259; I175065
Uptake Efficiency	NUpE_4	5 H	I160288	I230675
Nitrogen Utilisation Efficiency total	NUtEt_3	5 H	I160288	I230675
Stem Nitrogen	StemN_1	2 H	I177375	A10685; I127347; I1502; I222769; I4969
Stems	Stems_3	5 H	B30975	B30977
Stems	Stems_4	5 H	I148402	I237494
Sterile Spikelets	StS_2	2H	A10287	I91810; I196853; I88704; B30897; B30901; I160616; I179213; I211281; I211281; A10475; I138463; I156090; I159462; I160833; I172648; I180028; I185505; I188339; I198603; I235221; B10936; B31424; I157236; I174800; I176114; I182039; I211894; A21351;
	StS_5	6H	I131992	I217187; I144862; I154582
Thousand Grain Weight	TGW_3	1 H	A20810	A21000; I170542
	TGW_5	2 H	A10733	B30259; I175065
Thousand Grain Weight from grab samples	TGW-GS_3	1 H	A20810	A21000; I170542
	TGW-GS_6	2 H	A10602	A10796
Tillering from Grab samples	till.GS_1	4 H	A20732	I107010; I172072
	till.GS_2	5 H	B30975	B30977
	till.GS_5	7 H	A10550	B31325
Tillering from yield	till.Yld_1	4 H	A21385	I188190
	till.Yld_2	4 H	A20732	I107010; I172072
UPOV date of inscription	UPOV_1	2 H	I195164	A10138; A11307; A21175; B10649
	UPOV_3	4 H	I128723	I155554
	UPOV_5	5 H	I213753	B30745
	UPOV_7	6 H	I136897	I182195
	UPOV_8	7 H	A11222	B31411; I175756; I230060; I230149
Yield treated	Yld_1	1 H	I154646	I236160
	Yld_6	5 H	A20236	A21133

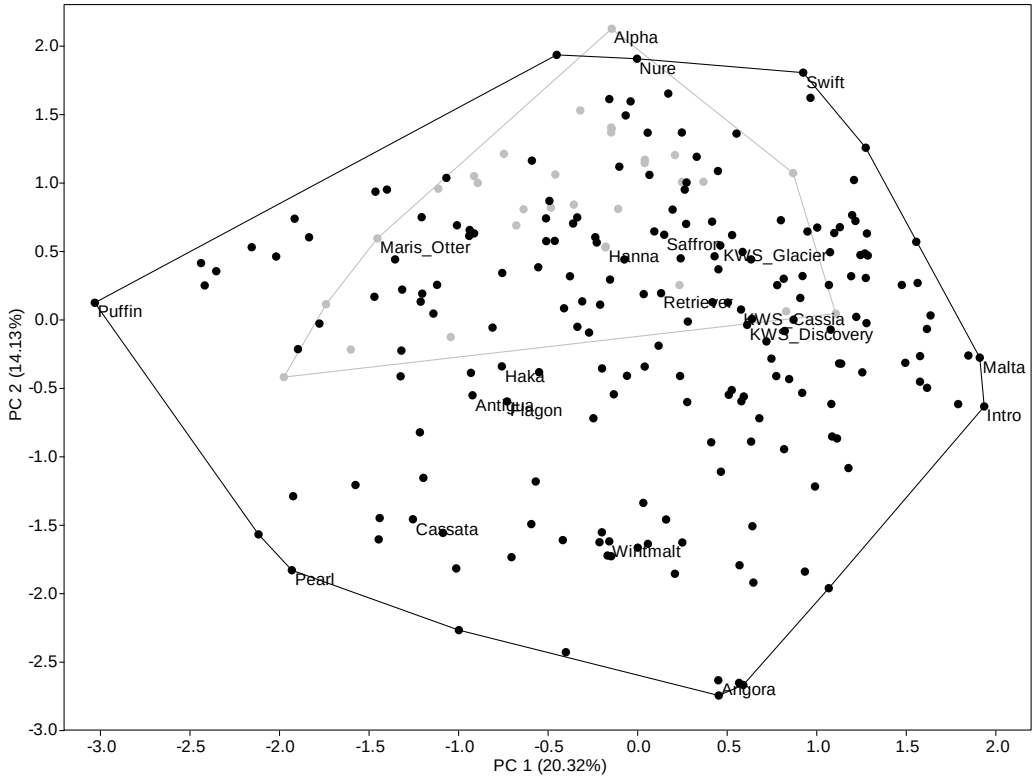
Appendix 3. 7 List of additional significant peak SNP for AGOUEB QTL.

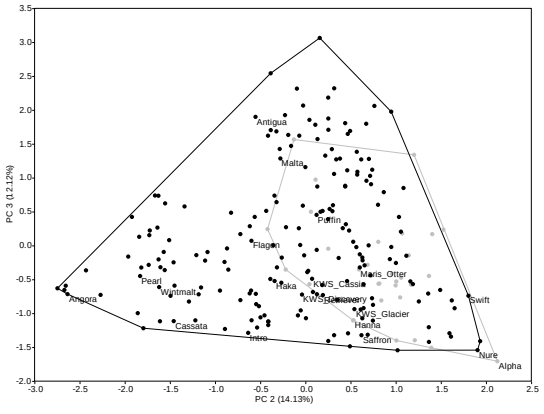
The table lists the significant markers sharing identical map position and association test result across the variety panel with the peak marker reported at the NUE CROPs QTL in Table 3.3. The shared position are indicative of markers with identical allele distribution across the panel of varieties used for the mapping and can be illustrated by r^2 between marker of 1.

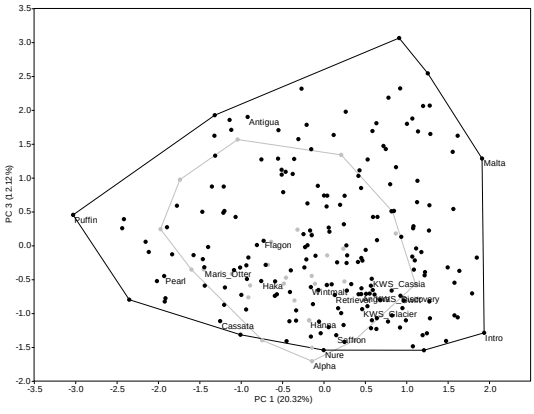
Trait	QTL	Chrom	Main reported marker	Other marker in complete R2
Aleurone	aleurone_2	4 H	A21087	A20453; A21273; A21296
Brown Rust	BR_1	2 H	I146785	I145381
Ear Glaucosity	EAR-G_1	1 H	I120059	I120053; A10419
	EAR-G_3	6 H	B11455	B30783; I219061
Grain Nitrogen	GrainN_4	5 H	B30975	B30977
	GrainN_5	5 H	B30400	I198008; I208686
Hot Water Extract	HWE_2	1 H	I165338	I165476
LLHS	LLHLS_2	5 H	A21355	I152365
Mildew	mild-(U)_1	4 H	I128147	I152166
Net Blotch	NB(U)_1	2 H	I16024	I16995
	NB(U)_3	6 H	I128460	A20707; B30317; I146235; I238855
Straw Length Treated	SL_T_3	5 H	A10236	I167103; I195241
Straw Length Untreated	SL_U_1	2 H	I177375	A10685; I1502; I222769; I4969
	SL_U_3	5 H	B30975	B30977
	SL_U_5	5 H	A10236	I167103; I195241
Straw Strenght Untreated	SS_U_2	2 H	A10358	A20251; A20631
Sterile Spikelets	StS_1	1 H	A21333	I156506; I56976
	StS_3	2 H	A10287	I91810; I196853; I88704; B30897; B30901
	StS_5	5 H	A10236	I167103; I195241
Thousand Grain weight	TGW_2	2 H	I110647	I192657
Winter Hardiness	WintHard_1	2 H	A21261	A21265; A21366; I7026
	WintHard_2	4 H	I110333	I147712; I229658; I239145
Yield Treated	Yld_T_1	4 H	I150603	I157396

Appendix 3. 8 Distribution of the allele of SNP B30265 (HvCEN) on the genetic diversity of NUE-CROPS and AGOUEB two-row winter barley panels.

The complete set of varieties used in the association panel is presented. For clarity, only a few varieties have been presented. The correlation based PCA uses a correlation matrix made from the 1284 common SNP markers between M_{AG} and M_{NUE}. The varieties represented by a back dot carry the "Nure" late allele at B30265 co-segregating with HvCEN (Comadran et al., 2012). The varieties in grey dots carry the early allele at that same SNP. Axis are in Eigenvalue scale.







Appendix 4. 1 Effect size and direction of QTL peak marker effect across the set phenotypes in NUE-CROPs GWAS.

Each QTL is presented with its peak marker (most significant SNP) with the associated marker map position, –log10(*p*) resulting from the test for association, the peak SNP alleles with its minor allele (MA) and minor allele frequency (MAF) in the NUE-CROPS panel. SNP effect size and direction are reported across the 20 agronomic traits analysed. The negative effects of the allele 1 on another trait are highlighted in grey. The effects associated with a significant marker association for the trait considered (i.e. a QTL) are highlighted in bold (e.g. SNP I195164 of Antho_2 is a QTL for both traits Anthocyanin and UPOV).

Trait	QTL	Chorm	Marker	Dist	-log10(p)	alleles 1 / 2	MA	MAF	Effect of marker allele 1																			
									Antho	GNYld	GrainN	Grains	Hd	HI	Ht	NUE	NupE	NutEg	NutEt	StemN	Stems	StS	TGW	TGW-GS	till.GS	till.Yld	UPOV	Yld
Anthocyanin color	Antho_1	2 H	A10326	6.45	2.92	G / A	A	0.18	**-0.39**	1.18	0.02	0.39	-0.17	-0.65	3.15	-0.08	0.01	-0.72	0.50	0.00	-0.14	0.02	1.12	0.88	-40.67	-36.64	0.59	0.00
	Antho_2	2 H	I195164	96.8	15.80	C / T	T	0.48	**-1.04**	-0.67	0.06	-0.17	0.01	-0.80	1.78	-1.00	0.00	-1.26	-0.65	0.00	-0.12	-0.14	-0.79	-0.98	-16.74	-4.80	**-5.86**	-0.21
	Antho_3	7 H	I174285	112.46	2.67	A / G	A	0.11	**-0.49**	-1.16	0.01	-0.73	0.21	-0.24	0.95	-0.52	0.00	-0.55	-0.62	0.01	0.02	0.04	0.47	0.46	29.98	19.22	0.93	-0.12
Grain Nitrogen Yield	GNYld_1	1 H	I154646	100.7	2.59	A / C	C	0.11	-0.30	**3.88**	-0.07	0.87	1.53	1.88	-0.27	**2.54**	0.01	2.23	2.14	0.00	0.08	-0.04	1.12	0.10	20.37	16.07	1.67	**0.53**
	GNYld_2	2 H	I151535	52.47	3.39	G / A	A	0.14	0.04	**4.39**	0.00	-1.09	-0.95	0.59	-3.26	1.04	0.02	-0.38	-0.01	0.01	-0.03	0.19	**3.43**	3.10	0.79	-4.72	-0.36	0.26
	GNYld_3	3 H	I115045	39.45	2.56	G / A	A	0.30	0.00	**2.85**	0.03	-0.44	-1.85	0.34	-0.22	0.39	0.01	-0.22	-0.45	0.00	0.04	0.22	0.93	0.76	7.43	-1.19	1.08	0.11
	GNYld_4	3 H	I164290	120.59	3.52	T / C	C	0.14	-0.03	**4.37**	8.69	1.41	0.58	1.59	-0.88	1.28	0.01	0.79	0.16	0.00	-0.08	-0.03	0.53	-0.65	-7.22	-30.76	1.74	0.27
	GNYld_5	4 H	A20482	59.37	3.50	G / A	A	0.15	0.15	**4.63**	-0.03	-0.22	0.55	1.14	-2.92	1.80	0.01	1.15	0.94	0.00	0.04	0.03	0.68	-0.14	9.05	24.28	4.73	0.40
	GNYld_6	5 H	A20553	2.81	4.06	A / G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
	GNYld_7	5 H	A21508	60.74	4.88	A / G	G	0.11	-0.09	**5.46**	-0.02	0.08	-0.32	1.40	-1.63	**2.19**	**0.02**	1.18	1.75	-0.01	0.06	-0.05	0.15	-0.44	9.84	19.72	4.57	**0.47**
	GNYld_8	5 H	A20236	80.61	2.61	C / A	A	0.26	-0.08	**2.90**	-0.04	0.51	-0.54	1.16	-0.59	1.45	0.01	1.24	0.57	0.00	-0.07	0.07	0.44	0.16	-7.50	-5.86	2.73	**0.33**
	GNYld_9	6 H	I123065	1.34	2.69	C / T	T	0.13	0.05	**3.66**	0.00	0.36	-0.31	0.29	-0.60	1.03	0.01	0.03	0.60	0.00	-0.16	0.10	0.80	0.64	-8.11	-28.92	-2.06	0.26
	GNYld_10	7 H	I186187	14.96	2.65	T / G	G	0.22	0.34	**3.07**	-0.07	0.73	0.27	1.57	-1.43	**1.91**	0.01	1.82	1.29	0.00	0.05	0.16	0.61	0.94	7.50	14.85	4.52	**0.44**
	GNYld_11	7 H	I138457	34.82	3.04	C / A	A	0.22	0.09	**3.18**	**-0.07**	-0.14	1.22	1.32	-2.82	**2.10**	0.01	**2.22**	1.83	0.00	0.04	0.00	0.83	1.00	6.77	26.61	2.58	**0.43**
Gain Nitrogen	GrainN_1	2 H	A20862	63.5	3.53	T / A	A	0.42	0.03	0.37	**0.08**	**-3.01**	**-2.43**	-1.35	**-3.93**	-1.52	0.00	**-2.53**	-1.68	0.01	0.07	0.13	0.66	0.60	33.72	46.37	0.87	-0.31
	GrainN_2	4 H	I168399	92.4	4.09	C / A	C	0.39	0.04	0.10	**-0.08**	-0.33	0.84	**1.65**	-1.36	**1.46**	0.00	**2.52**	0.79	0.00	**0.20**	0.07	-0.51	-0.83	**45.48**	**49.82**	3.97	0.28
	GrainN_3	5 H	A21121	68.35	3.46	G / A	A	0.35	0.01	0.88	**-0.08**	0.05	-0.37	**1.64**	-0.95	**1.46**	0.00	**2.09**	0.64	0.00	0.06	0.02	0.88	0.56	16.41	15.74	**5.05**	**0.30**
	GrainN_4	6 H	B30120	52.75	2.75	C / A	C	0.44	0.02	0.53	**-0.08**	-0.36	0.47	1.18	-2.85	1.47	0.00	**2.26**	0.70	0.00	0.12	0.04	1.27	1.34	29.06	24.34	4.69	0.32
	GrainN_5	7 H	I138457	34.82	2.60	C / A	A	0.22	0.09	**3.18**	**-0.07**	-0.14	1.22	1.32	-2.82	**2.10**	0.01	**2.22**	1.83	0.00	0.04	0.00	0.83	1.00	6.77	26.61	2.58	**0.43**
Grains	Grains_1	2 H	A20862	63.5	7.20	T / A	A	0.42	0.03	0.37	**0.08**	**-3.01**	**-2.43**	-1.35	**-3.93**	-1.52	0.00	**-2.53**	-1.68	0.01	0.07	0.13	0.66	0.60	33.72	46.37	0.87	-0.31
	Grains_2	6 H	I138716	88.9	2.83	G / A	A	0.43	0.00	-1.27	0.04	**-1.74**	0.02	-0.85	-0.31	-0.90	0.00	-1.56	-0.41	0.01	0.06	-0.09	0.82	1.12	24.65	26.03	-1.48	-0.21
Heading date	Hd_1	1 H	B30241	20.82	3.80	C / A	C	0.29	-0.01	-1.64	-0.02	0.25	**2.67**	0.17	1.28	-0.16	-0.01	0.70	-0.06	-0.01	0.13	-0.11	-1.59	-2.19	27.88	11.42	1.74	-0.06
	Hd_2	1 H	A21384	135.56	3.02	A / G	G	0.27	-0.02	0.77	0.03	-0.54	**-2.19**	0.09	-1.64	-0.33	0.00	-0.71	-0.96	0.00	-0.01	0.15	0.43	1.06	-9.63	-4.00	1.39	-0.05
	Hd_3	2 H	B30871	26.57	4.96	A / G	A	0.39	-0.21	-0.16	-0.02	0.49	**3.02**	-0.01	1.29	0.46	0.00	0.34	0.41	0.00	0.04	-0.04	-1.22	-1.78	10.90	6.68	-3.96	0.08
	Hd_4	2 H	A10191	63.53	4.56	C / A	A	0.14	0.35	2.96	0.01	**-2.80**	**-4.09**	0.87	**-8.18**	0.12	0.01	-0.17	-1.46	0.01	0.15	0.06	1.75	1.64	48.67	40.69	4.67	0.08
	Hd_5	5 H	B30867	136.43	2.86	C / A	A	0.23	-0.04	-0.40	0.01	0.17	**-2.30**	0.97	-0.13	-0.32	-0.01	-0.21	-1.58	0.00	-0.07	0.12	1.40	2.16	-30.42	-44.53	3.39	-0.03
Harvest Index	HI_1	2 H	I10398	54.95	3.56	C / T	T	0.17	-0.27	-2.20	**0.11**	-0.99	-0.10	**-2.79**	2.79	**-2.35**	-0.01	**-3.06**	-0.98	-0.01	-0.06	-0.09	-1.90	-1.70	1.54	7.85	-1.73	**-0.54**
	HI_2	3 H	I204057	51.7	2.68	C / T	T	0.10	0.32	2.67	-0.07	-0.19	-0.21	**2.46**	-5.52	**2.37**	0.01	2.18	0.19	0.01	0.04	0.00	1.46	2.82	-4.08	12.31	1.14	0.48
	HI_3	3 H	I103215	126.27	2.68	A / G	G	0.14	0.04	3.61	-0.04	0.37	0.61	**2.07**	-4.46	**2.05**	0.01	1.92	0.46	0.00	0.02	0.03	0.65	0.04	40.00	8.18	2.77	0.41
	HI_4	4 H	I129218	92.4	3.50	C / A	C	0.42	-0.01	0.46	**-0.08**	-0.29	0.93	**1.71**	-1.60	**1.65**	0.00	**2.51**	0.96	0.00	**0.22**	0.11	-0.54	-0.99	**49.82**	**52.49**	3.83	**0.31**
	HI_5	5 H	A21121	68.35	3.02	G / A	A	0.35	0.01	0.88	**-0.08**	0.05	-0.37	**1.64**	-0.95	**1.46**	0.00	**2.09**	0.64	0.00	0.06	0.02	0.88	0.56	16.41	15.74	**5.05**	**0.30**
	HI_6	5 H	A10183	80.02	2.63	G / A	G	0.46	0.11	-1.12	0.04	-0.26	0.78	**-1.53**	0.51	-0.87	0.00	-1.24	0.33	0.00	0.00	0.05	-0.68	-0.36	-4.81	1.75	-1.63	-0.22
Height	Ht_1	1 H	A10338	117.8	2.66	C / A	A	0.21	0.12	1.47	-0.03	-0.50	-0.58	0.91	**-4.32**	0.87	0.01	0.99	0.53	0.00	0.02	-0.12	0.56	0.62	5.94	14.69	-1.22	0.16
	Ht_2	2 H	B30265	63.53	5.48	A / G	G	0.14	0.35	3.31	0.01	**-2.71**	**-3.69**	0.82	**-8.90**	0.44	0.01	-0.11	-0.95	0.01	0.09	0.07	1.65	1.44	35.64	38.96	5.43	0.13
	Ht_3	3 H	A11016	58.64	4.08	G / C	C	0.17	0.22	0.98	-0.05	-1.76	-0.45	1.27	**-6.00**	1.34	0.00	1.51	0.11	0.01	0.16	0.03	0.73	1.37	43.86	47.81	2.57	0.26
	Ht_4	3 H	A21163	80.89	2.54	A / G	G	0.30	0.07	-0.26	-0.02	-1.47	-0.08	0.67	**-4.03**	0.07	-0.01	0.67	-0.68	-5.04	0.03	0.01	1.75	2.01	21.34	21.96	0.52	0.01
	Ht_5	4 H	I190401	48.72	3.41	A / G	G	0.32	0.02	1.01	-0.01	-0.99	-0.19	0.56	**-4.81**	0.71	0.00	0.34	0.22	0.00	-0.02	-0.03	0.68	0.87	17.63	23.49	-0.74	0.13
	Ht_6	5 H	I4717	34.25	2.90	G / A	A	0.19	0.21	0.66	-0.03	-1.11	-0.71	0.90	**-5.18**	0.85	0.00	0.78	0.11	0.00	0.10	-0.04	0.02	0.58	22.23	45.09	0.31	0.20
	Ht_7	5 H	B31257	48.11	5.40	A / T	T	0.14	0.18	2.97	-0.05	-1.76	-1.59	1.89	**-8.56**	1.94	0.01	1.62	0.50	0.01	0.07	0.13	2.05	2.53	49.25	58.91	5.73	0.40
	Ht_8	6 H	I129756	80.52	3.07	T / G	G	0.11	-0.04	2.66	-0.04	-0.78	-0.15	1.87	**-5.98**	1.65	0.01	1.27	-0.09	0.01	0.10	0.12	0.02	0.27	40.31	47.36	0.56	0.36

Appendix 4. 1 cont.

Trait	QTL	Chorm	Marker	Dist	-log10(p)	alleles 1	alleles 2	MA	MAF	Effect of marker allele 1																		
										Antho	GNYld	GrainN	Grains	Hd	HI	Ht	NUE	NupE	NutEg	NutEt	StemN	Stems	StS	TGW	TGW-GS	till.GS	till.Yld	UPOV	Yld
Nitrogen Use Efficiency	**NUE_1**	1 H	I154646	100.7	3.51	A	C	C	0.11	-0.30	**3.88**	-0.07	0.87	1.53	1.88	-0.27	**2.54**	0.01	2.23	2.14	0.00	0.08	-0.04	1.12	0.10	20.37	16.07	1.67	**0.53**
	NUE_2	2 H	A21304	33.74	2.89	A	G	G	0.30	-0.08	-2.53	0.06	0.14	-0.22	-1.16	2.11	**-1.83**	-0.01	-1.84	-1.27	0.00	0.03	-0.11	-1.81	-2.01	2.65	-14.92	-3.94	**-0.37**
	NUE_3	2 H	A10358	59.21	2.87	C	A	A	0.41	0.05	-0.30	**0.08**	**-2.59**	-1.84	-1.49	-2.20	**-1.65**	-3.58	**-2.39**	-1.50	0.01	0.04	0.12	0.91	0.93	26.15	33.05	1.66	**-0.35**
	NUE_4	3 H	I204057	51.7	2.60	C	T	T	0.10	0.32	2.67	-0.07	-0.19	-0.21	**2.46**	-5.52	**2.37**	0.01	2.18	0.19	0.01	0.04	0.00	1.46	2.82	-4.08	12.31	1.14	0.48
	NUE_5	3 H	B31242	69.6	2.67	A	C	A	0.28	0.02	0.98	-0.06	-0.06	0.24	0.84	-1.94	**1.79**	0.00	1.48	1.51	0.00	0.03	0.05	0.99	0.99	17.34	11.73	2.68	0.33
	NUE_6	3 H	I103215	126.27	2.72	A	G	G	0.14	0.04	3.61	-0.04	0.37	0.61	**2.07**	-4.46	**2.05**	0.01	1.92	0.46	0.00	0.02	0.03	0.65	0.04	40.00	8.18	2.77	0.41
	NUE_7	4 H	I129218	92.4	3.40	C	A	C	0.42	-0.01	0.46	**-0.08**	-0.29	0.93	**1.71**	-1.60	**1.65**	0.00	**2.51**	0.96	0.00	**0.22**	0.11	-0.54	-0.99	**49.82**	**52.49**	3.83	**0.31**
	NUE_8	5 H	I231238	63.31	3.62	T	C	C	0.21	0.12	**3.69**	-0.05	0.30	-0.47	**1.81**	-2.80	**2.19**	0.01	1.71	1.53	0.00	0.01	0.08	1.40	0.72	8.69	1.25	5.10	**0.45**
	NUE_9	5 H	B31427	90.84	2.58	G	C	C	0.30	0.18	1.68	**-0.07**	0.07	-0.92	1.37	-0.90	**1.51**	0.01	1.43	0.77	0.00	0.03	-0.01	0.86	1.38	7.60	6.94	3.40	0.30
	NUE_10	5 H	A10080	151.36	2.55	G	A	A	0.24	0.11	2.43	-0.06	-0.69	0.07	1.44	-3.88	**1.77**	0.01	1.77	1.17	0.00	0.06	0.07	1.65	1.24	10.57	22.21	0.12	0.36
	NUE_11	6 H	I118381	54.6	2.58	C	T	C	0.47	-0.01	1.20	-0.07	-0.37	0.58	1.00	-1.96	**1.62**	0.00	1.92	1.04	0.00	0.10	0.04	1.68	1.69	28.22	26.87	4.68	**0.34**
	NUE_12	7 H	I186187	14.96	3.12	T	G	G	0.22	0.34	**3.07**	-0.07	0.73	0.27	1.57	-1.43	**1.91**	0.01	1.82	1.29	0.00	0.05	0.16	0.61	0.94	7.50	14.85	4.52	**0.44**
	NUE_13	7 H	I138457	34.82	4.14	C	A	A	0.22	0.09	**3.18**	**-0.07**	-0.14	1.22	1.32	-2.82	**2.10**	0.01	**2.22**	1.83	0.00	0.04	0.00	0.83	1.00	6.77	26.61	2.58	**0.43**
Nitrogen Uptake Efficiency	**NUpE_1**	2 H	A10733	54.95	2.58	G	C	C	0.14	0.05	**3.82**	0.01	-1.60	-1.08	0.42	-3.72	0.86	**0.02**	-0.70	-0.02	0.01	-0.05	0.22	**4.03**	**3.76**	0.77	-1.81	-0.95	0.20
	NUpE_2	5 H	A20553	2.81	3.18	A	G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
	NUpE_3	5 H	A21508	60.74	2.61	A	G	G	0.11	-0.09	**5.46**	-0.02	0.08	-0.32	1.40	-1.63	**2.19**	**0.02**	1.18	1.75	-0.01	0.06	-0.05	0.15	-0.44	9.84	19.72	4.57	**0.47**
	NUpE_4	5 H	I160288	129.41	3.13	G	A	A	0.11	0.08	3.55	-0.01	0.70	-0.15	0.05	1.66	2.12	**0.02**	0.75	**3.18**	-0.01	-0.09	0.18	1.97	2.24	-6.75	-1.34	2.04	0.37
Nitrogen Utilisation Efficiency in Grain	**NUtEg_1**	2 H	A20862	63.5	4.19	T	A	A	0.42	0.03	0.37	**0.08**	**-3.01**	**-2.43**	-1.35	**-3.93**	-1.52	0.00	**-2.53**	-1.68	0.01	0.07	0.13	0.66	0.60	33.72	46.37	0.87	-0.31
	NUtEg_2	4 H	I129218	92.4	5.04	C	A	C	0.42	-0.01	0.46	**-0.08**	-0.29	0.93	**1.71**	-1.60	**1.65**	0.00	**2.51**	0.96	0.00	**0.22**	0.11	-0.54	-0.99	**49.82**	**52.49**	3.83	**0.31**
	NUtEg_3	5 H	I49958	68.35	3.44	A	G	G	0.38	0.09	1.06	**-0.08**	0.43	0.62	1.50	-0.20	**1.59**	0.00	**2.21**	1.18	0.00	-0.03	0.05	1.78	1.40	-4.13	-5.54	4.32	**0.32**
	NUtEg_4	6 H	I124850	52.7	3.25	T	C	T	0.43	0.05	0.74	**-0.08**	-0.25	0.86	1.35	-2.56	1.62	0.00	**2.43**	0.73	-0.01	0.13	0.03	1.36	1.29	30.48	23.61	4.57	0.34
	NUtEg_5	7 H	I138457	34.82	3.33	C	A	A	0.22	0.09	**3.18**	**-0.07**	-0.14	1.22	1.32	-2.82	**2.10**	0.01	**2.22**	1.83	0.00	0.04	0.00	0.83	1.00	6.77	26.61	2.58	**0.43**
	NUtEg_6	7 H	I14119	161.4	2.54	A	G	A	0.33	0.02	1.39	-0.07	0.49	0.33	1.39	-1.09	1.32	0.00	**2.02**	0.70	0.00	-0.06	0.01	0.90	0.58	0.18	-3.03	2.24	0.26
Nitrogen Utilisation Efficiency total	**NUtEt_1**	3 H	I165444	99.89	2.68	A	G	G	0.44	-0.02	-1.73	0.02	-0.58	-0.60	0.18	-1.45	-1.09	-0.01	-0.48	**-1.74**	0.01	0.07	-0.01	-1.24	-0.64	1.53	10.18	0.53	-0.23
	NUtEt_2	3 H	I154449	155.9	3.26	A	C	C	0.29	-0.14	1.53	-0.02	1.19	1.24	0.20	2.13	1.28	0.01	1.03	**2.15**	-0.01	-0.10	-0.04	0.25	0.32	-17.82	-10.64	-1.88	0.22
	NUtEt_3	5 H	I160288	129.41	3.50	G	A	A	0.11	0.08	3.55	-0.01	0.70	-0.15	0.05	1.66	2.12	**0.02**	0.75	**3.18**	-0.01	-0.09	0.18	1.97	2.24	-6.75	-1.34	2.04	0.37
	NUtEt_4	5 H	I156273	176.62	3.15	A	G	G	0.19	0.16	2.17	-0.03	0.14	-0.55	-0.38	0.91	1.39	0.01	0.91	**2.62**	-0.01	-0.10	0.03	1.99	2.05	-14.23	-3.36	1.45	0.29
Stem Nitrogen	**StemN_1**	2 H	I177375	63.5	3.88	C	T	T	0.24	0.20	2.24	0.05	**-3.25**	**-2.84**	-0.27	**-6.23**	-0.56	0.01	-1.59	-1.23	**0.01**	0.12	0.11	1.99	1.84	38.98	53.73	2.57	-0.06
	StemN_2	5 H	A21318	53.18	2.64	G	A	A	0.15	0.02	0.63	0.02	0.66	1.79	-0.67	1.83	0.13	0.00	0.15	0.96	**-0.01**	-0.07	0.07	1.11	0.40	-18.91	-33.19	-0.79	-0.02
Stems	**Stems_1**	1 H	I182656	11.4	3.19	A	G	G	0.13	0.26	2.32	-0.03	1.33	0.55	0.19	1.46	1.10	0.01	1.00	1.47	-0.01	**-0.28**	0.09	2.68	**3.37**	-40.23	-56.42	0.02	0.23
	Stems_2	4 H	I129218	92.4	3.71	C	A	C	0.42	-0.01	0.46	**-0.08**	-0.29	0.93	**1.71**	-1.60	**1.65**	0.00	**2.51**	0.96	0.00	**0.22**	0.11	-0.54	-0.99	**49.82**	**52.49**	3.83	**0.31**
	Stems_3	5 H	B30975	6.4	2.55	A	C	C	0.17	-0.03	**3.56**	0.04	1.41	-0.05	-0.69	3.59	0.53	**0.02**	-1.04	1.24	-0.01	**-0.22**	0.06	1.37	1.62	**-66.86**	**-63.32**	0.82	0.13
	Stems_4	5 H	I148402	135.72	2.57	G	A	A	0.25	0.07	1.68	0.01	1.18	0.91	-0.22	2.62	0.18	0.00	0.08	0.80	-0.01	**-0.20**	0.08	1.46	1.63	-41.81	-50.02	1.64	0.04
Sterile Spikelets	**StS_1**	2 H	A10823	46.98	3.78	A	G	A	0.37	0.33	0.99	0.00	-0.78	-0.18	-0.04	-3.27	0.22	0.00	-0.03	0.20	0.00	0.03	**0.30**	0.91	0.70	-0.42	10.75	1.65	0.08
	StS_2	2 H	A10287	85.92	16.51	A	G	A	0.20	0.25	1.11	-0.04	0.21	-0.44	1.05	-2.16	0.86	0.00	1.22	-0.31	0.00	0.02	**0.93**	1.88	2.10	2.69	-19.36	5.69	0.16
	StS_3	4 H	I138835	87.5	3.15	G	A	A	0.26	0.18	1.28	0.00	0.08	-0.07	0.76	-1.58	0.80	0.00	0.74	0.09	0.00	-0.05	**0.28**	1.51	1.58	-25.96	-24.45	1.13	0.13
	StS_4	5 H	I12887	89.4	2.79	C	T	T	0.20	-0.11	-2.91	0.01	0.07	0.35	-0.33	2.05	-1.16	-0.01	-0.21	-1.01	0.00	0.10	**-0.32**	-1.38	-1.82	16.87	26.19	-2.73	-0.24
	StS_5	6 H	I131992	81.88	3.86	A	G	A	0.42	0.10	0.37	-0.02	0.14	-0.92	0.27	-0.65	0.21	0.00	0.48	0.37	-3.95	-0.12	**0.32**	1.01	1.46	-18.08	-15.97	2.59	0.07

Appendix 4.1 cont.

Trait	QTL	Chorm	Marker	Dist	-log10(p)	alleles 1	alleles 2	MA	MAF	Effect of marker allele 1																	
										Antho	GNYld	GrainN	Grains	Hd	HI	Ht	NUE	NupE	NutEg	NutEt	StemN	Stems	StS	TGW	TGW-GS	till.GS	till.Yld	UPOV	Yld
Thousand Grain Weight	TGW_1	1 H	I232660	18.05	3.24	C	T	C	0.36	0.09	-1.43	-0.02	0.30	0.48	-0.32	0.87	-0.19	-0.01	0.15	0.13	0.00	0.15	0.06	**-2.75**	**-3.25**	30.18	24.80	1.35	-0.03
	TGW_2	1 H	I128285	31.15	3.45	T	C	T	0.36	0.06	-1.26	-0.02	0.32	0.49	-0.24	0.38	-0.21	-0.01	0.17	0.08	0.00	0.10	0.06	**-2.64**	**-2.88**	22.25	20.90	0.73	-0.03
	TGW_3	1 H	A20810	52.46	3.56	A	G	A	0.46	-0.02	-1.62	0.00	0.47	0.11	-0.29	1.00	-0.49	-0.01	-0.18	-0.54	0.00	0.09	0.08	**-2.76**	**-3.20**	17.85	8.64	-0.07	-0.11
	TGW_4	2 H	I146936	6.4	3.33	T	C	C	0.18	0.11	-1.34	0.01	0.10	-0.09	0.02	-0.92	-0.71	0.00	-0.48	-1.07	0.00	0.13	-0.18	**-3.53**	**-4.01**	2.08	13.14	-0.69	-0.15
	TGW_5	2 H	A10733	54.95	4.33	G	C	C	0.14	0.05	**3.82**	0.01	-1.60	-1.08	0.42	-3.72	0.86	**0.02**	-0.70	-0.02	0.01	-0.05	0.22	**4.03**	**3.76**	0.77	-1.81	-0.95	0.20
	TGW_6	2 H	I195051	156.72	2.76	T	C	C	0.40	0.10	2.33	0.00	0.28	0.41	0.58	1.85	0.86	0.01	0.63	1.01	-0.01	-0.06	0.05	**2.71**	2.29	-11.73	-25.63	-0.96	0.16
	TGW_7	4 H	B30427	53.5	2.61	T	A	A	0.25	0.15	0.86	0.00	-0.57	0.03	0.16	-1.08	0.46	0.01	-0.06	0.66	0.00	-0.05	0.14	**2.45**	2.46	-2.91	-19.02	1.59	0.08
	TGW_8	5 H	A20553	2.81	3.44	A	G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
	TGW_9	5 H	I194030	166.63	2.83	A	G	A	0.38	-0.17	-1.36	0.03	-0.55	0.44	-0.92	0.02	-1.20	0.00	-1.15	-0.69	0.00	0.13	0.03	**-2.62**	-2.61	26.24	27.98	-3.63	-0.24
Thousand Grain Weight from grab samples	TGW-GS_1	1 H	I232660	18.05	3.42	C	T	C	0.36	0.09	-1.43	-0.02	0.30	0.48	-0.32	0.87	-0.19	-0.01	0.15	0.13	0.00	0.15	0.06	**-2.75**	**-3.25**	30.18	24.80	1.35	-0.03
	TGW-GS_2	1 H	I128285	31.15	3.19	T	C	T	0.36	0.06	-1.26	-0.02	0.32	0.49	-0.24	0.38	-0.21	-0.01	0.17	0.08	0.00	0.10	0.06	**-2.64**	**-2.88**	22.25	20.90	0.73	-0.03
	TGW-GS_3	1 H	A20810	52.46	3.63	A	G	A	0.46	-0.02	-1.62	0.00	0.47	0.11	-0.29	1.00	-0.49	-0.01	-0.18	-0.54	0.00	0.09	0.08	**-2.76**	**-3.20**	17.85	8.64	-0.07	-0.11
	TGW-GS_4	2 H	I213799	8.57	3.76	A	C	C	0.24	0.10	-0.75	0.01	-0.77	0.62	-0.69	-2.05	-0.30	0.00	-0.54	-0.01	0.00	0.05	-0.06	**-3.09**	**-3.94**	33.59	48.61	-0.11	-0.12
	TGW-GS_5	2 H	I143250	27.3	3.05	A	G	A	0.47	0.14	1.31	-0.01	0.46	0.03	0.19	1.77	0.61	0.00	0.52	0.56	-0.01	-0.10	0.03	2.16	**2.83**	-19.36	-23.52	3.30	0.12
	TGW-GS_6	2 H	A10602	58.24	3.09	A	C	C	0.26	-0.29	-1.12	0.05	-0.14	-1.15	-1.02	-0.96	-0.79	0.00	-1.34	-0.72	0.00	0.10	-0.16	**-2.92**	**-3.51**	21.60	25.85	-2.06	-0.20
	TGW-GS_7	5 H	A20553	2.81	3.75	A	G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
Tillering from Grab samples	till.GS_1	4 H	A20732	92.38	3.75	G	A	G	0.42	-0.03	0.31	**-0.08**	-0.51	0.65	1.45	-2.03	**1.56**	0.00	**2.47**	0.95	0.00	**0.22**	0.05	-0.86	-1.28	**55.67**	**57.53**	3.53	0.29
	till.GS_2	5 H	B30975	6.4	3.59	A	C	C	0.17	-0.03	**3.56**	0.04	1.41	-0.05	-0.69	3.59	0.53	**0.02**	-1.04	1.24	-0.01	**-0.22**	0.06	1.37	1.62	**-66.86**	**-63.32**	0.82	0.13
	till.GS_3	5 H	I147762	109.56	2.62	C	T	T	0.26	-0.11	1.29	-0.03	1.36	1.13	0.43	2.00	0.95	0.01	1.16	1.29	-0.01	-0.11	0.06	1.68	1.87	**-48.35**	-45.97	0.78	0.18
	till.GS_4	5 H	I720	159.8	3.26	A	G	A	0.37	-0.04	-0.58	0.01	-1.32	0.03	-0.88	-1.66	-0.33	0.00	-0.79	0.07	0.00	0.13	0.03	-1.30	-1.28	**55.52**	49.56	-1.91	-0.08
	till.GS_5	7 H	A10550	143.68	2.63	G	A	A	0.25	-0.18	-0.03	0.02	0.70	0.66	-0.38	1.61	-0.61	0.00	-0.68	-0.24	0.00	-0.13	-0.05	0.05	0.24	**-51.15**	-40.98	-1.60	-0.12
Tillering from yield	till.Yld_1	4 H	A21385	23.1	2.73	G	C	C	0.26	0.06	-0.46	-0.01	0.68	0.96	-0.37	1.99	-0.66	-0.01	-0.14	-0.71	-0.01	-0.09	0.00	0.69	0.52	-36.20	**-58.15**	-2.71	-0.12
	till.Yld_2	4 H	A20732	92.38	3.31	G	A	G	0.42	-0.03	0.31	**-0.08**	-0.51	0.65	1.45	-2.03	**1.56**	0.00	**2.47**	0.95	0.00	**0.22**	0.05	-0.86	-1.28	**55.67**	**57.53**	3.53	0.29
	till.Yld_3	5 H	A20553	2.81	3.05	A	G	G	0.20	0.02	**4.15**	0.01	1.07	-0.66	-0.15	2.66	1.17	**0.02**	-0.10	1.48	-0.01	-0.20	0.11	**3.04**	**3.68**	**-62.83**	**-65.67**	2.02	0.28
	till.Yld_4	6 H	I4707	81.2	2.81	C	T	T	0.33	-0.19	-0.65	-0.02	-1.10	0.83	0.39	-2.00	0.10	0.00	0.52	-0.62	0.00	0.12	-0.07	-1.53	-1.48	41.69	**59.19**	-1.19	-0.03
UPOV date of inscription	UPOV_1	2 H	I195164	96.8	3.80	C	T	T	0.48	**-1.04**	-0.67	0.06	-0.17	0.01	-0.80	1.78	-1.00	0.00	-1.26	-0.65	0.00	-0.12	-0.14	-0.79	-0.98	-16.74	-4.80	**-5.86**	-0.21
	UPOV_2	3 H	A10767	172.42	2.93	G	A	A	0.32	-0.14	-1.18	0.02	0.27	0.01	-0.80	3.01	-0.93	0.00	-0.84	-0.33	0.00	-0.02	-0.07	-1.26	-1.13	-14.42	-0.33	**-4.93**	-0.21
	UPOV_3	4 H	I128723	54.98	2.66	A	G	G	0.18	0.24	0.54	-0.07	0.36	2.13	0.71	0.26	0.99	0.00	1.54	1.05	0.00	0.00	0.10	1.29	1.41	1.51	3.53	**5.98**	0.17
	UPOV_4	5 H	I192396	19.4	2.96	T	A	T	0.46	0.30	1.15	-0.07	-0.13	-0.23	1.59	-2.98	1.41	0.00	1.99	0.07	0.00	0.04	0.02	0.55	0.27	25.76	15.74	**5.93**	0.31
	UPOV_5	5 H	I213753	64	2.98	C	A	A	0.25	0.04	0.63	-0.04	-0.14	-0.80	0.87	-0.61	0.81	0.00	0.99	0.28	0.00	0.00	0.03	0.52	0.38	-3.70	-3.38	**5.45**	0.15
	UPOV_6	6 H	I230959	4.9	3.12	G	T	T	0.23	0.26	1.28	-0.03	-0.06	0.76	0.60	-1.94	0.99	0.00	0.94	0.88	0.00	-0.09	0.11	2.00	2.01	-5.11	-5.74	**5.89**	0.19
	UPOV_7	6 H	I136897	53.29	4.21	A	G	G	0.23	0.14	-0.48	-0.04	-0.10	0.78	0.78	-0.51	0.90	0.00	1.74	0.74	-0.01	0.07	0.01	1.15	1.34	18.80	13.85	**7.92**	0.14
	UPOV_8	7 H	A11222	4.9	3.12	G	C	C	0.23	0.26	1.28	-0.03	-0.06	0.76	0.60	-1.94	0.99	0.00	0.94	0.88	0.00	-0.09	0.11	2.00	2.01	-5.11	-5.74	**5.89**	0.19
Yield treated	Yld_1	1 H	I154646	100.7	3.49	A	C	C	0.11	-0.30	**3.88**	-0.07	0.87	1.53	1.88	-0.27	**2.54**	0.01	2.23	2.14	0.00	0.08	-0.04	1.12	0.10	20.37	16.07	1.67	**0.53**
	Yld_2	2 H	A21304	33.74	2.73	A	G	G	0.30	-0.08	-2.53	0.06	0.14	-0.22	-1.16	2.11	**-1.83**	-0.01	-1.84	-1.27	0.00	0.03	-0.11	-1.81	-2.01	2.65	-14.92	-3.94	**-0.37**
	Yld_3	2 H	I10398	54.95	3.33	C	T	T	0.17	-0.27	-2.20	**0.11**	-0.99	-0.10	**-2.79**	2.79	**-2.35**	-0.01	**-3.06**	-0.98	-0.01	-0.06	-0.09	-1.90	-1.70	1.54	7.85	-1.73	**-0.54**
	Yld_4	4 H	I182626	96.6	3.08	T	G	G	0.31	0.22	1.05	**-0.08**	-0.27	0.93	0.95	-3.04	**1.88**	0.01	1.65	1.65	0.00	0.16	0.16	0.02	-0.10	43.44	45.13	3.17	**0.37**
	Yld_5	5 H	I231238	63.31	3.55	T	C	C	0.21	0.12	**3.69**	-0.05	0.30	-0.47	**1.81**	-2.80	**2.19**	0.01	1.71	1.53	0.00	0.01	0.08	1.40	0.72	8.69	1.25	5.10	**0.45**
	Yld_6	5 H	A20236	80.61	2.60	C	A	A	0.26	-0.08	**2.90**	-0.04	0.51	-0.54	1.16	-0.59	1.45	0.01	1.24	0.57	0.00	-0.07	0.07	0.44	0.16	-7.50	-5.86	2.73	**0.33**
	Yld_7	6 H	I118381	54.6	2.69	C	T	C	0.47	-0.01	1.20	-0.07	-0.37	0.58	1.00	-1.96	**1.62**	0.00	1.92	1.04	0.00	0.10	0.04	1.68	1.69	28.22	26.87	4.68	**0.34**
	Yld_8	7 H	I186187	14.96	3.66	T	G	G	0.22	0.34	**3.07**	-0.07	0.73	0.27	1.57	-1.43	**1.91**	0.01	1.82	1.29	0.00	0.05	0.16	0.61	0.94	7.50	14.85	4.52	**0.44**
	Yld_9	7 H	I138457	34.82	3.94	C	A	A	0.22	0.09	**3.18**	**-0.07**	-0.14	1.22	1.32	-2.82	**2.10**	0.01	**2.22**	1.83	0.00	0.04	0.00	0.83	1.00	6.77	26.61	2.58	**0.43**

Appendix 4. 2 Effect size and direction of QTL peak marker effect across the set phenotypes in AGOUEB GWAS.

Each QTL is presented with its peak marker (most significant SNP) with the associated marker map position, –log10(*p*) resulting from the test for association, the peak SNP alleles with its minor allele (MA) and minor allele frequency (MAF) in the AGOUEB panel. SNP effect size and direction are reported across the 20 agronomic traits analysed. The negative effects of the allele 1 on another trait are highlighted in grey. The effects associated with a significant marker association for the trait considered (i.e. a QTL) are highlighted in bold (e.g. SNP A10823 of Antho_1 is a QTL for both traits Anthocyanin and sterile spikelets).

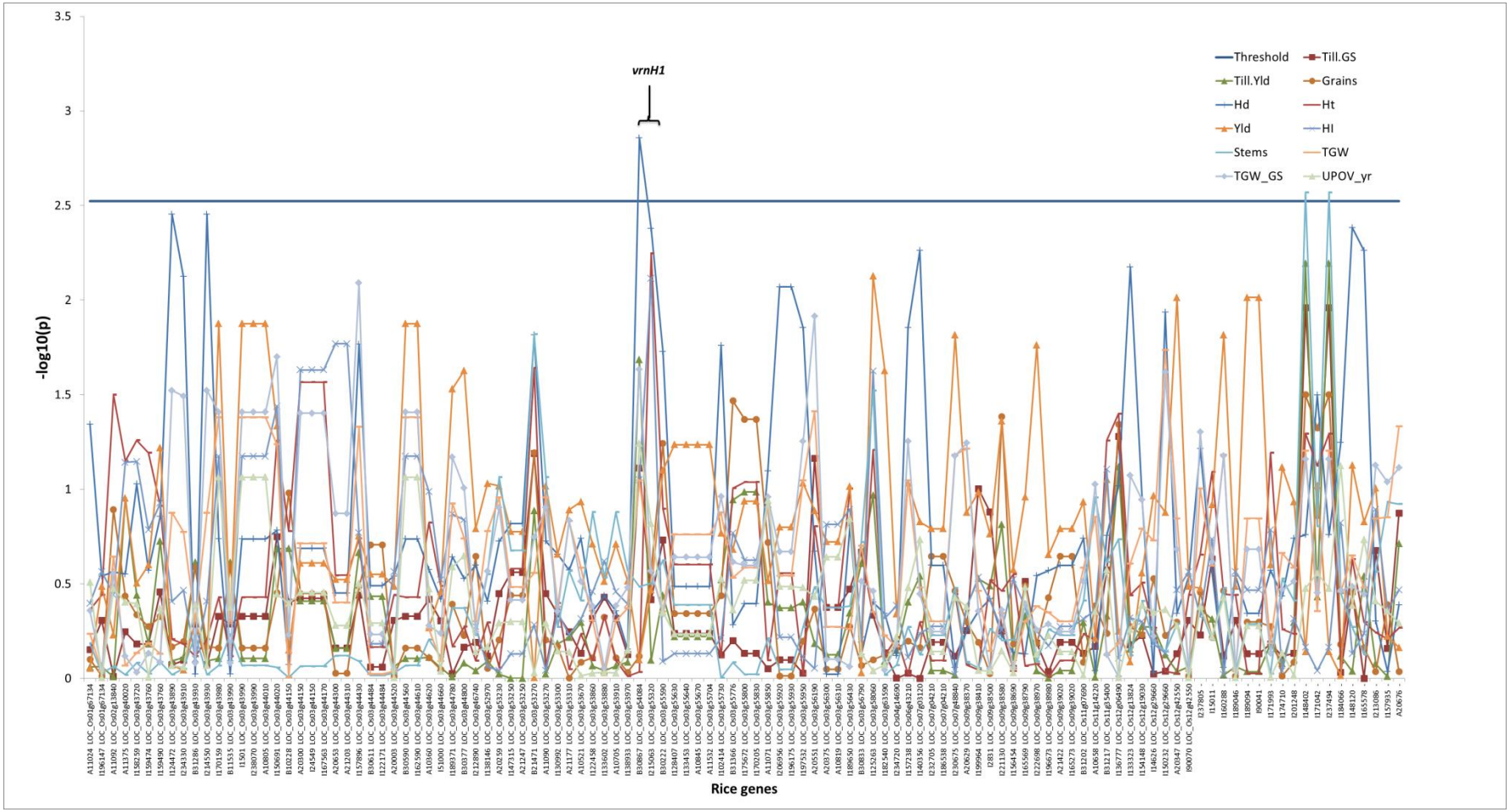
Trait	QTL	Chrom	Marker	Dist	-log10(p)	alleles 1	alleles 2	MA	MAF	Effect of marker allele 1																		
										aleurone	Antho	BR	EAR-G	Grain-N	Hd	HLW	HWE	LLHS	mild	NB	ryncho	SL_T	SL_U	SS_U	StS	TGW	VFH	WH	Yld_T
Aleurone	aleurone_1	2 H	I171032	83.82	**2.87**	G	A	A	0.28	**-0.28**	0.06	0.00	0.70	0.00	0.35	0.11	1.64	0.01	-0.12	0.33	-1.76	0.39	1.31	-1.11	-0.57	0.56	-0.04	0.00	0.14
	aleurone_2	4 H	A21087	62.10	**21.61**	G	A	G	0.41	**-0.93**	0.05	0.23	0.09	0.00	0.13	-0.21	-0.04	0.10	-0.15	0.16	0.06	0.01	-0.29	-0.73	-0.15	0.11	-0.01	0.02	0.00
Anthocyanine	Antho_1	2 H	A10823	46.98	**3.15**	G	A	A	0.29	-0.03	**-0.34**	-0.20	-0.45	0.00	0.14	-0.08	-0.22	0.01	0.20	0.30	1.48	0.81	0.57	1.79	**1.25**	-1.49	-0.10	-0.01	-0.15
Color	Antho_2	2 H	A10138	96.82	**20.83**	G	A	A	0.45	-0.04	**-0.93**	-1.22	0.28	0.01	0.22	-0.07	-0.37	0.06	-0.85	0.03	0.44	0.74	1.14	-0.09	0.23	0.07	-0.12	-0.05	-0.06
Brown Rust	BR_1	2 H	I146785	70.50	**4.20**	G	C	G	0.13	-0.13	-0.07	**3.21**	0.04	-7.94	-1.01	-0.01	-1.07	-0.05	-0.17	0.19	-1.14	-0.17	0.11	-0.04	-0.29	0.53	0.00	0.01	0.02
Ear Glaucosity	EAR-G_1	1 H	I120059	0.75	**3.01**	T	C	T	0.30	-0.02	0.03	-0.11	**1.05**	0.00	0.12	0.27	0.43	0.10	0.01	0.32	-0.24	0.63	0.97	0.46	0.26	0.05	-0.03	-0.04	-0.05
	EAR-G_2	5 H	A10524	93.70	**2.54**	C	A	A	0.11	-0.09	0.02	-0.88	**1.33**	0.00	0.83	0.00	1.37	0.11	0.18	0.28	0.32	0.27	0.52	-1.17	-0.21	0.33	-0.04	0.03	0.11
	EAR-G_3	6 H	B11455	42.36	**2.69**	G	A	A	0.40	0.12	-0.12	0.08	**1.08**	0.02	-0.48	0.01	-1.64	-0.02	0.37	0.63	-0.41	0.40	0.62	-0.80	0.06	0.83	0.12	-0.01	-0.09
	EAR-G_4	7 H	I1347	116.33	**3.44**	G	T	T	0.21	-0.10	-0.12	-1.02	**1.31**	0.00	0.71	-0.08	0.37	0.10	-0.01	-0.12	-1.19	1.76	1.74	2.87	0.23	-0.56	-0.06	0.01	0.02
Grain Nitrogen	GrainN_1	2 H	A11384	60.68	**4.23**	C	G	G	0.42	0.03	0.03	0.55	-0.35	**0.04**	-0.72	0.52	-0.53	-0.03	0.63	-0.46	0.05	-1.78	**-3.16**	**-4.08**	-0.71	1.21	-0.09	0.03	0.02
	GrainN_2	4 H	I149873	0.74	**3.11**	G	A	A	0.29	0.10	0.06	0.90	0.31	**-0.03**	-0.47	0.00	1.40	-0.04	-0.18	-0.02	-0.10	-1.03	-1.13	-0.96	-0.28	0.74	-0.02	0.02	0.16
	GrainN_3	4 H	B31362	73.57	**2.67**	C	A	A	0.29	-0.14	-0.11	0.39	0.09	**0.03**	-0.10	0.44	-0.47	0.05	0.16	-0.17	-0.34	-1.32	0.07	-1.16	0.12	0.21	-0.06	0.01	0.08
	GrainN_4	5 H	B30975	6.40	**4.86**	A	C	C	0.11	-0.03	-0.11	-0.42	-0.04	**0.05**	-0.07	0.52	-0.73	-0.01	-0.87	0.32	-0.89	3.09	**4.63**	-1.14	0.22	2.13	-0.07	0.06	0.06
	GrainN_5	5 H	B30400	149.10	**2.57**	A	C	A	0.37	-0.03	-0.05	0.96	-0.07	**0.03**	-0.77	0.38	-0.55	-0.03	0.09	-0.05	-0.42	-0.24	-0.01	-1.83	-0.12	1.71	-0.01	0.03	-0.04
Heading date	Hd_1	2 H	B30042	59.21	**2.70**	A	G	G	0.20	-0.06	-0.01	1.17	-0.06	0.03	**-1.48**	0.29	-0.07	-0.02	0.42	-0.37	-0.95	**-3.60**	-2.91	-2.97	0.02	**2.54**	-0.05	-0.01	0.02
	Hd_2	2 H	B10937	152.79	**3.13**	G	C	G	0.15	0.00	0.12	1.37	0.38	-0.01	**-1.69**	0.24	-0.95	0.12	-0.06	0.31	0.65	-1.11	-1.77	-1.45	0.11	-0.08	-0.09	0.03	0.03
Hectoliter weight	HLW_1	6 H	B30025	117.68	**2.60**	A	G	G	0.40	0.03	-0.02	0.53	-0.06	0.00	-0.47	**-0.72**	-0.05	0.06	-0.88	0.97	-0.05	-1.15	-1.11	-0.26	-0.02	-0.35	0.04	-0.03	-0.02
Hot Water	HWE_1	1 H	A10985	52.46	**4.75**	A	C	A	0.49	0.03	0.05	0.48	0.11	-0.01	-0.40	0.15	**3.89**	0.06	-0.03	0.00	-0.43	0.71	0.75	-0.05	-0.32	-1.24	-0.01	0.04	0.03
Extract	HWE_2	1 H	I165338	131.15	**2.53**	G	A	A	0.11	0.11	0.13	-0.23	-0.07	-0.01	-0.06	0.43	**3.41**	0.06	-0.30	-0.45	-0.34	0.25	1.27	0.28	-0.29	-0.59	-0.07	0.02	0.05
	HWE_3	2 H	I118168	9.28	**2.67**	A	T	T	0.50	0.15	-0.04	0.30	0.02	-0.01	-0.35	-0.08	**2.84**	0.00	0.85	0.43	0.24	0.32	-0.38	-0.48	0.28	-1.08	0.08	0.02	-0.02
LLHS	LLHLS_1	4 H	A10611	114.66	**8.71**	C	A	A	0.12	0.01	-0.05	0.29	0.37	-0.02	0.17	-0.11	2.23	**0.52**	0.10	0.18	0.80	-1.24	-1.40	2.48	0.44	-0.22	-0.06	0.02	-0.04
	LLHLS_2	5 H	A21355	153.50	**6.53**	A	G	G	0.11	-0.16	-0.09	-0.07	0.28	-0.01	-0.23	-0.23	2.84	**0.44**	-0.55	0.32	-1.29	-0.24	-1.67	1.71	0.65	1.60	-0.09	0.00	-0.05
	LLHLS_3	7 H	A20365	166.56	**3.20**	C	G	G	0.18	-0.05	0.13	-0.10	-0.35	0.00	-0.22	0.38	1.59	**0.26**	0.68	0.22	0.57	0.59	0.05	0.18	0.40	0.08	-0.18	-0.03	0.01
Mildew	mild-(U)_1	4 H	I128147	86.27	**2.56**	T	C	C	0.32	-0.13	-0.02	-1.35	0.72	0.00	0.28	-0.30	0.69	0.03	**-1.70**	0.53	0.02	0.99	0.90	0.40	-0.02	0.00	0.00	-0.01	-0.05
Untreated	mild-(U)_2	5 H	I108541	19.40	**5.21**	C	G	G	0.18	0.18	0.09	-1.00	-0.17	0.00	0.34	-0.26	0.40	0.03	**-3.07**	0.45	0.59	-1.12	-0.68	-0.93	0.36	-0.07	0.02	0.01	0.09
	mild-(U)_3	5 H	I204494	51.30	**2.74**	C	A	A	0.13	-0.02	0.08	0.94	-0.01	-0.02	0.25	-0.41	0.98	-0.07	**-2.36**	0.31	-1.34	-0.06	0.08	-0.29	-0.21	-1.22	-0.04	0.09	0.03
	mild-(U)_4	6 H	I147090	33.74	**2.72**	T	C	C	0.23	-0.08	0.04	-1.12	0.05	0.02	0.01	0.05	0.46	0.06	**-1.89**	0.13	-0.37	0.81	1.94	-0.03	0.04	-0.09	-0.17	0.00	0.06
	mild-(U)_5	6 H	I164156	90.15	**2.82**	T	C	C	0.39	-0.09	-0.10	-0.40	-0.21	-0.01	-0.13	-0.27	-0.38	0.03	**-1.80**	0.22	0.53	-0.78	-0.04	-0.68	-0.03	-0.03	-0.01	0.00	0.08
	mild-(U)_6	7 H	I163976	29.82	**3.72**	T	C	C	0.13	-0.14	0.23	0.57	-0.43	-0.01	0.04	-0.08	-0.82	0.06	**-2.87**	-0.45	-0.52	1.01	1.54	0.51	-0.14	-0.28	-0.01	0.06	0.01
Net Blotch	NB(U)_1	2 H	I16024	71.12	**2.95**	C	T	C	0.48	0.01	0.04	0.30	-0.04	0.00	0.68	-0.32	-0.08	-0.07	-0.55	**1.19**	0.01	1.32	1.24	1.58	0.03	-1.11	-0.04	0.02	-0.09
Untreated	NB(U)_2	4 H	B10063	40.36	**3.05**	G	A	A	0.16	0.06	-0.10	0.70	-0.29	-0.01	-0.30	-0.41	-1.44	0.02	0.69	**-1.67**	0.66	-0.07	-0.43	0.50	-0.11	0.07	0.16	-0.04	0.13
	NB(U)_3	6 H	I128460	45.40	**5.01**	T	C	C	0.30	0.01	-0.05	0.72	0.20	0.00	-0.20	-0.05	-0.26	-0.08	-0.30	**1.79**	0.87	0.13	-0.70	-0.62	0.20	-0.85	0.10	0.04	-0.06
Rynchosporium	ryncho(U)_1	2 H	I129821	158.15	**4.99**	T	G	G	0.19	-0.03	0.06	0.61	-0.33	-0.01	-0.30	-0.51	0.63	-0.03	0.99	0.02	**-3.57**	-1.01	-0.30	2.39	0.07	-0.92	0.01	0.01	-0.04
Untreated	ryncho(U)_2	3 H	A20252	6.03	**3.52**	G	A	A	0.31	-0.02	-0.20	0.57	0.17	-0.01	0.27	-0.41	-0.13	-0.03	0.47	-0.24	**-2.45**	0.15	-0.14	-0.98	0.11	-0.16	-0.01	0.00	0.11
	ryncho(U)_3	5 H	B30456	113.11	**3.21**	A	G	G	0.26	-0.10	0.05	-0.21	-0.23	0.00	-0.36	0.02	-0.94	-0.08	0.59	-0.06	**2.42**	-0.20	-1.28	-1.08	0.31	0.20	-0.02	0.06	-0.05

Appendix 4. 2 cont.

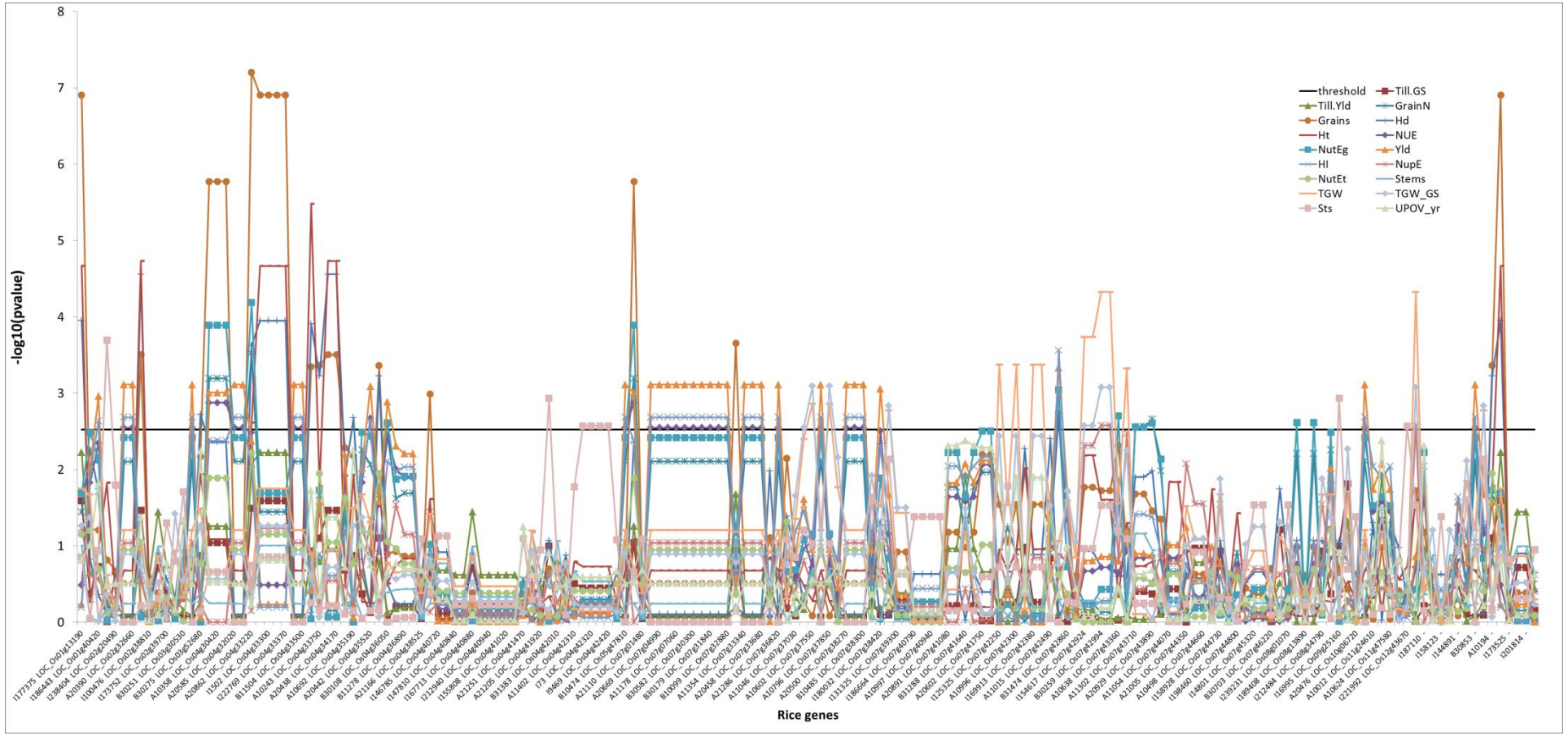
Trait	QTL	Chrom	Marker	Dist	-log10(p)	alleles 1	alleles 2	MA	MAF	Effect of marker allele 1																			
										aleurone	Antho	BR	EAR-G	Grain-N	Hd	HLW	HWE	LLHS	mild	NB	ryncho	SL_T	SL_U	SS_U	StS	TGW	VFH	WH	Yld_T
Straw Length	SL_T_1	2 H	I127347	63.53	**4.07**	T	C	C	0.28	-0.07	0.07	0.90	-0.21	0.03	-1.24	0.27	-0.42	-0.01	0.52	-0.54	0.39	**-3.74**	**-4.45**	**-3.88**	-0.26	**2.26**	-0.11	0.02	0.03
Treated	SL_T_2	5 H	I136777	123.08	**2.69**	C	T	T	0.32	-0.07	-0.08	0.78	0.68	0.00	0.89	0.04	0.99	0.08	-0.08	0.39	-0.84	**2.42**	**2.70**	1.88	0.14	-0.48	0.03	-0.01	0.01
	SL_T_3	5 H	A10236	181.43	**3.44**	A	G	A	0.14	-0.23	0.15	0.44	-0.24	-0.01	-0.58	-0.25	1.65	-0.04	0.21	0.82	0.26	**-4.21**	**-4.90**	-2.88	**-1.25**	0.61	-0.01	0.02	0.10
Straw Length Untreated	SL_U_1	2 H	I177375	63.50	**4.57**	C	T	T	0.28	-0.11	0.11	0.76	-0.27	**0.03**	-1.34	0.32	-0.05	0.00	0.53	-0.51	0.28	**-3.68**	**-4.87**	**-4.04**	-0.34	**2.20**	-0.07	0.03	0.04
	SL_U_2	4 H	I129218	92.40	**3.95**	A	C	C	0.45	-0.02	-0.13	-0.46	0.47	0.00	0.07	0.47	-0.01	-0.05	-0.56	0.28	0.39	1.90	**3.45**	-0.74	0.03	1.34	-0.05	0.00	**-0.17**
	SL_U_3	5 H	B30975	6.40	**3.45**	A	C	C	0.11	-0.03	-0.11	-0.42	-0.04	**0.05**	-0.07	0.52	-0.73	-0.01	-0.87	0.32	-0.89	3.09	**4.63**	-1.14	0.22	2.13	-0.07	0.06	0.06
	SL_U_4	5 H	I136777	123.08	**2.58**	C	T	T	0.32	-0.07	-0.08	0.78	0.68	0.00	0.89	0.04	0.99	0.08	-0.08	0.39	-0.84	**2.42**	**2.70**	1.88	0.14	-0.48	0.03	-0.01	0.01
	SL_U_5	5 H	A10236	181.43	**3.31**	A	G	A	0.14	-0.23	0.15	0.44	-0.24	-0.01	-0.58	-0.25	1.65	-0.04	0.21	0.82	0.26	**-4.21**	**-4.90**	-2.88	**-1.25**	0.61	-0.01	0.02	0.10
	SL_U_6	6 H	A20745	28.39	**2.62**	C	A	C	0.39	0.03	0.00	0.79	0.12	0.00	-0.29	-0.12	0.92	0.08	0.55	-0.07	0.54	-1.57	**-2.86**	0.17	0.20	-0.38	0.10	-0.01	-0.05
Straw Strenght	SS_U_1	1 H	I3336	15.74	**2.82**	A	C	C	0.23	-0.10	-0.01	0.17	-0.11	-0.01	-0.09	-0.12	0.31	-0.01	-0.75	-0.23	0.26	-0.36	-0.34	**-3.26**	-0.10	-0.02	-0.03	-0.04	0.03
Untreated	SS_U_2	2 H	A10358	59.21	**4.14**	C	A	A	0.42	0.02	0.02	0.57	-0.35	**0.04**	-0.79	0.50	-0.36	0.01	0.65	-0.45	0.16	-1.92	**-3.30**	**-4.06**	-0.58	1.17	-0.08	0.03	0.00
	SS_U_3	3 H	I155763	83.23	**3.34**	A	G	G	0.18	-0.05	0.05	0.27	-0.08	-0.01	0.86	-0.30	2.35	0.20	-0.52	0.28	-1.28	0.66	0.68	**4.25**	-0.17	0.39	0.01	0.01	-0.04
	SS_U_4	4 H	A21035	113.92	**2.81**	A	G	A	0.21	0.07	0.07	-0.38	-0.17	0.01	-0.24	-0.05	-0.24	**-0.29**	-0.25	0.31	-0.23	-1.38	-0.77	**-3.40**	-0.39	0.25	0.01	-0.02	0.03
	SS_U_5	7 H	I138111	58.57	**2.55**	C	G	G	0.28	0.03	-0.06	-0.03	-0.15	-0.01	-0.30	0.07	-0.48	-0.03	-0.15	0.01	0.94	-1.11	-0.76	**-2.94**	-0.12	-0.25	0.02	0.04	0.06
	SS_U_6	7 H	I150049	104.78	**3.37**	T	C	C	0.37	0.01	0.04	0.07	-0.22	-0.01	-0.02	-0.15	1.31	-0.07	-0.31	0.08	0.40	-1.02	-1.74	**-3.50**	0.18	0.14	0.03	0.02	0.06
Sterile	StS_1	1 H	A21333	59.71	**3.13**	G	C	C	0.46	-0.08	0.13	-0.31	-0.12	0.01	0.26	-0.18	-1.13	-0.01	-0.52	-0.22	-0.50	-0.50	-0.20	-0.58	**-0.93**	1.06	0.04	0.01	0.08
Spikelets	StS_2	2 H	A10823	46.98	**3.60**	G	A	A	0.29	-0.03	**-0.34**	-0.20	-0.45	0.00	0.14	-0.08	-0.22	0.01	0.20	0.30	1.48	0.81	0.57	1.79	**1.25**	-1.49	-0.10	-0.01	-0.15
	StS_3	2 H	A10287	85.92	**13.98**	G	A	A	0.17	0.05	-0.27	-0.49	0.41	-0.01	0.52	-0.11	-0.57	0.04	-0.12	0.43	2.47	0.25	0.25	2.36	**3.23**	-0.80	0.02	-0.03	-0.21
	StS_4	4 H	A10319	8.25	**2.68**	A	G	G	0.24	0.11	0.04	-0.55	-0.07	-0.01	0.40	-0.35	-0.30	0.05	-0.60	0.36	0.19	0.10	-0.12	1.89	**0.87**	-0.32	-0.08	0.04	-0.03
	StS_5	5 H	A10236	181.43	**2.59**	A	G	A	0.14	-0.23	0.15	0.44	-0.24	-0.01	-0.58	-0.25	1.65	-0.04	0.21	0.82	0.26	**-4.21**	**-4.90**	-2.88	**-1.25**	0.61	-0.01	0.02	0.10
	StS_6	6 H	I207933	4.41	**2.76**	C	G	C	0.32	0.09	-0.09	0.20	0.27	0.00	0.51	-0.12	0.71	-0.08	0.39	-0.08	0.27	0.09	0.42	0.00	**0.83**	-0.40	0.20	-0.03	-0.05
	StS_7	6 H	I204148	60.23	**2.72**	T	C	C	0.28	-0.11	0.02	-0.63	-0.05	0.01	-0.16	0.50	0.44	-0.06	0.66	0.50	1.15	0.38	-0.19	-0.92	**-0.97**	0.15	-0.09	0.07	0.00
Thousand	TGW_1	1 H	I184784	40.99	**2.64**	T	C	T	0.44	0.08	0.03	0.60	0.04	-0.01	-0.38	0.15	**3.49**	0.10	-0.38	0.32	-0.02	0.37	-0.27	0.08	0.06	**-1.86**	0.06	0.06	-0.01
Grain Weight	TGW_2	2 H	I110647	31.00	**2.58**	T	C	C	0.18	-0.13	-0.12	-0.73	0.19	0.01	-0.23	-0.01	1.39	0.09	-0.35	0.63	-1.02	-0.84	-1.00	-1.39	-0.05	**2.23**	0.01	0.06	0.05
	TGW_3	2 H	B30042	59.21	**3.32**	A	G	G	0.20	-0.06	-0.01	1.17	-0.06	0.03	**-1.48**	0.29	-0.07	-0.02	0.42	-0.37	-0.95	**-3.60**	-2.91	-2.97	0.02	**2.54**	-0.05	-0.01	0.02
	TGW_4	4 H	I160461	103.10	**2.68**	T	C	C	0.46	0.04	0.01	-0.68	-0.29	0.01	0.31	0.20	-0.66	-0.07	-0.26	-0.62	0.34	0.13	0.93	-0.84	-0.07	**1.77**	-0.07	0.01	0.03
	TGW_5	5 H	A20553	2.81	**3.90**	A	G	G	0.22	0.07	-0.01	0.30	-0.33	**0.03**	-0.50	0.29	-1.17	-0.12	0.38	-0.47	-0.91	1.76	2.01	-2.43	0.13	**2.53**	-0.04	0.03	0.09
Ventral Furrow Hairs	VFH_1	2 H	I152485	101.78	**2.61**	G	A	A	0.27	-0.02	0.18	0.34	0.69	9.33	0.11	0.12	0.57	-0.04	1.39	0.54	0.23	0.60	-0.02	0.65	0.50	-0.20	**0.23**	0.06	-0.06
	VFH_2	6 H	I194036	6.07	**12.50**	A	C	C	0.35	0.08	-0.01	0.45	-0.04	0.00	0.10	-0.58	0.34	-0.06	1.10	-0.14	0.08	1.36	0.97	1.49	0.56	0.27	**0.55**	0.01	-0.02
Winter	WintHard_1	2 H	A21261	28.44	**3.01**	G	A	A	0.19	-0.18	-0.03	-0.76	-0.12	0.00	-0.14	0.13	0.66	0.16	0.73	0.61	-0.31	0.76	0.44	0.04	-0.35	1.46	-0.10	**0.12**	0.02
Hardiness (WH)	WintHard_2	4 H	I110333	65.80	**2.80**	T	C	T	0.25	-0.08	-0.07	0.51	-0.51	0.00	0.65	-0.45	0.28	-0.08	0.72	-0.21	-1.74	0.25	0.25	-0.12	-0.20	1.07	-0.02	**0.10**	0.01
	WintHard_3	5 H	I214760	18.72	**2.93**	G	A	A	0.42	-0.02	0.04	-0.81	-0.27	0.01	0.34	0.31	-1.18	-0.09	-0.04	0.19	-0.64	-0.55	-0.88	-1.30	-0.14	-0.02	-0.05	**0.08**	-0.01
	WintHard_4	6 H	I114351	56.48	**2.84**	C	T	T	0.32	0.09	-0.03	0.97	0.25	0.01	0.03	-0.51	-1.32	-0.09	0.23	-0.02	-1.09	0.92	0.02	0.23	0.09	-0.43	0.02	**-0.10**	-0.05
Yield	Yld_T_1	4 H	I150603	48.72	**2.73**	G	A	A	0.14	-0.29	0.11	-0.56	-0.44	-0.01	0.20	-0.74	-0.02	0.02	-0.81	-0.93	0.67	0.28	-0.09	-1.20	-0.24	-0.39	0.02	-0.01	**0.23**
Treated	Yld_T_2	4 H	I129218	92.40	**2.89**	A	C	C	0.44	-0.02	-0.13	-0.46	0.47	0.00	0.07	0.47	-0.01	-0.05	-0.56	0.28	0.39	1.90	**3.45**	-0.74	0.03	1.34	-0.05	0.00	**-0.17**
	Yld_T_3	5 H	I192396	19.40	**4.17**	A	T	A	0.29	-0.13	-0.13	1.67	0.27	0.01	-0.71	0.19	-0.38	-0.01	**2.57**	-0.36	-0.56	1.48	1.35	1.65	0.11	0.67	0.00	-0.02	**-0.23**
	Yld_T_4	5 H	I205853	86.63	**2.58**	T	G	T	0.16	-0.03	0.00	0.46	-0.12	0.02	-0.21	0.30	-1.41	-0.13	-1.03	0.61	-0.44	-0.27	0.38	0.08	0.62	0.50	-0.07	-0.02	**-0.20**
	Yld_T_5	6 H	I115369	55.90	**2.73**	T	C	C	0.45	0.12	-0.12	0.80	0.62	0.03	-0.19	0.02	-0.98	0.05	0.27	0.17	-0.48	1.58	2.04	1.53	0.20	-0.34	0.01	-0.03	**-0.20**
	Yld_T_6	7 H	B30380	138.17	**3.28**	G	A	G	0.42	0.08	-0.02	0.18	-0.36	-0.01	0.11	-0.20	0.58	0.08	-0.58	0.06	0.47	0.14	0.48	-1.24	-0.19	0.23	-0.08	0.01	**0.19**

Appendix 4. 3 Patterns of MTA for the QTL cluster at *vrn-H1* locus on chromosome 5HL identified from NUE-CROPs GWAS.

Patterns of marker trait associations for all 20 traits of the NUE-CROPS GWAS at the QTL cluster identified on 5HL at the *vrn-H1* candidate gene locus. The X-axis plots the barley SNP ordered according the physical position of their homologous rice locus. (e.g. SNP A11024 is homologous of rice locus LOC_Os01g67134 found at the rice gene 67134 on Os01). The Y-axis indicates the magnitude of the association –log10(*p*).



Appendix 4. 4 Patterns of MTA for the QTL cluster at *eam6* locus on the centromere of chromosome 2H identified from NUE-CROPs GWAS.
Patterns of marker trait associations for all 20 traits of the NUE-CROPS GWAS at the QTL cluster identified on 2H centromere at the *eam6* and *HvCEN* candidate gene locus. The X-axis plots the barley SNP ordered according the physical position of their homologous rice locus. (e.g. SNP A177375 is homologous of rice locus LOC_Os01g0113190 found at the rice gene 0113190 on Os01). The Y-axis indicates the magnitude of the association as –log10(*p*).

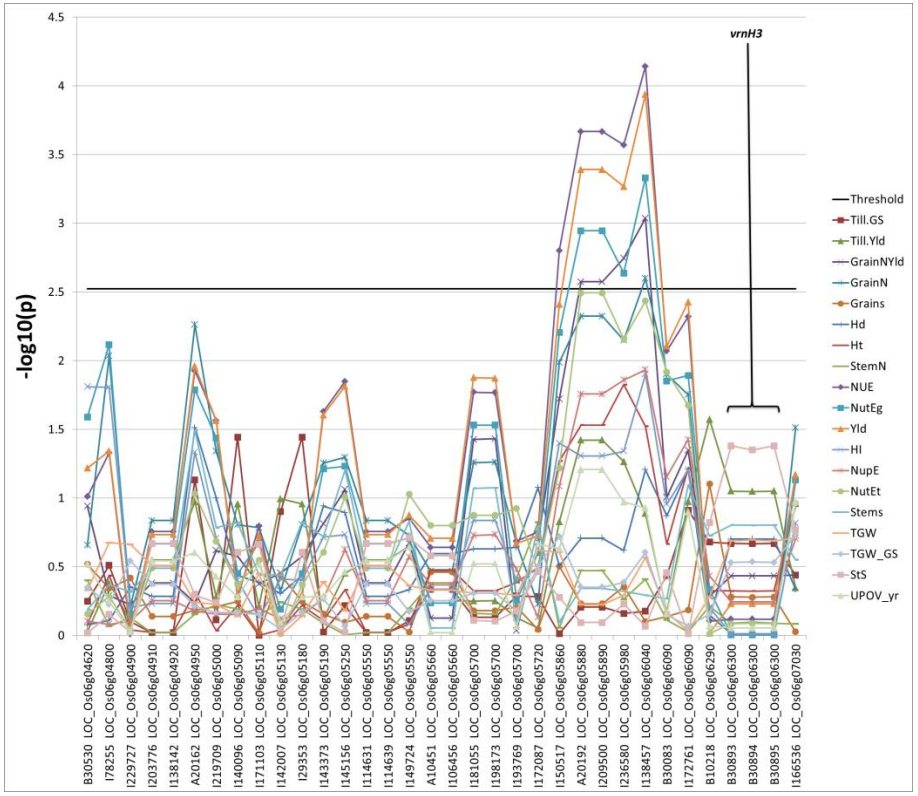


Appendix 4. 5 Patterns of MTA for the QTL cluster at the *vrn-H3* locus on chromosome 7H identified from NUE-CROPs GWAS and the corresponding putative rice homologous genes.

(a) Patterns of marker trait associations for all 20 traits of the NUE-CROPS GWAS at the QTL cluster identified on 7H at the *vrn-H3* candidate gene locus. The X-axis plots the barley SNP ordered according the physical position of their homologous rice locus. The Y-axis indicates the magnitude of the association.

(b) List of the putative rice homologous genes within the chromosome segment delimited by the significant barley SNP of the cluster.

a)

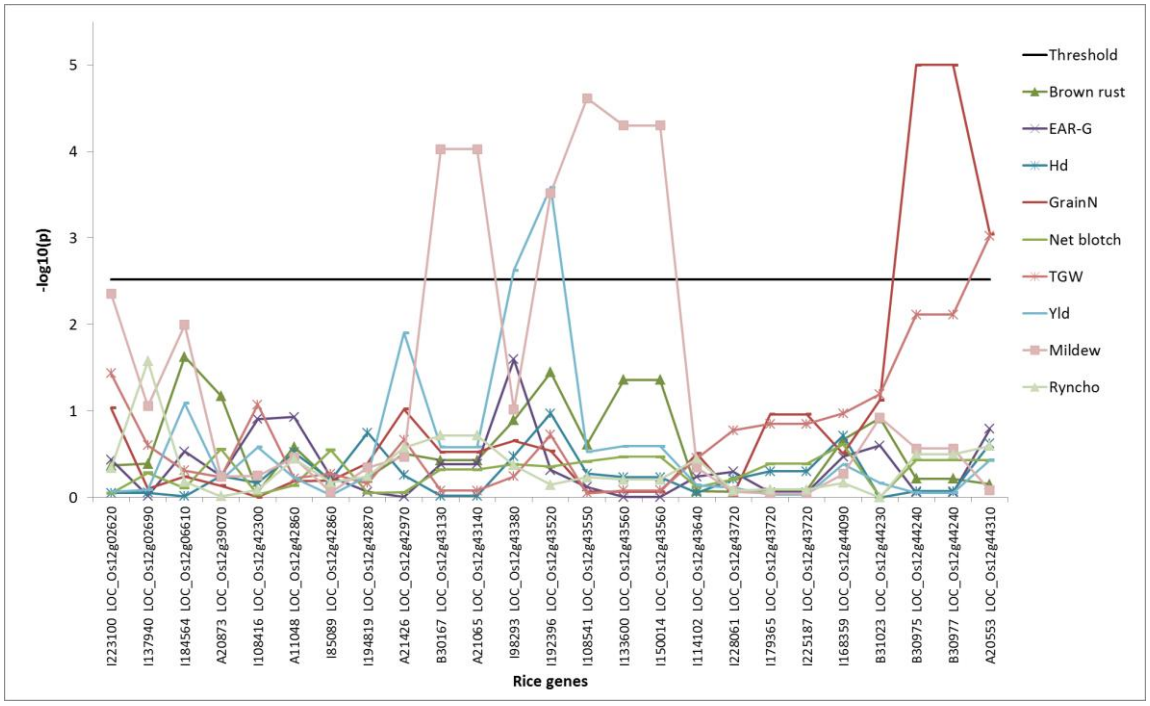


b)

RiceLocus	Gene product name	Barley SNP	SNP distance
LOC_Os06g05720	expressed protein	I172087	29.82
LOC_Os06g05730	expressed protein		
LOC_Os06g05740	expressed protein		
LOC_Os06g05750	transferase family domain containing protein, expressed		
LOC_Os06g05760	ubiquitin family protein, putative, expressed		
LOC_Os06g05770	expressed protein		
LOC_Os06g05790	transferase family domain containing protein, expressed		
LOC_Os06g05800	RNA recognition motif containing protein, putative, expressed		
LOC_Os06g05804	trafficking protein particle complex subunit 4, putative, expressed		
LOC_Os06g05820	OsLonP2 - Putative Lon protease homologue, expressed		
LOC_Os06g05830	protein kinase domain containing protein, expressed		
LOC_Os06g05860	6-phosphofructokinase, putative, expressed	I150517	34.82
LOC_Os06g05870	dual specificity protein phosphatase, putative, expressed		
LOC_Os06g05880	profilin domain containing protein, expressed	A20192	34.82
LOC_Os06g05890	B-box zinc finger family protein, putative, expressed	I209500	34.8
LOC_Os06g05900	methyltransferase, putative, expressed		
LOC_Os06g05910	methyltransferase domain containing protein, expressed		
LOC_Os06g05920	Rf1, mitochondrial precursor, putative, expressed		
LOC_Os06g05930	expressed protein		
LOC_Os06g05940	DTA2, putative, expressed		
LOC_Os06g05950	expressed protein		
LOC_Os06g05960	expressed protein		
LOC_Os06g05970	DUF581 domain containing protein, expressed		
LOC_Os06g05980	transporter family protein, putative, expressed	I236580	34.8
LOC_Os06g05990	zinc finger family protein, putative, expressed		
LOC_Os06g06000	expressed protein		
LOC_Os06g06014	expressed protein		
LOC_Os06g06030	peptidase, T1 family, putative, expressed		
LOC_Os06g06040	expressed protein	I138457	34.82
LOC_Os06g06050	OsFBL27 - F-box domain and LRR containing protein, expressed		
LOC_Os06g06080	serine esterase family protein, putative, expressed		
LOC_Os06g06090	CGMC_MAPKCMGC_2_ERK.12 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases, expressed	B30083	34.82
LOC_Os06g06100	dihydroneopterin aldolase, putative, expressed		
LOC_Os06g06115	expressed protein		
LOC_Os06g06120	expressed protein		
LOC_Os06g06130	glutamate receptor, putative, expressed		
LOC_Os06g06140	hypothetical protein		
LOC_Os06g06150	zinc finger, C3HC4 type domain containing protein, expressed		
LOC_Os06g06160	IQ calmodulin-binding motif domain containing protein, expressed		
LOC_Os06g06170	expressed protein		
LOC_Os06g06180	transferase family protein, putative, expressed		
LOC_Os06g06190	uncharacterized protein ycf45, putative, expressed		
LOC_Os06g06210	expressed protein		
LOC_Os06g06220	hypothetical protein		
LOC_Os06g06230	GDSL-like lipase/acylhydrolase, putative, expressed		
LOC_Os06g06240	hypothetical protein		
LOC_Os06g06250	GDSL-like lipase/acylhydrolase, putative, expressed		
LOC_Os06g06260	GDSL-like lipase/acylhydrolase, putative, expressed		
LOC_Os06g06270	transcription elongation factor 1, putative, expressed		
LOC_Os06g06280	galactosyltransferase family protein, putative, expressed		
LOC_Os06g06290	GDSL-like lipase/acylhydrolase, putative, expressed	B10218	39.04
LOC_Os06g06300	osFTL3 FT-Like3 homologous to Flowering Locus T gene%3B contains Pfam profile PF01161: Phosphatidylethanolamine-binding protein, expressed	B30893	37.55
LOC_Os06g06310	expressed protein		
LOC_Os06g06320	osFTL2 FT-Like2 homologous to Flowering Locus T gene%3B contains Pfam profile PF01161: Phosphatidylethanolamine-binding protein, expressed		

Appendix 4. 6 Patterns of MTA for the QTL cluster on chromosome 5HS identified from AGOUEB GWAS.

Patterns of marker trait associations for all 20 traits of the AGOUEB GWAS at the QTL cluster identified on 5H short arm with significant QTL for mildew (mild-(U)_2), yield (Yld_T_3), winter hardiness (WintH_3), straw length (SL_U_3) and Grain Nitrogen (GrainN_4) (see Table 3.3 and Figure 4.1e)). The X-axis plots the barley SNP ordered according the physical position of their homologous rice locus. The Y-axis indicates the magnitude of the association. The homologous rice segment has been identified on chromosome 12.



Appendix 4. 7 List of putative homologous rice genes predicted as equivalent to the QTL cluster on chromosome 5HS identified from AGOUEB GWAS.

List of putative rice gene models homologous to the barley chromosome segment comprised between SNP markers B30167 and A20553 including the QTL cluster presented in Appendix 4. 6. This chromosome segment contains potential candidate genes involved in the mildew resistance.

Rice Locus	Gene product name	Barley SNP	SNP distance
LOC_Os12g43130	phytoene synthase, chloroplast precursor, putative, expressed	B30167	26.28
LOC_Os12g43140	late embryogenesis abundant protein D-34, putative, expressed	A21065	26.28
LOC_Os12g43150	expressed protein		
LOC_Os12g43300	expressed protein		
LOC_Os12g43310	hypothetical protein		
LOC_Os12g43320	expressed protein		
LOC_Os12g43330	expressed protein		
LOC_Os12g43340	actin-depolymerizing factor, putative, expressed		
LOC_Os12g43350	expressed protein		
LOC_Os12g43363	fatty acid hydroxylase, putative, expressed		
LOC_Os12g43370	peptidase, M24 family protein, putative, expressed		
LOC_Os12g43380	thaumatin, putative, expressed	I98293	19.40
LOC_Os12g43390	thaumatin, putative, expressed		
LOC_Os12g43400	expressed protein		
LOC_Os12g43410	thaumatin, putative, expressed		
LOC_Os12g43430	thaumatin, putative, expressed		
LOC_Os12g43440	thaumatin, putative, expressed		
LOC_Os12g43450	thaumatin family domain containing protein, expressed		
LOC_Os12g43490	thaumatin, putative, expressed		
LOC_Os12g43500	expressed protein		
LOC_Os12g43510	expressed protein		
LOC_Os12g43520	cbbY protein-related, putative, expressed	I192396	19.40
LOC_Os12g43530	no apical meristem protein, putative, expressed		
LOC_Os12g43540	expressed protein		
LOC_Os12g43550	ras-related protein, putative, expressed	I108541	19.40
LOC_Os12g43560	zinc finger, putative, expressed	I133600	19.40
LOC_Os12g43564	expressed protein		
LOC_Os12g43570	expressed protein		
LOC_Os12g43580	expressed protein		
LOC_Os12g43590	FAD dependent oxidoreductase domain containing protein, expressed		
LOC_Os12g43600	RNA recognition motif containing protein, expressed		
LOC_Os12g43610	expressed protein		
LOC_Os12g43620	helix-loop-helix DNA-binding domain containing protein, expressed		
LOC_Os12g43630	lactate/malate dehydrogenase, putative, expressed		
LOC_Os12g43640	receptor-like protein kinase HAIKU2 precursor, putative, expressed	I114102	2.10
LOC_Os12g43660	receptor-like protein kinase HAIKU2 precursor, putative, expressed		
LOC_Os12g43664	FGGY family of carbohydrate kinases, putative, expressed		
LOC_Os12g43670	ergosterol biosynthetic protein 28, putative, expressed		
LOC_Os12g43700	SCP-like extracellular protein, expressed		
LOC_Os12g43710	expressed protein		
LOC_Os12g43720	early-responsive to dehydration protein-related, putative, expressed	I228061	7.48
LOC_Os12g43730	expressed protein		
LOC_Os12g43740	oxidoreductase, short chain dehydrogenase/reductase family, putative, expressed		
LOC_Os12g43750	expressed protein		
LOC_Os12g43770	OsFBX465 - F-box domain containing protein, expressed		
LOC_Os12g43780	expressed protein		
LOC_Os12g43790	bZIP transcription factor domain containing protein, expressed		
LOC_Os12g43810	expressed protein		
LOC_Os12g43820	GCRP5 - Glycine and cysteine rich family protein precursor, expressed		
LOC_Os12g43830	NUC189 domain containing protein, expressed		
LOC_Os12g43840	ankyrin repeat domain-containing protein, putative, expressed		
LOC_Os12g43870	expressed protein		
LOC_Os12g43880	DNA binding protein, putative, expressed		
LOC_Os12g43890	GNS1/SUR4 membrane family protein, putative, expressed		
LOC_Os12g43930	zinc finger, C3HC4 type domain containing protein, expressed		
LOC_Os12g43940	ankyrin repeat domain-containing protein, putative, expressed		
LOC_Os12g43950	homeobox domain containing protein, expressed		
LOC_Os12g43960	hypothetical protein		
LOC_Os12g43970	hydrolase, alpha/beta fold family domain containing protein, expressed		
LOC_Os12g43990	expressed protein		
LOC_Os12g44000	ubiquitin-conjugating enzyme E2 W, putative, expressed		
LOC_Os12g44010	purple acid phosphatase precursor, putative, expressed		
LOC_Os12g44020	Ser/Thr protein phosphatase family protein, putative, expressed		
LOC_Os12g44030	purple acid phosphatase precursor, putative, expressed		
LOC_Os12g44050	purple acid phosphatase precursor, putative, expressed		
LOC_Os12g44060	nodulin, putative, expressed		
LOC_Os12g44070	nodulin, putative, expressed		
LOC_Os12g44080	CHCH domain containing protein, expressed		
LOC_Os12g44090	leucine-rich repeat family protein, putative, expressed	I168359	5.70
LOC_Os12g44100	peptide transporter PTR2, putative, expressed		
LOC_Os12g44110	ligA, putative, expressed		
LOC_Os12g44130	expressed protein		
LOC_Os12g44140	expressed protein		
LOC_Os12g44150	plasma membrane ATPase, putative, expressed		
LOC_Os12g44160	oxidoreductase, putative, expressed		
LOC_Os12g44170	pentatricopeptide, putative, expressed		
LOC_Os12g44180	nodulin, putative, expressed		
LOC_Os12g44190	ATPase 3, putative, expressed		
LOC_Os12g44210	ATPase, AAA family domain containing protein, expressed		
LOC_Os12g44220	ATPase 2, putative, expressed		
LOC_Os12g44230	expressed protein	B31023	4.96
LOC_Os12g44240	N-acetylglucosaminyltransferase, putative, expressed	B30975	6.40
LOC_Os12g44250	vesicle-associated membrane protein, putative, expressed		
LOC_Os12g44260	heat shock protein DnaJ, putative, expressed		
LOC_Os12g44270	glycine-rich protein, putative, expressed		
LOC_Os12g44280	subtilase, putative, expressed		
LOC_Os12g44290	cytochrome P450, putative, expressed		
LOC_Os12g44300	CHX28, putative, expressed		
LOC_Os12g44310	carotenoid cleavage dioxygenase, putative, expressed	A20553	2.81
LOC_Os12g44320	carbohydrate binding protein, putative, expressed		
LOC_Os12g44330	serine/threonine-protein kinase PRP4, putative, expressed		
LOC_Os12g44340	ATMAP70 protein, putative, expressed		
LOC_Os12g44350	actin, putative, expressed		
LOC_Os12g44360	sodium/hydrogen exchanger 7, putative, expressed		
LOC_Os12g44370	expressed protein		
LOC_Os12g44380	sucrose transporter, putativ, expressed		
LOC_Os12g44390	RecF/RecN/SMC N terminal domain containing protein, expressed		

Appendix 5. 1 Drilling plan of the 2012 NIL multiplication experiment.

The NILs were grown as miniplots made of six rows of one meter long. Each bed was made of six miniplots. Plots were grown in untreated field conditions with standard fertilisation regime (see 5.3.2).

Book 1

bed\run	1	2	3	4	5	6
1	SAFFRON	SAFFRON	SAFFRON	SAFFRON	SAFFRON	SAFFRON
2	RETRIEVER	RETRIEVER	RETRIEVER	RETRIEVER	RETRIEVER	RETRIEVER
3	B1041 G9/1	B1041 G9/8	SAFFRON	B1041 D9	B1041 H8	B1041 G7
4	B1041 G9/6	B1041 G9/4	RETRIEVER	B1041 H8	B1041 D9	B1041 H8
5	B1041 E8/6	SAFFRON	B1041 E8/2	B1041 E8/5	B1041 E8/7	SAFFRON
6	b1041 E8/1	RETRIEVER	B1041 E8/6	B1041 E8/4	B1041 E8/5	RETRIEVER
7	B1041 G11/5	B1041 H12	B1041 B12/7	SAFFRON	B1041 F11	B1041 A12
8	B1041 B12/7	B1041 F12	B1041 A12	RETRIEVER	B1041 H12	B1041 F11
9	SAFFRON	B1041 B12/4	B1041 B12/1	B1041 G11/3	SAFFRON	B1041 G11/5
10	RETRIEVER	B1041 B12/1	B1041 B12/4	B1041 G11/4	RETRIEVER	B1041 B12/7
11	B1041 A10	B1041 B10	SAFFRON	B1041 A11	B1041 B10	B1041 C10
12	B1041 B10	B1041 A10	RETRIEVER	B1041 C11	B1041 A11	B1041 B10
13	B1042 D1	SAFFRON	B1042 A3	B1042 D2	B1041 C11	SAFFRON
14	B1042 E1	RETRIEVER	B1042 E2	B1042 A3	B1041 C10	RETRIEVER
15	B1042 G1	B1042 F3	B1042 A4	SAFFRON	B1042 E3	B1042 H3
16	B1042 C2	B1042 H3	B1042 E3	RETRIEVER	B1042 A4	B1042 F3
17	SAFFRON	B1042 H5	B1042 C5	B1042 D5/3 C6	SAFFRON	B1042 D5/5 E6
18	RETRIEVER	B1042 C5	B1042 H5	B1042 D5/5 E6	RETRIEVER	B1042 D5/3 C6
19	B1042 E6	B1042 C7	SAFFRON	B1042 G7	B1042 E7	SAFFRON
20	B1042 C7	B1042 E6	RETRIEVER	B1042 E7	B1042 G7	RETRIEVER

Book 2

21	SAFFRON	B1042 D8/1 D4	B1042 F10/5 &/8	B1042 D8/4 G4	SAFFRON	B1042 F10/2 E3
22	RETRIEVER	B1042 D8/4 G4	B1042 F10/2 E3	B1042 D8/1 D4	RETRIEVER	B1042 F10/5 &/8
23	B1042 G11	B1042 E11	B1042 A11	B1042 D12	B1042 A12	B1042 E12
24	B1042 D12	B1042 A12	B1042 E12	B1042 E11	B1042 G11	B1042 A11
25	4042 A2	SAFFRON	4042 D2	4042 G1	4042 F1	SAFFRON
26	4042 G1	RETRIEVER	4042 F1	4042 A2	4042 D2	RETRIEVER
27	4190 A8	4190 G7/2	4190 A7	SAFFRON	4188 G6	4188 D6
28	4190 A7	4190 A8	4190 G7/7	RETRIEVER	4188 D6	4188 G6
29	SAFFRON	4188 C5	4188 B4	4188 E5	SAFFRON	SAFFRON
30	RETRIEVER	4188 E5	4188 C5	4188 B4	RETRIEVER	RETRIEVER
31	4187 H3/4 & H3/6	4187 C3/6 & E3/5	SAFFRON	4187 H3/6 & E3/5	4187 E3/4 & C3/6	SAFFRON
32	4187 F3/9 & E3/4	4187 F3/8 & C3/1	RETRIEVER	4187 C3/1 & F3/9	4187 H3/4 & F3/8	RETRIEVER
33	SAFFRON	SAFFRON	4045 E8/4	4045 D9	4045 E8/8	SAFFRON
34	RETRIEVER	RETRIEVER	4045 E8/2	4045 E8/6	4045 H8	RETRIEVER
35	6505-16 D1	6505-7 G1	6506 A2	SAFFRON	6506 C2	SAFFRON
36	6505-7 G1	6505-16 D1	6506 C2	RETRIEVER	6506 A2	RETRIEVER
37	SAFFRON	6507-3	6507-6	6513-3 C4	SAFFRON	6513-11
38	RETRIEVER	6507-6	6507-3	6513-11 A4	RETRIEVER	6513-3
39	SAFFRON	SAFFRON	SAFFRON	SAFFRON	SAFFRON	SAFFRON
40	RETRIEVER	RETRIEVER	RETRIEVER	RETRIEVER	RETRIEVER	RETRIEVER

Appendix 5. 2 Experimental layout of the 2013 NIL yield plot experiment

The replicated blocks are identified by a blue and red quadrat. The brown quadrats delineate the 7 Heterogeneous Inbred Families (HIF) and 2 controls (Saffron and Retriever) within each of the blocks. Additional replication of the NILs was made within Rep.

Rep 2 (block 2)	B-4 F3/8	B-4 F3/9	B-4 H3/4	B-4 C3/1	B-3 E8/4	B-3 H8	A-8 D8/1	A-8 F10/8
	B-4 C3/1	B-4 H3/4	B-4 F3/8	B-4 F3/9	B-3 H8	B-3 E8/4	A-8 F10/8	A-8 D8/1
	A-3 A12	A-3 H12	A-3 F12	Saffron	Retriever	A-5 E3	A-5 H3	A-5 F3
	A-3 F12	A-3 A12	A-3 H12	Retriever	Saffron	A-5 F3	A-5 E3	A-5 H3
	A-2 A11	A-2 A10	A-2 B10	A-2 C10	Saffron	Retriever	B-2 A8	B-2 A7
	A-2 B10	A-2 A10	A-2 C10	A-2 A11	Retriever	Saffron	B-2 A7	B-2 A8
Rep 1 (block 1)	Saffron	Retriever	A-3 A12	A-3 F12	A-3 H12	A-5 H3	A-5 F3	A-5 E3
	Retriever	Saffron	A-3 F12	A-3 H12	A-3 A12	A-5 E3	A-5 H3	A-5 F3
	B-3 E8/4	B-3 H8	A-2 A11	A-2 B10	A-2 A10	A-2 C10	Saffron	Retriever
	B-3 H8	B-3 E8/4	A-2 C10	A-2 A10	A-2 A11	A-2 B10	Retriever	Saffron
	B-2 A8	B-2 A7	B-4 F3/8	B-4 H3/4	B-4 F3/9	B-4 C3/1	A-8 D8/1	A-8 F10/8
	B-2 A7	B-2 A8	B-4 F3/9	B-4 H3/4	B-4 C3/1	B-4 F3/8	A-8 F10/8	A-8 D8/1

Photography of the 2013 field based NILs. For simplicity, only NILs of family B-4 of Rep 1 have been detailed. The layout describe above correspond to the photography. In B-4, the late emergence of NILs 4187 H3/4 is clearly visible.

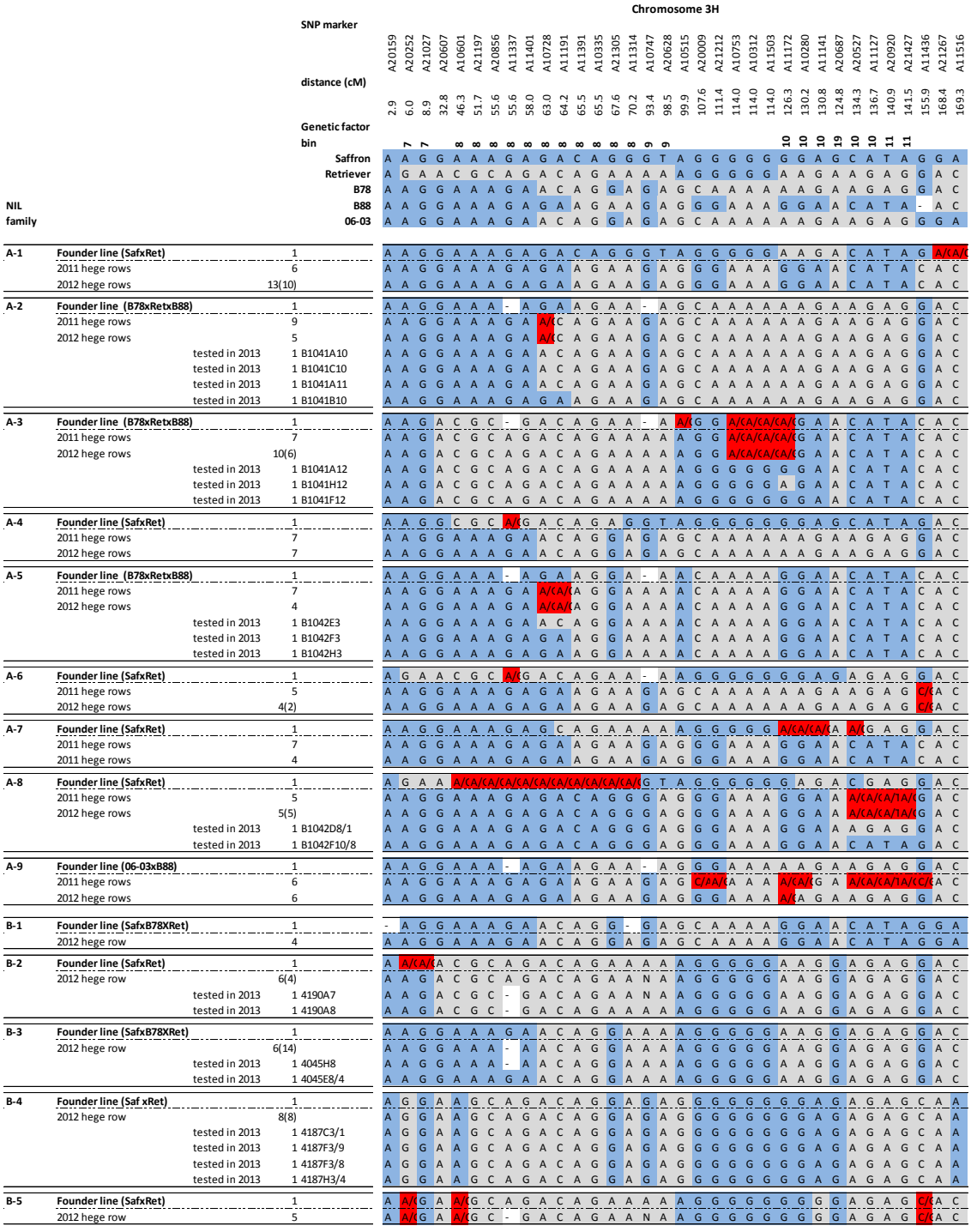
Appendix 5. 3 HIF founders and NILs genotypes for chromosomes 1H, 3H, 4H, 5H, 6H and 7H.

Only informative markers across the 14 HIF and their founder lines are presented (polymorphic markers of the Bx384 genotyping platform across the whole set of HIF). Markers have been ordered from left to right based on the OPA1 consensus genetic distance. Genotype of NILs multiplied in hege row in 2011 and 2012 are presented as a consensus haplotype (homozygous haplotypes within HIF could also be present at these same multiplication stages). Heterozygous markers are highlighted in red. The genetic factor number (bin numbers) correspond to results presented in (Table 2.9)

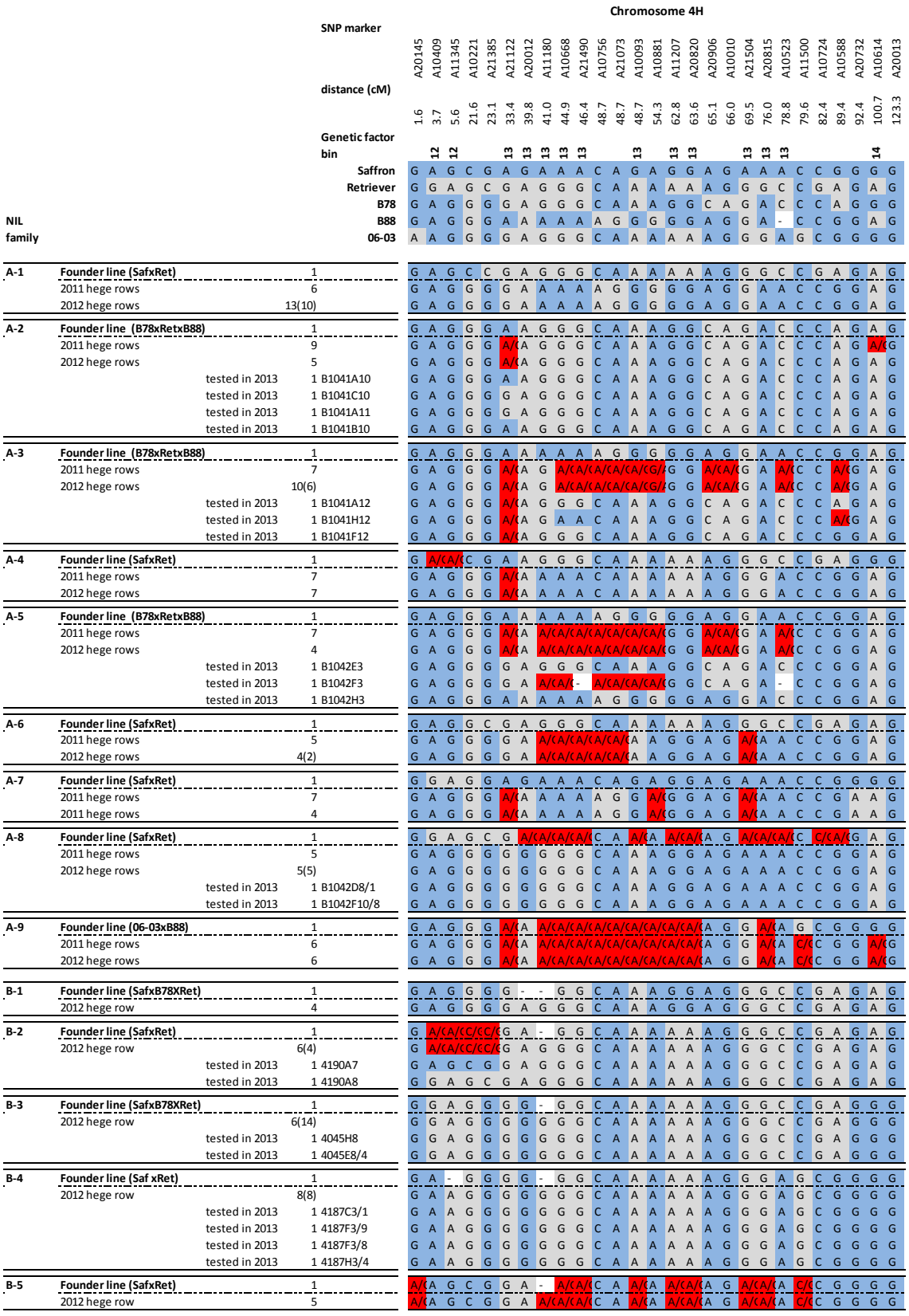
Table 2.9 and were associated with the Bx384 SNP based on their colocation with the OPA1 SNP used for mapping in Chapter 2.

				Chromosome 1H									
			SNP marker	A21354	A10419	A21226	A10760	A20609	A21431	A21392	A10644	A10782	A10590
			distance (cM)	0.8	3.8	8.8	34.8	51.9	64.9	114.8	127.1	131.9	138.3
			Genetic factor bin	1	1	1							
			Saffron	C	A	A	G	G	A	G	A	A	G
			Retriever	G	G	G	A	G	A	A/G	G	A	G
			B78	G	G	G	G	A	A	G	G	G	G
NIL family			B88	G	G	A	G	G	A	G	A/G	A	G
			06-03	G	G	G	A	G	A	G	A	G	G
A-1	Founder line (SafxRet)		1	G	G	G	A	G	A/-	-	A/G	A	G
	2011 hege rows		6	C	A	A	G	G	A	G	A	A	G
	2012 hege rows		13(10)	C	A	A	G	G	A	G	A	A	G
A-2	Founder line (B78xRetxB88)		1	G	G	A	A	G	A/-	G	A	A	G
	2011 hege rows		9	G	G	A/G	A	G	A	A/G	A	A	G
	2012 hege rows		5	G	G	A/G	A	G	A	A/G	A	A	G
		tested in 2013	1 B1041A10	G	G	A/G	A	G	A	A/G	A	A	G
		tested in 2013	1 B1041C10	G	G	G	A	G	A	A/G	A	A	G
		tested in 2013	1 B1041A11	G	G	A/G	A	G	A	A/G	A	A	G
		tested in 2013	1 B1041B10	G	G	A	A	G	A	G	A	A	G
A-3	Founder line (B78xRetxB88)		1	G	G	A/G	A	A	A/-	-	A	A	G
	2011 hege rows		7	G	G	A/G	A	A	A	-	A/G	A	G
	2012 hege rows		10(6)	G	G	A/G	A	A	A	-	A/G	A	G
		tested in 2013	1 B1041A12	G	G	G	A	A	A	-	A	A	G
		tested in 2013	1 B1041H12	G	G	A	A	A	A	-	A	A	G
		tested in 2013	1 B1041F12	G	G	G	A	A	A	-	A	A	G
A-4	Founder line (SafxRet)		1	G	A/G	A/G	A/G	G	A/-	-	G	A	G
	2011 hege rows		7	G	G	A/G	G	G	A	G	A	A/G	G
	2012 hege rows		7	G	G	A/G	G	G	A	G	A	A/G	G
A-5	Founder line (B78xRetxB88)		1	G	G	G	G	A	A/-	A/G	A	A	G
	2011 hege rows		7	G	G	G	G	A/G	A	-	A	A	G
	2012 hege rows		4	G	G	G	G	A/G	A	-	A	A	G
		tested in 2013	1 B1042E3	G	G	G	G	A/G	A	-	A	A	G
		tested in 2013	1 B1042F3	G	G	G	G	G	A	-	A	A	G
		tested in 2013	1 B1042H3	G	G	G	G	A	A	-	A	A	G
A-6	Founder line (SafxRet)		1	C	A	A	G	G	A/-	-	G	A	G
	2011 hege rows		5	G	G	A	G	G	A	G	A	A	G
	2012 hege rows		4(2)	G	G	A	G	G	A	G	A	A	G
A-7	Founder line (SafxRet)		1	G	G	G	A	G	A	-	G	A	G
	2011 hege rows		7	G	A	A	G	G	A	G	G	A	G
	2011 hege rows		4	G	A	A	G	G	A	G	G	A	G
A-8	Founder line (SafxRet)		1	C	A	A	G	G	A	-	A	A	G
	2011 hege rows		5	C	A	A	G	G	-	G	G	A	G
	2012 hege rows		5(5)	C	A	A	G	G	-	G	G	A	G
		tested in 2013	1 B1042D8/1	C	A	A	G	G	-	G	G	A	G
		tested in 2013	1 B1042F10/8	C	A	A	G	G	-	G	G	A	G
A-9	Founder line (06-03xB88)		1	G	G	A	A	G	A	G	A	G	G
	2011 hege rows		6	G	G	A/G	A	G	A	G	A	A/G	G
	2012 hege rows		6	G	G	A/G	A	G	A	G	A	A/G	G
B-1	Founder line (SafxB78XRet)		1	G	-	G	G	A	-	G	G	G	G
	2012 hege row		4	G	G	G	G	A	A	G	G	G	G
B-2	Founder line (SafxRet)		1	C	-	A	G	G	-	A/G	G	A	G
	2012 hege row		6(4)	C	A	A	G	G	A	-	G	A	G
		tested in 2013	1 4190A7	C	A	A	G	G	A	-	G	A	G
		tested in 2013	1 4190A8	C	A	A	G	G	A	-	G	A	G
B-3	Founder line (SafxB78XRet)		1	C	-	A	A/G	A	-	G	G	A	G
	2012 hege row		6(14)	C	A	A	A/G	A	A	-	G	A	G
		tested in 2013	1 4045H8	C	A	A	A/G	A	A	-	G	A	G
		tested in 2013	1 4045E8/4	C	A	A	A	A	-	G	G	A	G
B-4	Founder line (Saf xRet)		1	C	-	A	G	A	-	G	G	A	G
	2012 hege row		8(8)	C	A	A	G	A	-	G	G	A	G
		tested in 2013	1 4187C3/1	C	A	A	G	A	-	G	G	A	G
		tested in 2013	1 4187F3/9	C	A	A	G	A	-	G	G	A	G
		tested in 2013	1 4187F3/8	C	A	A	G	A	-	G	G	A	G
		tested in 2013	1 4187H3/4	C	A	A	G	A	-	G	G	A	G
B-5	Founder line (SafxRet)		1	C	-	A	A	G	-	A/G	G	A	G
	2012 hege row		5	C	A	A	A	G	A	-	G	A	G

Appendix 5. 3 cont : Chromosome 3H



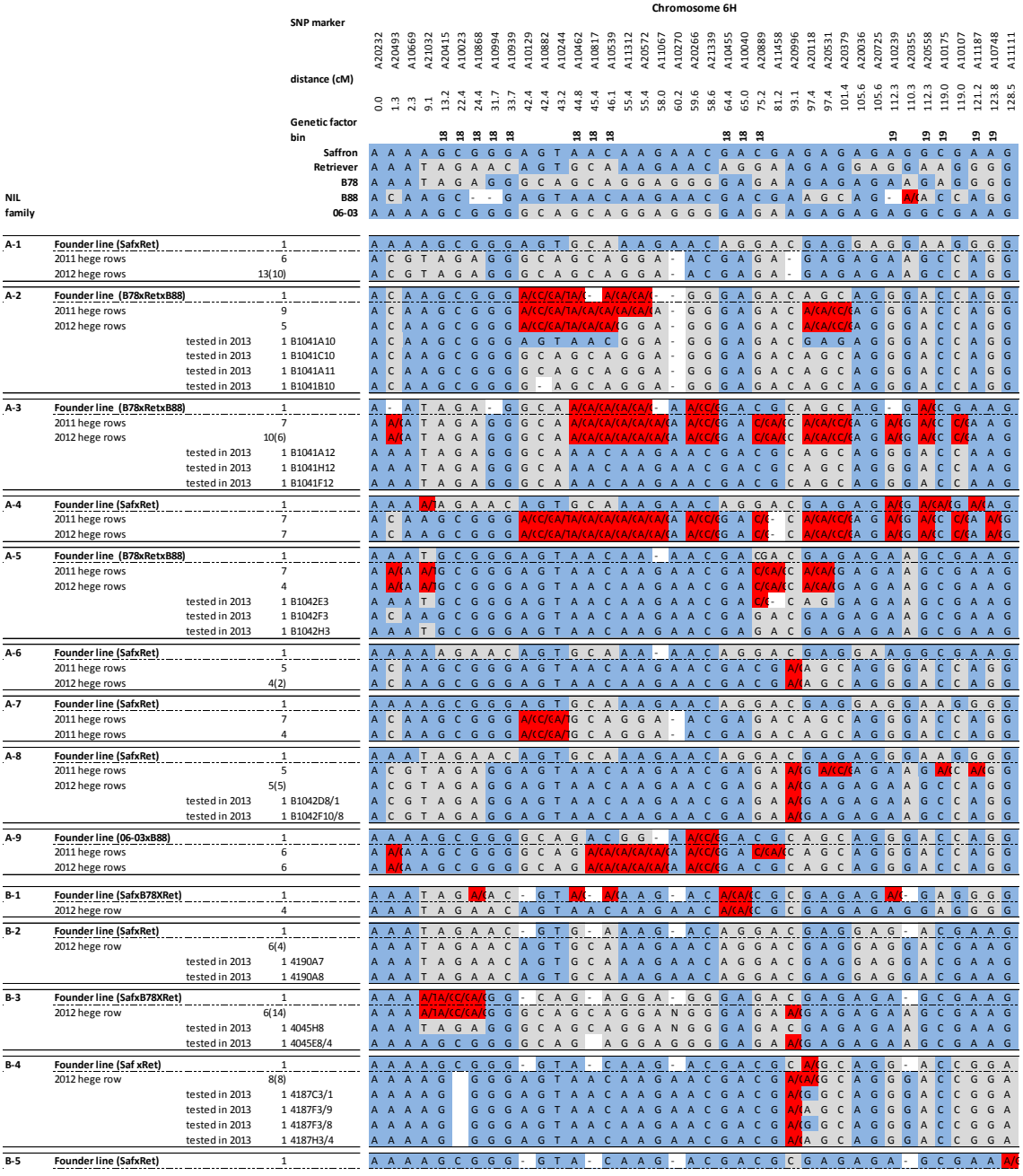
Appendix 5. 3 cont : Chromosome 4H



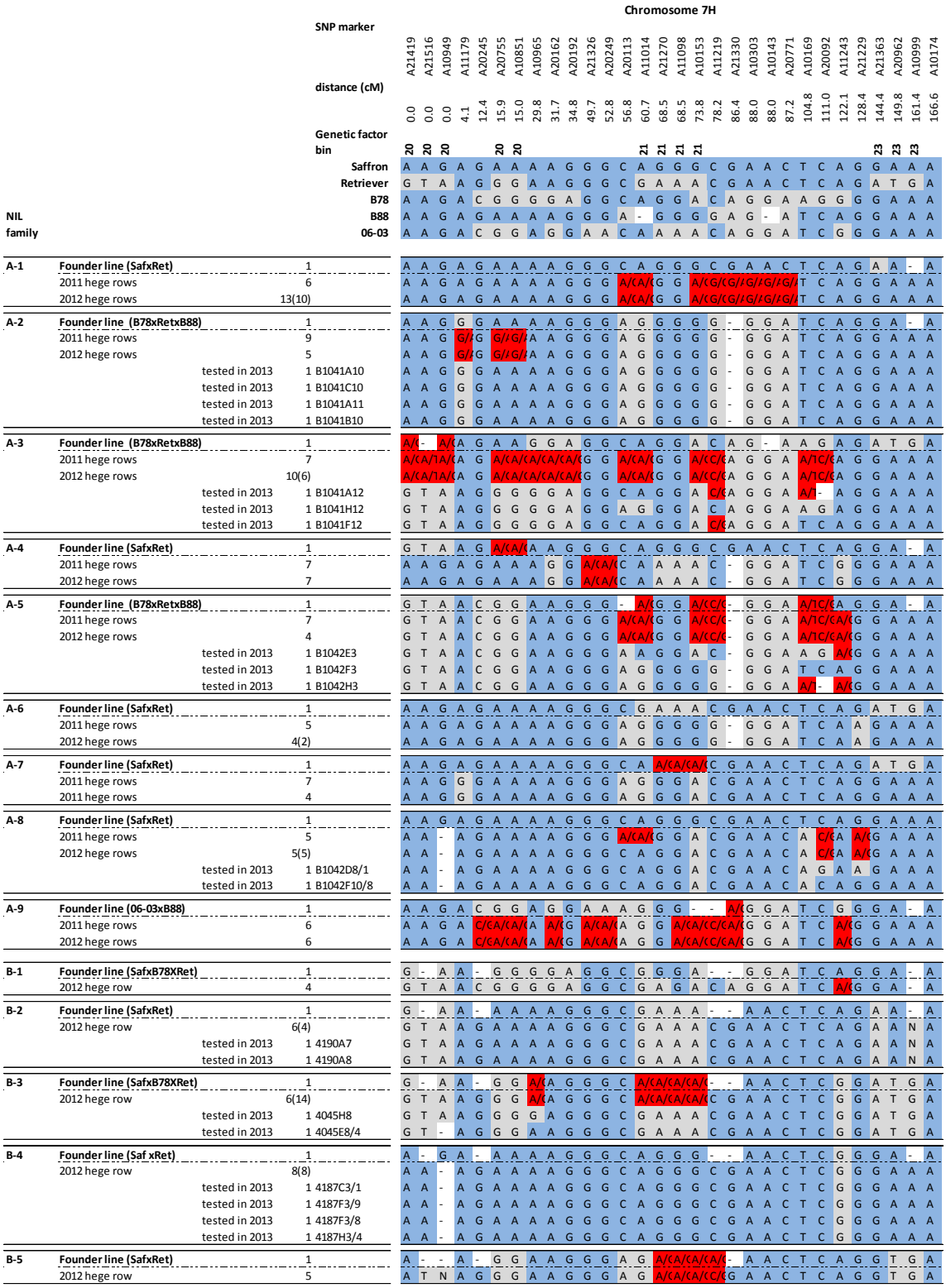
Appendix 5. 3 cont : Chromosome 5H

NIL family			SNP marker	A20553	A21221	A20533	A20010	A20873	A21426	A10688	A10621	A20845	A11273	A20127	A21203	A10805	A10819	A11092	A20676	A11441	A11216	A21141	A10736	A20022	A20897
			Chromosome 5H																						
			distance (cM)	2.8	9.3	17.4	18.7	26.3	27.0	34.3	37.1	40.0	111.7	123.5	129.4	130.1	143.9	145.4	146.0	151.4	171.7	177.1	180.7	181.4	182.9
			Genetic factor bin				15		15																
			Saffron	A	G	A	G	G	G	G	G	A	G	A	C	A	G	A	A	A	A	G	A	G	A
			Retriever	A	G	A	A	G	A	G	G	A	G	A	C	A	G	A	A	A	A	A	A	G	A
			B78	A	G	A	A	A	A	G	A	G	G	T	A	G	G	G	C	G	G	A	G	A	G
			B88	A	-	-	A	A	A	G	G	A	G	A	C	A	G	A	A	A	A	G	A	G	A
			06-03	A	G	A	A	A	A	A	A	G	G	A	A	G	G	G	C	G	A	A	A	G	A
A-1	Founder line (SafxRet)	1		A	-	-	G	G	G	G	G	A	G	A	C	A	G	A	A	A	-	G	A	G	A
	2011 hege rows	6		A	A/(	A/(	A	G/	G/	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	2012 hege rows	13(10)		A	A/(	G	A	G/	G/	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
A-2	Founder line (B78xRetxB88)	1		A	-	-	A	A	A	G	G	A	G	A/T	-	-	G	A	A	A	-	A	G	A	G
	2011 hege rows	9		A	A/(	A	A	A/(	A	G	A/(	A/(	G	A/(	-	G	G	A	A	A	A/(	G	A	G	A
	2012 hege rows	5		A	A/(	A	A	A/(	A	G	A/(	A/(	G	A/(	-	G	G	A	A	A	A/(	G	A	G	A
	tested in 2013	1	B1041A10	A	A/(	A	A	G	A	G	A	G	G	T	-	G	G	A	A	A	A/(	G	A	G	A
	tested in 2013	1	B1041C10	A	A/(	A	A	A	A	G	A	G	G	A	-	A	G	A	A	A	A/(	A	G	A	G
	tested in 2013	1	B1041A11	A	A/(	A	A	A	A	G	A	G	G	A	-	A	G	A	A	A	G	A	G	A	G
	tested in 2013	1	B1041B10	A	G	A	A	A	A	G	G	A	G	T	-	G	G	A	A	A	G	G	A	G	A
A-3	Founder line (B78xRetxB88)	1		A	-	-	A	A	A	G	A/(	A/(	G	A	-	A	G	A	A	A	-	A	A	G	A
	2011 hege rows	7		A	G	A	A	A	A	G	A/(	A/(	G	A	-	A	G	A	A	A	A	A	A	G	A
	2012 hege rows	10(6)		A	G	A	A	A	A	G	A	G	G	A	-	A	G	A	A	A	A	A	A	G	A
	tested in 2013	1	B1041A12	A	G	A	A	A	A	G	A	G	G	A	-	A	G	A	A	A	A	A	A	G	A
	tested in 2013	1	B1041H12	A	G	A	A	A	A	G	A	G	G	A	-	A	G	A	A	A	A	A	A	G	A
	tested in 2013	1	B1041F12	A	G	A	A	A	A	G	A	G	G	A	-	A	G	A	A	A	A	A	A	G	A
A-4	Founder line (SafxRet)	1		A	-	-	A	G	A	G	G	A	G	A	C	A	G	A	A	A	-	G	A	G	A
	2011 hege rows	7		A	G	A	A	A	A	G	G	A	G	A	-	G	G	A/(	A/(	A/(	A	G	A	G	A
	2012 hege rows	7		A	G	A	A	A	A	G	G	A	G	A	-	G	G	A/(	A/(	A/(	A	G	A	G	A
A-5	Founder line (B78xRetxB88)	1		A	-	-	A	A	A	G	G	A	G	A	-	A	G	A	A	A	-	G	A	G	A
	2011 hege rows	7		A	G	A	A	A	A	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	2012 hege rows	4		A	G	A	A	A	A	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	tested in 2013	1	B1042E3	A	G	A	A	A	A	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	tested in 2013	1	B1042F3	A	G	A	A	A	A	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	tested in 2013	1	B1042H3	A	G	A	A	A	A	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
A-6	Founder line (SafxRet)	1		A	-	-	G	G	G	G	G	A	G	A	C	A	G	A	A	A	-	G	A	G	A
	2011 hege rows	5		A	-	G	A	G	G	G	G	A	G	A	-	-	G	A	A	A	A	G	A	G	A
	2012 hege rows	4(2)		A	-	G	A	G	G	G	G	A	G	A	-	-	G	A	A	A	A	G	A	G	A
A-7	Founder line (SafxRet)	1		A	-	-	A/(	G	G	G	G	A	G	A	C	A	G	A	A	A	-	G	A	G	A
	2011 hege rows	7		A	G	A	A	A	A	G	G	A	G	A	-	G	G	A	A	A	A	G	A	G	A
	2011 hege rows	4		A	G	A	A	A	A	G	G	A	G	A	-	G	G	A	A	A	A	G	A	G	A
A-8	Founder line (SafxRet)	1		A	-	-	G	G	G	G	G	A	G	A	C	A	G	A	A	A	-	A	A	G	A
	2011 hege rows	5		A	A/(	G	A	G	G	G	G	A	G	A	C	A	G	A	A	A	A	G	A	G	A
	2012 hege rows	5(5)		A	-	G	A	G	G	G	G	A	G	A	C	A	G	A	A	A	A	G	A	G	A
	tested in 2013	1	B1042D8/1	A	-	G	A	G	G	G	G	A	G	A	C	A	G	A	A	A	A	G	A	G	A
	tested in 2013	1	B1042F10/8	A	-	G	A	G	G	G	G	A	G	A	C	A	G	A	A	A	A	G	A	G	A
A-9	Founder line (06-03xB88)	1		A	-	-	A	A	A	G	G	A	G	A	-	G	G	G	C	G	-	A	A	G	A
	2011 hege rows	6		A	G	A/(	A	A	A	A/(	A/(	A/(	G	A	-	G	G	A/(	A/(	A/(	A	A/(	A	G	A
	2012 hege rows	6		A	G	A/(	A	A	A	A/(	A/(	A/(	G	A	-	G	G	G	C	G	A	A	A	G	A
B-1	Founder line (SafxB78XRet)	1		A	-	A	G	G	G	G	G	A	G	T	-	G	G	A	A	A	A	A	A	G	A
	2012 hege row	4		A	-	A	G	G	G	G	G	A	G	T	-	G	G	A	A	A	A	A	A	G	A
B-2	Founder line (SafxRet)	1		A	-	-	A	G	A/(	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	2012 hege row	6(4)		A	-	N	A	G	A/(	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	tested in 2013	1	4190A7	A	-	N	A	G	A	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
	tested in 2013	1	4190A8	A	-	N	A	G	G	G	G	A	G	A	-	A	G	A	A	A	A	G	A	G	A
B-3	Founder line (SafxB78XRet)	1		A	-	-	A	G	G	G	G	A	G	T	-	G	G	A	A	A	A	A	A	G	A
	2012 hege row	6(14)		A	-	N	A	G	G	G	G	A	G	T	A	G	G	A	A	A	A	A	A	G	A
	tested in 2013	1	4045H8	A	-	N	A	G	G	G	G	A	G	T	-	G	G	A	A	A	A	A	A	G	A
	tested in 2013	1	4045E8/4	A	-		A	G	G	G	G	A	G	T	A	G	G	A	A	A	A	A	A	G	A
B-4	Founder line (Saf xRet)	1		A	-	A	G	G	G	G	G	A	A/(	A	-	A	G	A	A	A	A	A	A	G	A
	2012 hege row	8(8)		A	-	A	G	G	G	G	G	A	A/(	A	C	A	G	A	A	A	A	A	A	G	A
	tested in 2013	1	4187C3/1	A	-	A	G	G	G	G	G	A	A	A	C	A	G	A	A	A	A	A	A	G	A
	tested in 2013	1	4187F3/9	A	-	A	G	G	G	G	G	A	G	A		A	G	A	A	A	A	A	A	G	A
	tested in 2013	1	4187F3/8	A	-	A	G	G	G	G	G	A	G	A	C	A	G	A	A	A	A	A	A	G	A
	tested in 2013	1	4187H3/4	A	-	A	G	G	G	G	G	A	A	A	C	A	G	A	A	A	A	A	A	G	A
B-5	Founder line (SafxRet)	1		A	-	A	G	G	G	G	G	A	G	A	-	A/(	A/(	A	A	A	A	G	A	G	A
	2012 hege row	5		A	-	A	G	G	G	G	G	A	G	A	-	A/(	A/(	A	A	A	A	G	A	G	A

Appendix 5. 3 cont: Cromosome 6H



Appendix 5. 3 cont: Chromosome 7H



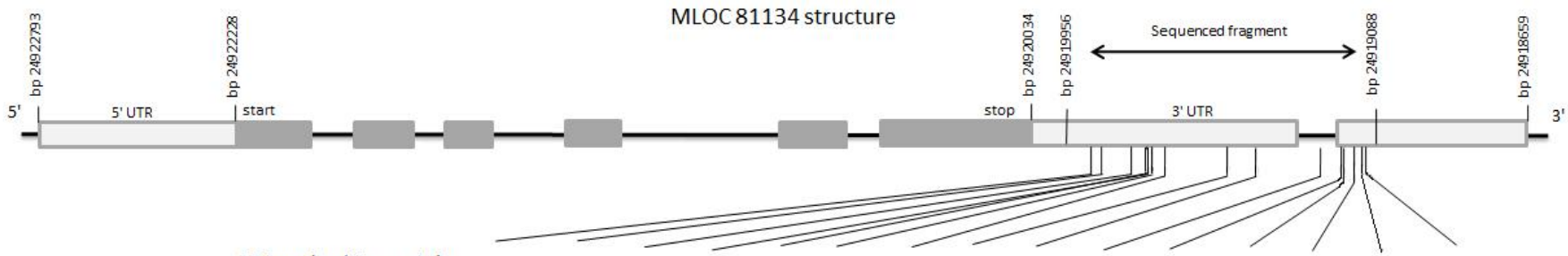
Appendix 5. 4 Preliminary work on *Ppd-H1* sequence polymorphisms in winter barley.

Sequencing of barley *Ppd-H1*

For *Ppd-H1*, the primer pair (Forward: PP04 CCGTTTTCATTCTTTGCAAGGT and reverse: PP05 AGGTTATCTCTCCACGGTCG) were developed and optimised to sequence a segment of 884 base pairs corresponding to the 3'UTR of the gene sequence. This segment overlaps the fragment amplified by primers HvF14 and HvR8 in Turner et al., (2005) in which specific SNP have been identified. The sequencing of *Ppd-H1* was carried out by at the James Hutton Institute

Ppd-H1 haplotypes in winter barley

The variation of flowering time in temperate cereals is partly controlled by genes involved in photoperiod response (Cockram et al., 2007). In wheat, a deletion 2kb deletion on the in the *Ppd* gene sequence of D genome was associated with insensitivity to photoperiod (Beales et al., 2007). In barley the *Ppd-H1* gene mapped on 2H (Laurie, 1997) is coding for a protein member of the PPR family (Turner et al., 2005) has been characterised with different haplotypes that could be grouped to differentiate between the spring and winter types (Stracke et al., 2009; Turner et al., 2005). Turner et al., (2005) identified a SNP marker (SNP22) as causal mutation inducing a change in the coding sequence from Glycine to Tryptophan in the CCT domain of 8 exons gene structure. In the gene structure of 7 exons, the SNP22, monomorphic in winter barley is located in the 3'UTR region which would indicate that it is not translated but might influence the post transcriptional gene expression.



SNP number (Turner et al., 2005)	-	16	-	17	-	18	-	19	20	-	-	21	22	-	23
Chromosome 2H base pair number	24919895	24919871	24919785	24919736	24919726	24919703	24919685	24919507	24919426	24919240	24919180	24919174	24919146	24919121	24919114
Ensembl variant	**no**	yes	yes	yes	yes	yes	yes	yes	yes	yes	**no**	yes	**no**	yes	yes
forward strand alleles	C/T	C/T	G/A	C/A	G/A	C/T	G/A	T/C	C/T	AGA/A	A/T	G/A	A/C	C/T	T/C
Igri (winter)	C	T	G	C	G	T	G	C	T	AGA	T	A	C	C	C
Clarine(winter)	C	T	G	C	G	T	G	C	T	AGA	T	A	C	C	C
Flagon (winter)	C	T	G	C	G	T	G	C	T	AGA	T	A	C	C	C
Puffin (winter)	C	T	G	C	G	T	G	C	T	AGA	T	A	C	C	C
Retriever (winter)	C	T	G	C	G	T	G	C	T	AGA	T	A	C	C	C
Saffron (winter)	C	C	G	C	A	C	A	C	C	AGA	T	G	C	C	C
carat (winter)	T	C	G	C	A	C	A	C	C	AGA	A	G	C	C	C
Morex (spring)	C	C	G	C	G	C	G	T	C	AGA	T	G	A	C	T
Optic (spring)	C	C	G	A	G	C	G	T	C	AGA	T	G	A	C	T

Figure Appendix 5.4. SNP based haplotypes found in winter barley in the 3'UTR region of the *Ppd-H1* gene (MLOC 81134).

The SNP are identidied from the 907 bp long contig made of the sequences alignment of the 3'UTR of *Ppd-H1* in a panel of winter barley varieties and NILs. The SNP position on the ensembl *Hordeum vulgare* reference sequence information is indicated as base pair number. The SNP markers (sequence variants) present in the Ensembl database are indicated as well as the reference to the SNP from Turner et al., (2005). Alleles of the SNP presented correspond to the forward stand.

An initial attempt to investigate the genetic diversity of *Ppd-H1* in winter barley was done by sequencing the 3' UTR sequence of the gene in nine two-row barley varieties. The sequence alignment enabled to retrieve 8 SNP identified by Turner et al., (2005) while additional ensemblplants variants were also identified (Figure Appendix 5. 4) [4].. The spring and winter haplotypes could easily be distinguished in the set of sequenced varieties by 3 SNP (19, 22, 23). In the winter varieties, 4 haplotypes were found and both Saffron and Carat showed atypical haplotypes. The winter barley Saffron and Retriever, parents of the DH population used for QTL mapping in Chapter 2, were monomorphic for the SNP 22 associated with the spring/winter alleles but they differed for 6 other the SNP 17bis, 18, 18bis, 20, 20.2 and 21.

Based on the haplotype signature on marker traits association in GWAS (Chapter 4,Figure 4.2 a) an independent control of heading date and TGW for the QTL cluster was suggested at that locus on 2HS. It could underpin the TGW and tillering effects found in Chapter 2. Because only subset of SNP belonging to the *Ppd-H1* gene sequence were present on the 9K Illumina chip, any additional polymorphic SNP identified from the sequencing have not been tested for the GWAS scans of Chapter 3. They may correspond to the haplotypes observed in (Figure 4.2a) which could reinstate *Ppd-H1* as a putative candidate gene for the all traits at that QTL cluster. It also needs to be pointed out that the gene was only partially sequenced and the haplotypes differences observed suggest that additional polymorphisms in the rest of the sequence.

Further work

The significant differences in heading date found in the NIL experiment (Chapter 5) and HIF A-2 and B-4 (Table 5.1 and Table 5.4) reinforced the presence of a genetic control of heading date in winter are associated with polymorphisms on 2H short arm. Indeed, the effects of A20394, a very close SNP to the *Ppd-H1* gene, suggest that the photoperiod controlling gene is a strong candidate underpinning the trait variation in those NIL pairs. In addition, the locus was also significant for the variation in TGW in HIF B-4, although no significant effects were found in A-2. It would be interesting to analyse and compare the *Ppd-H1* SNP haplotypes of the NILs with the set of winter barley varieties, especially Saffron and Retriever, in order to help to discriminate the genetic controls of photoperiod and TGW variation at that locus. The analysis can be extended to the whole diversity of haplotypes found in the GWAS panels at that 2HS locus.

[4] The sequence alignments suggest that SNP22 identified by Turner et al., (2005) is not included in the ensemblplants database despite its alleles were significantly associated with heading.

List of supplementary data

The supplementary documents are available on the CD attached to the thesis.

Supplementary data 1: Barley SNP markers information.
The document presents the SNP markers available on the Illumina iSelect 9K chip (Comadran et al., 2012) and associated information when available: the corresponding oligo pool assay 1 name (OPA1); the barley OPA identification code (bOPA1 SNP id), the SNP name on the Illumina BeadXpress platform (Bx384); the corresponding SNP number in the GWAS of Pasam et al. (2012); the barley OPA 2009 consensus chromosome and position (Close et al., 2009); The distances of the S×R map (2.2.3); The map position based on Linkage disequilibrium mapping used by Comadran et al. (2012); the BLAST results of the barley Unigene35 library on the *sorghum*, *rice* and *brachipodium* genomes downloadable from http://pgsb.helmholtz-muenchen.de/plant/barley/ (Mayer et al., 2009).

Supplementary data 2: Genome wide association mapping scans of NUE CROPS and AGOUEB two-row winter barley varieties.
The document is composed of two spreadsheets presenting the genome wide association results for NUE CROPS and AGOUEB two-row winter barley panels. For each experiment, the tests of association was carried out using the Tassel-MLM_K model for 4319 SNPs in NUE-CROPs and 3982 SNPs in AGOUEB (see 3.3.3). The SNPs are the presented with chromosome and position, the alleles and distribution across the panel. For each of the traits analysed the Tassel-MLM_K results have been summarised to SNP effect, the test of association F and probability result *p*, the –log10(*p*) reported in the Manhattan plots, the error degree of freedom and the marker R square (markerR2).