Assessing wild barley germplasm in multiparent and advanced backcross populations for mapping, gene discovery, and improvement of malting barley

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Liana Nice

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Gary J. Muehlbauer

May 2015

Acknowledgements

I would like to thank my advisor Dr. Gary J. Muehlbauer for his support and guidance over the course of my graduate research and studies. His ideas, encouragement, and mentoring have helped me to develop both personally and professionally.

I thank my committee members: Dr. Kevin P. Smith, Dr. Brian J. Steffenson, Dr. Peter L. Morrell, and Dr. Ruth G. Shaw for their suggestions and guidance. I would like to acknowledge Dr. Rex N. Bernardo for many thoughtful discussions. Additionally, I thank collaborators across the country for producing genotypic data, running field trials, and performing malting quality analyses, including: Dr. Gina Brown-Guedira, Dr. Eduard Akhunov, Dr. Thomas K. Blake, Dr. Richard D. Horsley, and Dr. Paul Schwarz.

This work could not have been accomplished without the help of many talented and knowledgeable individuals who helped me through the nitty-gritty aspects of agricultural science, particularly: Shane Heinen for his greenhouse, lab, and logistical support; Ed Schiefelbein, Guillermo Velasquez, and Martin Hochhalter for their field trial support; Karen Beaubien and Kim Howell for their genotyping support; and Tom Kono for his sequence analysis support.

I also thank past and present members of the Muehlbauer lab for their day to day interactions, ideas, help, and companionship: Addie Thompson, Maria Muñoz-Amatriaín, Anna Hofstad, Ron Okagaki, Allison Haaning, Lin Li, Yadong Huang, Xin Li, Sanghyun Shin, Suma Sreekanta, Nick Ames, Summer St. Pierre, Catherine Springer, Michael Van De Weghe, Emmi Klarer, and others.

In addition to those who have already been listed, I thank my fellow grad students who have been integral to my success over these past years, particularly: Emily Combs,

Kathryn Turner, Brian Seda, Godwin Macharia, Vikas Vikram, Amy Jacobson, Celeste Falcon, Ilene Jones, Chris Schaefer, Carol Powers, Stephanie Navara, Tyler Tiede, Adria Fernandez, and many others.

I thank my family for their constant support, particularly Keith who has been my 24-7 computer programming tutor and who has encouraged me through every step of this process.

Funding for this work was provided by the USDA National Needs Fellowship and by the USDA-NIFA funded Triticeae Coordinated Agricultural Project.

Abstract

Wild barley (*Hordeum vulgare* ssp. *spontaneum*), the progenitor of cultivated barley (*Hordeum vulgare* ssp. *vulgare*), is a rich source of genetic diversity. This diversity is not easily exploited due to challenges inherent in identifying and extracting beneficial alleles from widely unadapted germplasm. Wild barley germplasm has been shown to contain valuable alleles for several disease resistance and abiotic stress tolerance traits, and due to its extensive diversity, it is a key target for expanding the limited diversity of cultivated breeding germplasm. Therefore, our aim was to create and characterize populations that incorporate diverse sources of wild barley germplasm into cultivated barley backgrounds using advanced backcross (AB) and nested association mapping (NAM) techniques. To do this, we created a wild barley AB-NAM population by backcrossing 25 wild barley accessions to the 6-rowed malting barley cultivar *Rasmusson*. The 25 wild barley parents were selected to capture approximately ~90% of the allelic content of the wild barley diversity collection (WBDC), which contains 318 accessions sourced from across wild barley's native range. The resulting set of 796 $BC_2F_{4:6}$ lines were genetically characterized and analyzed in augmented field trials for agronomic, yield, grain protein, and wax production traits. Additionally, we analyzed the *Harrington* (2-rowed malting barley cultivar) x *OUH-602* (wild barley accession), biparental AB population for yield and malting quality characteristics. The AB-NAM population was genotyped with 384 SNP markers and 263,531 markers were imputed onto the population from exome capture sequence of the parents. Linkage disequlibrium in the AB-NAM was significantly lower than the HOUH biparental mapping population, indicating a higher potential mapping resolution. Qualitative traits were mapped to

candidate gene resolution and beneficial alleles were identified for several quantitative traits. Ultimately, the AB-NAM population will serve as a community resource for barley breeders and geneticists to explore a large proportion of wild barley germplasm in a single, relatively adapted mapping population.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 2

Chapter 4

List of Supplemental Datasets

Chapter 2

Chapter 3

Chapter 1: Literature Review

Introduction

Barley (*Hordeum vulgare* ssp. *vulgare*) production in Minnesota has decreased by an order of magnitude since the early 1990s. Due to a complex set of economic, agro-ecological, and disease pressures, this decline reflects the demands placed on plant breeders to continue improvement despite mounting challenges, some of which are unpredictable. To decrease the vulnerability of barley to such devastating disease pressures or climatic changes, efforts must be taken to expand the genetic diversity of the barley breeding germplasm. The progenitor of cultivated barley, wild barley (*Hordeum vulgare* ssp. *spontaneum*), is a rich source of genetic and phenotypic diversity. But, accessing this diversity in the context of a breeding program can be difficult. The malting industry in the United States has stringent quality requirements, and breeders have been reluctant to introduce exotic germplasm that might disrupt the beneficial alleles that have accumulated over many cycles of breeding. Nevertheless, with advances in molecular marker technology, statistical methodologies, and novel breeding designs, exotic germplasm is becoming more accessible to breeders, and studies are showing that useful variation is obtainable from exotic sources.

Barley Production and Breeding

Barley is an economically important crop, grown in most temperate areas of the world, particularly in those environments with limited rainfall and short growing seasons that are less conducive to other major cereal crops. Used for animal feed, malt production, and human consumption, barley is the fourth most produced cereal crop

worldwide (Ullrich 2010). Cultivated barley is often categorized according to three major criteria: spike morphology (six-rowed or two-rowed), seasonal growth habit (spring or winter), and end use (animal feed or malting). Of the barley produced worldwide, approximately 75% is used for animal feed, 20% for malt production, and the remaining 5% is for human consumption or seed production (Blake et al. 2010). Of the US barley growing regions, the Red River Valley area of northwestern Minnesota has climate and soil particularly conducive to growing high quality, six-rowed malting barley. Until the early 1990s, farmers in this region produced approximately 10% of the US barley crop (USDA-NASS 2015) and garnered high premiums by selling their harvest to the local malting industry (Windels 2000). However, barley production in the US has steadily decreased over the past two decades, due to a variety of economic, ecological and cropping systems shifts. In Minnesota and other parts of the North American barley growing region, this decrease has been accelerated by the emergence of the devastating cereal disease Fusarium Head Blight (FHB). Caused chiefly by the fungus *Fusarium graminearum*, FHB can severely reduce grain yield and contaminate harvested grain with mycotoxins such as deoxynivalenol (DON). Even low levels of DON can result in outright rejection or severe discount of the grain, causing economic hardship for growers (Windels 2000). Besides FHB, barley breeders must be cognizant of several other economically important disease threats, changes in climate that affect growing conditions, pressures to reduce nutrient inputs, and a series of seed quality characteristics that must be maintained.

In Minnesota, barley breeding efforts have focused on producing high quality six-rowed malting varieties to support the malt barley industry of the Upper Midwest. The

University of Minnesota has maintained a barley breeding program since the early 1900s, and has used an advanced cycle breeding technique since 1958. Over the past 50 years, the Minnesota barley breeding program has produced many highly successful six-rowed malting varieties using this technique. Advanced cycle breeding is defined as crossing elite lines within a closed breeding population, over many generations, without the introduction of diverse germplasm (Bernardo 2002). Because the US malting industry has such stringent quality requirements, breeders have been reluctant to use exotic germplasm, which might disrupt the beneficial alleles that have accumulated over many cycles of breeding (Rasmusson and Phillips 1997). Still, studies have shown that using an advanced cycle breeding strategy can lead to a reduction of genetic diversity in the breeding program (Condón et al. 2008; Fu and Somers 2009). This balance between maintaining sufficient diversity for improvement and imposing strict selection for favorable alleles is integral to the breeding process.

In recent years, the availability of inexpensive, high-throughput molecular markers has revolutionized the breeding process and provided insight into the nature of selection within the breeding program. Condón et al. (2008) assessed the genetic diversity of a set of historic barley breeding lines from the University of Minnesota breeding program using molecular markers and found a nonuniform, genome-wide reduction in diversity. Transcriptome analysis of the breeding program found similar results. Muñoz-Amatriaín et al. (2010) reported that the number of differentially expressed genes decreased as the breeding program progressed. The most notable decrease in both genetic diversity and differential expression occurred in regions of the genome subject to selection for disease resistance or malt quality characteristics. Still,

only a relatively small portion of the loci sampled exhibited this reduction. Both Muñoz-Amatriaín et al. (2010) and Condón et al. (2008) concluded that despite the reductions in diversity, there is still enough remnant genetic diversity for improvement in the University of Minnesota barley breeding program, with the caveat that introduction of diversity is recommended for long-term improvement.

Barley Genome and Molecular Genetic Resources

Barley is a predominantly self-pollinating species in the Triticeae tribe of grasses. Barley maintains a genome size of approximately 5.1Gb spread across 7 chromosomes. While substantially less complex than the allo-hexaploid wheat genome, the barley genome contains a large amount of repetitive sequence that has inhibited whole-genome sequencing efforts. In 2012, the International Barley Genome Sequencing Consortium (IBSC 2012) released a barley sequence assembly which included a 4.98 Gb physical map, of which 3.90 Gb are anchored to the barley genetic map. They found that over 84% of the genome is composed of mobile or repetitive elements and that pericentromeric and centromeric regions with low recombination comprised approximately 37% of the genome. The sequencing efforts identified 26,159 high confidence gene calls, with higher densities of genes located towards the ends of chromosomes. Additionally, when sequences of four cultivars and a wild barley accession were compared, the frequency of single nucleotide variants (SNVs) was reduced in centromeric regions, particularly in the cultivated barleys.

Resources related to the barley genome sequence are quickly expanding, and new methods of sequence-based genotyping technologies are becoming more accessible to

barley breeders and geneticists. Close et al. (2009) produced the first widely available, high-density single nucleotide polymorphism (SNP) genotype assay for barley. The Barley Oligo Pool Arrays 1 & 2 (BOPA1 & BOPA2) are a set of two 1,536-SNP Illumina GoldenGate assays which were developed from expressed sequence tags (ESTs) and sequence amplicons derived primarily from cultivated barleys. Close et al. (2009) developed a genetic map using this platform that was updated by Muñoz-Amatriaín et al. (2011) to include additional populations and higher map resolution. The Muñoz-Amatriaín et al. (2011) map (M11) spans 1,137.3 cM with 2,994 SNP loci mapped to 1,163 unique positions. The subsequent generation of barley SNP-genotyping platform used the 9,000-SNP Illumina iSelect technology. Of these 9,000 SNPs, 2,832 were represented on the BOPA assays (Comadran et al. 2012). The Muñoz-Amatriaín et al. (2014) iSelect consensus map (M14) combined the original *Morex* x *Barke* Comadran et al. (2012) map with the maps used to generate the M11 consensus map. The resulting map spans 1,113 cM with 5,665 SNP loci mapped to 2,032 unique positions. The M14 map and the iSelect assay improved the marker density from an average of one marker bin every 0.99 cM to one marker bin every 0.55 cM.

Because these platforms survey a preselected set of biallelic markers, developed primarily from cultivated barley genotypes, using these biased markers to assess wild or distantly related germplasm can affect the conclusions drawn regarding diversity and other population genetic parameters. This phenomenon, termed ascertainment bias, results from a biased sampling of the variants, causing an over-emphasis on alleles common to the sampled group and an underrepresentation of rare alleles that may be present at higher frequencies in different populations (Nielsen 2000). The effect of

ascertainment bias on assessing diverse barley germplasm was examined by Moragues et al. (2010) who found that estimates of barley landrace diversity were lower than estimates of barley cultivar diversity using the BOPA1 SNP assay, despite evidence to the contrary (Russell et al. 2004; Jilal et al. 2008; Russell et al. 2011; Khodayari et al. 2012).

Two strategies can circumvent the effects of ascertainment bias: using markers which are designed based on the relevant germplasm, and using markers which are intrinsically less subject to bias. To develop a genetic map and marker platform directly relevant to wild barley germplasm, Alsop et al. (2011) used a Diversity Arrays Technology (DArT) genotyping platform based on wild barley germplasm. They developed a 1,159 cM map composed of 2,935 DArT markers. The DArT platform is an effective genotyping system that produces high-throughput markers with little sequence information (Wenzl et al. 2004). Now that sequencing technologies are becoming more affordable, researchers are looking to sequence-based marker technologies, which can capture thousands or hundreds of thousands of variants, without bias and without restriction to biallelic SNP variation. Currently, the two sequence based technologies employed in the barley genome are Genotyping by Sequencing (GBS) and exome capture sequencing. These strategies produce large quantities of genomic sequence by reducing the representation of the genome prior to sequencing, thereby minimizing the quantity of repetitive DNA in the sequenced pool. Poland et al. (2012) demonstrated the utility of GBS in barley by mapping over 34,000 SNPs with a genotyping cost less than other whole genome SNP platforms. Exome capture sequencing produces even greater sequence coverage than GBS, targeting 61.6 Mb of coding sequence, and approximately

86% of the genes in the barley genome (Mascher et al. 2013b). Additionally, this technology has been shown to be effective in sequencing wild barley as well as more distantly related barley species (Wendler et al. 2014). The variants identified from these sequencing technologies, in addition to the SNP markers represented on the BOPA and iSelect assays, can be aligned to the IBSC (2012) *Morex* genome, which has been anchored to a population sequencing (POPSEQ) genetic map that was developed by sequencing 90 individuals of the *Morex* x *Barke* recombinant inbred line (RIL) and 82 individuals of the Oregon Wolfe Barley mapping population (Mascher et al. 2013a).

Barley Domestication

Wild barley can be found throughout the Fertile Crescent region of the Middle East, extending east into Central Asia, north into the Caucasus region, and west into North Africa (von Bothmer 1995). Barley is one of the oldest grains used by humans, with cultivation beginning approximately 10,000 years ago (Zohary et al. 2012). The domestication history of a crop can influence the accessibility and population dynamics of germplasm for modern crop improvement. Considerable research has been conducted surrounding the domestication history of modern crop species. Archaeological evidence, coupled with genetic signatures of domestication, reveals information about the origin and spread of domesticated barley. But, differing germplasm and analysis techniques have led to debate regarding the timing and location of barley domestication. While Badr et al. (2000) concluded that domesticated barley had a single monophyletic origin near Israel or Jordan with additional diversification after domestication in the Himalayan region, Morrell and Clegg (2007) claim that haplotype diversity of wild and cultivated

barleys indicate two domestication events, one in the Fertile Crescent and one farther east. Furthermore, they showed that these domestication events appear to have differential contributions to modern crop germplasm, with the majority of US and European cultivars derived from the Fertile Crescent germplasm, and Central Asian cultivars were more closely related to the wild barley accessions in the eastern domestication region. This difference has implications for sampling and preserving diversity across the range.

The defining characteristics of domestication vary depending on the species, but there is often a suite of characteristics that influenced the success of the crop under cultivation, sometimes referred to as the domestication syndrome (Doebley et al. 2006; Brown et al. 2009). In barley, the key domestication traits include: a non-brittle rachis which confers a non-shattering phenotype, a six-rowed spike morphology, and the naked caryopsis or free-threshing hull type (Pourkheirandish and Komatsuda 2007). The non-shattering phenotype is the most important domestication trait, allowing farmers to easily harvest grain that is retained on the spike. The responsible non-brittle rachis morphology is controlled by two tightly linked genes *Btr1* and *Btr2* that have been fine mapped to chromosome 3H (Komatsuda et al. 2004). Further analysis of these loci suggests that there are two distinct origins of the non-shattering phenotype, corresponding to the eastern and western areas of domestication, corroborating the two independent domestication events (Azhaguvel and Komatsuda 2007). Although the trait originated only twice, the domestication history is not simple. Archaeological evidence suggests that it took approximately 2,000 years for the non-shattering phenotype to become fixed in the cultivated barley population (Purugganan and Fuller 2009).

The six-rowed phenotype and the naked caryopsis phenotype originated later in evolutionary history. The six-rowed phenotype appears in the archaeological evidence around 8,800-8,000 years ago, and dominated the cultivation until around 1,000 years ago when two-rowed types reemerged in Europe (Komatsuda et al. 2007). The *Vrs1* gene on chromosome 2H has been cloned, and like the non-brittle rachis phenotype, multiple independent mutational events led to the six-rowed phenotype (Komatsuda et al. 2007). The *Nud* gene on chromosome 7H is responsible for the naked caryopsis phenotype that confers a non-adhering hull more suited for human consumption (Taketa et al. 2008). Unlike *btr1*/*btr2* and *vrs1*, there appears to be a single origin of the naked barley phenotype (Taketa et al. 2008). In modern cultivation, this phenotype is less represented because the hull is favored for malting and animal feed uses.

Besides these morphological traits that define boundaries between wild and cultivated barley, diversification of genes involved in reduced photoperiod sensitivity and reduced vernalization requirements allowed early barley domesticates to spread beyond wild barley's native range (Pourkheirandish and Komatsuda 2007). There are two main photoperiod genes: *Ppd-H1* on chromosome 2H that controls flowering time under long-day conditions, and *Ppd-H2* on chromosome 1H that controls flowering time under short-day conditions (Laurie et al. 1994). The *Ppd-H1* gene is involved in circadian clock function, and individuals expressing the dominant, photoperiod sensitive allele flower much earlier than those with the insensitive allele under long-day conditions (Turner et al. 2005). Besides differences in flowering time, *Ppd-H1* has pleiotropic effects on plant height, yield, and yield components (Laurie et al. 1994). Jones et al. (2008) sequenced *Ppd-H1* in 194 landrace and 72 wild barley accessions and identified a reduction of

diversity in landraces that is characteristic of a domestication bottleneck. Furthermore, they found that the causative SNP and associated haplotypes are strongly correlated with geographic distribution, providing evidence that this trait influenced local adaptation and domestication. In addition to the direct affect on flowering, photoperiod can also interact with vernalization, allowing for more specific adaptation to regions where winter conditions are too harsh for fall-planted barley (Trevaskis et al. 2006).

Barley Germplasm Resources

There are extensive barley germplasm collections that researchers can explore for cultivar improvement and gene discovery. Barley's primary gene pool includes current and historic elite germplasm, early domesticates and landraces, wild barley, and extensive induced and natural mutant collections (Lundqvist and Franckowiak 2003; van Hintum and Menting 2003). These germplasm collections are maintained throughout the world in private and public breeding programs, as well as in local, national, and international germplasm repositories. In the United States, the US Department of Agriculture-Agricultural Research Service (USDA-ARS) maintains the National Small Grains Collection (NSGC), which houses over 32,000 cultivated barley accessions, as well as approximately 1,600 wild barley accessions (http://www.ars-grin.gov/npgs/; Bockelman and Valkoun 2010). Beyond barley's primary gene pool resources, a single species, *Hordeum bulbosum*, makes up the secondary gene pool, and an additional 30 species in the genus *Hordeum* comprise barley's tertiary gene pool (von Bothmer 1995). Because of the vast quantity of accessions present in germplasm collections, efforts to genetically and phenotypically characterize these populations are laborious.

To make germplasm collections more manageable, curators have developed core collections with fewer accessions that act as a reduced representation of the whole collection. These lines are often selected as a function of geographic origin or genotypic characterization, and these smaller collections are able to be more extensively characterized. Two notable collections are the NSGC Barley Core (Bockelman and Valkoun 2010; Muñoz-Amatriaín et al. 2014), and the Wild Barley Diversity Collection (WBDC) (Steffenson et al. 2007). The NSGC core collection was genotyped with the barley iSelect 9K SNP platform and narrowed from 2,417 accessions to an informative set of 1,860 unique lines comprising 815 landraces, 781 cultivars/breeding lines, 21 genetic stocks, and 243 undefined accessions (Muñoz-Amatriaín et al. 2014). The WBDC comprises 318 wild barley accessions selected from the International Center for Agricultural Research in the Dry Areas (ICARDA) germplasm collection based on ecogeographical diversity. The accessions are primarily from the Fertile Crescent region of the Middle East, with fewer lines from Central Asia, North Africa, and the Caucasus region (Steffenson et al. 2007). These accessions have been genotyped with DArT, DArTseq, BOPA1 and BOPA2 markers (Roy et al. 2010; Ames et al. 2015). Genetic characterization of these and other core collections presents opportunities for exploring marker-trait associations, population dynamics, and it sets the basis for developing mini-core collections which can be even more extensively examined.

Linkage Disequilibrium

Linkage disequilibrium (LD) is the nonrandom association of alleles at two or more loci. The distribution of LD throughout the genome depends on several population-

dependent factors. In traditional biparental linkage mapping populations, the primary contributor to LD is direct linkage of adjacent polymorphisms. This linkage allows for high power to detect marker trait associations, with the resolution of mapping dictated by the number and distribution of recombination events present in the population (Zhu et al. 2008). In contrast, LD in natural populations has a more complex basis. The history of the population, with regard to selection, mating, genetic drift, and mutation, influences the measure of LD. Because of this researchers must consider LD in the design of mapping studies (Slatkin 2008). Linkage mapping studies generally require a relatively low number of markers to adequately cover the genome because LD extends to relatively large distances in the genome. Unlike mapping in traditional linkage mapping populations, Genome-wide association studies (GWAS) do not require crossing and controlled population development. Instead, existing germplasm resources are analyzed directly for traits of interest (Rafalski 2002a). While the development of these populations requires less time for crossing, the selection of lines to include in the study can have implications for the resolution and power of mapping.

Association mapping studies have been widely used in human genetics research, and only recently have these techniques been adopted by plant genetics researchers (Rafalski 2002a). GWAS relies on lower LD within the natural populations to associate markers with traits at a higher resolution due to a more extensive history of recombination events than what is present in linkage mapping populations (Zhu et al. 2008). The use of these populations presents two main statistical difficulties. First, cryptic relatedness or population structure within the mapping population can confound marker-trait associations, leading to false positives if not properly controlled (Yu et al.

2006; Vilhjálmsson and Nordborg 2013), and second, small effect loci and low frequency alleles are difficult to detect (Korte and Farlow 2013). Statistical methodologies have addressed some of these difficulties. Mixed-model algorithms can account for kinship and/or population structure, minimizing the number of false positives (Yu et al. 2006; Price et al. 2010). Despite these drawbacks, GWAS has become a routine mapping strategy, thanks to the ease of population development, and the potential for high resolution mapping.

In barley, GWAS has been performed in current and historical breeding germplasm (Rostoks et al. 2006; Cockram et al. 2010; Massman et al. 2011; Comadran et al. 2011; Pasam et al. 2012), as well as landrace collections (Muñoz-Amatriaín et al. 2014) and wild barley collections (Steffenson et al. 2007; Roy et al. 2010; Ames et al. 2015) to identify a variety of marker-trait associations. Studies have shown that LD tends to be higher in elite barley breeding populations than in wild barley populations (Morrell et al. 2005; Caldwell et al. 2006; Morrell et al. 2014), but both types of populations have significant population structure that must be accounted for (Steffenson et al. 2007; Waugh et al. 2009; Cockram et al. 2010; Zhou et al. 2012). Because of these differences in LD, Caldwell et al. (2006) suggested a combined approach that involves mapping in wild barley populations to enhance the resolution of mapping over cultivated barley populations. While this strategy is promising because of the low LD and high allelic diversity found in wild barley populations, phenotyping wild barley accessions for traits of interest to barley breeders can be logistically difficult. Furthermore, strong selection on loci may influence local LD and consequently affect mapping results (Caldwell et al. 2006; Stracke et al. 2007).

Utilizing Exotic Germplasm for Crop Improvement

Utilizing exotic germplasm in a breeding program is difficult because deleterious, unadapted traits can mask beneficial genetic variation, and once such beneficial variation is identified, it can be challenging to integrate the beneficial alleles without bringing along deleterious traits, a challenge known as linkage drag (Tanksley and Nelson 1996). Over the course of domestication and cultivar improvement, genetic bottlenecks lead to the loss of alleles, some of which may have value to modern crop breeders (Tanksley and McCouch 1997; McCouch et al. 2013). Ellis et al. (2000) estimated a mere 40% of the alleles present in wild barley are found in cultivated barley. These alleles that have been lost include beneficial alleles for economically important traits such as disease resistance (Fetch et al. 2003; Steffenson et al. 2007; Schmalenbach et al. 2008), abiotic stress tolerance (Ellis et al. 2000; Lakew et al. 2013b; Kalladan et al. 2013), and quality traits (Tanksley and McCouch 1997; von Korff et al. 2008). Genome-wide molecular markers can aid in both the discovery and integration of beneficial traits from exotic germplasm. Two commonly used strategies to understand the genetic architecture of traits in exotic germplasm are advanced backcross QTL (AB-QTL) mapping and association mapping. Once beneficial alleles are identified, breeders can perform marker assisted backcrossing to integrate alleles, or newer methods such as genomewide selection, which attempts to integrate diversity for more complex traits (Bernardo 2009; Combs and Bernardo 2013).

The AB-QTL technique developed by Tanksley and Nelson (1996), involves backcrossing an exotic donor to an adapted recurrent parent two or more times and selfing to develop a linkage mapping population of individuals that are primarily adapted, with smaller portions of unadapted genome. This approach is beneficial, because it

minimizes the amount of unadapted alleles in individuals, allowing for a more accurate assessment of their genetic value. By containing fewer exotic alleles, the lines can be screened in target environments, and the effects of background or epistatic interactions developed from a wide cross are minimized. Additionally, the time needed to introgress beneficial alleles is lessened, because the introgression process has already begun (Tanksley and Nelson 1996). This technique has been used in many crop species to identify valuable alleles from unadapted germplasm (Wang and Chee 2010). Several wild barley x cultivated barley advanced backcross populations have been developed and analyzed for a wide variety of traits, including: yield and agronomic traits (Li et al. 2005; von Korff et al. 2006; Gyenis et al. 2007; Eshghi et al. 2013), malting quality (von Korff et al. 2008; Schmalenbach and Pillen 2009; March et al. 2012), disease resistance (von Korff et al. 2005; Yun et al. 2006; Li et al. 2006), abiotic stress tolerance (Saal et al. 2011; Hoffmann et al. 2012; Schnaithmann and Pillen 2013; Kalladan et al. 2013), and root traits (Naz et al. 2014).

Multiparent Populations

Researchers have begun to explore next generation population design as the next step towards understanding the genetic control of complex quantitative traits (Cavanagh et al. 2008; Myles et al. 2009; Morrell et al. 2012). These multiparent populations involve more intricate mating designs than biparental populations. Designs vary with regard to particular aims, as well as practical considerations of mating habit and generation time of the species involved. Collaborative efforts have been undertaken to develop and analyze these large populations in several model organisms: the Drosophila

Synthetic Population Resource (DSPR) (King et al. 2012a; King et al. 2012b), the mouse Collaborative Cross (Churchill et al. 2004; Aylor et al. 2012), the Arabidopsis multiparent RIL (AMPRIL) (Huang et al. 2011), and the Arabidopsis Multiparent Advanced Generation Inter-Cross (MAGIC) (Kover et al. 2009). Each of these populations uses a set of diverse founder lines that are intermated in a single generation or over multiple generations, then selfed or sib-mated to create inbred lines available for extensive phenotypic analysis. MAGIC populations have also been created for several crop species including: barley (Sannemann et al. 2015), rice (Bandillo et al. 2013), tomato (Pascual et al. 2014), and bread wheat (Verbyla et al. 2014). The maize Nested Association Mapping (NAM) population was the first multiparent population developed in a crop species that utilized modern genetic tools for high resolution genetic mapping. Unlike the intercross populations described above, the NAM design involves crossing several diverse lines to a reference genotype, without intercrossing the founding lines (Yu et al. 2008). The NAM design has also been adopted by several other crops, including: wheat and barley (http://www.triticeaecap.org/, Schnaithmann et al. 2014), soybean (Stupar and Specht 2013), and sorghum (Mace et al. 2013).

These population designs share a few notable benefits over biparental linkage mapping and over association mapping. By using multiple parents in the mating design, multiparent populations capture a larger sample of allelic diversity than their biparental counterparts (King et al. 2012a). While they are unlikely to capture all of the causal rare variants that are contributing to variation in an association mapping panel, the crossing design increases the allele frequencies of the rare alleles present in the founding lines to detectable levels. This results in more statistical power to detect rare variants than would

be possible in association analysis (Yu et al. 2008). Additionally, the multiparent design allows for evaluation of multi-allelic series. When multiple alleles at a single locus contribute to variation for a trait, both linkage mapping and association mapping can fail to attribute the appropriate genetic variation to the locus (Long et al. 2014). Because these rare alleles and multi-allele loci are difficult to detect using association mapping, some have suggested that multiparent populations may minimize the missing heritability problem by accounting for additional sources of genetic variation (Long et al. 2014).

In addition to the added benefits of sampling multiple alleles and increasing the detection of rare alleles, the crossing schemes in multiparent populations create relatively uniform genetic backgrounds that minimize population structure (Yu et al. 2008). In association mapping studies, population structure can lead to false negatives when variation for a trait is confounded with population structure or to false positives when population structure is not appropriately controlled (Larsson et al. 2013). One of the major advantages of association mapping over linkage mapping is the added resolution enabled by low levels of LD in association mapping panels (Rafalski 2002a). While the multiparent populations retain large linkage blocks from founder parents, the different haplotypes of the founders nested within those linkage blocks creates lower LD and consequently, higher resolution in multiparent populations (Yu et al. 2008). These linkage blocks also enable direct projection of parental haplotypes onto progeny, allowing for high density marker data to be projected onto the lower density genotyping of the progeny (Guo and Beavis 2011).

The maize NAM population was developed from 25 diverse lines crossed to a single reference line, *B73*. From each of the 25 crosses, 200 RIL lines were developed,

for a total of 5,000 NAM lines (Yu et al. 2008). The population has been genotyped with 1,106 SNP markers, the parents were sequenced, and from the sequence data, an additional 1.6 million markers were imputed onto the population (Tian et al. 2011). Traits that have been measured include: flowering time (Buckler et al. 2009), disease traits (Kump et al. 2011; Poland et al. 2011), and agronomic traits (Tian et al. 2011; Peiffer et al. 2013; Peiffer et al. 2014). The most recent maize NAM analyses have involved a joint linkage analysis (Buckler et al. 2009), followed by a GWAS using a bootstrapping approach that sampled 80% of each family 100 times. Background effects were controlled by analyzing each chromosome while correcting for effects of other chromosomes (Tian et al. 2011). By coupling a linkage mapping approach with an association mapping approach, the maize NAM approach is able to identify many significant marker trait associations, minimize false positives, and assign genotypic effects of alleles at significant markers for each parent (Poland et al. 2011). The results of these studies in maize show that many traits are controlled by complex genetic architecture, with many small effect loci contributing to the phenotypes observed. Similar analyses need to be conducted to determine if inbreeding species such as barley will reveal the same trends.

In sorghum and barley, researchers have altered the maize NAM design to include a backcross generation. The sorghum BC-NAM was a repurposing of previously developed populations aimed at moving diverse sorghum germplasm into an adapted Australian sorghum breeding line (Jordan et al. 2011; Mace et al. 2013). A total of 56 accessions, including elite breeding lines, landraces, and weedy sorghums were each backcrossed to a single sorghum line to produce 30-90 BC_1F_4 lines per family. Selection

for height, maturity, and other characteristics was used to maintain a relatively adapted population. This selection led to significant deviations from the expected 25% diverse parent allele frequency in some areas of the genome. Mace et al. (2013) used over 1,400 lines in 24 of the families to analyze the genetic architecture of sorghum flowering time. Using multi-populational QTL analysis and mixed model association mapping, they identified 40 QTL of relatively small and additive effect, a result similar to mapping in the maize NAM (Buckler et al. 2009).

Schnaithmann et al. (2014) developed a wild barley backcross – NAM population named Halle Exotic Barley (HEB). The parents of the HEB population were selected from the wild barley collection described by Badr et al. (2000), based on their origin across the Fertile Crescent. Each of the wild barley parents were crossed to the German spring barley *Barke* and the resulting F_1 plants were backcrossed once to *Barke*, giving rise to 20 BC_1 plants and 50-71 BC_1S_1 lines, making up the 295 HEB-5 mapping lines. Lines were genetically characterized with the BOPA1 marker platform, and phenotypically characterized for seedling leaf rust resistance. In two cases, a higher than expected wild barley genotype frequency was observed on chromosome 1H in HEB-F08 and on chromosome 5H in HEB-F14. In another family, HEB-F23, lower than expected wild barley frequency was identified on chromosomes 1H, 6H and 7H. To identify marker-trait associations, Schnaithmann et al. (2014) conducted two mixed model analyses, one across all families, and a second analysis which was performed within each family. Eight QTL were detected, four across families, and four within a single family. The HEB-5 is a subset of a larger HEB-25 population comprising 1,420 BC_1S_3 individuals in 25 families (Maurer et al. 2015). This larger population was genotyped

with the 9K iSelect marker platform and scored for flowering time in three environments. Heritability for heading date was estimated on a line-mean basis of 91.6%. Using stepwise multiple linear regression, eight major QTL were identified that collectively explained 60% of the phenotypic variance. Of these QTL, a SNP located in the *Ppd-H1* gene on chromosome 2H explained 36% of the variance. Other loci detected colocalized with vernalization and flowering time regulation genes *Vrn-H1*, *Vrn-H2*, *Vrn-H3*, *HvCEN*, *denso*, and *HvELF3*. Maurer et al. (2015) compared prediction accuracies of various flowering time models. Using *Ppd-H1* haplotype classifications slightly improved the prediction accuracy from 36% to 38% over the single marker *Ppd-H1* model. The 8-SNP model explained 64% of the variance, and genome-wide marker models using RR-BLUP and BayesCπ improved the model to 71% and 74% respectively. Lastly, adding epistatic interactions between the significant main effect SNPs increased the model to 77%. They concluded that unlike maize and sorghum, large effect loci are responsible for most of the variation in flowering time in barley.

Ultimately, multiparent populations incorporate many favorable characteristics of association mapping and linkage mapping, without many of their drawbacks. The use of these populations is still relatively recent, and it is likely that population designs and analysis techniques will evolve as empirical data demonstrates specific strengths and weaknesses. Among the questions which remain unanswered are: which and how many founding genotypes to use, how large the populations should be, which crossing design is most effective, and how best to sample from within populations for smaller trials. Many of these questions are bounded by logistical considerations such as difficulty of crossing, size and availability of field plots, and genotypic capacity.

Chapter 2: Development and genetic characterization of an Advanced Backcross – Nested Association Mapping (AB-NAM) population of wild × cultivated barley

Overview

To increase the accessibility of wild barley germplasm to barley breeders and geneticists, we developed a wild barley advanced backcross-nested association mapping (AB-NAM) population. This population was created by backcrossing 25 wild barley accessions to the 6-rowed malting barley cultivar *Rasmusson*. The 25 wild barley parents were selected from the 318 accession Wild Barley Diversity Collection (WBDC) to maximize the allelic diversity of the chosen set at 1,402 SNP markers, 556 DArT markers, and 46 SSR markers. The resulting 796 $BC_2F_{4:6}$ lines were genotyped with 384-SNP markers, and an additional 4,022 SNPs and 263,531 sequence variants were imputed onto the population using 9K iSelect SNP genotypes and exome capture sequence of the parents. On average, 96% of each wild parent was introgressed into the *Rasmusson* background. To compare mapping resolution potential, genomewide linkage disequilibrium (LD) was calculated for the AB-NAM, the WBDC, and two biparental wild x cultivated barley mapping populations: the Harrington x OUH-602 advanced backcross (AB) and recombinant inbred line (RIL) populations. While LD decay (r^2 = 0.2) was lowest in the WBDC (0.36 cM), the AB-NAM (9.2 cM) exhibited quicker LD decay than the AB (28.6 cM) and RIL (32.3 cM) populations. Three qualitative traits: glossy spike, glossy sheath, and blue hull color were mapped with high resolution to loci corresponding to known barley mutants for these traits. Additionally, a total of 10 QTL were identified for the quantitative trait grain protein content. The combination of low

LD, negligible population structure, and high diversity in an adapted background make the AB-NAM an important tool for exploring wild barley germplasm.

Introduction

Diverse germplasm collections are valuable resources for continued crop improvement. However, using these resources is a challenge that is often avoided by breeders due to the time and effort required to identify and deploy beneficial exotic alleles. Breeding for complex traits requires balancing the introduction of genetic diversity with maintaining the selective progress obtained over many cycles of breeding (Bernardo 2002). Due to the stringent malting quality requirements imposed by North American malting and brewing industries, barley (*Hordeum vulgare* ssp. *vulgare*) breeding has been restricted to a narrow germplasm base and focused on elite by elite crosses (Rasmusson and Phillips 1997). Over many cycles of breeding, extensive genome-wide linkage disequilibrium (LD) develops in closed breeding populations (Fang et al. 2013), and the genetic diversity of these populations becomes reduced (Condón et al. 2008; Fu and Somers 2009; Muñoz-Amatriaín et al. 2010). The need to expand the genetic diversity of the breeding pool has become evident as breeders face disease and environmental pressures which are threatening crop production. Today, genomics technologies are advancing our ability to understand the genetic basis of variation across barley germplasm, and providing the opportunity to look beyond conventional sources of genetic diversity for sustained crop improvement.

Wild barley (*Hordeum vulgare* ssp. s*pontaneum)*, the progenitor of cultivated barley, is a rich source of genetic diversity for genetic studies and breeding programs. Genetic analysis of cultivated barley has shown that the crop has retained only 40% of the diversity existing in the wild populations (Ellis et al. 2000). Despite low levels of outcrossing in both wild and cultivated barleys (~0-2%) (Abdel-Ghani et al. 2004), wild barley populations exhibit much lower LD than typical breeding populations (Morrell et al. 2005; Caldwell et al. 2006; Hamblin et al. 2010). The heightened level of LD present in cultivated barley is likely due to population parameters such as geographic structure, selection, and reduced effective population size under cultivation. The combination of low LD and high diversity in wild barley germplasm collections presents an opportunity for high resolution association mapping, which has been used successfully to identify loci associated with barley disease resistance (Steffenson et al. 2007; Roy et al. 2010). Still, association mapping within wild barley populations presents challenges when assessing agronomic traits of interest. For example, wild barley has a brittle rachis that contributes to seed shattering (Pourkheirandish and Komatsuda 2007), and many accessions are not well adapted to agronomic growing environments. These characteristics make assessing wild barley in field trials difficult, particularly for agronomic traits of interest.

The advanced backcross (AB) technique was developed to address the difficulties of using unadapted germplasm for trait mapping and cultivar improvement (Tanksley and Nelson 1996). Instead of developing typical F_2-derived recombinant inbred line (RIL) mapping populations, AB populations are composed of multiple-backcross derived RILs, with an exotic donor parent crossed to an adapted recurrent parent. With a much smaller portion of the exotic genome present in each of the lines, the effects of deleterious alleles

are reduced, allowing researchers to perform field experiments to analyze the value of exotic alleles in the context of cultivated germplasm. Furthermore, hybrid or epistatic effects are minimized allowing for more accurate measures of additive genetic variation, and since at least two rounds of backcrossing have already been performed, the process of introgressing traits has already been initiated should beneficial alleles be identified (Tanksley and Nelson 1996). AB populations have been developed and used successfully to identify beneficial alleles in several crops, including tomato, rice, wheat, maize, cotton (reviewed in Wang and Chee 2010), and barley (Pillen et al. 2003; Matus et al. 2003; von Korff et al. 2004; Yun et al. 2005; 2006; Li et al. 2006). The most extensively characterized wild × cultivated barley AB population, *ISR42-8* × *Scarlett*, has been analyzed for many traits including disease resistance, agronomic characteristics, malting quality, nitrogen use, drought stress response, and root architecture (von Korff et al. 2005; von Korff et al. 2006; von Korff et al. 2008; Saal et al. 2011; Sayed et al. 2012; Naz et al. 2014).

While AB analysis has identified quantitative trait loci (QTL) for valuable traits, it is limited by the genetic content of the parents selected, and the controlled crossing scheme leads to extensive LD, minimal recombination events, and unbalanced allele frequencies (Tanksley and Nelson 1996). Traditional QTL mapping techniques suffer similar limitations, resulting in relatively low resolution mapping. To achieve higher resolution mapping using more diverse germplasm, association mapping has become a popular technique for mapping quantitative traits in crop species (Rafalski 2002a). But, cryptic relatedness or population structure within association mapping populations can confound marker-trait associations, leading to false positives if not properly controlled

(Yu et al. 2006; Vilhjálmsson and Nordborg 2013), and small effect loci and low frequency alleles are difficult to detect (Korte and Farlow 2013). To address these limitations of linkage and association mapping techniques, researchers are analyzing populations developed from more complex crossing schemes that involve multiple parental lines and a family based association mapping approach (Guo et al. 2013). The nested association mapping (NAM) approach developed by Yu et al. (2008) combines the power of linkage mapping with the resolution of association mapping by crossing a diverse set of lines to a single reference genotype. This population structure nests ancestral LD within novel recombination events, allowing for imputation of high density genotypic data from parental lines, high power and high resolution mapping, and utilization of diverse germplasm (Yu et al. 2008).

Genetic characterization of the maize NAM population revealed patterns of recombination and segregation distortion within the maize genome (McMullen et al. 2009), and trait mapping has identified numerous QTL for flowering time, northern leaf blight resistance, leaf architecture, and plant height (Buckler et al. 2009; Tian et al. 2011; Poland et al. 2011; Peiffer et al. 2014). The NAM design is one of several multi-parent mapping strategies that have been developed to dissect complex trait architecture. Population design varies based on considerations such as mating system and resource availability. Multi-parent inter-cross populations that utilize more complex crossing designs have been developed in mouse (Churchill et al. 2004), Arabidopsis (Kover et al. 2009; Huang et al. 2011), rice (Bandillo et al. 2013), wheat (Rebetzke et al. 2014), and barley (Sannemann et al. 2015), and backcross-NAM designs have been developed in sorghum (Jordan et al. 2011) and barley (Schnaithmann et al. 2014).

Here, we describe the development and genetic characterization of a resource for barley breeders and geneticists that combines the population development schematics of AB and NAM populations. Our objectives were: (1) to develop a barley population that incorporates a large amount of exotic germplasm, but is adapted enough to be analyzed in standard field trials; (2) to identify regions of the barley genome that are subject to segregation distortion during wild barley introgression; (3) to compare the linkage disequilibrium structure of four different populations that utilize exotic germplasm for mapping: a wild barley advanced backcross - nested association mapping (AB-NAM) population, a diverse wild barley association mapping panel, a wild x cultivated barley AB population, and a wild x cultivated barley RIL population; and (4) to map three qualitative traits with varying segregation patterns in the population and a quantitative trait using the AB-NAM population. This genetic characterization will inform our efforts to perform trait mapping within the AB-NAM population and improve our understanding of the challenges surrounding trait introgression from wild barley.

Materials and Methods

Plant materials

Twenty-five wild barley parents were chosen from the Wild Barley Diversity Collection (WBDC), a germplasm collection of 318 accessions that represent wild barley's range in the Middle East, Central Asia, North Africa, and Caucasas region (Table 2.1, Figure 2.1a, Steffenson et al. 2007). Using the Power Marker software's Core Set function (Liu and Muse 2005), parental accessions were selected by maximizing the allele content of 25 chosen parents at 1,402 SNP Barley Oligo Probe Assay 1 (BOPA1,

Close et al. 2009), 556 DArT (Alsop et al. 2011), and 46 SSR markers that were polymorphic in the WBDC. The recurrent parent *Rasmusson* is a high-yielding, high-quality, six-rowed malting barley cultivar (Smith et al. 2010).

Population development

The wild barley AB-NAM population was developed by crossing each of the 25 wild barley donor parents to *Rasmusson*. F_1 plants, and subsequently unique BC_1F_1 plants, were backcrossed to *Rasmusson*. A total of 796 individual BC_2F_2 plants were selected for six-rowed spike morphology and selfed by single seed descent to the BC_2F_4 generation. Six-rowed spike morphology, the preferred type in the Minnesota production region at the time, was selected to reduce the confounding effect of spike morphology on trait analysis. *Rasmusson* served as the female parent in all crosses. Individuals being crossed were vernalized as seedlings in a cold chamber for 4-6 weeks to ensure consistent flowering. Population development occurred in controlled greenhouse and growth chamber environments to minimize inadvertent selection.

Parental genotyping and exome capture sequencing

The 25 wild barley parents and *Rasmusson* were genotyped with the barley 9K-SNP Illumina iSelect platform (Comadran et al. 2012) and exome capture sequenced (Mascher et al. 2013b). Genomic libraries were constructed using the barley Roche NimbleGen SeqCap EZ Developer probe pool and Illumina HiSeq sequenced at the University of Kansas Medical Center Genome Sequencing Facility. The exome capture assay provided the opportunity to capture 61.6 Mb of barley gene space, 34.6 Mb of

which corresponds to high confidence gene calls (Mascher et al. 2013b). Sequence variants were called using the HaplotypeCaller function in the GATK 3.3 software package (https://www.broadinstitute.org/gatk/). Exome capture sequences for parent WBDC103 were discarded due to low quality sequence alignment, and subsequent analyses were performed on the 761 individuals derived from 24 wild barley parents. Sequence variant calls were filtered by GATK genotyping quality (GQ) score > 10 and read depth (DP) > 10 reads. Sequence variants with heterozygous *Rasmusson* genotype calls, greater than 2 alleles, missing calls, or calls located on contigs without genetic map information were excluded from the marker set. Sequence variants that mapped to the same cM location and that contained identical parental genotype calls were binned because they were in perfect LD, and only one variant was used for further analysis.

Population genotyping and marker imputation

Tissue from six to eight seedlings per $BC_2F_{4:5}$ line was bulk harvested, and DNA was extracted at the USDA-ARS Eastern Regional Small Grains Genotyping Laboratory. Each of the 796 lines was genotyped using a custom 384-SNP Illumina VeraCode assay that contained markers selected from 2,994 mapped SNPs on the Barley Oligo Probe Assays 1 & 2 (BOPA1 & 2; Close et al. 2009; Muñoz-Amatriaín et al. 2011). SNP selection was based on even distribution throughout the genome according to map locations from the Muñoz-Amatriaín et al. (2011) consensus map, and based on their ability to distinguish *Rasmusson* from the 25 wild parents. Five of the 384 SNPs did not meet quality standards and were excluded from further analysis.

For each family, the subset of segregating genotyped markers (Supplemental Table 2.1) was used as a scaffold to impute the higher density iSelect and exome capture sequence variants of each respective wild parent. The probabilities of SNP calls were calculated based on the genetic map locations of typed scaffold marker calls (Guo and Beavis 2001). For each line, when two typed flanking markers were derived from the same parent, the imputed marker took the value of the flanking markers. When the markers indicated a recombination event, the imputed marker was given an intermediate value as a function of the genetic distance between the imputed marker and each of the two flanking markers. The Muñoz-Amatriaín et al. (2014) consensus map and the Mascher et al. (2013a) PopSeq *Morex* x *Barke* sequence contig map were used to impute scores for iSelect and exome capture markers, respectively.

Additionally, a set of artificial, interpolated marker values were constructed for each 1cM in the consensus map. Marker calls were generated by interpolating the presence or absence of wild barley introgression using the appropriate flanking, segregating genotyped markers.

Segregation distortion and recombination

Regions of introgression were determined based on the set of segregating markers for each family, with recombination events predicted halfway between two flanking markers. For each line, the proportion of wild barley introgression was calculated as the number of 1 cM intervals with wild introgression divided by the total map size. To test for segregation distortion from the expected 12.5% BC_2 wild allele frequency, chi-square tests of significance were calculated at each marker, across the entire population (df = 1)

and within families (df = 24). Number of recombination events were identified for 10 cM intervals and tested for significant deviations from the mean population recombination rate.

Population structure and linkage disequilibrium

Neighbor joining trees were calculated using the nj function in the ape R package using 2,878 BOPA1 and BOPA2 SNP markers for the WBDC and 6,976 iSelect markers for the AB-NAM parents. Tree diagrams were constructed in FigTree 1.2.0 (http://tree.bio.ed.ac.uk/software/figtree/). Principal component analysis was performed using the R eigen function for the 4,022 imputed iSelect markers across the entire AB-NAM, with and without the parental genotypes included.

To compare genome-wide measures of LD in wild barley-derived mapping populations, genome-wide pairwise marker correlations were calculated using 967 typed or imputed BOPA1 SNP markers that were segregating in each of four populations: (1) the 796 line wild barley AB-NAM described here, (2) the 318 accession WBDC association mapping population from which the AB-NAM parents were chosen (Steffenson et al. 2007), and (3) 98 BC_2F_8, AB lines (HOUH-AB) and (4) 92 RIL lines (HOUH-RIL) derived from the wild barley x 2-row malting barley cross *OUH-602* x *Harrington* (Yun et al 2005; 2006). The HOUH-AB and HOUH-RIL populations were genotyped with the same 384-SNP platform used to genotype the AB-NAM.

To eliminate the effect of population size confounding LD comparisons, LD was also calculated after sampling 90 individuals from each population, 100 times. LD decay was measured as the genetic distance (cM) at which the squared correlation coefficient r^2

decayed to 0.2 on a logarithmic regression. Background levels of LD were calculated as the average r^2 between all pairwise combinations of markers on different chromosomes. The average pairwise r^2 was calculated for 40 cM sliding windows shifting by 2 cM, which contained an average of 102 markers per window.

Trait phenotyping

The traits glossy spike and glossy sheath were visually scored as present/absent in augmented field experiments in Crookston, MN in the summers of 2012 and 2013 (CR12, CR13). Blue hull color was scored as present/absent in the remnant seed from the CR12 and St. Paul, MN 2012 (SP12) field trials. The mean trait value was calculated across environments for each line. Grain protein levels were obtained using a Perten (Hägersten, Sweden) diode array near infrared spectroscopy (NIRS) instrument on 22 mL samples of cleaned grain from augmented field experiments in CR12, SP12, and Bozeman, MT and Fargo, ND in 2013 (MT12, ND13). Grain protein was analyzed in each environment and best linear unbiased predictors (BLUPs) were calculated for each line based on phenotypic data from all four environments using ASReml-R (VSN International).

Trait mapping

Genome-wide association analysis was performed for each trait using the rrBLUP R package GWAS function (Endelman 2011), which implements an EMMAX variance component model (Kang et al. 2010). Blue hull was mapped using all individuals with genotypic information, and results were filtered with a minor allele frequency (MAF)

threshold = 0.001 due to the low frequency of the trait in the population. For grain protein, a bootstrap approach was taken where 25 individuals from each family (600 total individuals) were sampled 100 times. Results for glossy sheath, glossy spike, and grain protein were filtered on a MAF = 0.013 threshold, corresponding to 10 individuals containing a wild barley allele.

Marker-trait associations were deemed significant when they fell above a 0.05 false discovery rate (FDR) threshold. To obtain estimates of allele substitution effect, marker effects were calculated by passing a single marker to the mixed.solve function in the rrBLUP R package (Endelman 2011; Mohammadi et al. 2015). For protein, the frequency of detection of marker-trait associations was calculated across sampling subsets, and those associations observed in less than 5% of samples are not reported. The reported marker effects and significance (-log(p-value)) values are averages across significant bootstrapped samples. Marker-trait associations were deemed independent loci if there was a 5cM or greater gap between significant markers. Markers with maximum significance in each direction are reported.

Mapping was performed using 4 different marker sets: (1) 379 SNPs genotyped across the entire population; (2) 4,022 SNPs (3,520 unique bins) imputed from the parental iSelect SNP marker calls; (3) 263,531 variants (126,303 unique bins) imputed from exome capture sequencing of the parents; and (4) 1,148 interpolated biparental calls at each 1 cM of the barley genome consensus map.

Results

Development of a wild barley AB-NAM population

To develop an advanced backcross nested association mapping (AB-NAM) population in barley, we first selected a highly diverse set of 25 wild barley accessions from the 318 accessions in the WBDC. The selected 25 accessions contained 92% of the genetic variation assayed across all marker sets. In particular, the 25 wild barley parents captured 96% of the BOPA1 & 2 SNP alleles, 99% of DArT alleles, and 57% of SSR alleles present in the 318 accessions of the WBDC. The lower percent of SSR alleles captured was due to an average of 19 alleles per SSR locus. The selected wild barley parents were representative of the geographic and genetic distribution of the WBDC (Figure 2.1a and Figure 2.1b), but ascertainment bias introduced by using the BOPA1 marker platform led us to select 8 individuals that subsequent analysis revealed contain allele frequencies characteristic of feral barley accessions or recent introgression from landrace or cultivated barley (Fang et al. 2014).

Using these 25 WBDC accessions, we developed the AB-NAM population by backcrossing each of the selected parents twice to a common recurrent cultivar *Rasmusson* (Figure 2.1c). *Rasmusson* was chosen to ensure that the wild barley alleles could be evaluated in an adapted genetic background in the field. Between 27 and 39 BC_1 individuals were derived from each wild barley accession and backcrossed to *Rasmusson* to produce the BC_2 individuals. Each BC_2 was advanced via single seed descent for four generations to develop 25 BC_2F_4 derived families. At the F_2 stage in the single seed descent process, individuals with six-rowed spike morphology, the preferred morphology in the Midwestern US at the time of development, were selected. Therefore,

the 25 BC_2F_4 family sizes ranged from 27-39 individuals per wild parent, for a total of 796 unique BC_2F_4 AB-NAM spring six-row lines.

Genetic characterization

To determine the genomic regions of wild parent introgression in the *Rasmusson* background, the AB-NAM population was genotyped with 379 SNP markers distributed throughout the genome. The number of genotyped, segregating markers ranged from 233-326 per family and were distributed approximately every 3.57 cM throughout the genome (Supplemental Table 2.1). Areas with lower density of segregating markers were found in each family. The maximum gap between segregating markers ranged from 13.43 - 23.42 cM (Supplemental Table 2.1), not including the line WBDC020 which shows a significant region on chromosome 3HS that appears to be monomorphic with *Rasmusson*. The parents were genotyped using the barley 9K iSelect SNP platform and exome capture sequenced. We imputed 4,022 SNP (3,520 unique bins) and 263,531 sequence variants (126,303 unique bins) onto the population. Principal component analysis revealed minimal population structure with the first two components accounting for less than 6% of the genetic variation, indicating that population structure was successfully controlled by the crossing scheme (Figure 2.2).

On average, 96% of each wild parent genome was introgressed into the *Rasmusson* background (Table 2.1). Individual lines contain 0.79% - 37.4% wild barley introgression with a mean of 13.5% wild barley introgression per line (Table 2.2). The average introgression size is 27.98 cM, and the average number of introgressions per line is 5.9 on an average of 4.3 chromosomes (Table 2.2). Selection for six-row spike

morphology in the BC_2F_2 led to a significant decrease in wild barley allele frequency surrounding the six-rowed spike morphology *vrs1* locus on chromosome 2H. Additional genomic regions of segregation distortion were found on chromosomes 1H and 6H, where wild barley introgression was more frequent than expected, both within and across families (Figure 2.3). To determine whether it is appropriate to apply the previously developed genetic maps to this population, we compared the recombination frequency across the genome in the AB-NAM with the Muñoz-Amatriaín et al (2014) consensus map. When the AB-NAM recombination events were binned in 10 cM intervals, we observed no areas of the genome where recombination frequency was significantly different from the genome-wide average of 67.1 recombination events per 10 cM (Figure 2.4). Recombination frequency deviated slightly across chromosomes with the lowest average recombination rate of 61.1 recombinations/10 cM on chromosome 4H and the highest rate of 76.8 on chromosome 1H (Supplemental Table 2.2).

Patterns of LD in wild barley derived populations

To assess the mapping utility of the AB-NAM compared to other populations, we compared the linkage disequilibrium (LD) in the AB-NAM, the WBDC, a RIL population and an AB population. The parents of the RIL and AB populations were the wild barley accession *OUH602* and the two-rowed spring cultivar *Harrington* (Yun et al 2005; 2006). LD was calculated using 967 SNP markers that are segregating in each of four wild barley populations. To reduce the confounding factor of population size, LD was calculated from a random sample of 90 individuals in each population. In the WBDC population, LD decayed to an $r^2 = 0.2$ within 1 cM (Supplemental Figure 2.1b).

In contrast, LD decay in the biparental HOUH-AB and HOUH-RIL populations extended to 28.6 cM and 32.3 cM, respectively (Supplemental Figure 2.1c and 2.1d). The AB-NAM had an intermediate level of LD decay of 9.2 cM when 90 individuals were sampled (Supplemental Figure 2.1a), but the LD decay was more rapid when sampled across the entire 796 individual AB-NAM population (Supplemental Figure 2.2a). In the biparental populations, LD appears to slightly increase in intervals containing the pericentromeric region. This trend is less distinct in the AB-NAM and WBDC populations (Figure 2.5). Interchromosomal measures of LD were highest in the WBDC population (0.0216), but similar in the AB-NAM (0.0121), AB (0.0137) and RIL (0.0117) populations (Supplemental Table 2.3, Supplemental Figure 2.3, Supplemental Figure 2.4), suggesting that the crossing design of the AB-NAM minimizes LD caused by factors other than linkage, such as population structure.

Glossy traits, hull color and grain protein in the AB-NAM

Glossy spike, glossy sheath and blue hull color are all highly heritable qualitative traits which were scored visually in the AB-NAM population, and grain protein content is a quantitative trait scored using near infrared spectroscopy (NIRS). These traits were selected because they are easily scored and because they exhibit different segregation patterns in the population. The glossy spike and glossy sheath phenotypes are characterized by reduced or absent wax that leads to a shiny, bright green appearance of the spike or spike and sheath, respectively. The glossy spike phenotype is common in wild barley populations, and is segregating in all families except WBDC173, with a total of 109 glossy spike AB-NAM lines. Glossy sheath is segregating in 3 families derived

from parents WBDC032, WBDC035, and WBDC348 and is expressed in 12 individual AB-NAM lines. Because glossy sheath exhibits a glossy spike phenotype, individuals expressing the glossy sheath phenotype were called as missing data for the glossy spike phenotype. Dark blue coloration of the mature hull is segregating in a single family, WBDC042 and is expressed in 4 individual AB-NAM lines.

Grain protein content exhibited transgressive segregation in the AB-NAM population (Figure 2.6). The BLUP values for grain protein (%) across the four measured environments (CR12, SP12, MT13, and ND13) ranged from 10.19 to 13.85 with a mean of 12.24. The population mean was higher than the *Rasmusson* BLUP value of 11.83% protein. The estimated broad sense heritability calculated on a line mean basis for protein in the AB-NAM was 0.42.

Mapping in the wild barley AB-NAM

To determine the effectiveness of the wild barley AB-NAM population for mapping, we mapped four traits: glossy spike, glossy sheath, blue hull color, and grain protein content with four marker sets: low density genotyped SNP markers, medium density imputed SNP markers, high density imputed sequence variants, and biparental interpolated markers. For all three qualitative traits, marker-trait associations were detected in all marker sets, with the most significant results found using the high-density sequence imputed markers. Associations for these traits with high density sequence variants across 761 individuals in 24 AB-NAM families are shown in Figure 2.7. For each of these traits, a single highly significant QTL was identified, along with one or more loci that appear to be false positives. In the cases of glossy sheath and blue hull,

sequence variants were detected that are perfectly segregating with the trait. Assuming these highly significant loci are the causative loci, the traits with fewer individuals expressing the phenotype have more false positive associations (Figure 2.7).

For the quantitative trait grain protein, we detected a single variant on chromosome 6H in the three lower density marker sets (Figure 8a, 8b, 8d). When mapped with high density sequence imputed variants, several other peaks appear to be significant (Figure 8c). To provide an additional measure of confidence to the detected associations, we used a bootstrap mapping approach to determine the frequency over 100 replicates that marker-trait associations were detected when subsets of AB-NAM individuals were used in the analysis. This value along with the minor allele frequency of the associated variant, the average non-*Rasmusson* relative effect size, and the average significance level is reported in Table 2.3 and Supplemental Dataset 2.1.

Glossy spike

The glossy spike trait was mapped to the distal region of chromosome 1HS. Significant variants were found between position 0.11cM and 9.92cM on the Mascher et al. (2013a) POPSEQ map, with the maximum significance of -log(p) = 101.69 occurring at 0.11cM (Table 2.3). An additional significant locus was detected on chromosome 3H with the peak significance of -log(p) = 9.21 occurring at 96.6 cM (Supplemental Dataset 2.1). This association falls in the region of the most significant glossy sheath locus, and is likely due to the interaction between glossy spike and glossy sheath. The highly significant region on chromosome 1HS is coincident with the location of the *Eceriferum*-

yy (*Cer-yy*) barley mutants, which exhibit a reduction of wax production on the barley spike (Lundqvist and Wettstein-Knowles 1982; Druka et al. 2011).

Glossy sheath

A total of three significant loci were identified for the glossy sheath phenotype. Two small regions of chromosome 1H at 46.5cM (-log(p) = 7.32) and 97.5cM (-log(p) = 6.90) were identified, in addition to 5 variants on chromosome 3HL (96.6 cM) which cosegregate with the trait (Supplemental Dataset 2.1). These variants are located in a single gene on *Morex* contig 41718 (Table 2.3). This map location corresponds to the *glossy sheath 2* (*gsh2*) mutant that exhibits similar lack of wax on the leaf sheath (Wettstein-Knowles 1990; Druka et al. 2011).

Blue hull

The blue hull phenotype was only identified in 4 individuals. Because of this, there appear to be many spurious associations, including peaks on chromosomes 2H, 3H, 4H and 6H with significance ranging from -log(p) = 7.03 to -log(p) = 30.41 (Supplemental Dataset 2.1). On chromosome 1HL, nine bins that include 43 variants on 15 contigs cosegregated with the blue hull phenotype. These contigs are located in a 2.9 cM region between 116.5 cM and 119.4 cM. This region corresponds to the *Black lemma and pericarp* (*Blp1*) mutant (Druka et al. 2011).

Grain protein

Ten QTL were detected for grain protein content (Table 2.3). These QTL were located on chromosomes 1H, 3H, 4H and 6H. The wild barley alleles for the QTL on chromosomes 1H and 3H conferred lower grain protein content, a beneficial quality for malting and brewing. Four significant loci were identified on chromosome 4H, each of which had both positive and negative wild barley effects, indicating that there may be multiple haplotypes segregating at these loci. The most frequently detected QTL was located on chromosome 6H at 50.0 cM. The wild barley alleles at this locus conferred higher (deleterious for malting) grain protein content (Supplemental Figure 2.5).

Discussion

We developed a population that incorporates a large amount of diversity from wild barley into a six-rowed, cultivated barley background, which will serve as a resource for barley breeders and geneticists to efficiently screen a large amount of genetic diversity for many traits of interest. To evaluate the effectiveness of this design, we examined the genetic composition of the population for factors which influence mapping: segregation distortion, distribution of recombination, population stratification, linkage disequilibrium, population size, and marker set. Additionally, three qualitative traits: glossy spike, glossy sheath, and blue hull color, and a quantitative trait: grain protein content were mapped in the population.

The wild barley AB-NAM increases the genetic diversity available to barley breeders

Because we aimed to maximize the amount of genetic diversity in the population, the parents of the AB-NAM are a set of 25 wild barley accessions selected solely on genetic data, without considering ecogeographical or phenotypic characteristics. While this set of diverse lines approaches an ideal mini core of the larger wild barley population, the selection of lines was subject to ascertainment bias. The identification of SNPs for the BOPA marker platforms was based on ESTs and sequences primarily from cultivated barley (Close et al. 2009). Because of this, a sampling bias introduced the tendency to overestimate the diversity of germplasm more closely related to the sampled individuals (Nielsen 2000). In this case, those individuals that contain landrace or cultivated barley germplasm were selected because they appeared more diverse than the majority of the wild barley accessions. Moragues et al. (2010) demonstrated this phenomenon by showing that estimates of diversity were higher in cultivated barley germplasm than in landrace germplasm when assayed with BOPA1 markers.

Since the original selection of parental lines, the genetic signatures that differentiate wild and landrace barleys has been examined more extensively (Russell et al. 2011), and it is likely that ascertainment bias led to the inclusion of 8 accessions which appear to contain varying levels of landrace and/or cultivated barley admixture (Fang et al. 2014). These eight accessions are those with larger genetic distances from the majority of the WBDC, and can be seen at the bottom right in the WBDC dendrogram in Figure 2.1b. These eight accessions appear in between the AB-NAM population and the majority of the wild parents in the PCA plot in Figure 2.2. Still, the phenotypic and geographic distribution of the selected parents is only slightly narrower than the WBDC

range (Figure 2.1a). In addition to the BOPA SNP markers, the use of SSR markers, which have been shown in maize to be less influenced by ascertainment bias (Hamblin et al. 2007), and DArT markers, which were developed from wild-barley germplasm (Alsop et al. 2011), may have helped mitigate the bias. Nevertheless, the presence of landraces in the population should not influence the effectiveness of mapping and gene discovery in this population.

During population development, we attempted to minimize unintentional selection, and therefore, maximize the amount of wild barley genome introgressed. At the same time, we imposed selection for the *Rasmusson* six-rowed spike morphology. As expected, wild barley introgression is significantly reduced in the region of 2H surrounding the *vrs1* locus responsible for the six-rowed phenotype (Figure 2.3). This selection will limit our ability to map loci on chromosome 2H, but it will also minimize the confounding effects of phenotyping a population that is segregating for a major morphological trait. This is especially true for assessment of Fusarium head blight resistance and malting quality, which can be markedly affected by row type (Marquez-Cedillo et al. 2000; Choo et al. 2004). No other regions of the genome show reduced introgression frequency, but regions on chromosomes 1H and 6H show an overrepresentation of wild barley introgression across the population. Notably, *vrs3*, a known determinant of spike morphology, is located in the region of chromosome 1H that has an overrepresentation of wild barley introgression (Muñoz-Amatriaín et al. 2014). The *vrs3* phenotype appears 6-rowed at the top of the spike (Lundqvist and Franckowiak 1997), a characteristic that may have influenced the selection for 6-rowed morphology in the population, and consequently, affected the allele frequency in this region.

Schnaithmann et al. (2014) also observed some degree of segregation distortion in the Halle Exotic Barley (HEB) population, but generally it was restricted to specific crosses and none of the regions identified in the HEB population correspond with those identified here.

Wide crosses can be subject to factors such as segregation distortion and deviations in recombination rate across the genome (Xu et al. 1997; Bauer et al. 2013). To determine if it is appropriate to use a previously published consensus map for mapping in the AB-NAM, we compared the recombination rate across the genome in the AB-NAM using the Muñoz-Amatriaín et al (2014) consensus map (Figure 2.4). We found no regions of the genome that deviate significantly from the mean of 61.1 recombinations per 10 cM window, suggesting that fluctuations in recombination throughout the AB-NAM population are consistent with the genetic distances represented by the consensus map. We also observed that the percentage of introgressed regions ranged widely, with individuals containing 0.79-37.45% wild barley introgression, but the average introgression frequency of 13.55% was relatively consistent with the expectation of 12.5%. Collectively, these results corroborate previous studies (Yun et al. 2006; Schnaithmann et al. 2014) that have show only minor fluctuations in allele frequency occur when introgressing wild barley genomes, and breeders and geneticists can expect crosses of wild barley to cultivated barley such as the AB-NAM to behave generally as expected.

The wild barley AB-NAM exhibits low population structure and low LD

Population stratification present in association mapping panels can lead to spurious associations if not appropriately controlled, and effective control of population structure can lower the power to detect marker-trait associations (Larsson et al. 2013). The NAM design attempts to eliminate this problem by nesting segments of diverse blocks in a controlled crossing structure (Yu et al. 2008). Principal component analysis shows only minimal population stratification in the wild barley AB-NAM (Figure 2.2), with no consistent trends among families. Because of this, we did not include measures of population stratification (Q) in our mapping, but we did use a kinship (K) model to control for cryptic relatedness among individuals.

Linkage disequilibrium is an important consideration for mapping. Because wild barley populations exhibit lower LD than cultivated barley populations, using wild barley populations presents an opportunity for high-resolution mapping. But, phenotyping wild barley for adult plant traits can be extremely difficult and population structure can confound mapping results. One of the aims of developing the AB-NAM was to develop a population which mitigates the difficulties of working directly with wild barley, while maintaining the high genetic diversity and high mapping resolution possible with wild barley populations. To understand how the AB-NAM compares to other wild barley-derived populations, we compared genome-wide LD in the AB-NAM, the WBDC association mapping panel, and the biparental wild-barley derived biparental mapping populations HOUH-AB and HOUH-RIL. As expected, the rate of LD decay is much higher in the WBDC population than in the biparental populations (Supplemental Figure 2.1). Because differences in the parental genotypes can be imputed onto the segregating

families of the AB-NAM, LD in the AB-NAM approaches the low level of the WBDC (Figure 2.5).

The level of LD present in the mapping population is an indicator of potential mapping resolution, but it also an indication of what level of marker coverage is necessary for high power trait mapping. When working with low LD populations, a higher density of markers is required to tag QTL with markers. Simulations have shown that power increases as markers are added to multifamily mapping experiments, and being able to impute high density markers onto a lower density genotyped panel appears to be an effective means to increasing marker density (Liu et al. 2013). Ultimately, the AB-NAM optimizes the tradeoff between the practical difficulties of phenotyping unadapted material and high resolution mapping in a diverse germplasm set.

The wild barley AB-NAM provides a new resource for trait mapping in barley

The NAM design provides a straightforward means to projecting high-density marker data onto a segregating population. We showed that the addition of medium and high density imputed sequence variants can improve the power of detecting marker-trait associations (Figure 2.8), in some cases allowing for the identification of variants that cosegregate with traits of interest (Table 2.3). These highly significant associations were identified even when only a small number of individuals exhibited the trait. In the case of glossy sheath, the variants with the highest significance were localized to a single contig on the genome sequence. This indicates that there is a benefit to the NAM design for both power and resolution as compared to a biparental population. Furthermore, the

size of the population allows for a bootstrapping approach that can provide additional information to support or filter marker-trait associations.

A major advantage of the NAM design is the increase in minor allele frequency of uncommon alleles to detectable levels as compared to association mapping panels. In the NAM design, minor allele frequency is a function of the number of parents containing the allele and the expected minor allele frequency in the individual crosses. In RIL-derived populations, the expected MAF is 0.5. In BC_2 derived populations, the expected MAF is 0.125. This means that when private alleles are present in the AB-NAM population with 25 parents and 796 individuals, the expected MAF of that allele is 0.125/25 = 0.005, approximately 4 individuals in the 796 individual wild barley AB-NAM. In contrast, a similarly sized RIL-derived NAM would have a private allele MAF of 0.5/25 = 0.02 or approximately 16 individuals with the allele. This means that, except for very large effect traits, like blue hull, causative alleles segregating in a single family in the wild barley AB-NAM may not be detected.

To examine a quantitative trait with lower heritability, we used the wild barley AB-NAM population to map grain protein content. Grain protein is an important trait for both malting (requiring low protein) and animal feed (requiring high protein) purposes (Blake et al. 2010). A single QTL was identified in each of the marker sets, corresponding to the grain protein content (GPC) QTL on chromosome 6HS (Distelfeld et al. 2008; Lacerenza et al. 2010). When mapped with higher density, sequence imputed markers, additional QTL were detected throughout the genome (Table 2.3). QTL identified on chromosome 1H exhibited negative marker effects, indicating that wild barley alleles may have beneficial effects for malting quality. Additionally, QTL

identified on chromosome 4H exhibit both negative and positive effects of wild barley alleles, indicating that either multiple haplotypes or multiple genes are influencing protein at these loci. Additional analyses that take into account marker haplotypes may improve the interpretability of these results.

Additional studies are being performed to analyze the effectiveness of mapping complex agronomic and disease traits in this population. Due to smaller family size, fewer effective recombination events, and lower minor allele frequencies, this population is unlikely to be as powerful as a NAM derived from RILs. But, the effort expended to produce advanced backcross design is balanced by the added ease and accuracy of phenotyping. Utilization of exotic germplasm resources continues to be a challenge for breeders, and this resource bridges the gap between the germplasm resources of gene banks and elite breeding programs.

Table 2.1 Wild barley parental accessions, ICARDA genebank or alternate designation, collection location, number of BC_2-derived AB-NAM lines, and percentage of parental genome covered by introgression in BC_2-derived individuals.

Parent accession[a]	ICARDA genebank designator	Country of origin	Total BC_2 individuals	Percentage of parent genome introgressed
WBDC 016	38661	Iran	34	94.9
WBDC 020	38672	Turkey	36	92.0
WBDC 028	38840	Israel	26	96.4
WBDC 032	38869	Israel	33	96.0
WBDC 035	38981	Israel	34	98.0
WBDC 042	39673	Israel	27	89.3
WBDC 061	39910	Syria	29	98.3
WBDC 082	40009	Jordan	32	98.3
WBDC 092	40034	Jordan	34	95.1
WBDC 103	40071	Jordan	35	95.9
WBDC 115	40104	Turkmenistan	39	100.0
WBDC 142	40188	Lebanon	26	92.4
WBDC 150	40200	Iran	30	95.6
WBDC 172	112673	Iran	32	90.8
WBDC 173	112674	Iran	30	97.4
WBDC 182	115781	Jordan	31	98.9
WBDC 227	132552	Azerbaijan	36	99.1
WBDC 234	39884	Cyprus	30	96.6
WBDC 255	115792	Jordan	30	94.1
WBDC 292	38926	Israel	30	97.9
WBDC 302	38635	Syria	30	97.7
WBDC 336	126406	Turkmenistan	33	94.0
WBDC 340	116116	Turkey	34	99.0
WBDC 348	Damon 11-11(B)[b]	Israel	35	97.6
WBDC 350	41-1[c]	Israel	30	97.0

[a] Steffenson et al. (2007)
[b] Obtained from University of Haifa collection, Fetch et al. (2003)
[c] Unknown ICARDA germplasm number, Baum et al. (2003)

Table 2.2 Summary of introgression frequency and size in AB-NAM lines.

	Percent of genome containing introgressions	Number of introgressions	Size of introgressions (cM)	Number of chromosomes without introgressions
mean	13.55	5.9	27.98	2.7
sd	6.47	3.3	11.60	1.3
max	37.45	43.0	79.47	6.0
min	0.79	1.0	2.65	0.0

Table 2.3 Summary of marker-trait associations of maximum significance. All significant marker-trait associations can be found in Supplemental Dataset 2.1. QTL on the same chromosome were considered unique loci if there was a greater than 5 cM gap between significant markers.

Trait	Variant[a]	Chrom[b]	cM[b]	MAF[c]	-log(p)[d]	Effect[e]	Freq[f]
Glossy spike	morex_contig_40051:2172_A/G*	1H	0.1	0.12	101.7	0.46	
Glossy sheath	morex_contig_41718:9777_CG/C*	3H	96.6	0.02	Inf[g]	0.50	
Blue hull	morex_contig_1573652:70_A/G	1H	116.5	0.01	Inf	0.51	
	morex_contig_38907:1672_G/A*	1H	116.6	0.01	Inf	0.50	
	morex_contig_53289:2443_G/A*	1H	116.8	0.01	Inf	0.50	
	morex_contig_42987:7765_T/C*	1H	117.7	0.01	Inf	0.50	
	morex_contig_5603:2020_T/C*	1H	117.8	0.01	Inf	0.50	
	morex_contig_243581:952_C/G*	1H	118.1	0.01	Inf	0.50	
	morex_contig_1576759:813_C/T*	1H	118.3	0.01	Inf	0.50	
	morex_contig_5976:324_G/A*	1H	119.0	0.01	Inf	0.50	
	morex_contig_39431:11917_G/A*	1H	119.4	0.01	Inf	0.50	

Table 2.3 continued

Trait	Variant[a]	Chrom[b]	cM[b]	MAF[c]	-log(p)[d]	Effect[e]	Freq[f]
Protein	morex_contig_43675:8617_T/C	1H	50.6	0.17	4.9	-0.14	0.13
	morex_contig_47454:1292_C/T	1H	119.7	0.01	6.8	-0.37	0.95
	morex_contig_44650:4992_C/T	3H	25.3	0.02	4.8	-0.18	0.17
	morex_contig_41236:9829_G/A	3H	37.4	0.02	4.9	-0.10	0.06
	morex_contig_99201:2611_C/T	4H	27.5	0.10	5.6	0.17	0.90
	morex_contig_46131:3154_TA/T	4H	27.8	0.01	4.8	-0.05	0.08
	morex_contig_47914:307_C/T	4H	43.5	0.02	4.9	0.36	0.09
	morex_contig_246786:1453_AACGTACGC/A	4H	44.9	0.03	5.0	-0.20	0.59
	morex_contig_156722:1830_G/A	4H	51.6	0.03	4.8	0.33	0.11
	morex_contig_45564:4922_A/G	4H	52.2	0.07	4.5	-0.07	0.06
	morex_contig_52783:3343_C/T	4H	63.4	0.02	4.5	-0.27	0.06
	morex_contig_1575414:1899_A/G	4H	67.1	0.03	4.5	0.29	0.13
	morex_contig_146580:705_T/A	4H	76.3	0.01	5.0	-0.30	0.56
	morex_contig_42454:4904_G/A	4H	78.7	0.04	4.8	0.23	0.11
	morex_contig_7108:5074_G/T*	6H	50.0	0.17	5.8	0.37	0.96

[a] Variants significantly associated with trait. Association determined by an FDR > 0.05 significance threshold, and a frequency of detection > 0.05 in 100 bootstrap samplings of 25 individuals from each family for the trait protein.
[b] Chromosome and cM positions of *Morex* contig containing the variant as determined by the Mascher et al. (2013a) POPSEQ map
[c] Minor allele frequency of variant in the AB-NAM population
[d] -log(p-value) averaged across significant bootstrap tests
[e] Relative effect of non-Rasmusson allele, averaged across significant bootstrap tests
[f] Frequency of significant association with variant detected in 100 bootstrap samples
[g] Significance labeled Inf refers to markers which cosegregate with the trait values
[*] Additional variants in the same bin with identical segregation pattern can be found in Supplemental Dataset 2.1

Figure 2.1 (A) Geographic distribution of the wild barley diversity collection (WBDC) with selected parents in red; (B) neighbor joining tree diagram of 318 WBDC accessions based on BOPA 1 & 2 SNP genotypes; (C) population development scheme.

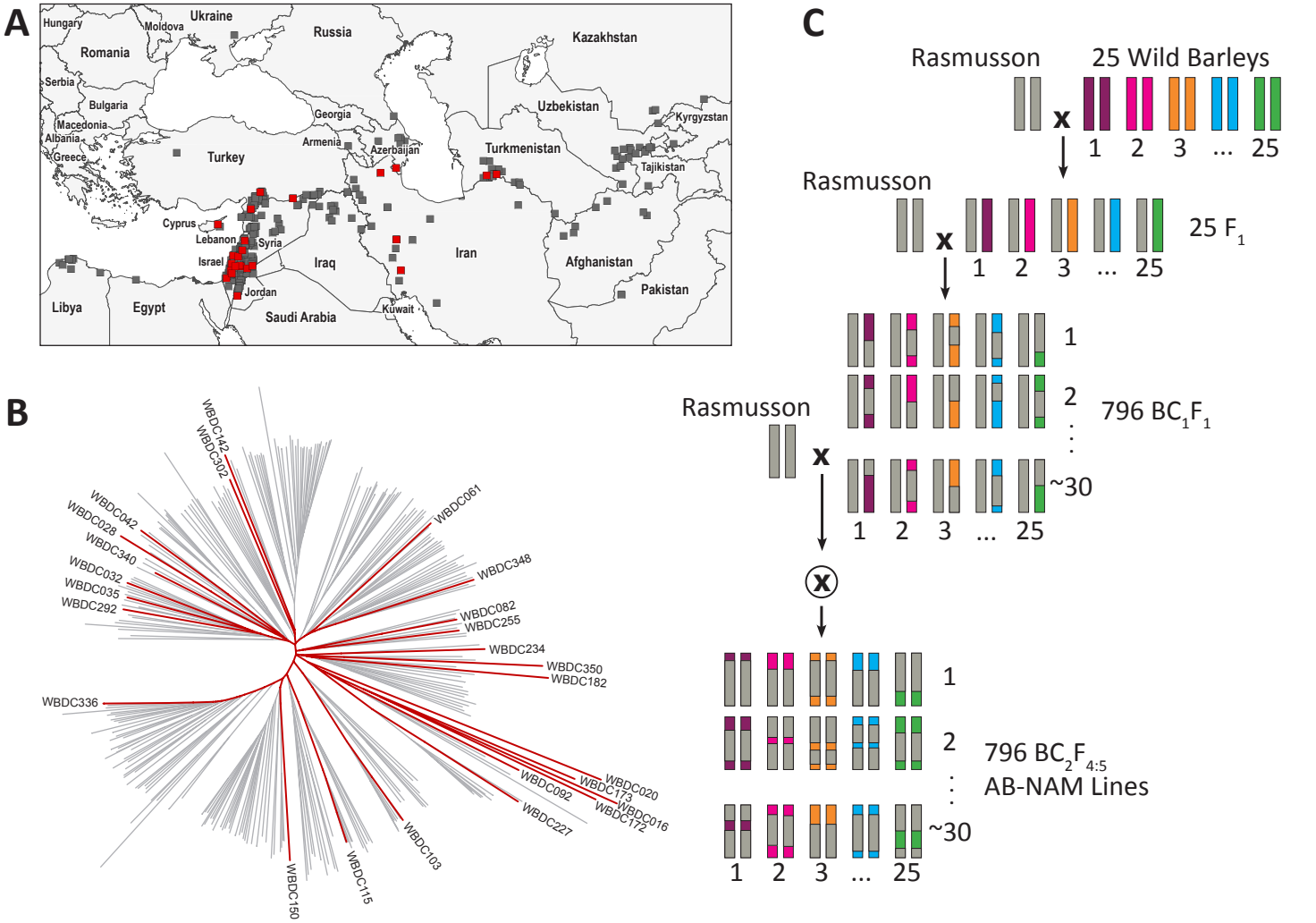


Figure 2.2 Principal component analysis of 796 AB-NAM lines (grey), wild barley parents (red), and *Rasmusson* (blue) based on 3,520 unique imputed SNPs.

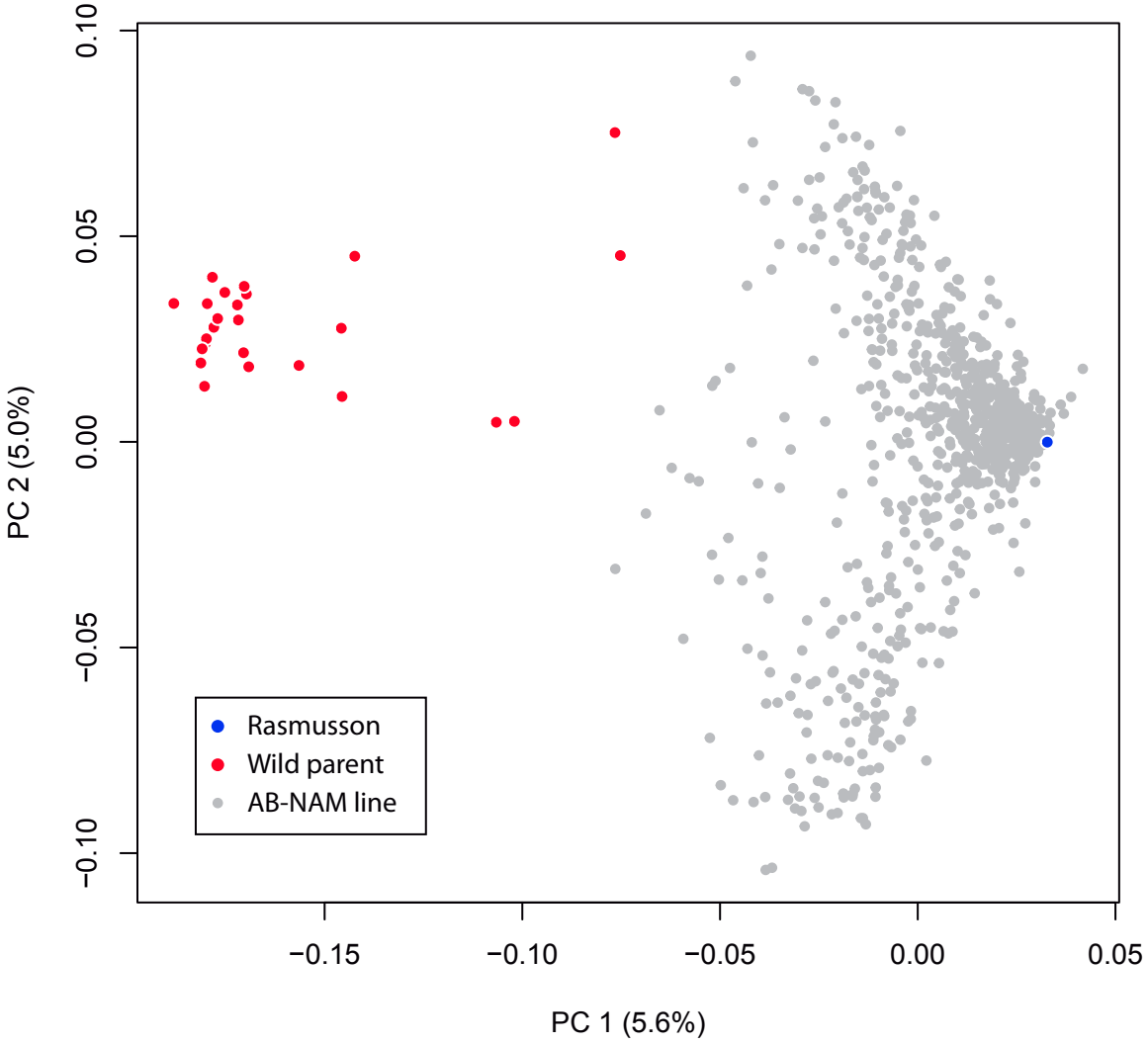


Figure 2.3 Wild barley introgression frequency across the genome. (A) Within families and (B) whole population. Introgression frequency was determined using the appropriate set of segregating markers within each family.

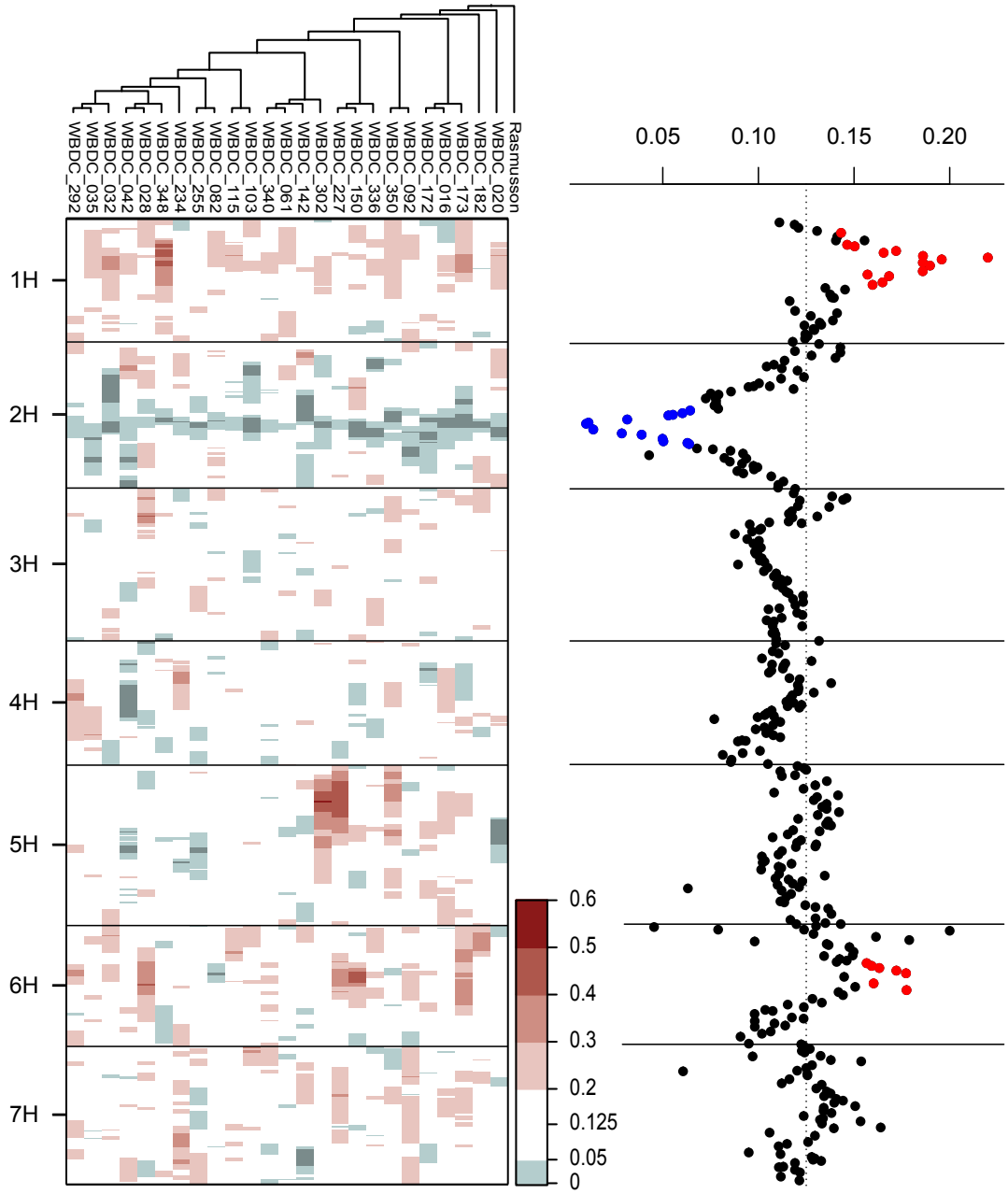


Figure 2.4 Recombination events in the AB-NAM population, identified in 10 cM windows across the genome. Horizontal dashed line indicates genome-wide mean of 61.1 recombination events per 10 cM interval.

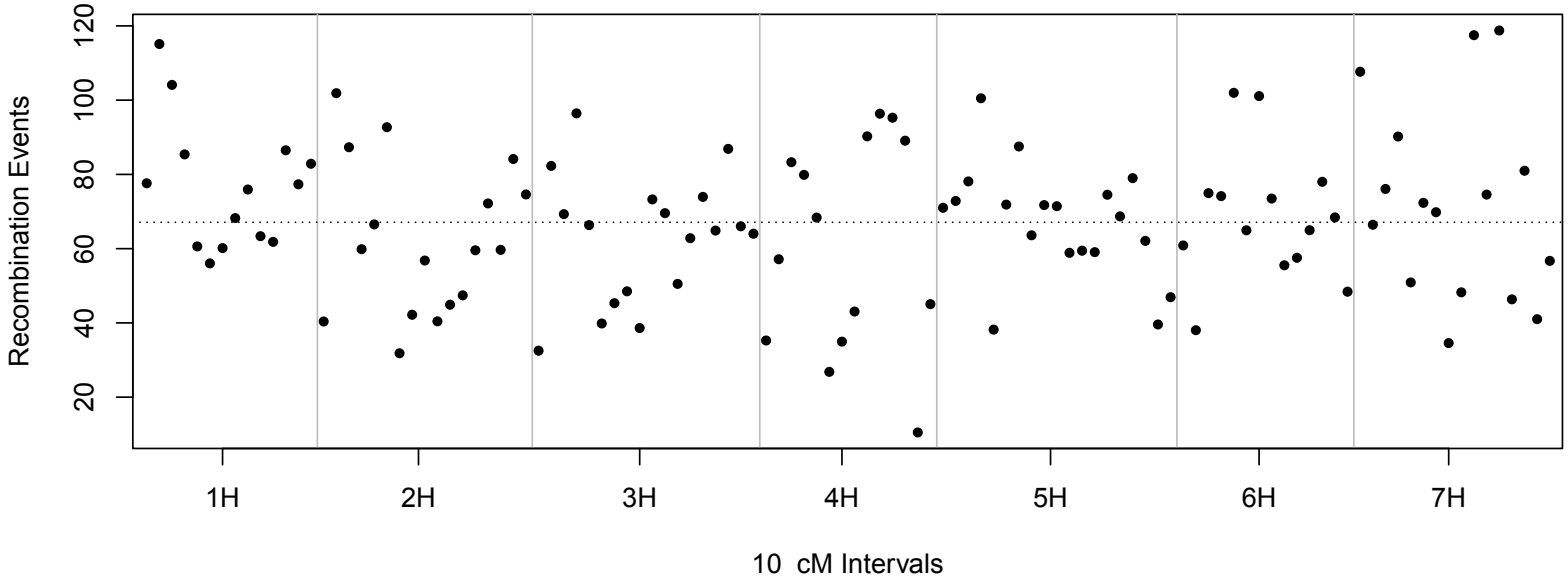


Figure 2.5 Average pairwise LD (r^2) calculated for all segregating BOPA1 markers in 40 cM interval moving windows, with windows calculated every 2 cM across each chromosome 1H-7H in the wild barley Advanced Backcross-Nested Association Mapping (AB-NAM) population, the Wild Barley Diversity Collection (WBDC) association mapping panel, the *OUH-602* x *Harrington* Advanced Backcross (HOUH-AB) mapping population, and the *OUH-602* x *Harrington* Recombinant Inbred Line (HOUH-RIL) mapping population. Pericentromeric regions as identified in Muñoz-Amatriaín et al. (2011) are denoted by vertical grey bars.

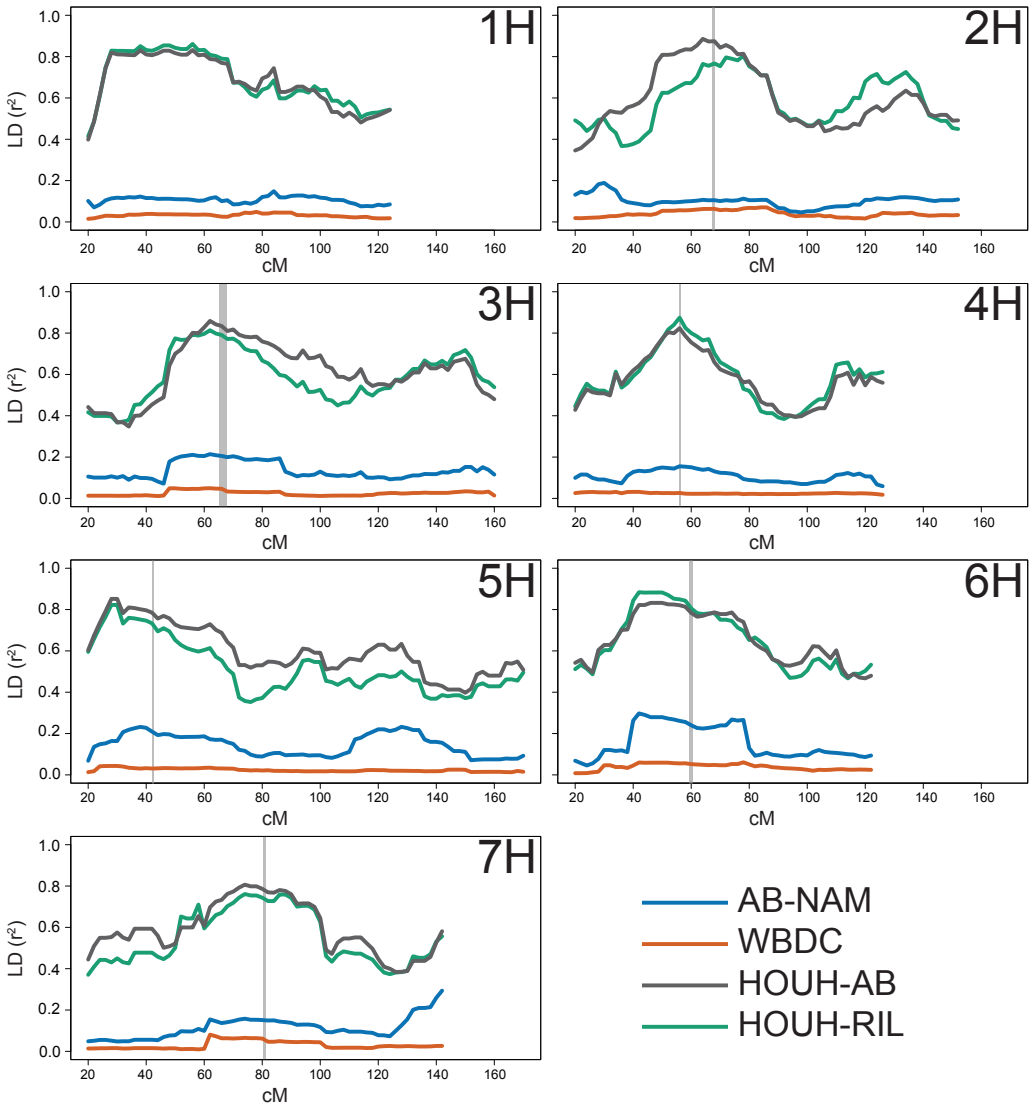


Figure 2.6 Boxplots of percent grain protein values for each AB-NAM family. The population-wide mean is depicted as a solid horizontal line and the *Rasmusson* BLUP value is a horizontal dotted line.

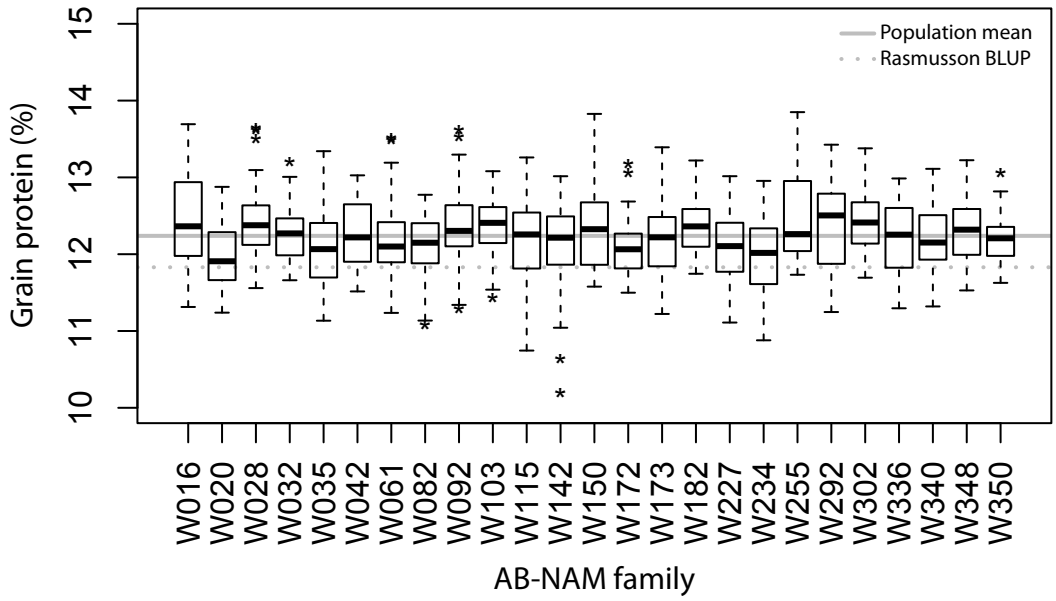


Figure 2.7 Manhattan plots of marker trait associations for traits (A) glossy spike, (B) glossy sheath, and (C) blue hull color in the AB-NAM population using 263,531 exome capture sequence variant markers. Horizontal line indicates a false discovery rate (FDR) = 0.05 significance threshold. Traits glossy sheath and blue hull color had additional variants which perfectly segregated with the trait. These associations were not plotted due to an infinite -log(p-value) result.

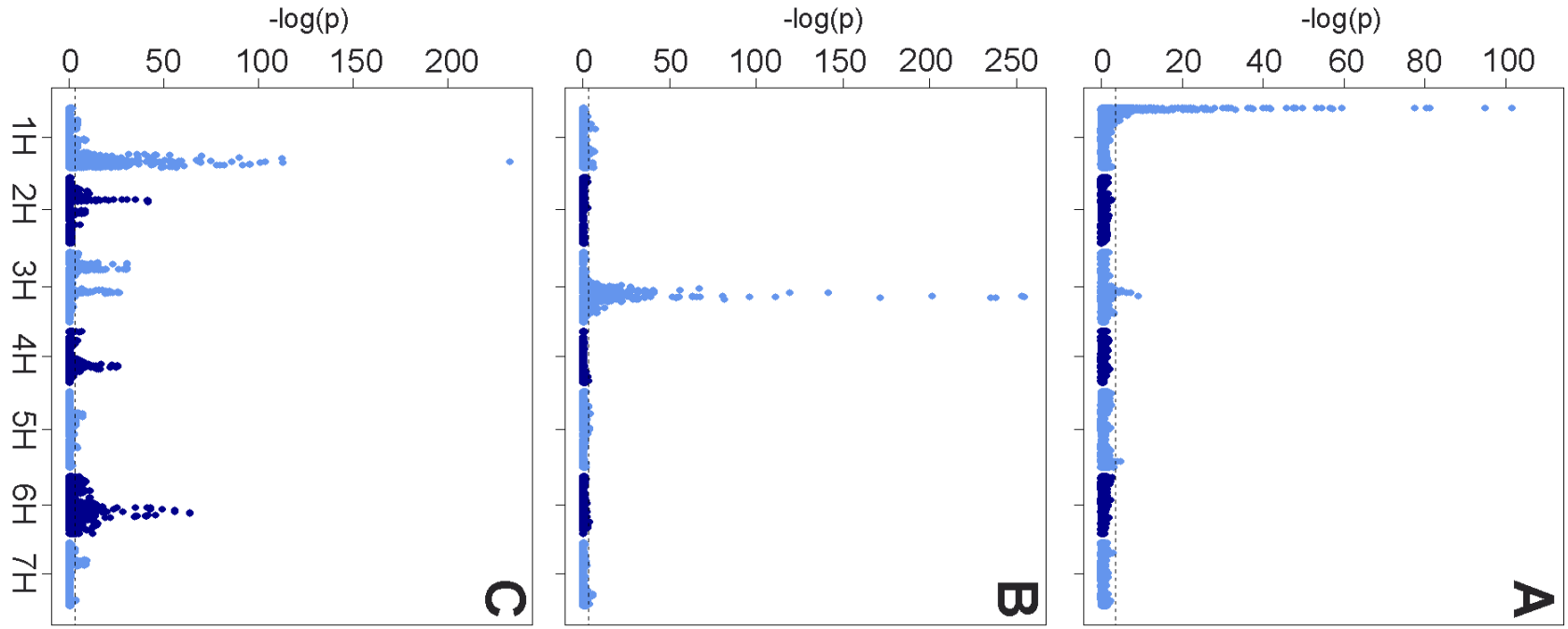
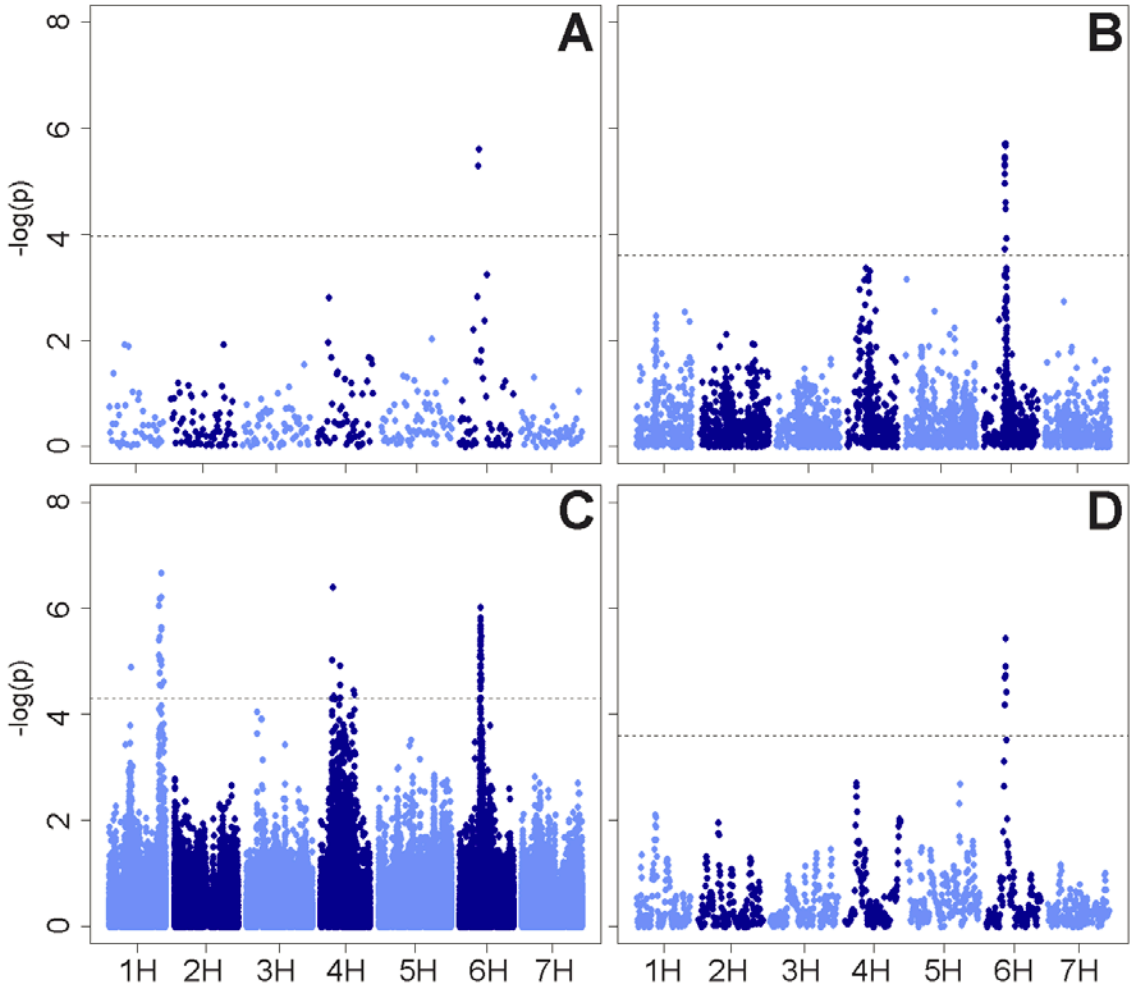


Figure 2.8 Manhattan plots of marker trait association for grain protein content using various marker sets: (A) 379 genotyped SNP markers, (B) 4,022 imputed SNP markers, (C) 263,531 imputed sequence variants, and (D) 1,148 interpolated markers. Dashed horizontal line indicates a false discovery rate (FDR) = 0.05 significance threshold.



Supplemental Table 2.1 Segregating marker calls between *Rasmusson* and each wild barley parent, number of private alleles for each wild barley parent, and average and maximum distance between segregating 384 markers for each family.

	Segregating markers by genotyping platform			Private exome capture alleles	Distance (cM) between segregating scaffold markers	
	384	9K iSelect	Exome capture		Mean	Maximum
Population	379	4,022	263,531	109,033		
Unique Bins	379	3,506	126,303	26,312		
WBDC_016	304	1,973	50,658	3,462	3.63	15.09
WBDC_020	233	1,298	40,399	3,540	4.45	36.64
WBDC_028	301	1,789	52,781	4,624	3.63	15.09
WBDC_032	326	1,774	54,870	4,358	3.40	15.09
WBDC_035	315	1,794	55,096	3,926	3.48	13.54
WBDC_042	326	1,814	55,864	6,618	3.41	15.66
WBDC_061	317	1,837	52,769	4,131	3.49	23.42
WBDC_082	321	1,728	54,973	6,089	3.44	13.43
WBDC_092	306	1,914	51,577	2,847	3.63	15.66
WBDC_103	319	1,738	NA	NA	3.50	15.77
WBDC_115	322	1,797	54,808	6,094	3.44	15.09
WBDC_142	324	1,836	53,653	4,637	3.43	22.96
WBDC_150	317	1,863	53,431	3,252	3.45	13.63
WBDC_172	298	2,088	50,509	3,172	3.69	17.22
WBDC_173	290	2,107	51,127	2,813	3.77	17.24
WBDC_182	302	1,695	49,407	4,973	3.70	21.02
WBDC_227	287	1,734	50,254	2,981	3.81	20.62
WBDC_234	315	1,759	54,700	6,636	3.47	13.51
WBDC_255	326	1,781	55,268	6,250	3.38	17.95
WBDC_292	318	1,782	55,120	4,128	3.46	17.42
WBDC_302	323	1,810	53,993	6,171	3.47	15.09
WBDC_336	321	1,951	55,395	4,269	3.43	15.09
WBDC_340	321	1,753	54,039	4,420	3.44	15.09
WBDC_348	325	1,871	54,637	5,181	3.49	15.66
WBDC_350	295	1,788	50,528	4,461	3.73	17.07
Average	315	1,956	52,744	4,543	3.57	17.36
Min	233	1,298	40,399	2,813	3.38	13.43
Max	326	2,107	55,864	6,636	4.45	36.64

Supplemental Table 2.2 Segregating markers and average recombination rate within 10 cM windows on each chromosome.

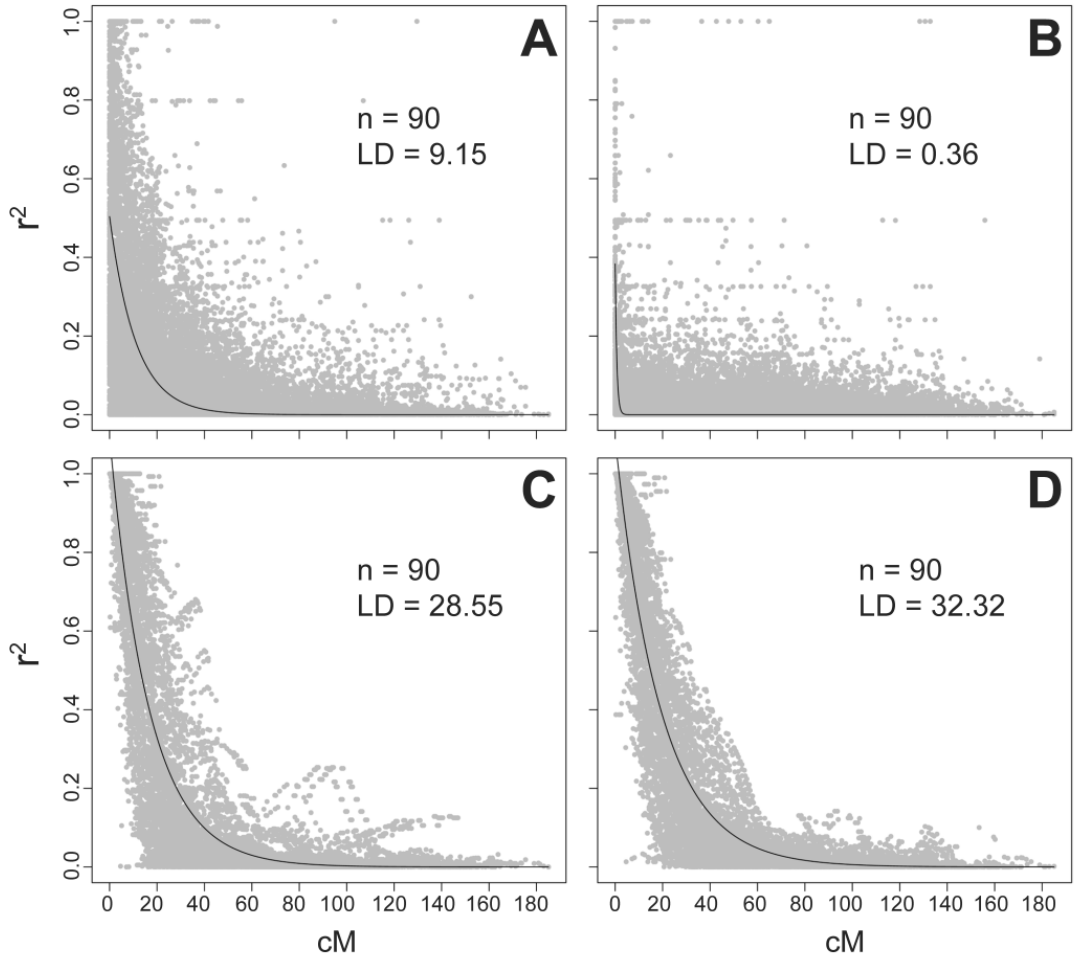
Chromosome	384	9K	Exome capture	Recombination events[a]
1H	41	378	15,852	76.8
2H	61	659	23,557	62.5
3H	59	524	18,973	62.9
4H	52	397	10,980	61.1
5H	63	670	21,150	67.1
6H	49	452	15,274	68.8
7H	54	440	20,517	72.0

[a]Average number of recombination events per 10 cM interval across each chromosome

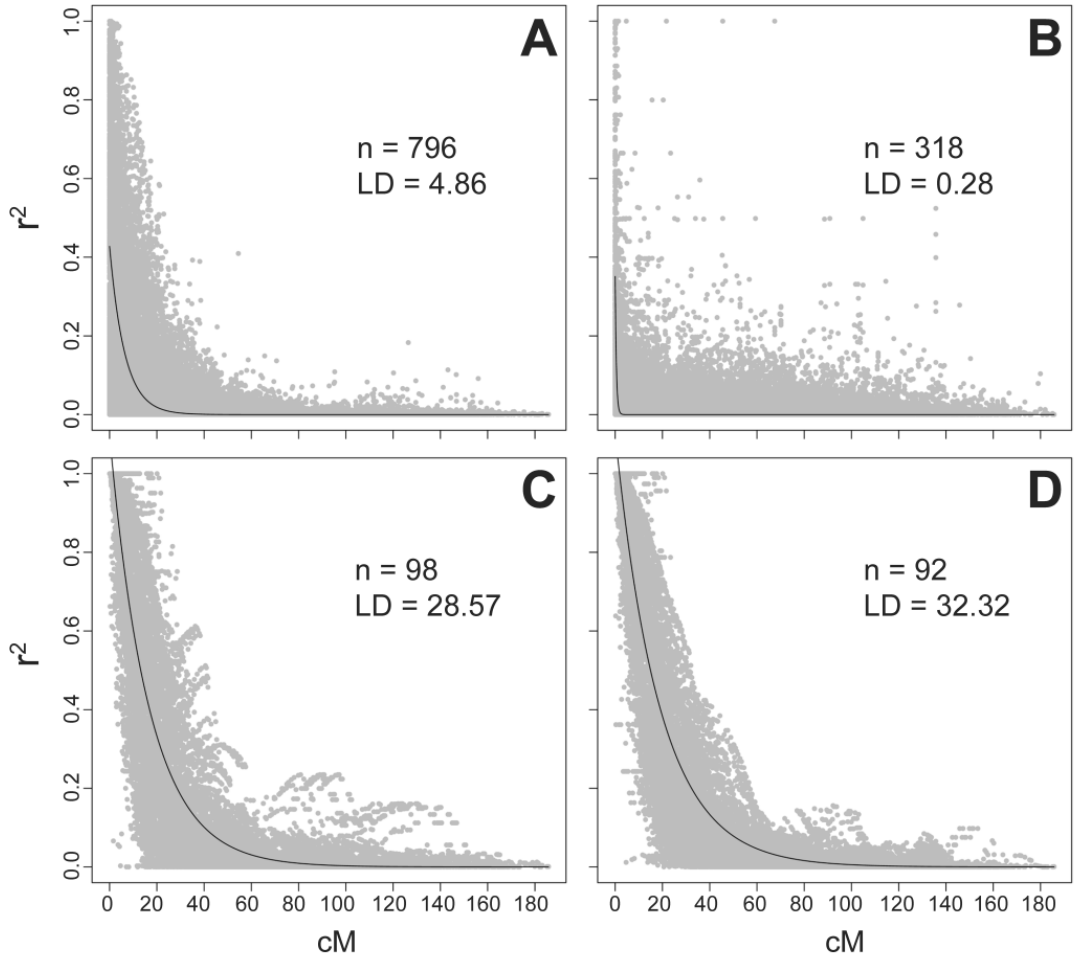
Supplemental Table 2.3 Average pairwise LD (r^2) by chromosome, genome-wide, and interchromosomal.

	AB-NAM	WBDC	HOUH_AB	HOUH_RIL	AB-NAM	WBDC	HOUH_AB	HOUH_RIL
Individuals	796	318	98	92	90	90	90	90
1H	3.32	0.21	34.59	33.47	7.18	0.08	35.54	34.39
2H	3.25	0.38	27.42	33.41	8.57	0.49	27.27	35.18
3H	5.56	0.28	32.26	35.68	10.26	0.26	32.45	36.02
4H	4.60	0.15	30.91	30.42	10.46	0.17	31.22	29.74
5H	7.10	0.23	21.85	29.39	10.42	0.34	22.43	30.65
6H	4.89	0.29	32.70	32.04	6.95	0.40	31.40	32.00
7H	3.51	0.25	24.31	30.44	7.74	0.25	24.17	30.05
Genome-wide	4.86	0.28	28.57	32.32	9.15	0.36	28.55	32.71
Interchromosomal	0.0018	0.0121	0.0124	0.0117	0.0121	0.0216	0.0137	0.0117

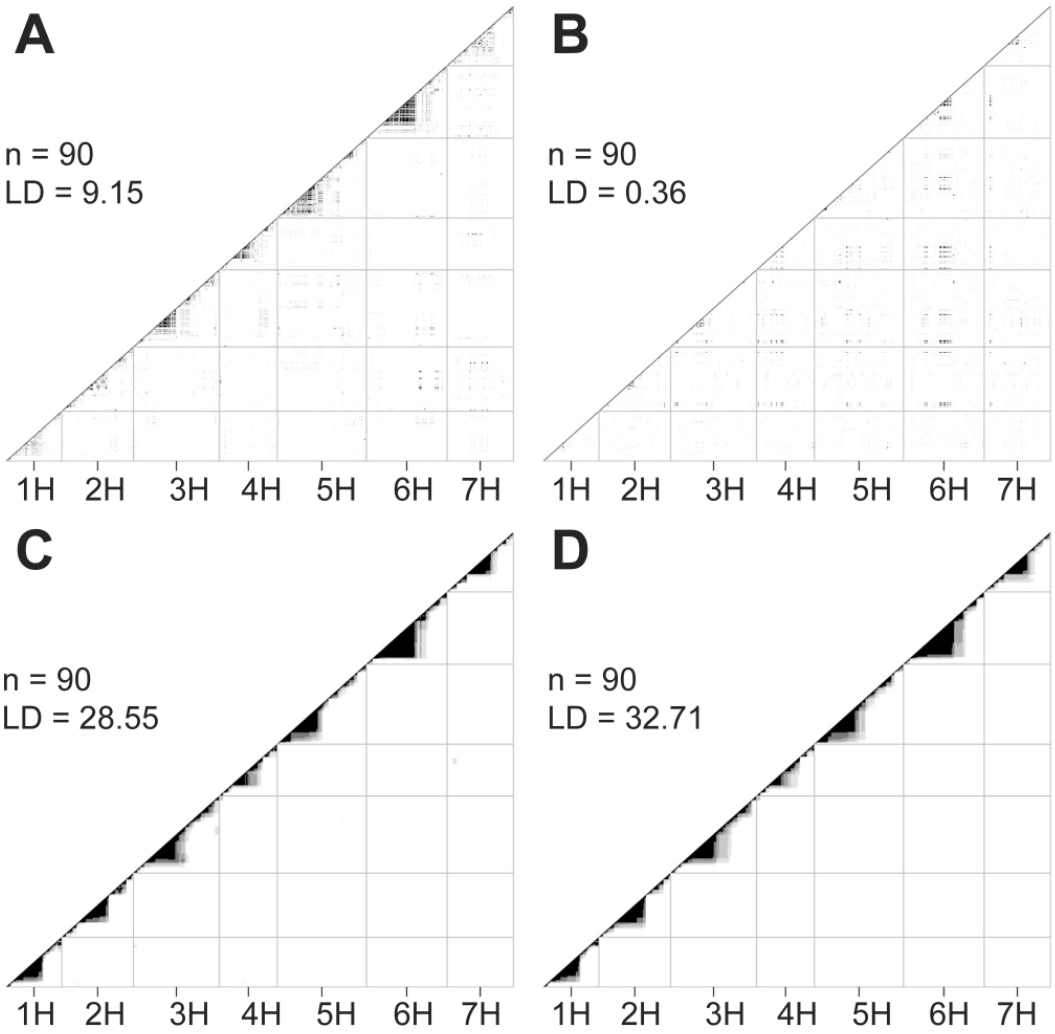
Supplemental Figure 2.1 Genome-wide pair-wise linkage disequilibrium decay. LD calculated as the genetic distance (cM) at which the r^2 decays to 0.2. Populations were randomly sampled to a common population size of 90 individuals. (A) wild barley advanced backcross-nested association mapping population, (B) wild barley diversity collection association mapping panel, (C) *OUH-602* x *Harrington* advanced backcross mapping population, (D) *OUH-602* x *Harrington* recombinant inbred line mapping population.



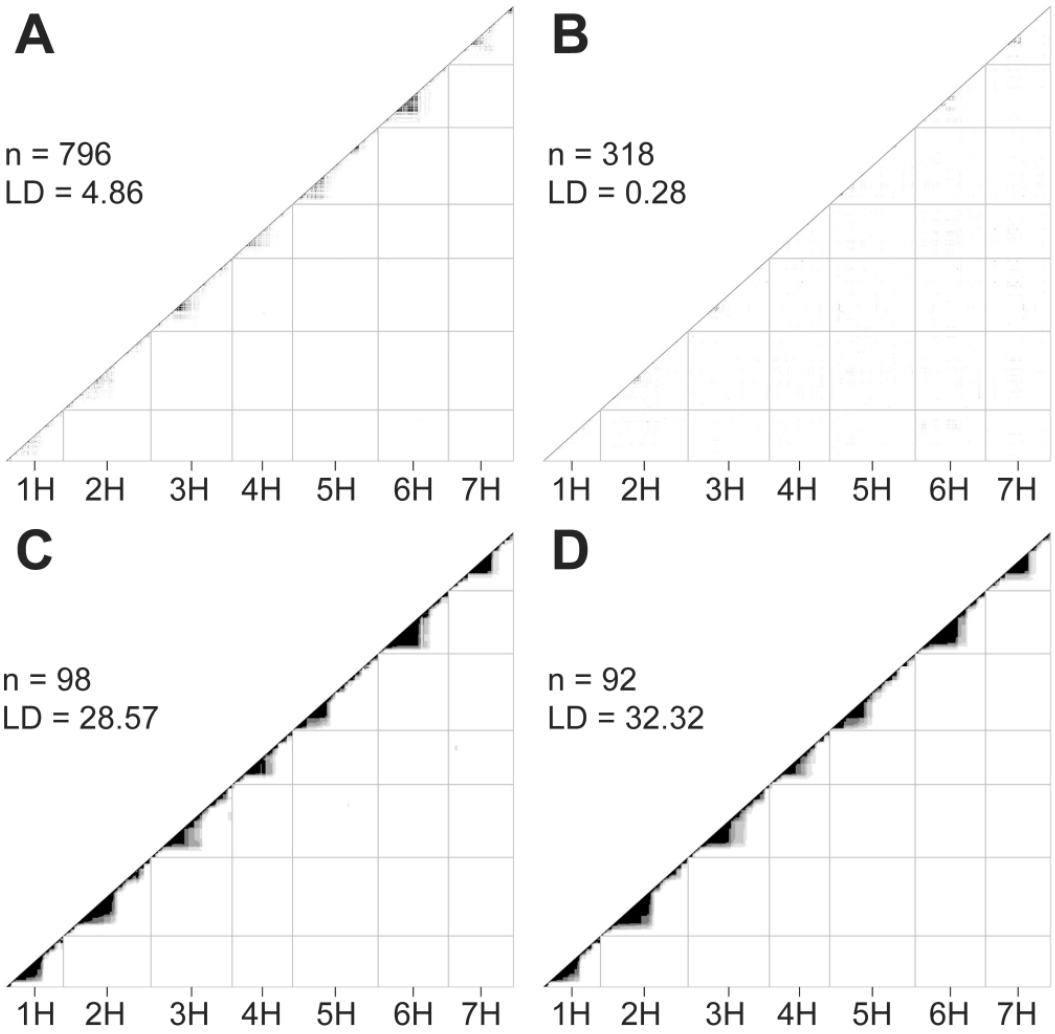
Supplemental Figure 2.2 Genome-wide pair-wise linkage disequilibrium decay. LD calculated as the genetic distance (cM) at which the r^2 decays to 0.2. (A) wild barley advanced backcross-nested association mapping population, (B) wild barley diversity collection association mapping panel, (C) *OUH-602* x *Harrington* advanced backcross mapping population, (D) *OUH-602* x *Harrington* recombinant inbred line mapping population.



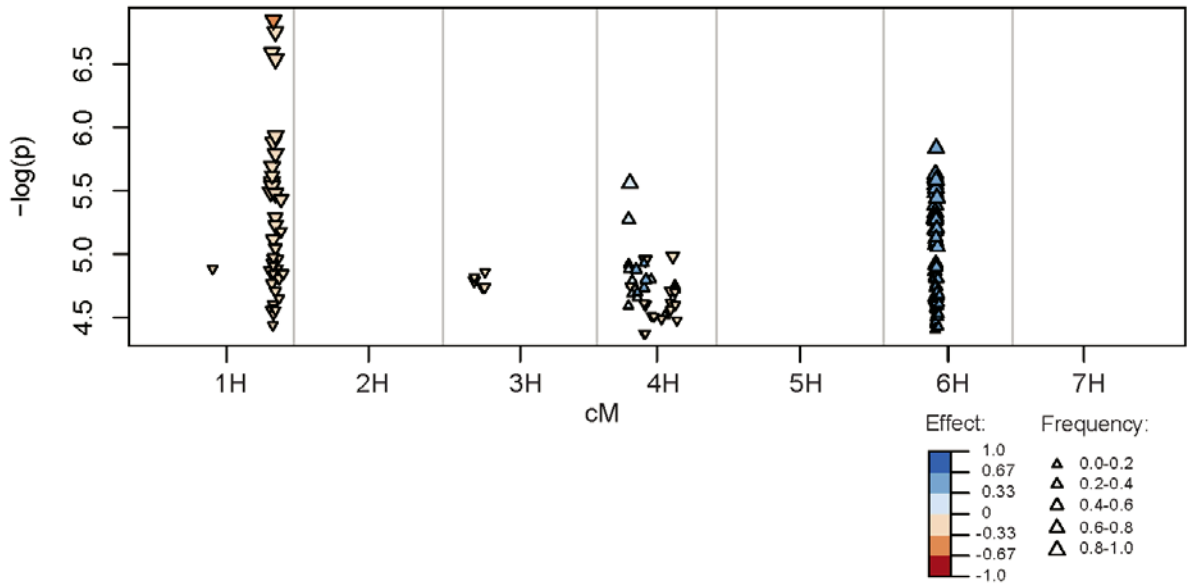
Supplemental Figure 2.3 Heatmap of genome-wide pair-wise linkage disequilibrium (r^2) for populations randomly sampled to a common population size of 90 individuals. LD as calculated in Supplemental Figure 2.1. (A) wild barley advanced backcross-nested association mapping population, (B) wild barley diversity collection association mapping panel, (C) *OUH-602* x *Harrington* advanced backcross mapping population, (D) *OUH-602* x *Harrington* recombinant inbred line mapping population.



Supplemental Figure 2.4 Heatmap of genome-wide pair-wise linkage disequilibrium (r^2) for full population sizes. LD as calculated in Supplemental Figure 2.2. (A) wild barley advanced backcross-nested association mapping population, (B) wild barley diversity collection association mapping panel, (C) *OUH-602* x *Harrington* advanced backcross mapping population, (D) *OUH-602* x *Harrington* recombinant inbred line mapping population.



Supplemental Figure 2.5 Significant marker-trait associations for grain protein content. Associations shown are those detected above a FDR = 0.05 threshold in greater than 5% of 100 bootstrap samplings of 25 individuals per AB-NAM family. Triangles pointing up indicate a positive effect of the non-*Rasmusson* allele. Triangles pointing down indicate a negative effect of the non-*Rasmusson* allele. Color indicates the magnitude of effects. Size of the triangle indicates the frequency of detection among bootstrap samples.



Supplemental Dataset 2.1 All significant marker trait associations for traits: grain protein content, glossy spike, glossy sheath, and blue hull color.

Chapter 3: Mapping agronomic traits in a wild barley advanced backcross nested association mapping (AB-NAM) population

Overview

The wild barley advanced backcross – nested association mapping population (AB-NAM) was developed to assess wild barley germplasm in an elite 6-rowed malting barley background. The 796 $BC_2F_{4:6}$ lines were derived from 25 wild barley accessions backcrossed to the cultivar *Rasmusson*. The population was grown in augmented field trials in Crookston, MN, and St. Paul, MN in 2012, and in Crookston, MN, Bozeman, MT, and Fargo, ND in 2013. Because the AB-NAM design minimizes the deleterious, unadapted traits of the wild barley parents, the population was able to be screened for agronomic traits: days to heading, height, tiller number, test weight, and yield in standard field trial conditions. The number of loci identified ranged from 2 (tillering, the trait with the lowest heritability) to 19 (height). Domestication related genes were identified for each trait, particularly the photoperiod gene *Ppd-H1* for days to heading, height, and tiller count, and the *Btr1*/*Btr2* locus for test weight and yield. Only a single variant showed beneficial variation for yield from a wild barley allele, but positive effects were identified for the other traits examined, including the yield component test weight. Overall, the AB-NAM population provides a means to map agronomically important traits in a population that contains a large amount of wild barley diversity.

Introduction

Populations of wild barley (*Hordeum vulgare* ssp. *spontaneum*), the progenitor of cultivated barley (*Hordeum vulgare* ssp. *vulgare*), are known to be rich sources of genetic

and phenotypic diversity. Over the course of crop domestication and improvement, the diversity of cultivated barley has been eroded by forces such as selection and drift (Rostoks et al. 2006). Today, cultivated barley populations retain an estimated 40% of the alleles present in wild barley populations (Ellis et al. 2000). Strong selection and inbreeding without the introduction of new sources of diversity have led to reduced genetic diversity in breeding populations (Condón et al. 2008; Fu and Somers 2009; Muñoz-Amatriaín et al. 2010). This limited diversity exposes these breeding programs to the risk of insufficient genetic resources available to overcome future disease threats, environmental changes, or resource limitations.

Barley breeders have several areas of focus, including yield, malting quality, disease resistance, and abiotic stress resistance. Historically, exotic sources of germplasm have been used primarily for introgressing large effect traits such as qualitative disease resistance. Large effect alleles for important traits have a clear economic value, justifying the time and resources needed to identify and introgress these beneficial alleles from unadapted germplasm. The use of exotic material for more complex traits including yield and quality characteristics has been more limited (Fischbeck 2003). Yield is a classic complex trait, influenced by many genetic and environmental factors. Furthermore, yield is not a trait that has necessarily undergone selective pressure in wild populations. Because yield is influenced by many genes and because there is large genetic diversity present in wild populations, it is likely that alleles for improved yield do exist in otherwise unadapted germplasm. With the cost of genotyping decreasing and with the improvement in analysis techniques, the prospects for identifying and using diverse germplasm for improving complex traits are expanding.

Identifying alleles for complex traits can be challenging, particularly when working with unadapted germplasm. Besides having seed substantially smaller than cultivated barley, wild barley seed is difficult to harvest and process because of its brittle rachis and tough awns (Pourkheirandish and Komatsuda 2007). The history of domestication and improvement in cultivated barley has removed many of these deleterious traits from the breeding pool, and a large improvement gap exists between wild and cultivated barley. Because of this, it is difficult to characterize the genetic value of unadapted germplasm directly. To minimize the confounding effects of these deleterious traits, advanced backcross (AB) quantitative trait loci (QTL) analysis was developed (Tanksley and Nelson 1996). This technique involves QTL mapping in a population that has been created by repeatedly backcrossing an unadapted accession to an adapted variety, and then selfing to create recombinant inbred lines (RILs). The resulting AB population is a set of lines that have a relatively small amount of unadapted genome in an adapted background. The backcross generations minimize the number of deleterious traits in any single individual, but given sufficient individuals in the population, the exotic alleles can be adequately sampled and analyzed. Furthermore, if beneficial alleles are identified, these alleles are already partially introgressed into an adapted background. This technique has been used by barley researchers to identify beneficial alleles from wild barley sources for many traits of interest, including: yield and agronomic traits (Li et al. 2005; von Korff et al. 2006; Gyenis et al. 2007; Eshghi et al. 2013), malting quality (von Korff et al. 2008; Schmalenbach and Pillen 2009; March et al. 2012), disease resistance (von Korff et al. 2005; Yun et al. 2006; Li et al. 2006),

abiotic stress tolerance (Saal et al. 2011; Hoffmann et al. 2012; Schnaithmann and Pillen 2013; Kalladan et al. 2013), and root traits (Naz et al. 2014).

Two major drawbacks to the AB strategy are that alleles from only one unadapted germplasm source are sampled, and that relatively few recombination events are present in these mapping populations, resulting in relatively low mapping resolution. Association mapping is a technique that generally circumvents these limitations by using the historic linkage disequilibrium in diverse populations to identify marker trait associations (Rafalski 2002b). But, association mapping with widely unadapted germplasm can be logistically difficult, and it is often limited to traits such as disease resistance that can be scored at the seedling stage (Steffenson et al. 2007; Roy et al. 2010). Recently, multiparent approaches that involve more intricate mating designs than biparental population have been developed and implemented in several model organisms (Churchill et al. 2004; Kover et al. 2009; Huang et al. 2011; King et al. 2012b) and crop species (Yu et al. 2008; Stupar and Specht 2013; Mace et al. 2013; Schnaithmann et al. 2014). While the designs of these populations vary, they all aim to incorporate a larger amount of diversity than biparental mapping populations, while minimizing the downsides of association mapping techniques: low power due to low allele frequencies and confounding effects of cryptic population structure (Myles et al. 2009; Morrell et al. 2012).

The resolution and power of multiparent populations allow for the assessment of complex traits like yield in diverse germplasm. To explore the genetic architecture and availability of beneficial alleles that contribute to yield and yield-related traits in wild barley germplasm, we analyzed an advanced backcross – nested association mapping

(AB-NAM) panel derived from 25 wild barley parents each backcrossed twice to a single cultivated barley cultivar *Rasmusson*. Our objectives were (1) to assess the genetic architecture of yield and yield-related traits in a wild barley derived mapping population, and (2) to identify loci conveying beneficial effects from wild barley germplasm.

Materials and Methods

Plant materials

The wild barley AB-NAM population is composed of 796 $BC_2F_{4:6}$ lines derived from backcrossing 25 diverse wild barley lines to the cultivar *Rasmusson*. Families ranged in size from 27-39 individuals. *Rasmusson* is a 6-rowed malting barley that was selected as the recurrent parent because of its favorable agronomic and malting characteristics (Smith et al. 2010). The wild parents were selected from the 318 wild barley accessions in the Wild Barley Diversity Collection (WBDC; Steffenson et al. 2007) to maximize the allelic content of the selected individuals at 1,402 barley oligo pool assay (BOPA1) SNP markers (Close et al. 2009), 556 DArT markers (Alsop et al. 2011), and 46 SSR markers. Each BC_2 individual was derived from a unique BC_1 plant, and selfed for four generations via single seed descent, with selection for 6-rowed spike morphology in the BC_2F_2 generation.

Genotyping

Parental lines were genotyped with the Illumina iSelect 9K SNP platform (Comadran et al. 2012) and exome capture sequenced (Mascher et al. 2013b). The AB-NAM lines were genotyped with a custom 384 Illumina GoldenGate SNP assay. The

markers included on the assay were selected based on parental BOPA1 and BOPA2 markers to be evenly distributed throughout the genome and to distinguish *Rasmusson* from the 25 wild barley parents. Five markers were excluded due to low quality data. DNA was extracted from bulked tissue samples of 6-8 greenhouse grown $BC_2F_{4:5}$ seedlings. To test for seed purity in the germplasm when grown in the field, 5 randomly selected lines in each of the two 2012 field environments were genotyped using the same 384-SNP platform.

The higher density marker datasets (9K SNP platform and exome capture sequence) were projected onto the AB-NAM lines using the 379 genotyped markers as a scaffold. When flanking scaffold markers indicated that a recombination event occurred, the projected markers were assigned an average of the flanking markers, weighted by the relative genetic distance between the projected marker and the flanking markers (Tian et al. 2011). The Muñoz-Amatriaín et al. (2014) 9K iSelect consensus map was used for imputing SNP markers, and the Mascher et al. (2013a) *Morex* x *Barke* POPSEQ map was used for imputing sequence variants. A total of 4,022 iSelect SNP markers (3,520 unique bins) and 263,531 sequence variants (126,303 unique bins) were imputed after filtering variants that were non-segregating, called heterozygous in *Rasmusson*, unmapped, or containing missing genotype calls. The exome capture sequencing of accession WBDC103 did not produce high quality sequence alignments, and therefore this population was excluded from trait mapping analysis.

Field trials and trait measurements

The AB-NAM population was grown in an augmented, type-2 field design (Lin and Poushinsky 1985) in five environments over two years: Crookston, MN (CR12) and St. Paul, MN (SP12) in 2012, and Crookston, MN (CR13), Fargo, ND (ND13), and Bozeman, MT (MT13) in 2013. Plot sizes were as follows: Minnesota trials – 8g seeded in 0.9m x 0.6m plots, North Dakota trial – 55g seeded in 1.5m x 2.4m plots, Montana trial – 20g seeded in 1.8m x 0.6m plots. *Rasmusson* was grown as a repeated check in the center of each 3 plot x 5 plot block. In 2012, cvs. *Conlon* and *Harrington* were grown as secondary repeated checks in each of 8 randomly selected blocks. In 2013, cvs. *Robust* and *Tradition* were used as secondary checks. To fill the remaining plots in the 900 plot design, 21-33 AB-NAM lines were used as duplicated entries. Field designs in 2012 included blocking in columns by heading date to facilitate harvest of lines with large variation in maturity. These included early, mid, late, and unknown heading date blocks. Entries were assigned to blocks based on field-collected data in the summer of 2011. This design was not repeated in 2013.

Traits measured included heading date (all environments), productive tiller count (CR12, SP12, CR13), height (all environments), yield (all environments), and test weight (ND13, MT13). Heading date was measured as the number of days from planting where 50% of the plot exhibited 50% spike emergence from the boot. Productive tiller count was measured as the number of tillers which held grain producing spikes in a 0.5 m section of the middle of a single plant row. Height was measured on a representative plant in the plot as the cm from the base to the top of the spike, excluding the awns, approximately one week after heading. Plots were mechanically harvested after end

trimming. Threshed, cleaned seed from each plot was weighed and yield measurements were converted to kg/ha. Test weight was measured on 0.118 L samples and converted to g/L.

Estimating heritability and partitioning trait variance

When necessary, trait measurements were corrected for field spatial variability using a moving average correction approach. Moving average shape design and subsequent value adjustments were performed using the *mvngGrAd* R package (Technow 2011). After spatial correction within environments, a multivariate mixed model was fit in ASReml-R (VSN International), and best linear unbiased predictors (BLUPs) were obtained, resulting in a single trait value for each AB-NAM line. Broad-sense heritability for each trait was calculated on a line mean basis across environments. Pearson correlation coefficients were calculated for each trait pair.

Trait mapping

Genome-wide association analysis was performed for each trait using multi-environment BLUPs. The GWAS function in the rrBLUP R package (Endelman 2011) was used to calculate marker-trait associations using the EMMAX/P3D approximation method (Kang et al. 2010; Zhang et al. 2010). Because there was negligible population structure, a K mixed model was used to account for marker-based relatedness. A bootstrapping approach was taken whereby 25 individuals were sampled from each of the 24 families with imputed sequence variants, for a total of 600 individuals or 79% of the 761 individuals used for mapping. This sampling process and the subsequent trait

mapping were performed 100 times. For each mapping analysis, a false discovery rate (FDR) of 0.05 was used as a threshold for calling significant associations, and a minor allele frequency threshold of 0.013 (10 individuals) was used to minimize spurious associations due to low frequency variants. To obtain estimates of allele substitution effect, marker effects were calculated by passing a single marker to the mixed.solve function in the rrBLUP R package (Endelman 2011; Mohammadi et al. 2015). To determine an overall significance, the –log(p-value) results, along with the marker effect estimates for marker-trait associations were averaged across significant bootstrap sampling replications. Additionally, to obtain a measure of confidence, we report the proportion of samples in which each marker trait association was identified. Those associations that occurred in less than 5% of samples were not reported. Independent loci were assumed when a gap of greater than 5 cM appeared between significantly associated markers.

Results

Trait variation, heritability, and correlations

Agronomic traits: days to heading, height, test weight, tiller count, and yield were measured in 2-5 augmented field trials over 2 years. Analysis of variance indicated significant ($p < 0.0001$) variation was explained by lines, and significant genotype by environment interactions were observed when considering those individuals that were replicated within and across environments. Best linear unbiased predictors (BLUPs) were calculated from multi-environment phenotypic data. The population-wide distribution of BLUPs is summarized in Table 3.1. The distribution of phenotypic values

subset by wild barley donor parent is shown in Figure 3.1. For days to heading, height, and tiller count, the population mean lies close to the value of *Rasmusson*. Whereas, for test weight and yield, the *Rasmusson* phenotype outperforms a large majority of the individuals in the population.

Estimates of trait heritability ranged from 0.11 to 0.82 (Table 3.1). The heritability of tiller count was the lowest ($h^2 = 0.11$). This is a result of relatively low variation (CV = 2.75), and relatively high measurement error. The heritability of days to heading was high ($h^2 = 0.82$), an observation that is consistent with other barley populations including the Halle Exotic Barley (HEB-25) wild barley derived NAM population (Maurer et al. 2015). Height, test weight and yield each had intermediate heritability estimates that ranged from 0.61-0.70 (Table 3.1).

Heading date was significantly correlated with each of the other agronomic traits measured (Table 3.2). Heading date was most strongly, positively correlated with height, but it also had a moderate positive correlation with yield and moderate negative correlations with test weight and tiller number. In addition to the correlation with days to heading, height was negatively correlated with number of tillers and positively correlated with yield. Test weight was strongly, positively correlated with yield.

Distribution of trait values among AB-NAM families

Generally, the mean of individual families for each trait was similar to the mean of the full population (Table 3.1). For a few families trait means deviated significantly ($p < 0.01$) from the population mean. The mean number of days to heading for the family derived from WBDC302 was significantly higher than the population-wide mean of 51.3

days. The families derived from WBDC336 and WBDC348 were on average taller than the population mean, and the WBDC042 family mean was shorter than the population mean of 72.5 cm. Test weights for families WBDC016, WBDC092, and WBDC350 were higher than the population mean of 684 g/L, and test weights for families WBDC035 and WBDC302 were lower than the population mean. WBDC255 and WBDC348 had lower tiller counts and WBDC350 had higher tiller counts than the population mean of 58.7 tillers per 0.5m. WBDC020, and WBDC350 showed significantly higher yield. Among the families, WBDC350 performed quite well with regard to yield and test weight (Figure 3.1). Individuals from 20 of the 25 families had yield BLUPs greater than or equal to *Rasmusson*, and the highest yielding, WBDC173-derived individual out-yielded *Rasmusson* by 480 kg/ha (Table 3.1).

Marker – trait associations identified

Days to heading

A total of ten loci were identified that exhibited significant associations with days to heading (Table 3.3). These loci were found on chromosomes 2H, 3H, 5H, and 7H (Figure 3.2). Loci on chromosome 2H and 7H were identified in all bootstrap samples and were the most highly significant loci identified for heading date (Table 3.3). The most significant variants were located on *Morex* sequence contig 94710, which contains the *Ppd-H1* photoperiod sensitivity gene (Alqudah et al. 2014). These variants had an effect of reducing days to heading by 3.02 days. While a majority of the variants in this region had a negative effect, some variants had a small positive effect of increasing the number of days to heading by 0.01-0.40 days (Supplemental Dataset 3.1). Three

additional loci were identified on chromosome 2H, with those at 41.22 cM and 85.98 cM increasing the days to heading by 0.47 and 2.06 days respectively, and another locus on 2H exhibited both negative and positive effects with the positions of highest significance at 53.82 cM and 59.70 cM (Table 3.3).

Loci on chromosome 3H at 51.77 cM and 5H at positions 42.15 cM, 113.89 cM, and 119.24 cM exhibited positive effects of 0.14-1.43 days in 8-45% of the bootstrap samples. An infrequently detected association was identified at 13.60 cM on chromosome 7H with an effect of reducing days to heading by 0.37 days. The other 7H locus contained variants of both positive and negative effects. Located at 29.96 cM, the variants with the highest significance in the region had an effect of increasing days to heading by 2.10 days, while the other variants in the region, at 34.21 cM, had an effect of decreasing days to heading by 0.68 days.

Height

Nineteen loci were identified for height. These loci were found on all chromosomes except 6H (Figure 3.2), and exhibited effects that ranged from decreasing height by 4.63 cm to increasing height by 2.61 cm (Supplemental Dataset 3.1). The most consistently identified loci were found in all bootstrap samples and located on chromosomes 2H at 19.90 cM, and on 3H at 108.85 cM. The 2H association produced the largest effect of -4.63 cm and corresponds to the same *Ppd-H1* variants that are strongly associated with heading date. The locus on chromosome 7H at 29.96 cM was also coincident with a days to heading locus. In this case, the most significant variants/contigs were not the same for the heading date and height loci, although the

contigs in which they were identified are mapped to the same location (Supplemental Dataset 3.1).

Test weight

Test weight was mapped to nine loci found on chromosomes 1H, 2H, 3H, 5H, and 7H (Figure 3.2). The 1H locus at 94.90 cM was the most consistently identified locus with an effect of decreasing the test weight by 20.66 g/L. Additional positive effect variants were identified in this region at 97.73 cM (Supplemental Dataset 3.1). A majority of the other loci identified produced a negative effect, with the exception of the locus at 43.11 cM and 45.82 cM on chromosome 3H which contained variants with both positive and negative effects.

Tiller count

Tiller count had a large amount of environmental error and consequently low heritability. Only two loci were identified for tiller count, both with a positive effect of the non-*Rasmusson* alleles. These loci are located on chromosomes 2H at the *Ppd-H1* locus, and on chromosome 4H at 91.29 cM (Table 3.3).

Yield

Eleven significant marker-trait associations for yield were identified across all chromosomes except 4H (Figure 3.2). Only a single, infrequently identified locus on 3H at 57.08 cM exhibited positive effects of the non-*Rasmusson* allele (Supplemental Dataset 3.1). The significant associations for yield overlapped with those for other traits. The

locus on chromosome 1H at 94.90 was shared with test weight along with the most commonly identified locus on 3H at 44.83 cM. A QTL for yield was also identified on chromosome 2H, with significant associations to variants on the contig containing *Ppd-H1*, a major contributor to heading date, height, and tiller count (Supplemental Dataset 3.1).

Discussion

Wild barley germplasm for crop improvement

The wild barley AB-NAM exhibits a large amount of phenotypic and genotypic variation that may be useful for cultivated barley breeding. Here we described the characterization of the AB-NAM for five agronomic traits of interest to the breeding program. Among the traits analyzed, all exhibited large ranges in phenotypic values, with the exception of tiller number, which had a relatively narrow phenotypic range (Table 3.1). Correlations between traits were common, and heading date was significantly correlated with each of the other traits: height, test weight, tiller count, and yield. This correlation is reflected in the overlap of the large effect QTL localized to variants in the *Ppd-H1* gene on *Morex* contig 94710 (Supplemental Dataset 3.1). For each trait except test weight, these variants had highly significant and frequently identified associations. This reflects the importance of *Ppd-H1* in regulating plant growth and must be considered when working with germplasm that contains various alleles at this locus. While this major effect QTL was shared between several traits, relatively few of the smaller effect loci were shared among traits. Furthermore, all traits except tiller count had both positive and negative significant marker-trait associations,

and at one or more loci for each trait, variants exhibited both positive and negative effects (Table 3.3), suggesting the presence of multiple non-*Rasmusson* haplotypes in the population.

Mapping complex traits with the AB-NAM: heading date

The genetic control of heading date in barley has been studied extensively (Cockram et al. 2007; Alqudah et al. 2014; Maurer et al. 2015). Maurer et al. (2015) examined heading date in a similar population to the wild barley AB-NAM. The Halle Exotic Barley (HEB-25) population is composed of 25 wild barley accessions backcrossed to a single cultivated barley. In their examination of heading date, they found a very wide phenotypic range of almost 50 days, whereas here we observed a more modest range of approximately 17 days (Table 3.1). This difference is likely due to the additional backcross generation included in the development of the BC_2-derived AB-NAM population, which creates a more uniform genetic background than in the BC_1-derived HEB-25 lines.

Consistent with results found here, the largest effect association identified with the HEB-25 population was the *Ppd-H1* locus on chromosome 2H (Maurer et al. 2015). Besides *Ppd-H1*, other QTL corresponded to those identified in the HEB-25 population. The *HvCEN* locus (Comadran et al. 2012) on chromosome 2H at 58.00 cM, the *Vrn-H1* locus (Yan et al. 2003) on chromosome 5H at 125.76 cM, and the *Vrn-H3* locus (Yan et al. 2006) on chromosome 7H at 34.43 cM were identified in both studies (POPSEQ map positions per Alqudah et al. 2014). For loci *Ppd-H1*, *HvCEN*, and *Vrn-H3*, we identified variants in the region with both positive and negative effects. Loci found in the HEB-25,

but not found in the AB-NAM included *denso* (Jia et al. 2009) on chromosome 3H, a novel locus on chromosome 4H, and *Vrn-H2* (Yan et al. 2004) on chromosome 4H. In addition to the loci already mentioned, we identified loci that colocalize with *HvCO18* on chromosome 2H at 41.85 cM, *HvCO3* on chromosome 5H at 43.76 cM, and three novel QTL on chromosomes 2H, 5H, and 7H (Supplemental Dataset 3.1, Alqudah et al. 2014).

Mapping complex traits with the AB-NAM: height

The largest number of QTL were identified for height, several of which appear to correspond to QTL identified in previous mapping experiments (as queried in Grain Genes: http://wheat.pw.usda.gov/). The *Ppd-H1* locus had the largest effect of reducing height by 4.63 cm and this locus was identified in all bootstrap samples. The locus with the second highest significance, located on 3H at 108.85 cM, colocalizes with the semidwarf gene *denso* (Jia et al. 2009).

Mapping complex traits with the AB-NAM: tillering

The heritability of tillering in this study was low ($h^2 = 0.11$). Relatively small variation coupled with high measurement error contributed to this low measurement. In the field, variability in seeding and germination may have contributed to high measurement error. The photoperiod sensitive, non-*Rasmusson* alleles for variants in the *Ppd-H1* gene were identified as having an effect of increasing number of tillers by 0.65, and a novel QTL on chromosome 4H at 91.29 cM increased number of tillers by 1.04. Neither loci were identified in a large proportion of bootstrap samplings (Table 3.3). Variation at photoperiod and vernalization loci have been shown to affect tillering in

barley, particularly *Vrn-H2* on chromosome 4H (Karsai et al. 2006), but the 4H QTL for tillering found here at 91.29 cM does not correspond with the *Vrn-H2* location at 4H, 114.94 cM (Alqudah et al. 2014) nor does it correspond to the locations of previously identified QTL or known tillering mutants (Hussien et al. 2014; Naz et al. 2014).

Mapping complex traits with the AB-NAM: test weight

Test weight is a volumetric measurement that is influenced by seed shape and size. We selected for 6-rowed spike morphology in the AB-NAM population to minimize the confounding effects of spike morphology on other traits. This led to a severe reduction of wild barley introgression surrounding the *vrs-1* locus on chromosome 2H in the AB-NAM population. Because spike morphology affects seed shape and size, we might expect that modifiers of spike morphology would influence the mapping of test weight. But, known modifiers of spike morphology do not seem to correspond with the test weight QTL identified here (Muñoz-Amatriaín et al. 2014). Test weight did show a small association with the brittle rachis locus.

Mapping complex traits with the AB-NAM: yield

The brittle rachis morphology of wild barleys provides a means for dispersal of seeds. Because seed dispersal is disadvantageous for cultivated agriculture, a non-shattering spike was one of the primary domestication traits (Pourkheirandish and Komatsuda 2007). During the process of developing the AB-NAM population, we did not impose selection against seed shattering. Therefore, it is not surprising that the brittle rachis complex on chromosome 3H (Komatsuda et al. 2004) is the most influential yield

QTL identified. Besides the brittle rachis locus, the *Ppd-H1* locus was consistently associated with yield, and a region of chromosome 1H at 94.90 cM was shared with test weight.

Imputation introduces uncertainty in marker calls

The NAM design allows for direct projection of marker calls from high density parental genotyping onto progeny. This provides an opportunity to minimize the costs of genotyping such large populations, while still providing sufficient marker coverage to take advantage of the increased resolution of the NAM population design. The wild barley AB-NAM population was genotyped at a relatively low density, with an average of one marker every 3.57 cM. Given that this average includes regions of the genome with larger gaps in marker coverage, the imputation of markers does introduce a degree of uncertainty, particularly in regions of the genome where marker coverage is low. This limitation could be improved by using higher density markers or by selecting markers from sequence-based resources that can provide a more uniform distribution of segregating markers throughout the genome. To address this issue, the AB-NAM lines have been genotyped using a genotyping by sequencing approach (Poland et al. 2012) that will provide higher density genotyping for future studies.

Low minor allele frequencies of AB-NAM associations

Coupling the advanced backcross design with the NAM design presents a challenge for mapping the effects of alleles which segregate in only a small number of AB-NAM families. Unlike RIL-derived NAM populations where the expected minor

allele frequency is (0.50 x p_s)/ p_t , where p_s is the number of diverse parents segregating at the locus and p_t is the total number of diverse parents, the expected minor allele frequency in the AB-NAM is (0.125 x p_s)/ p_t. This means when a single parent is segregating for a locus in a 25 parent NAM, we expect the minor allele frequency to be 0.005 as compared to 0.02 in a comparably sized RIL-derived NAM. This is problematic because the power for detecting significant interactions decreases as MAF increases, and the frequency of false positives increases (Tabangin et al. 2009). The balance becomes the number of diverse accessions sampled versus the number of lines created for each family. Higher detection of significant loci, particularly alleles private to individual families may have been achieved if we had made larger populations of fewer parental lines.

Conclusions

We grew the phenotypically and genetically diverse wild barley AB-NAM in field trials in 2012 and 2013 in Minnesota, North Dakota, and/or Montana. The AB-NAM design allowed for screening in standard field trials, minimizing the unadapted traits present in the wild barley parental accessions, while maintaining large amounts of diversity. We identified several QTL for each trait measured. In each case, both previously identified and novel QTL were found. Transgressive segregation for test weight and yield is minimal, but there were some lines that had higher test weight and/or yield than *Rasmusson*. The AB-NAM population is a resource available for screening large amounts of genetic diversity in a genetic background that is adapted to the Upper Midwest United States.

Table 3.1 Trait values and heritabilities for multi-environment trait BLUPs.

Trait	Min	Max	Rasmusson	Mean	CV	Heritability
Heading Date (days after planting)	44.24	61.62	50.69	51.26	4.62	0.82
Height (cM)	49.53	92.51	71.94	72.50	7.03	0.63
Test Weight (g/L)	569.29	735.50	716.83	684.48	3.82	0.70
Tillers (tillers/0.5m)	53.56	63.71	59.09	58.66	2.75	0.11
Yield (kg/ha)	1799	5764	5283	4545	12.77	0.61

Table 3.2 Pearson correlations between traits

	Heading Date	Height	Test Weight	Tillers	Yield
Heading Date					
Height	0.43***				
Test Weight	-0.16***	0.00			
Tillers	-0.18***	-0.32***	-0.01		
Yield	0.15***	0.20***	0.56***	0.03	

*p < 0.01, **p < 0.001, ***p < 0.0001

Table 3.3 Summary of marker-trait associations of maximum significance. All significant marker-trait associations can be found in Supplemental Dataset 3.1. QTL on the same chromosome were considered unique loci if there was a greater than 5cM gap between significant markers.

Trait	Locus[a]	Variant[b]	Chrom[c]	cM[c]	MAF[d]	Freq[e]	-log(p)[f]	Effect[g]	Ras[h]
Days to heading	2H.1	morex_contig_61987:8917_A/G	2H	19.05	0.02	1.00	7.29	0.01	A
	2H.1	morex_contig_94710:1754_T/C*	2H	19.90	0.11	1.00	71.73	-3.02	T
	2H.2	morex_contig_135313:5170_G/A	2H	41.22	0.01	0.39	4.28	0.47	G
	2H.3	morex_contig_39365:4995_G/A	2H	53.82	0.02	0.07	3.91	-1.16	G
	2H.3	morex_contig_135238:1265_G/T	2H	59.70	0.04	0.08	4.13	0.47	G
	2H.4	morex_contig_1565994:1482_TG/T	2H	85.98	0.01	0.12	4.18	2.06	T
	3H.1	morex_contig_1578959:209_TC/T	3H	51.77	0.01	0.08	4.10	0.14	TC
	5H.1	morex_contig_1595050:226_GTA/G*	5H	42.15	0.02	0.11	3.87	1.21	GTA
	5H.2	morex_contig_2553463:241_GC/G*	5H	113.89	0.01	0.09	4.16	1.43	GC
	5H.3	morex_contig_1560492:6299_T/C*	5H	119.24	0.03	0.45	4.44	1.26	T
	7H.1	morex_contig_43906:1474_G/A	7H	13.60	0.02	0.08	3.85	-0.37	G
	7H.2	morex_contig_9006:1415_G/C**	7H	29.96	0.02	1.00	7.96	2.10	G
	7H.2	morex_contig_1561657:2925_G/A*	7H	34.21	0.07	1.00	6.60	-0.68	G
Height (cm)	1H.1	morex_contig_38760:17841_C/T*	1H	21.25	0.02	0.09	4.41	-2.12	C
	1H.2	morex_contig_42099:3456_G/A	1H	42.78	0.01	0.08	4.77	-2.62	G
	1H.3	morex_contig_6883:3404_A/G	1H	70.54	0.01	0.15	4.47	-2.97	A
	1H.3	morex_contig_139934:324_ATGAACTTGGCTAG/A	1H	70.72	0.01	0.15	4.47	-2.96	ATGAA…
	1H.4	morex_contig_158638:4096_G/A*	1H	126.06	0.02	0.07	4.46	-2.91	G
	2H.1	morex_contig_94710:1754_T/C*	2H	19.90	0.11	1.00	27.84	-4.63	T
	2H.2	morex_contig_2547325:563_C/G	2H	33.68	0.08	0.06	4.49	-3.66	C
	2H.3	morex_contig_2106916:1709_C/A	2H	40.37	0.01	0.56	5.44	-3.77	C
	2H.4	morex_contig_6288:2299_T/C*	2H	123.65	0.02	0.06	4.26	-2.38	T
	3H.1	morex_contig_39909:12902_A/C*	3H	68.13	0.01	0.08	4.16	-1.85	A
	3H.2	morex_contig_40861:9707_C/T**	3H	108.85	0.13	1.00	6.55	2.52	C

Table 3.3 continued

Trait	Locus[a]	Variant[b]	Chrom[c]	cM[c]	MAF[d]	Freq[e]	-log(p)[f]	Effect[g]	Ras[h]
	4H.1	morex_contig_98608:1030_T/C	4H	22.24	0.04	0.50	4.52	2.61	T
	4H.2	morex_contig_1581403:1626_C/G	4H	43.48	0.01	0.07	4.19	-1.89	C
	4H.3	morex_contig_1569145:1199_G/T	4H	114.94	0.03	0.92	5.53	-2.35	G
	5H.1	morex_contig_54778:1059_G/A	5H	59.72	0.02	0.25	4.24	-2.49	G
	7H.1	morex_contig_1568403:2016_A/G	7H	20.96	0.04	0.14	4.17	2.37	G
	7H.2	morex_contig_2547812:2781_A/T	7H	29.96	0.07	0.22	4.38	2.51	T
	7H.3	morex_contig_1579633:859_G/C	7H	42.28	0.02	0.06	4.50	-1.72	G
	7H.4	morex_contig_323672:471_C/T*	7H	48.16	0.03	0.40	4.59	-1.34	C
	7H.5	morex_contig_42961:5037_C/T*	7H	133.92	0.02	0.06	4.27	-1.48	C
Test weight (g/L)	1H.1	morex_contig_166809:1737_A/T	1H	94.90	0.11	1.00	10.55	-20.66	A
	1H.1	morex_contig_1561363:4513_C/G	1H	97.73	0.03	1.00	8.18	4.25	C
	2H.1	morex_contig_381927:297_A/C	2H	52.62	0.02	0.38	4.68	-17.19	A
	2H.2	morex_contig_47818:3505_C/T	2H	59.42	0.03	0.77	5.22	-16.49	C
	3H.1	morex_contig_45244:1965_G/GA	3H	43.13	0.03	0.31	4.51	-10.26	G
	3H.1	morex_contig_5708:4611_C/T*	3H	45.82	0.04	0.07	4.32	2.22	C
	5H.1	morex_contig_41843:6990_C/T	5H	0.14	0.03	0.51	4.50	-14.32	C
	5H.2	morex_contig_8004:481_T/C	5H	80.49	0.01	0.09	4.44	-43.64	T
	5H.3	morex_contig_138324:6222_T/C	5H	107.57	0.02	0.11	4.44	-20.43	T
	7H.1	morex_contig_79300:514_C/A*	7H	41.43	0.02	0.50	4.77	-22.82	C
	7H.2	morex_contig_52768:4574_C/A	7H	61.19	0.02	0.36	4.54	-23.88	C
Tillers (per 0.5m)	2H.1	morex_contig_94710:4178_G/A	2H	19.90	0.10	0.24	6.65	0.65	G
	4H.1	morex_contig_1562579:4864_G/C	4H	91.29	0.03	0.10	6.29	1.04	G
Yield (kg/ha)	1H.1	morex_contig_166809:1737_A/T	1H	94.90	0.11	0.52	4.31	-327	A
	2H.1	morex_contig_56215:381_T/C*	2H	19.48	0.11	0.98	5.56	-288	T
	3H.1	morex_contig_158632:4370_A/G	3H	44.83	0.10	1.00	19.25	-590	A

Table 3.3 continued

Trait	Locus[a]	Variant[b]	Chrom[c]	cM[c]	MAF[d]	Freq[e]	-log(p)[f]	Effect[g]	Ras[h]
	3H.2	morex_contig_2557470:544_T/C	3H	57.08	0.03	0.06	4.09	169	T
	3H.3	morex_contig_64931:1051_G/A	3H	76.35	0.06	0.07	4.32	-335	G
	3H.4	morex_contig_358142:901_C/T	3H	86.33	0.01	0.06	3.97	-281	C
	5H.1	morex_contig_54820:6729_T/A	5H	13.26	0.04	0.06	4.26	-224	T
	5H.2	morex_contig_135266:2648_T/C*	5H	47.50	0.02	0.89	5.14	-361	T
	6H.1	morex_contig_37935:10308_C/A	6H	113.21	0.01	0.08	4.60	-412	C
	6H.2	morex_contig_2546718:7710_A/G*	6H	126.49	0.02	0.16	4.60	-404	A
	7H.1	morex_contig_37512:13958_A/G	7H	140.65	0.02	0.08	4.05	-302	A

[a] QTL on the same chromosome were considered unique loci if there was a greater than 5cM gap between significant markers.
[b] Variants significantly associated with the trait. Associations were determined by an FDR < 0.05 significance threshold and a frequency of detection > 0.05 in 100 bootstrap samplings of 25 individuals from each family.
[c] Chromosome and cM positions of *Morex* contig contain the variant as determined by the Mascher et al. (2013a) POPSEQ map.
[d] Minor allele frequency of variant in the AB-NAM population
[e] Frequency of significant association with variant detected in 100 bootstrap samples
[f] –log(p-value) averaged across significant bootstrap tests
[g] Relative effect of non-Rasmusson allele, averaged across significant bootstrap tests
[h] Rasmusson allele at variant locus
* Variant is one of multiple variants with identical segregation patterns at this locus.
** Variant is one of multiple variants with identical segregation patterns at this locus, some of which are found on different contigs.

Figure 3.1 Box plots of multi-environment BLUP values for days to heading, height, test weight, tiller number, and yield, split by AB-NAM family. Population mean trait values are denoted by a solid horizontal gray line. Rasmusson trait values are denoted by a dashed horizontal gray line.

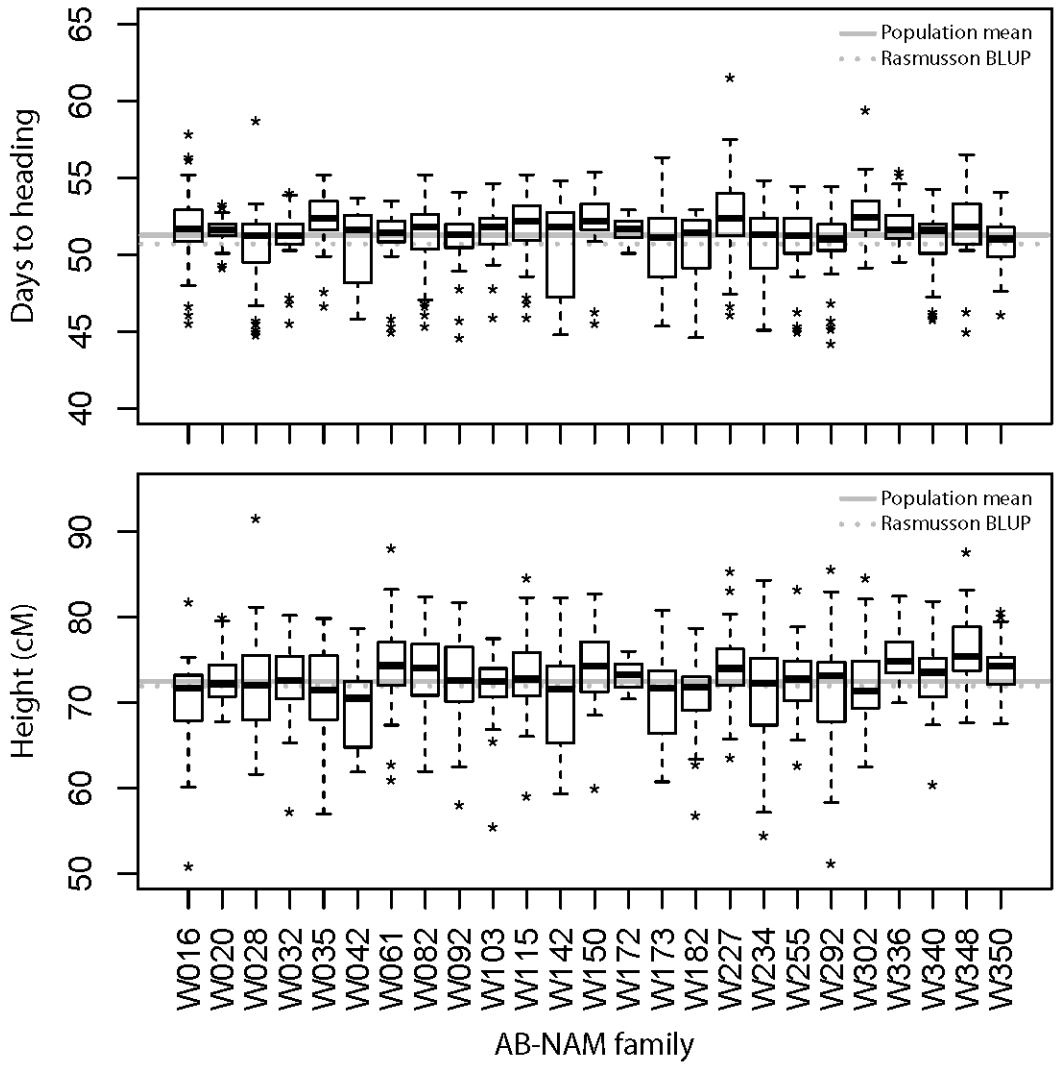


Figure 3.1 continued

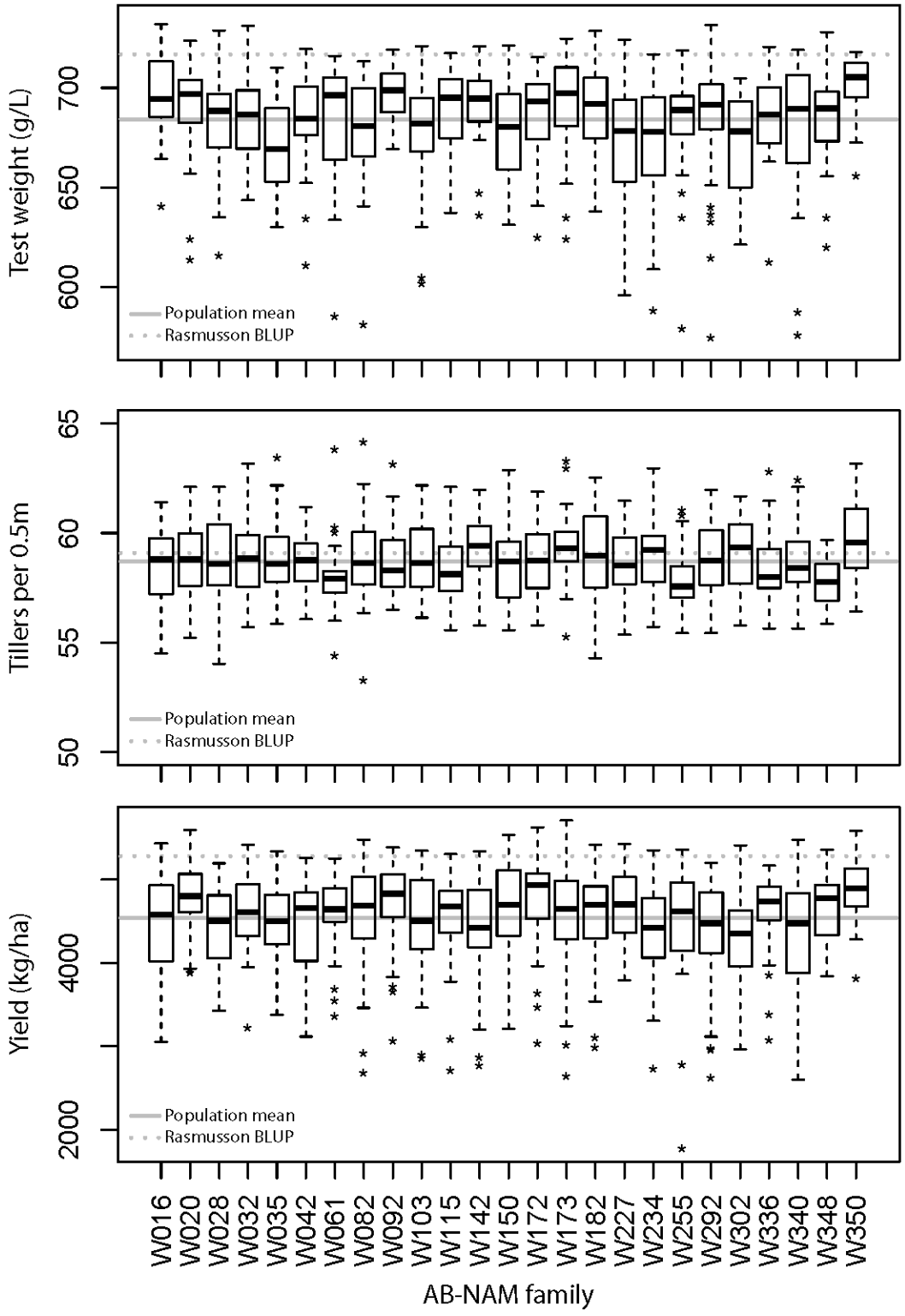


Figure 3.2 Significant marker-trait associations. Associations shown are those detected above a FDR = 0.05 threshold in greater than 5% of 100 bootstrap samplings of 25 individuals per AB-NAM family. Triangles pointing up indicate a positive effect of the non-*Rasmusson* allele. Triangles pointing down indicate a negative effect of the non-*Rasmusson* allele. Color indicates the magnitude of effects. Size of the triangle indicates the frequency of detection among bootstrap samples.

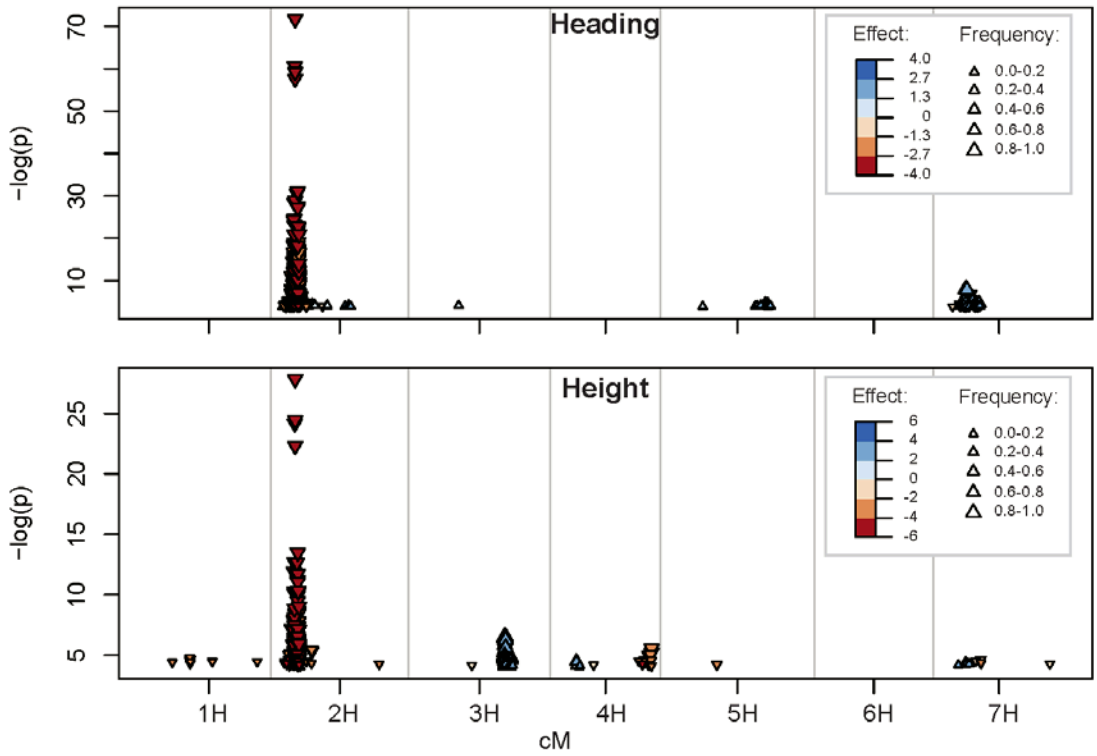
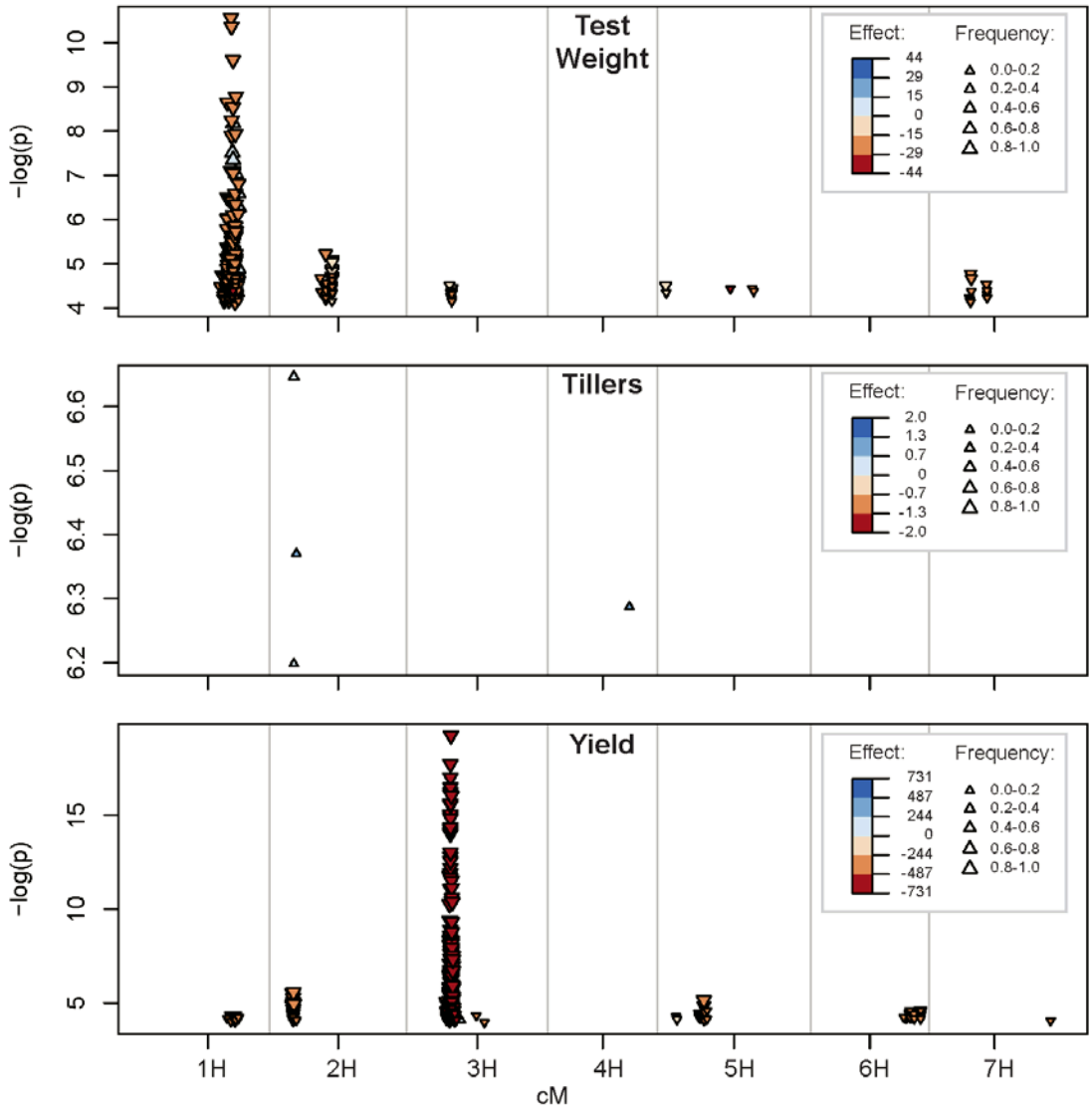


Figure 3.2 continued



Supplemental Dataset 3.1 All significant marker trait associations for traits: days to heading, height, test weight, tiller count, and yield.

Chapter 4: Mapping malting quality and yield characteristics in a North American 2-rowed malting barley × wild barley advanced backcross population

Overview

Due to the stringent quality requirements imposed by the US malting barley industry, barley breeders have been reluctant to introduce exotic germplasm into their breeding programs. To understand whether wild barley (*Hordeum vulgare* ssp. *spontaneum*) contains favorable alleles for yield and malting quality characteristics, we mapped QTL for heading date, height, lodging, yield, and nine grain characteristics important to the malting and brewing industry. Traits were mapped in a wild x cultivated barley, BC_2-derived advanced backcross mapping population. The recurrent parent *Harrington* is a North American 2-rowed malting barley cultivar, and the donor parent *OUH602* is a wild barley accession that exhibits resistance to multiple barley pathogens. Previous mapping studies using this population identified QTL for multiple disease resistance and agronomic traits. The 98 lines were grown in five replicated field trials at St. Paul and Crookston, MN in 2009-2011. One to four QTL were identified for each trait, for a total of 36 QTL identified. For all yield QTL identified, the *OUH-602* allele decreased the trait value. Wild barley alleles had both positive and negative trait effects on grain traits: diastatic power, free amino nitrogen, and soluble protein. For future validation of QTL identified, a set of pre-introgression lines composed of 29 BC_2-derived and 6 BC_3-derived lines were identified that collectively contain introgressions across the entire *OUH-602* genome, while minimizing background introgression.

Introduction

Worldwide, approximately 20% of the barley (*Hordeum vulgare* ssp. *vulgare*) crop is used for malting and brewing (Blake et al. 2010). Because of specific quality and flavor requirements imposed by the malting and brewing industries, malting barley is often grown under contract, subject to quality screenings, and incentivized with high premiums. Therefore, breeders are required to balance improvements in yield and agronomic performance with maintaining malting quality traits, which range from grain plumpness, to specific levels of protein content and enzyme activity (Fox et al. 2003). To maintain this balance, breeders have historically utilized techniques that focus on crossing elite cultivars from within highly inbred breeding programs (Rasmusson and Phillips 1997). But, over many cycles of breeding, the genetic and phenotypic diversity of these populations can become limited, and continued improvement necessitates addition of genetic diversity (Condón et al. 2008; Fu and Somers 2009; Muñoz-Amatriaín et al. 2010).

Wild barley (*Hordeum vulgare* ssp. *spontaneum*), the progenitor of cultivated barley, retains much of the genetic diversity that was lost during the process of domestication and cultivar improvement (Ellis et al. 2000). Natural populations have been subject to a broad range of biotic and abiotic stressors. Consequently, it is not surprising that wild barley germplasm has been shown to contain valuable alleles for disease resistance (Fetch et al. 2003; Steffenson et al. 2007; Schmalenbach et al. 2008) and abiotic stress tolerance (Ellis et al. 2000; Lakew et al. 2013a). To access these valuable alleles, breeders risk disrupting the favorable yield and quality traits that have been improved over the course of breeding. Additionally, wild barley germplasm

exhibits traits which are unadapted to modern agriculture, such as seed shattering, vernalization requirements, and photoperiod sensitivity (Pourkheirandish and Komatsuda 2007). Because of the logistical difficulties of working with unadapted barley germplasm, researchers have adopted the advanced backcross-quantitative trait loci (AB-QTL) mapping approach to explore the genetic architecture and utility of exotic germplasm.

The AB-QTL technique involves QTL mapping in populations derived from multiple backcrosses of an exotic donor parent to an adapted recurrent parent (Tanksley and Nelson 1996). This process results in lines which contain a small amount of exotic genome introgressed into an adapted background. By analyzing the exotic germplasm in the context of an adapted genome, researchers are able to minimize the confounding effects of epistasis and unadapted traits, allowing for a more accurate measurement of genetic value (Tanksley and Nelson 1996). This technique has been used in several crop systems to identify alleles for agronomic and quality traits (reviewed in Wang and Chee 2010). In barley, several studies have examined malting and yield in AB or introgression populations derived from wild barley sources (Li et al. 2005; von Korff et al. 2008). Despite malting quality and yield being unlikely selection targets in the wild, the complex nature of these traits provides many avenues for improvement. In the wild x cultivated barley mapping populations analyzed for yield and malting quality, a majority of the beneficial alleles are contributed by the cultivated parent, but transgressive segregation and beneficial wild barley alleles have been identified for malting and yield components (Matus et al. 2003; Li et al. 2005; von Korff et al. 2008).

Here we describe our efforts to map QTL for malting quality and yield traits in a BC_2-derived advanced backcross population from a highly disease resistant wild barley donor parent *OUH-602* and a North American 2-rowed malting barley cultivar *Harrington*. Our objectives were to (1) identify the genetic architecture of malting quality and yield traits in a wild barley advanced backcross population, (2) determine whether *OUH-602* contains beneficial alleles for malting quality and/or yield, and (3) genotypically characterize a set of BC_3 derived introgression lines for future validation experiments.

Materials and Methods

Plant material

The *Harrington* x *OUH-602* (HOUH) AB population was previously developed and analyzed for seedling disease resistance and morphological traits (Yun et al. 2006; Gyenis et al. 2007). *Harrington* is a North American 2-row malting barley, and *OUH-602* is a wild barley accession which exhibits multiple disease resistance. The population is composed of 98 $BC_2F_{6:8}$ lines. An additional 26 BC_3F_2 lines were derived from crossing 26 of the $BC_2F_{6:8}$ lines to *Harrington*. These 26 BC_3F_2 lines were genotyped, but they were not evaluated in any field experiments.

Genotyping

The HOUH population and the 26 BC_3F_2 lines were genotyped with a custom 384-SNP Illumina VeraCode assay developed for genotyping the wild barley advanced backcross – nested association mapping (AB-NAM) population. A total of 162 markers

were segregating between the parents, with an average of one marker every 7.0 cM. The Muñoz-Amatriaín et al. (2011) barley consensus map was used to order markers. Graphical genotypes were produced by linearly interpolating genotypes using 1 cM intervals across the consensus map. A set of introgression lines was selected from the 98 HOUH individuals and 26 BC_3F_2 lines by sequentially adding individuals containing homozygous *OUH-602* introgressions across the genome, while prioritizing individuals with low genome-wide introgression frequency.

Agronomic trait measurement

The HOUH population was grown in a randomized complete block design in St. Paul, MN (2009-2011; SP09, SP10, SP11) and Crookston, MN (2010-2011; CR10, CR11), for a total of 5 environments with 3 replications in each environment. Cultivars *Harrington*, *Conlon*, *Pinnacle*, *Robust*, and *Tradition* were grown as check lines in each block. Field plots consisted of two 1 meter long rows per plot. Agronomic traits included yield, days to heading, height and lodging. Yield was measured as the weight of harvested grain from each plot and converted to kg/hectare. Yield was not measured in SP09 due to spatial irregularity of field conditions. Days to heading was measured as the number of days from planting to the date when 50% of the row exhibited 50% spike emergence from the leaf sheath. Height was measured at maturity, from the base of a plant representative of the row, to the top of the spike, excluding the awns. Height was not measured in CR11. Lodging was scored, when environmental conditions allowed (SP10 and SP11), as a visual rating of stalk angle on a 1-9 scale where 1 is a completely vertical stalk and 9 is a horizontal stalk.

Malt quality trait measurement

A micromalt procedure was performed to characterize the malting quality characteristics of grain harvested from each of the five field trials. Seed was bulked across the three replicates of each line in each environment before analyzing. A pilot steep of 10g samples was used to calculate the time needed to increase the moisture content of the barley grain to 45%. Next, 80g samples were steeped in a 16°C water bath, with a 2h rest every 24 hours. At the end of steeping, samples were spread over paper towels to remove surface moisture. Samples were allowed to germinate in a 16C chamber with 85% relative humidity for four days. Samples were maintained at 45% moisture by weight and turned by hand to prevent matting. Kilning was performed using the schedule 3 temperature cycling described by Karababa et al. (1993).

To assess malt quality, α-amylase activity (DU), β-glucan content, protein content (%), diastatic power (°ASBC), free amino nitrogen, fine grind malt extract (%), soluble protein content (%), the ratio of soluble to total protein content, and wort viscosity (cP) were analyzed using standard methods described by Karababa et al. (1993).

Trait and QTL analysis

To obtain a single multi-environment reading for each trait, a multivariate mixed model was fit in ASReml-R (VSN International), and best linear unbiased predictors (BLUPs) were obtained. Broad-sense heritability for each trait was calculated on a line mean basis across environments. Pearson correlation coefficients were calculated for each trait pair. QTL analysis was performed using the BC_sF_t functions of the R/qtl package in the R statistical software environment (Broman et al. 2003; Shannon et al.

2013). Marker-trait associations were identified using a significance threshold based on the 95th percentile of 1000 permutations for each trait.

Results

Agronomic performance

Agronomic traits: heading date, height, lodging, and yield were measured in 2-5 replicated field trials over 3 years. Analysis of variance indicated significant ($p < 0.0001$) variation is explained by lines, environments, replications, and genotype × environment interaction. Table 4.1 shows the distribution of trait values for the AB lines and Harrington within individual environments and across environments as calculated by best linear unbiased predictors (BLUP). Trait values for *OUH-602* were not measured due to the difficulties of growing wild barley germplasm in field trials. The distribution of each trait showed a slight bimodal distribution, with the primary mode including the *Harrington* phenotype. Transgressive segregation was observed for all four agronomic traits, with 38.8% of the AB lines exhibiting equal to or greater yield than Harrington. Heritability calculated on a line-mean basis was high for each agronomic trait, ranging from 0.79 to 0.97 (Table 4.1).

Malt performance

Malting quality traits were measured using a micro-malting procedure on bulked samples of three replicates from each of five environments. One line, HOUH_A182 was removed as an outlier from the analysis of β-glucan and dynamic viscosity due to its high values for these traits in each environment. Analysis of variance indicated significant (p

< 0.0001) variation is explained by lines and environments. Heritability of malting quality traits ranged from 0.33 for protein to 0.90 for malt extract (Table 4.1). Transgressive segregation was identified for traits: α-amylase activity, barley grain protein, diastatic power, FAN, soluble protein, and dynamic viscosity, but not for β-glucan (with the exception of line HOUH_A182), malt extract, and the ratio of soluble to total protein. The *Harrington* phenotype was on the low end of the β-glucan distribution and on the high end of the malt extract and ST ratio distributions (Table 4.1).

Trait correlations identified

Correlations between agronomic traits, and malting quality traits were identified in the HOUH AB population (Table 4.2). Heading date was positively correlated with height, and both of these traits were negatively correlated with lodging. In this population, those lines which flowered earlier and/or were shorter statured tended to exhibit more severe lodging. These shorter statured plants appeared to have much weaker stalk strength, likely contributing to lodging susceptibility. Yield was not significantly correlated with any of the other measured agronomic traits. Each of the agronomic traits was also significantly correlated with one or more malting quality traits. Heading date was positively correlated with α-amylase activity and negatively correlated with viscosity. Height was positively correlated with grain protein and negatively correlated with malt extract, and yield was positively correlated with malt extract, and α-amylase activity, and negatively correlated with grain protein.

Nineteen correlations were detected among malting quality characteristics (Table 4.2). The ST ratio was significantly correlated with all malt quality traits except diastatic

power. All pairwise correlations between α-amylase activity, FAN, malt extract, soluble protein, and ST ratio were significant with the exception of malt extract with soluble protein. Besides this group of positive correlations, malt extract had a strong negative correlation with grain protein and a negative correlation with diastatic power. Grain protein was negatively correlated with the ST ratio and positively correlated with diastatic power. Viscosity was strongly correlated with β-glucan, which in turn was negatively correlated with α-amylase activity, FAN, soluble protein, and ST ratio.

Genotypic characterization of BC_2 and BC_3 derived lines

The 98 HOUH $BC_2F_{6:8}$ advanced backcross population lines, along with an additional set of 26 BC_3F_2 lines derived from crossing 26 of the HOUH lines to *Harrington*, were genotyped with a 384-SNP Illumina VeraCode assay. The platform was developed to distinguish a set of 25 wild barley accessions from the 6-rowed malting barley cultivar *Rasmusson*. This assay contained 162 markers that were segregating between *OUH-602* and *Harrington*. The markers were used to characterize the introgressions of the BC_3F_2 lines, and to improve the marker coverage over the previous 111 SSR markers used to genotype the HOUH population (Yun et al. 2006). Markers were distributed throughout the genome, with an average distance of 7.4 cM between markers (Supplemental Table 4.1). The lowest density marker coverage was on chromosome 1H, which has a gap of 38 cM in the central region of the chromosome (Figure 4.2).

Significant deviations from the expected 12.5% *OUH-602* SNP allele frequencies were not observed in the HOUH population (Supplemental Figure 4.1). The percent of

the *OUH-602* genome introgressed in each HOUH line ranged from 2.3% to 30.2% with a mean of 12.2%. The BC_3 lines ranged from 0% to 17.8% introgression with a mean of 8.0%. Collectively, the set of 98 HOUH lines contains full coverage of the *OUH-602* genome. Supplemental Figure 4.2 shows the graphical genotypes of all 98 HOUH lines (blue) and 26 BC_3F_2 lines (red), sorted based on percent *OUH-602* introgression. We identified a set of 35 lines, comprised of 29 BC_2-derived lines and 6 BC_3-derived lines that collectively contain homozygous introgressions spanning the entire *OUH-602* genome (Figure 4.1). Selected lines were chosen to minimize the percent introgression per line, and range from 3.1% to 17.3% introgression with an average of 8.9% *OUH-602* introgression.

Quantitative trait loci detected

Between one and four QTL were identified for each of the four agronomic traits and nine malting quality traits (Table 4.3). Of the 36 QTL identified, nine were identified in all measured environments as well as the multi-environment BLUPs. Nine QTL were identified in only a single environment, and three were identified only with the multi-environment BLUPs. Percentage of variation explained for individual QTL ranged from 5.7-70.2% and QTL were identified on every chromosome except 4H (Figure 4.2).

Heading date was largely explained by a single QTL on chromosome 2HS. This QTL explained 70.2% of the variation for heading date, with an *OUH-602* genotypic effect that reduced days to heading by 10.5 days. In a single environment, an additional QTL on chromosome 2HL reduced days to heading by an additional 3.3 days. Height and lodging were also strongly influenced by the 2HS QTL, where the *OUH-602*

genotype increased lodging severity by 1.7 and decreased height by 7.4 cm. QTL that increased height in individuals containing the wild barley genotype were identified on chromosomes 1H, 3H, and 7H. All three yield QTL identified had a negative effect from the *OUH-602* allele, particularly those located on chromosome 1H and 3H, explaining 31.0 and 13.9% of the yield variation across all environments. An additional yield QTL on chromosome 6H explained 17.4% of the variation in the St. Paul field trials.

Diastatic power, FAN, and soluble protein each had QTL with positive and negative effect of the wild barley alleles. For each of the malt extract, α-amylase activity and ST ratio QTL, the wild alleles had negative trait effects, and for each of the β-glucan, grain protein, and viscosity QTL, the wild alleles had positive trait effects (Table 4.3).

Several areas of the genome contain overlapping QTL intervals (Figure 4.2). Yield and malt extract QTL share a peak at 108.52 cM on chromosome 1H and have overlapping QTL intervals with diastatic power, height, and grain protein. Besides the region of 2H influencing heading date, height, and lodging, another region of 2H at 50.99 cM has overlapping QTL for diastatic power and soluble protein. On 3H, a locus at 26.71 cM influences α-amylase activity, grain protein, and malt extract, and farther down the chromosome QTL intervals overlap for yield, malt extract and height in the region 47.37 cM to 73.3 cM, although the peaks of these associations do not coincide. Chromosome 5H appears to have two clusters of QTL responsible for several malting quality characteristics. One at 67.66 cM influences α-amylase activity, FAN, soluble protein, and ST ratio, and another locus at the distal end of 5HL influences these same traits in addition to β-glucan. Lastly, QTL for β-glucan and dynamic viscosity are coincident on the distal end of chromosome 7HL.

Discussion

The *Harrington* × *OUH-602* advanced backcross population (HOUH) was developed to explore wild barley alleles in a modern cultivar (Yun et al. 2006). Because *OUH-602* contains resistance alleles for many agronomically-important disease traits, several seedling disease resistance traits were previously mapped in a *Harrington* × *OUH-602* RIL population and validated using the HOUH advanced backcross population (Yun et al., 2005; 2006). However, unlike disease traits, selection for malting quality and other agronomically desirable traits does not occur in the wild. Still, the genetic diversity present in wild barley populations provides an avenue for exploring natural variation for any trait of interest. The development of advanced backcross lines minimizes the presence of unadapted traits in individual HOUH lines, providing the opportunity to characterize exotic germplasm in standard field trials (Gyenis et al. 2007). By examining agronomic and malting quality traits in addition to the previously examined disease and morphological traits, we were able to construct a more complete picture of the genetic architecture of breeding-relevant traits that may be masked by the many deleterious traits found in unadapted germplasm. The results presented here indicate that some beneficial variation for yield-related and malting quality traits is present in the wild barley accession *OUH-602*, but for the most part we did not observe large improvement, particularly in valuable traits such as yield and malt extract.

QTL for agronomic and malting quality traits correspond with previous studies

In a previous analysis of the HOUH population, Gyenis et al. (2007) measured heading date and height as well as several morphological traits under field conditions in

St. Paul, MN and Cadriano, Italy and under greenhouse conditions in St. Paul. The QTL identified here correspond to those previously identified. Heading date QTL in both studies localize to two regions of chromosome 2H and height QTL localize to chromosomes 1H, 3H, and 7H with positive *OUH-602* effects and to chromosome 2H with negative effect. The SNP marker 12_30872 which tags the large effect heading date QTL, as well as the largest effect height and lodging QTL, is located in the *Ppd-H1* gene responsible for photoperiod sensitivity under long day conditions (Turner et al. 2005). The increased photoperiod sensitivity caused by the wild barley allele contributes to early flowering, shorter stature plants, and increased lodging susceptibility in the HOUH population. While generally, taller plants are found to have more severe lodging, many of the short statured plants in the HOUH population had thin stems and poor stalk strength. The *Ppd-H1* locus has been shown to have significant effects on heading date and other agronomic traits in several different wild x cultivated barley mapping populations (Li et al. 2005; von Korff et al. 2006), and further examinations have shown a wide range of epistatic and pleiotropic effects of wild barley *Ppd-H1* alleles (Wang et al. 2010; von Korff et al. 2010).

Yield was not measured in the previous Gyenis et al. (2007) study, but fragile rachis morphology was. The yield QTL identified here on chromosomes 1H and 3H correspond with the fragile rachis QTL that explained the most variation in the Gyenis et al. (2007) study. The locus on 3H corresponds to the brittle rachis locus (*Btr1/Btr2*), which is the primary domestication locus that controls seed shattering in barley (Azhaguvel and Komatsuda 2007). The yield QTL on 1H appears to correspond with *thresh-1* locus fine-mapped by Schmalenbach et al. (2011) using a set of wild barley

introgression lines. These results reflect that large effect, domestication-related loci are highly influential on the agronomic performance of wild × cultivated barley crosses. Because these two loci have such major effect on the yield measurement, identifying additional small effect QTL will require larger population sizes and/or alternative population designs that minimize these confounding effects even more than what was possible with the AB population.

Malting quality QTL analysis also revealed previously identified loci, found in a variety of mapping populations including wild barley as a parent (Matus et al. 2003; Li et al. 2005; von Korff et al. 2008), association mapping of U.S. barley breeding lines (Mohammadi et al. 2015) and biparental mapping (Mather et al. 1997; Marquez-Cedillo et al. 2000) populations. Malting quality trait QTL were often clustered in regions of the genome, indicating the likelihood of the pleiotropic effect of loci on multiple traits. As compared to the agronomic traits measured, there was a larger range of heritabilities observed across malting quality traits.

Utility of OUH-602 for breeding and future studies

The addition of SNP marker genotypes has increased the marker coverage in the HOUH population, allowing for higher resolution mapping than the previous 111 SSR marker data. Additionally, a set of 35 HOUH BC_2- and BC_3-derived lines were identified that contain complete coverage of the *OUH-602* genome in a *Harrington* genetic background. These lines are available for future validation studies.

The wild barley *OUH-602* did not contribute beneficial alleles for yield nor for a majority of the malting quality traits examined. Still, wild barley contributed positive

and negative alleles for three malting traits: diastatic power, free amino nitrogen, and soluble protein content, and transgressive segregation did exist for yield and many malting quality traits in the HOUH population. Understanding the genetic architecture for these traits in unadapted, diverse germplasm provides breeders with information that can aid in the introgression of valuable disease resistance traits while avoiding linkage drag of deleterious agronomic and malting characteristics.

Table 4.1 Trait values and heritabilities for agronomic and malting quality traits.

Trait	n	Mean	Min	Max	CV	Harr	h^2
Heading date (HD, days)[a]	98	57.2	47.8	61.9	15.1	59.6	0.97
CR10[b]	98	61.3	53.0	65.0	19.7	63.3	
CR11	98	55.6	47.0	61.0	14.1	58.3	
SP09	98	57.1	47.7	64.7	15.3	61.7	
SP10	98	56.8	46.0	62.7	12.6	58.0	
SP11	98	55.5	44.0	62.0	11.9	59.0	
Height (HT, cm)	98	78.1	66.9	87.3	19.0	75.1	0.84
CR10	98	83.2	66.0	98.0	12.6	80.3	
SP09	98	69.1	54.5	82.5	12.8	64.5	
SP10	98	84.1	72.3	95.3	17.4	81.5	
SP11	98	76.1	65.3	88.3	13.9	66.3	
Lodging (LG, 0-9)	98	5.9	4.2	8.0	8.3	5.5	0.79
SP10	98	5.6	3.7	8.0	5.9	5.5	
SP11	98	6.2	3.7	9.0	6.3	5.7	
Yield (YD, kg/ha)	98	5143	2638	6356	5.5	5677	0.87
CR10	98	6662	3566	8345	5.7	7349	
CR11	98	6037	2620	7818	5.8	5868	
SP10	98	4334	903	6554	3.2	5640	
SP11	98	3539	814	5159	3.6	3822	
α-amylose (AA, DU)	98	63.1	41.5	75.3	8.9	70.2	0.83
CR10	98	68.7	38.8	95.3	5.9	84.2	
CR11	98	60.2	37.5	81.8	6.0	70.9	
SP09	96	58.9	35.2	78.5	6.4	69.4	
SP10	98	60.9	32.0	79.0	6.7	69.1	
SP11	97	66.6	45.9	88.6	6.9	83.3	
β-glucan (BG)	97	378.6	208.5	676.4	3.5	223.3	0.79
CR10	97	358.0	103.0	781.0	2.4	337.0	
CR11	97	313.5	46.0	866.0	1.7	142.0	
SP09	95	438.0	88.0	681.0	2.6	106.0	
SP10	97	448.1	102.0	826.0	2.6	285.0	
SP11	96	332.3	76.0	850.0	1.9	165.0	
Barley protein (BP, %)	98	13.4	12.5	14.7	27.6	12.7	0.33
CR10	98	11.6	9.9	14.2	12.1	11.6	
CR11	98	11.8	9.8	14.7	12.6	10.7	
SP09	96	15.4	13.1	18.5	17.1	14.2	
SP10	98	14.2	12.7	16.3	19.0	13.4	
SP11	97	14.1	12.3	17.5	15.6	13.3	

Table 4.1 continued

Trait	n	Mean	Min	Max	CV	Harr	h2
Heading date (HD, days)[a]	98	57.2	47.8	61.9	15.1	59.6	0.97
CR10b	98	61.3	53.0	65.0	19.7	63.3	
CR11	98	55.6	47.0	61.0	14.1	58.3	
SP09	98	57.1	47.7	64.7	15.3	61.7	
SP10	98	56.8	46.0	62.7	12.6	58.0	
SP11	98	55.5	44.0	62.0	11.9	59.0	
Height (HT, cm)	98	78.1	66.9	87.3	19.0	75.1	0.84
CR10	98	83.2	66.0	98.0	12.6	80.3	
SP09	98	69.1	54.5	82.5	12.8	64.5	
SP10	98	84.1	72.3	95.3	17.4	81.5	
SP11	98	76.1	65.3	88.3	13.9	66.3	
Lodging (LG, 0-9)	98	5.9	4.2	8.0	8.3	5.5	0.79
SP10	98	5.6	3.7	8.0	5.9	5.5	
SP11	98	6.2	3.7	9.0	6.3	5.7	
Yield (YD, kg/ha)	98	5143	2638	6356	5.5	5677	0.87
CR10	98	6662	3566	8345	5.7	7349	
CR11	98	6037	2620	7818	5.8	5868	
SP10	98	4334	903	6554	3.2	5640	
SP11	98	3539	814	5159	3.6	3822	
α-amylose (AA, DU)	98	63.1	41.5	75.3	8.9	70.2	0.83
CR10	98	68.7	38.8	95.3	5.9	84.2	
CR11	98	60.2	37.5	81.8	6.0	70.9	
SP09	96	58.9	35.2	78.5	6.4	69.4	
SP10	98	60.9	32.0	79.0	6.7	69.1	
SP11	97	66.6	45.9	88.6	6.9	83.3	
β-glucan (BG)	97	378.6	208.5	676.4	3.5	223.3	0.79
CR10	97	358.0	103.0	781.0	2.4	337.0	
CR11	97	313.5	46.0	866.0	1.7	142.0	
SP09	95	438.0	88.0	681.0	2.6	106.0	
SP10	97	448.1	102.0	826.0	2.6	285.0	
SP11	96	332.3	76.0	850.0	1.9	165.0	
Barley protein (BP, %)	98	13.4	12.5	14.7	27.6	12.7	0.33
CR10	98	11.6	9.9	14.2	12.1	11.6	
CR11	98	11.8	9.8	14.7	12.6	10.7	
SP09	96	15.4	13.1	18.5	17.1	14.2	
SP10	98	14.2	12.7	16.3	19.0	13.4	
SP11	97	14.1	12.3	17.5	15.6	13.3	

Table 4.1 continued

Trait	n	Mean[a]	Min	Max	CV	Harr	h^2
Diastatic power (DP, °ASBC)	98	124.5	103.2	157.0	11.5	124.2	0.71
CR10	98	119.4	79.0	193.8	5.7	106.1	
CR11	98	118.1	74.9	169.8	5.8	112.0	
SP09	96	136.2	97.0	199.5	6.9	161.3	
SP10	98	131.0	88.2	204.5	6.1	126.9	
SP11	97	118.0	85.1	173.6	7.1	110.3	
Free amino nitrogen (FAN)	98	270.6	172.6	320.3	9.1	297.3	0.71
CR10	98	247.5	136.0	323.0	7.3	277.0	
CR11	98	240.6	124.0	353.0	5.4	388.0	
SP09	96	254.5	143.0	354.0	6.2	288.0	
SP10	98	323.8	194.0	392.0	7.7	349.0	
SP11	97	286.8	159.0	449.0	4.9	358.0	
Fine grind malt extract (ME, %)	98	77.6	72.0	79.9	46.3	79.9	0.90
CR10	98	78.7	72.4	81.2	42.6	79.8	
CR11	98	78.8	73.4	81.8	45.2	81.3	
SP09	96	77.1	71.5	80.4	45.0	79.9	
SP10	98	76.4	69.8	78.6	40.8	79.0	
SP11	97	76.9	68.7	79.6	41.5	79.5	
Soluble protein (SP, %)	98	6.0	4.3	7.2	9.6	6.7	0.77
CR10	98	5.7	3.7	7.1	8.6	6.6	
CR11	98	5.2	3.7	6.8	7.4	5.6	
SP09	96	5.8	3.9	7.6	7.3	6.6	
SP10	98	6.5	4.7	7.9	10.0	7.0	
SP11	97	6.9	4.4	9.2	6.5	8.3	
Soluble/total protein (ST)	98	45.2	29.9	54.3	9.3	53.3	0.79
CR10	98	49.4	25.9	59.0	7.8	56.6	
CR11	98	44.4	27.0	58.2	7.8	52.1	
SP09	96	37.5	24.4	50.0	7.6	46.4	
SP10	98	45.7	30.4	55.9	9.5	52.0	
SP11	97	49.1	33.8	68.1	6.2	62.2	
Viscosity (VIS, cP)	97	1.52	1.47	1.61	53.5	1.49	0.56
CR10	97	1.47	1.41	1.63	32.1	1.46	
CR11	97	1.50	1.43	1.68	28.3	1.48	
SP09	95	1.58	1.45	1.89	16.8	1.45	
SP10	97	1.50	1.41	1.68	31.6	1.45	
SP11	96	1.53	1.43	1.92	19.1	1.55	

[a]Values calculated for multi-environment best linear unbiased predictors
[b]Values calculated for individual environments: Crookston, MN 2010 and 2011 (CR10, CR11) and St. Paul, MN 2009-2011 (SP09, SP10, SP11).

Table 4.2 Pearson correlation coefficients between agronomic and malting quality traits.

	HD	HT	LG	YD	AA	BG	BP	DP	FAN	ME	SP	ST	VIS
HD													
HT	0.39***												
LG	-0.75***	-0.36**											
YD	ns	ns	ns										
AA	0.38**	ns	ns	0.28*									
BG	ns	ns	ns	ns	-0.33**								
BP	ns	0.28*	ns	-0.49***	ns	ns							
DP	ns	ns	ns	ns	ns	ns	0.52***						
FAN	ns	ns	ns	ns	0.58***	-0.39***	ns	ns					
ME	ns	-0.34**	ns	0.69***	0.46***	ns	-0.70***	-0.41***	0.29*				
SP	ns	ns	ns	ns	0.54***	-0.42***	ns	ns	0.90***	ns			
ST	ns	ns	ns	ns	0.63***	-0.52***	-0.28*	ns	0.88***	0.51***	0.92***		
VIS	-0.36**	ns	0.33*	ns	ns	0.79***	ns	ns	ns	ns	ns	-0.28*	

*p < 0.01, **p < 0.001, ***p < 0.0001

Table 4.3 Summary of significant QTL for agronomic and malting quality traits.

Trait	Marker[a]	Chr	Pos[b]	Range[c]	LOD[d]	R^{2} [e]	Environments[f]	Harr[g]	OUH[h]	RP[i]
HD	12_30872	2H	25.33	9.12-27.50	28.39	70.2	All[j]	58.5	49.0	-0.16
	11_11072	2H	77.76	57.29-94.25	3.26	14.2	*CR10	61.8	58.5	-0.05
HT	11_11037	1H	86.25	48.29-108.52	3.18	8.4	SP10	77.5	81.8	0.05
	12_30872	2H	25.33	9.12-27.50	9.59	28.8	All	79.1	71.7	-0.09
	11_21533	3H	47.37	26.71-82.62	3.2	5.1	CR10	77.5	81.7	0.05
	11_20485	7H	91.67	79.08-116.28	3.36	14.1	*Cr10	82.1	89.0	0.08
LG	12_30872	2H	25.33	9.12-27.5	21.27	63.3	All	5.7	7.4	0.30
YD	11_20780	1H	108.52	89.77-117.24	13.76	31.0	All	5372	3326	-0.38
	12_30467	3H	53.86	47.37-73.30	7.75	13.9	All	5327	3688	-0.31
	11_20379	6H	115.31	98.68-127.76	4.06	17.4	*SP10, SP11	4552	2905	-0.36
AA	11_10559	3H	26.71	14.00-32.92	2.95	5.7	CR10, SP09	64.2	56.5	-0.12
	11_21133	5H	67.66	56.77-79.68	3.3	14.5	*SP11	68.0	56.9	-0.16
	12_31123	5H	189.89	167.40-189.89	4.1	15.3	CR10, CR11, SP09	64.3	56.4	-0.12
	11_11147	6H	95.7	78.52-115.31	6.23	23.0	CR10, CR11, SP09, SP10	64.5	54.7	-0.15
BG	12_31123	5H	189.89	167.40-189.89	5.17	21.8	*SP11	302.7	560.0	0.85
	12_30826	7H	162.03	152.41-162.03	8.66	33.7	CR10, CR11, SP10, SP11	355.2	544.1	0.53
BP	11_10586	1H	123.1	117.24-133.47	5.53	10.6	CR10, SP10, SP11	13.3	14.1	0.06
	11_10559	3H	26.71	14.00-32.92	3.91	10.2		13.3	13.9	0.04
	12_30361	6H	43.38	20.62-78.52	3.17	13.9	*CR10	11.5	12.8	0.11
	11_10303	7H	89.78	64.98-105.07	3.64	11.1	CR10	13.3	13.9	0.04

Table 4.3 continued

Trait	Marker[a]	Chr	Pos[b]	Range[c]	LOD[d]	R^{2} [e]	Environments[f]	*Harr*[g]	*OUH*[h]	RP[i]
DP	12_10420	1H	0.54	0.54-22.04	3.68	7.8	SP10	126.0	112.4	-0.11
	11_10903	1H	133.47	89.77-133.47	3.72	10.4		123.1	137.3	0.12
	12_30703	2H	51	47.48-57.29	3.83	8.1		126.2	113.2	-0.10
FAN	11_21133	5H	67.66	33.55-71.93	3.85	16.5	*CR10	252.6	210.6	-0.17
	12_31123	5H	189.89	184.99-189.89	9.04	34.6	All	277.5	227.9	-0.18
	12_30368	7H	118.52	116.28-121.36	2.8	12.5	*SP11	282.0	308.9	0.10
ME	11_20780	1H	108.52	89.77-117.24	12.59	34.2	All	78.0	74.4	-0.05
	11_10559	3H	26.71	14.00-32.92	3.97	6.8	SP09	77.8	75.8	-0.03
	12_31281	3H	67.86	47.37-73.29	5.2	21.7	*SP10	76.7	73.7	-0.04
SP	12_30703	2H	51	47.48-57.29	3.81	16.5	*CR10	5.83	5.04	-0.14
	11_21133	5H	67.66	33.55-71.93	3.19	9.3	CR10, CR11	6.10	5.39	-0.12
	12_31123	5H	189.89	184.99-189.89	7.82	26.4	All	6.15	5.17	-0.16
	12_10605	7H	58.53	41.51-79.08	3.71	8.6	CR10, CR11, SP10, SP11	5.90	6.58	0.11
ST	11_21133	5H	67.66	56.77-71.93	3.23	13.0	CR10	45.9	40.3	-0.12
	12_31123	5H	189.89	184.99-189.89	7.69	29.2	All	46.3	38.7	-0.16
VIS	12_30826	7H	162.03	152.41-162.03	8.77	34.1	CR11, SP10	1.51	1.56	0.03

Table 4.3 continued

[a] Marker with highest significance
[b] cM position of marker with highest significance (Muñoz-Amatriaín et al. 2011)
[c] cM interval over which a 1.5 LOD decrease occurs on either side of the peak
[d] Likelihood odds ratio of peak marker as calculated by R/qtl BC_sF_t mapping function (Shannon et al. 2013) using multi-environment BLUP values
[e] Variation explained by peak marker in an ANOVA of a linear model containing each peak marker for the trait
[f] QTL identified in analysis of individual environment data in addition to BLUP, those marked with an * were only identified in individual environments, and not identified in the multi-environment analysis
[g] Mean phenotypic value of individuals containing the homozygous *Harrington* genotype at the given loci
[h] Mean phenotypic value of individuals containing the homozygous *OUH-602* genotype at the given loci
[i] Relative performance of the *OUH-602* genotype: (*Harr* – *OUH*)/*Harr* (von Korff et al. 2005)
[j] QTL identified in all measured environments in addition to the multi-environment analysis

Figure 4.1 Graphical genotype of 35 HOUH lines that collectively contain homozygous introgressions of the entire *OUH-602* genome. Darker regions are homozygous *OUH-602*, while lighter regions indicate heterozygosity or uncertainty in recombination breakpoint. Lines marked with a * are BC_3-derived lines.

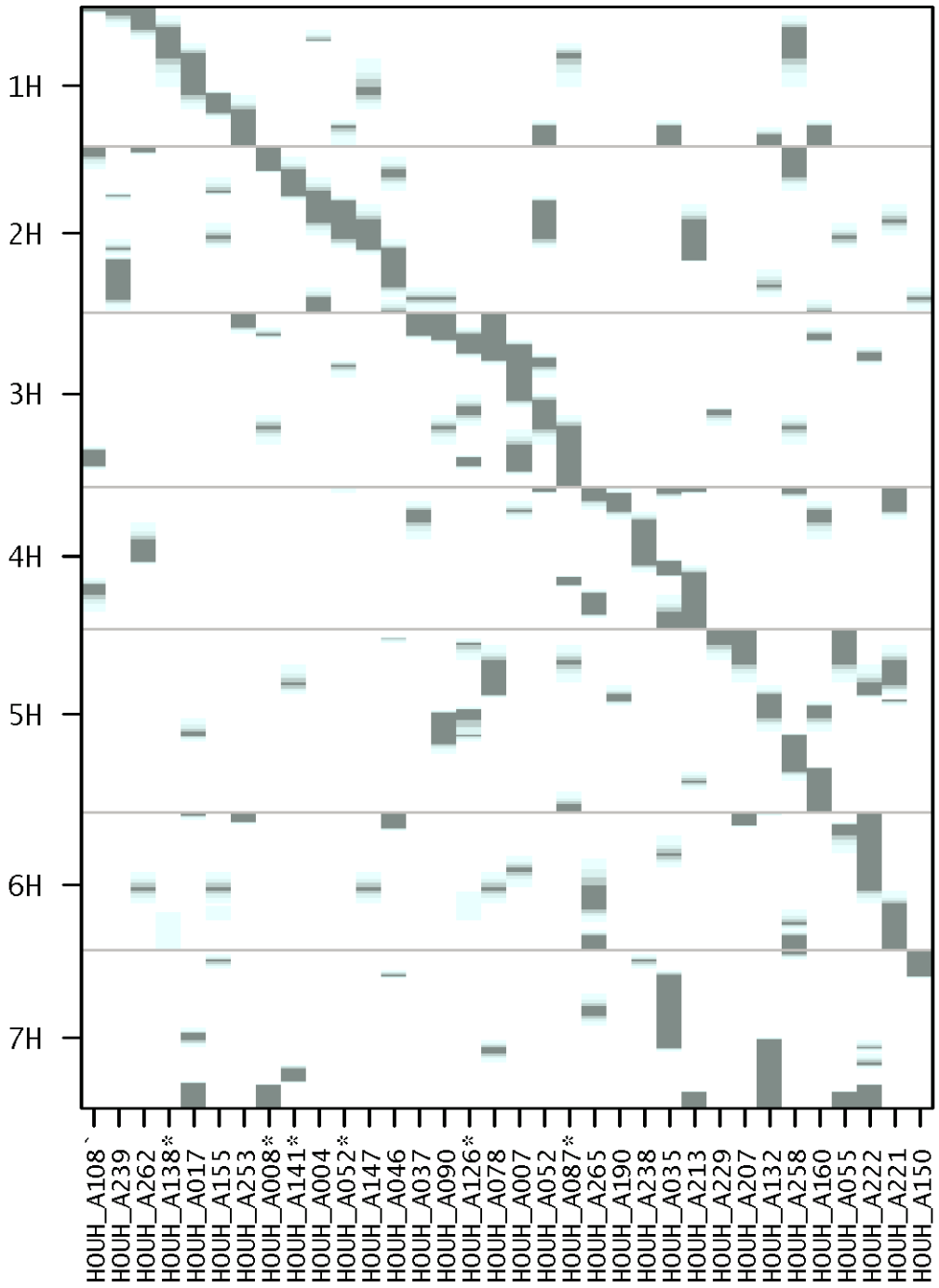
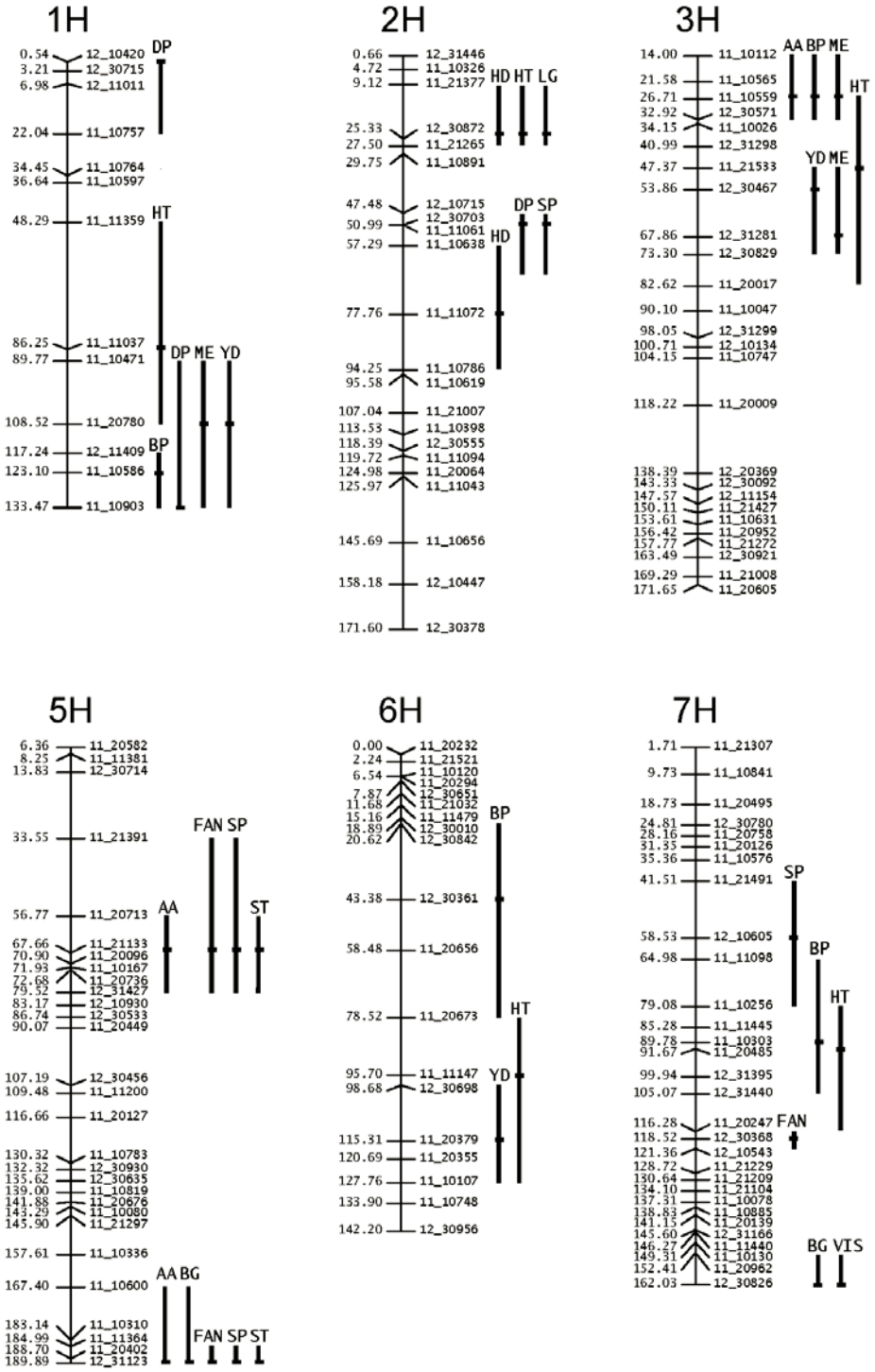


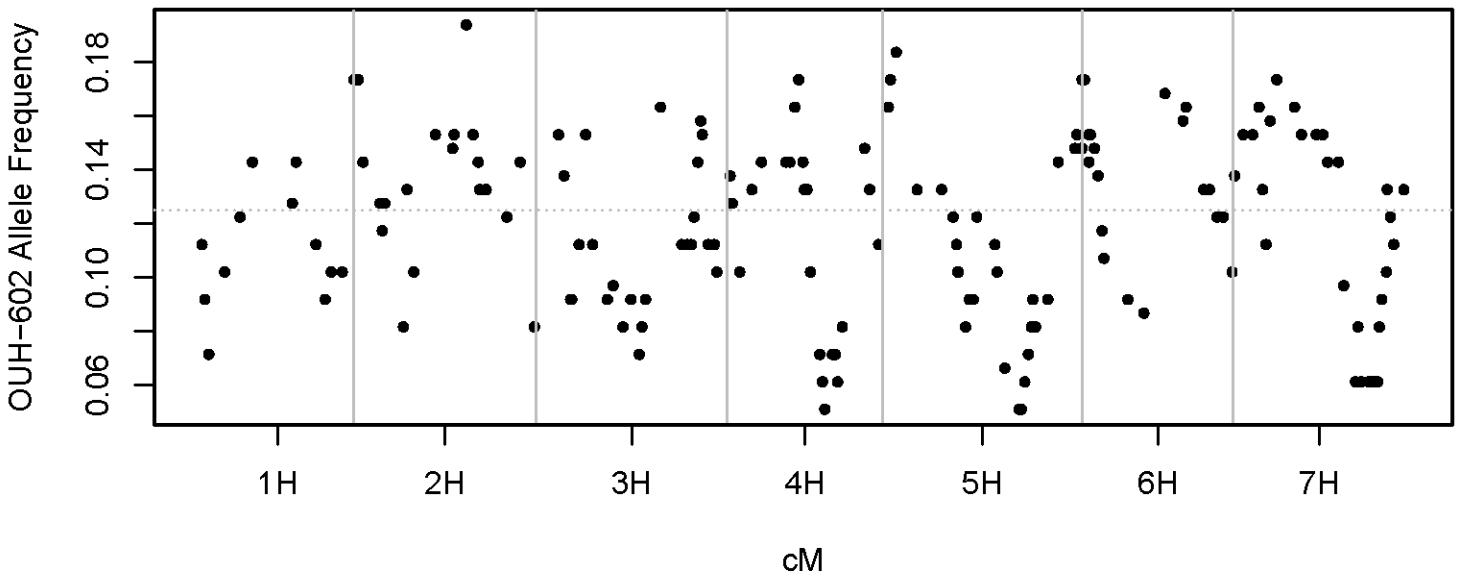
Figure 4.2 Genetic map positions (Muñoz-Amatriaín et al. 2011) of QTL intervals as indicated by vertical lines. Cross hatches on QTL intervals indicate the location of peak significance. No QTL were identified on chromosome 4H (not shown).



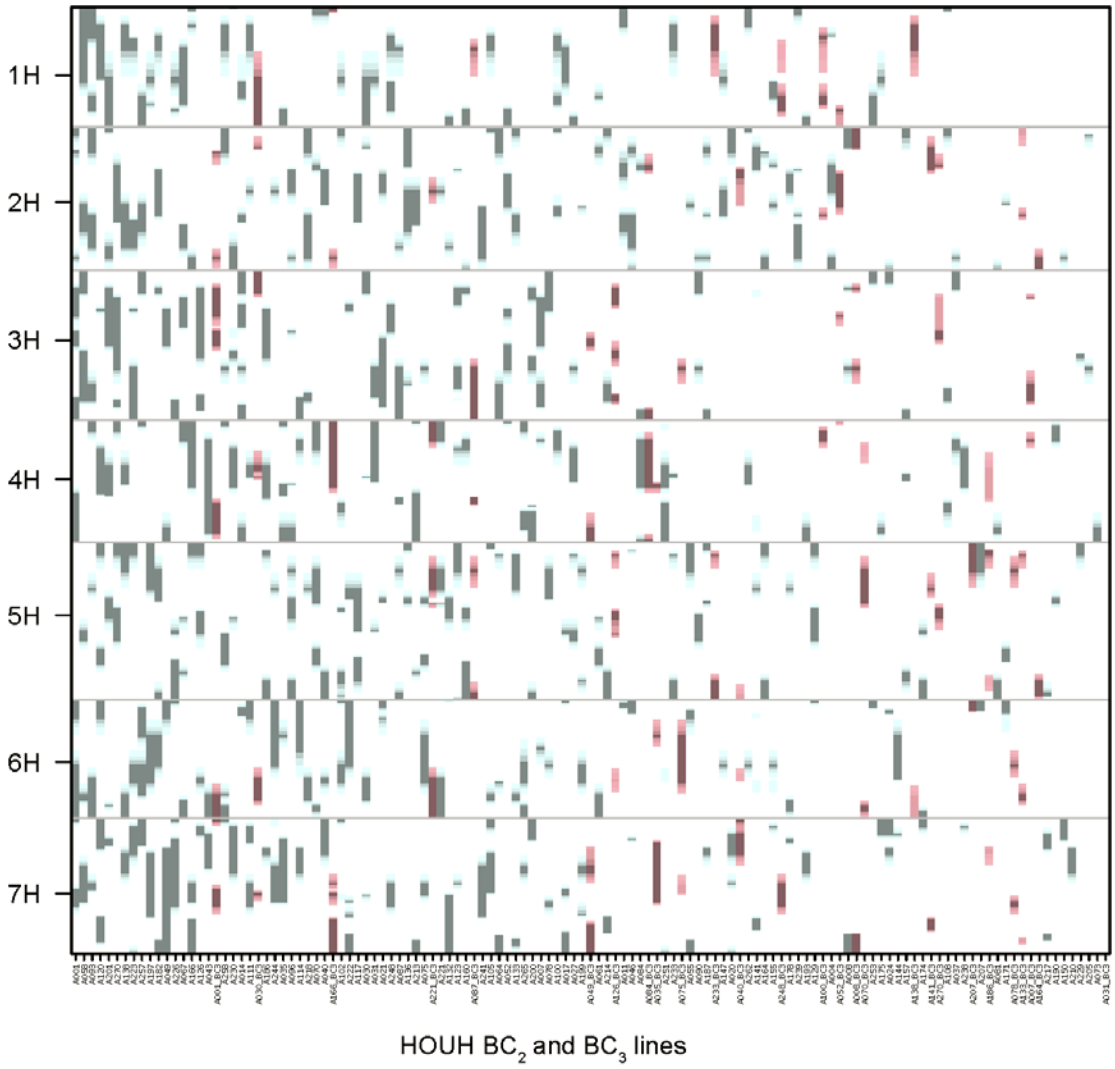
Supplemental Table 4.1 Marker coverage throughout the genome

Chromosome	N markers	Mean distance between markers (cM)	Max gap between markers (cM)
1H	13	11.1	38.0
2H	22	8.1	20.5
3H	26	6.3	20.2
4H	23	6.4	23.3
5H	29	6.6	23.2
6H	19	7.9	22.8
7H	30	5.5	17.0
Mean	23	7.4	23.6
Max	30	11.1	38.0
Min	13	5.5	17.0

Supplemental Figure 4.2 Allele frequency of 162 SNP markers in the *Harrington* x *OUH-602* advanced backcross population. Horizontal dashed line indicates expected allele frequency = 0.125 for BC_2 individuals.



Supplemental Figure 4.3 Graphical genotype of the 98 HOUH lines and 26 BC_3 lines. Lines are arranged in order from left to right by decreasing percent *OUH-602* introgression. HOUH BC_2 lines have introgressed regions highlighted in blue, and BC_3 lines have introgressed regions in red. Dark colors indicate homozygous *OUH-602* alleles. Light colors indicate heterozygosity or uncertainty in recombination breakpoint.



Bibliography

Abdel-Ghani AH, Parzies HK, Omary A, Geiger HH (2004) Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. Theor Appl Genet 109:588–595. doi: 10.1007/s00122-004-1657-1

Alqudah AM, Sharma R, Pasam RK, et al (2014) Genetic Dissection of Photoperiod Response Based on GWAS of Pre-Anthesis Phase Duration in Spring Barley. PLoS ONE 9:e113120. doi: 10.1371/journal.pone.0113120

Alsop BP, Farre A, Wenzl P, et al (2011) Development of wild barley-derived DArT markers and their integration into a barley consensus map. Mol Breed 27:77–92. doi: 10.1007/s11032-010-9415-3

Ames N, Dreiseitl A, Steffenson BJ, Muehlbauer GJ (2015) Mining wild barley for powdery mildew resistance. Plant Pathol n/a–n/a. doi: 10.1111/ppa.12384

Azhaguvel P, Komatsuda T (2007) A Phylogenetic Analysis Based on Nucleotide Sequence of a Marker Linked to the Brittle Rachis Locus Indicates a Diphyletic Origin of Barley. Ann Bot 100:1009–1015. doi: 10.1093/aob/mcm129

Badr A, M K, Sch R, et al (2000) On the Origin and Domestication History of Barley (Hordeum vulgare). Mol Biol Evol 17:499–510.

Bandillo N, Raghavan C, Muyco PA, et al (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6:1–15. doi: 10.1186/1939-8433-6-11

Bauer E, Falque M, Walter H, et al (2013) Intraspecific variation of recombination rate in maize. Genome Biol 14:R103. doi: 10.1186/gb-2013-14-9-r103

Baum M, Grando S, Backes G, et al (2003) QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross “Arta” × H. spontaneum 41-1. Theor Appl Genet 107:1215–1225. doi: 10.1007/s00122-003-1357-2

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49:419–425. doi:10.2135/cropsci2008.08.0452

Blake T, Blake VC, Bowman JGP, Abdel - Haleem H (2010) Barley Feed Uses and Quality Improvement. In: Ullrich SE (ed) Barley. Wiley-Blackwell, pp 522–531

Bockelman HE, Valkoun J (2010) Barley Germplasm Conservation and Resources. In: Ullrich SE (ed) Barley. Wiley-Blackwell, pp 144–159

Bothmer R von (1995) An Ecogeographical Study of the Genus Hordeum. IPGRI

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinforma Oxf Engl 19:889–890. doi: 10.1093/bioinformatics/btg112

Brown TA, Jones MK, Powell W, Allaby RG (2009) The complex origins of domesticated crops in the Fertile Crescent. Trends Ecol Evol 24:103–109. doi: 10.1016/j.tree.2008.09.008

Buckler ES, Holland JB, Bradbury PJ, et al (2009) The Genetic Architecture of Maize Flowering Time. Science 325:714–718. doi: 10.1126/science.1174276

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme Population-Dependent Linkage Disequilibrium Detected in an Inbreeding Plant Species, Hordeum vulgare. Genetics 172:557–567. doi: 10.1534/genetics.104.038489

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215–221. doi: 10.1016/j.pbi.2008.01.002

Choo TM, Vigier B, Shen QQ, et al (2004) Barley Traits Associated with Resistance to Fusarium Head Blight and Deoxynivalenol Accumulation. Phytopathology 94:1145–1150. doi: 10.1094/PHYTO.2004.94.10.1145

Churchill GA, Airey DC, Allayee H, et al (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 36:1133–1137. doi: 10.1038/ng1104-1133

Close TJ, Bhat PR, Lonardi S, et al (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582. doi: 10.1186/1471-2164-10-582

Cockram J, Jones H, Leigh FJ, et al (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J Exp Bot 58:1231–1244. doi: 10.1093/jxb/erm042

Cockram J, White J, Zuluaga DL, et al (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci 107:21611–21616. doi: 10.1073/pnas.1010179107

Comadran J, Kilian B, Russell J, et al (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. doi: 10.1038/ng.2447

Comadran J, Russell JR, Booth A, et al (2011) Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in Hordeum vulgare in Mediterranean environments. Theor Appl Genet 122:1363–1373. doi: 10.1007/s00122-011-1537-4

Combs E, Bernardo R (2013) Genomewide Selection to Introgress Semidwarf Maize Germplasm into U.S. Corn Belt Inbreds. Crop Sci 53:1427-1436. doi: 10.2135/cropsci2012.11.0666

Condón F, Gustus C, Rasmusson DC, Smith KP (2008) Effect of Advanced Cycle Breeding on Genetic Diversity in Barley Breeding Germplasm. Crop Sci 48:1027-1036. doi: 10.2135/cropsci2007.07.0415

Condón F, Rasmusson DC, Schiefelbein E, et al (2009) Effect of Advanced Cycle Breeding on Genetic Gain and Phenotypic Diversity in Barley Breeding Germplasm. Crop Science 49:1751-1761. doi: 10.2135/cropsci2008.10.0585

Consortium TIBGS (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716. doi: 10.1038/nature11543

Distelfeld A, Korol A, Dubcovsky J, et al (2008) Colinearity between the barley grain protein content (GPC) QTL on chromosome arm 6HS and the wheat Gpc-B1 region. Mol Breed 22:25–38. doi: 10.1007/s11032-007-9153-3

Doebley JF, Gaut BS, Smith BD (2006) The Molecular Genetics of Crop Domestication. Cell 127:1309–1321. doi: 10.1016/j.cell.2006.12.006

Druka A, Franckowiak J, Lundqvist U, et al (2011) Genetic Dissection of Barley Morphology and Development. Plant Physiol 155:617–627. doi: 10.1104/pp.110.166249

Ellis RP, Forster BP, Robinson D, et al (2000) Wild barley: a source of genes for crop improvement in the 21st century? J Exp Bot 51:9–17. doi: 10.1093/jexbot/51.342.9

Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome J 4:250-255. doi: 10.3835/plantgenome2011.08.0024

Eshghi R, Salayeva S, Ebrahimpour F, et al (2013) Advanced-backcross QTL analysis in hulless barley: I. Detection of exotic alleles for yield and yield components introgressed from Hordeum vulgare ssp. spontaneum. Intl J Agri Crop Sci 5:95-100

Fang Z, Eule-Nashoba A, Powers C, et al (2013) Comparative Analyses Identify the Contributions of Exotic Donors to Disease Resistance in a Barley Experimental Population. G3 Genes Genomes Genet 3:1945–1953. doi: 10.1534/g3.113.007294

Fang Z, Gonzales AM, Clegg MT, et al (2014) Two Genomic Regions Contribute Disproportionately to Geographic Differentiation in Wild Barley. G3 Genes Genomes Genet 4:1193-1203. doi: 10.1534/g3.114.010561

Fetch TG, Steffenson BJ, Nevo E (2003) Diversity and Sources of Multiple Disease Resistance in Hordeum spontaneum. Plant Dis 87:1439–1448. doi: 10.1094/PDIS.2003.87.12.1439

Fischbeck G (2003) Chapter 3 Diversification through breeding. In: Roland von Bothmer T van H, Helmut Knüpffer and Kazuhiro Sato (ed) Developments in Plant Genetics and Breeding. Elsevier, pp 29–52

Fox GP, Panozzo JF, Li CD, et al (2003) Molecular basis of barley quality. Aust J Agric Res 54:1081–1101. doi:10.1071/AR02237

Fu Y-B, Somers DJ (2009) Genome-Wide Reduction of Genetic Diversity in Wheat Breeding. Crop Sci 49:161-168. doi: 10.2135/cropsci2008.03.0125

Guo B, Beavis WD (2011) In silico genotyping of the maize nested association mapping population. Mol Breed 27:107–113. doi: 10.1007/s11032-010-9503-4

Guo B, Wang D, Guo Z, Beavis WD (2013) Family-based association mapping in crop species. Theor Appl Genet 126:1419–1430. doi: 10.1007/s00122-013-2100-2

Gyenis L, Yun SJ, Smith KP, et al (2007) Genetic architecture of quantitative trait loci associated with morphological and agronomic trait differences in a wild by cultivated barley cross. Genome 50:714–723. doi: 10.1139/g07-054

Hamblin MT, Close TJ, Bhat PR, et al (2010) Population Structure and Linkage Disequilibrium in U.S. Barley Germplasm: Implications for Association Mapping. Crop Sci 50:556-566. doi: 10.2135/cropsci2009.04.0198

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical Comparison of Simple Sequence Repeats and Single Nucleotide Polymorphisms in Assessment of Maize Diversity and Relatedness. PLoS ONE 2:e1367. doi: 10.1371/journal.pone.0001367

Hoffmann A, Maurer A, Pillen K (2012) Detection of nitrogen deficiency QTL in juvenile wild barley introgression lines growing in a hydroponic system. BMC Genet 13:88. doi: 10.1186/1471-2156-13-88

Huang X, Paulo M-J, Boer M, et al (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc Natl Acad Sci 108:4488–4493. doi: 10.1073/pnas.1100465108

Hussien A, Takavol E, Horner DS, et al (2014) Genetics of tillering in rice and barley. Plant Genome 7:1-20. doi:10.3835/plantgenome2013.10.0032

Jia Q, Zhang J, Westcott S, et al (2009) GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. Funct Integr Genomics 9:255–262. doi: 10.1007/s10142-009-0120-4

Jilal A, Grando S, Henry RJ, et al (2008) Genetic diversity of ICARDA's worldwide barley landrace collection. Genet Resour Crop Evol 55:1221–1230. doi: 10.1007/s10722-008-9322-1

Jones H, Leigh FJ, Mackay I, et al (2008) Population-Based Resequencing Reveals That the Flowering Time Adaptation of Cultivated Barley Originated East of the Fertile Crescent. Mol Biol Evol 25:2211–2219. doi: 10.1093/molbev/msn167

Jordan DR, Mace ES, Cruickshank AW, et al (2011) Exploring and Exploiting Genetic Variation from Unadapted Sorghum Germplasm in a Breeding Program. Crop Sci 51:1444-1457. doi: 10.2135/cropsci2010.06.0326

Kalladan R, Worch S, Rolletschek H, et al (2013) Identification of quantitative trait loci contributing to yield and seed quality parameters under terminal drought in barley advanced backcross lines. Mol Breed 32:71–90. doi: 10.1007/s11032-013-9853-9

Kang HM, Sul JH, Service SK, et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354. doi: 10.1038/ng.548

Karababa E, Schwarz P, Horsley R (1993) Effect of Kiln Schedule on Micromalt Quality Parameters. J Am Soc Brew Chem.51:163-167. doi: 10.1094/ASBCJ-51-0163

Karsai I, Mészáros K, Szűcs P, et al (2006) The influence of photoperiod on the Vrn-H2 locus (4H) which is a major determinant of plant development and reproductive fitness traits in a facultative × winter barley (Hordeum vulgare L.) mapping population. Plant Breed 125:468–472. doi: 10.1111/j.1439-0523.2006.01266.x

Khodayari H, Saeidi H, Roofigar AA, et al (2012) Genetic Diversity of Cultivated Barley Landraces in Iran Measured Using Microsatellites. Int J Biosci Biochem Bioinforma 2:287–290. doi: 10.7763/IJBBB.2012.V2.118

King EG, Macdonald SJ, Long AD (2012a) Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics 191:935–949. doi: 10.1534/genetics.112.138537

King EG, Merkes CM, McNeil CL, et al (2012b) Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res 22:1558–1566. doi: 10.1101/gr.134031.111

Komatsuda T, Maxim P, Senthil N, Mano Y (2004) High-density AFLP map of nonbrittle rachis 1 (btr1) and 2 (btr2) genes in barley (Hordeum vulgare L.). Theor Appl Genet 109:986–995. doi: 10.1007/s00122-004-1710-0

Komatsuda T, Pourkheirandish M, He C, et al (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci 104:1424–1429. doi: 10.1073/pnas.0608580104

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29. doi: 10.1186/1746-4811-9-29

Kover PX, Valdar W, Trakalo J, et al (2009) A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana. PLoS Genet 5:e1000551. doi: 10.1371/journal.pgen.1000551

Kump KL, Bradbury PJ, Wisser RJ, et al (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–168. doi: 10.1038/ng.747

Lacerenza JA, Parrott DL, Fischer AM (2010) A major grain protein content locus on barley (Hordeum vulgare L.) chromosome 6 influences flowering time and sequential leaf senescence. J Exp Bot 61:3137–3149. doi: 10.1093/jxb/erq139

Lakew B, Henry RJ, Ceccarelli S, et al (2013a) Genetic analysis and phenotypic associations for drought tolerance in Hordeum spontaneum introgression lines using SSR and SNP markers. Euphytica 189:9–29. doi: 10.1007/s10681-012-0674-4

Lakew B, Henry RJ, Eglinton J, et al (2013b) SSR analysis of introgression of drought tolerance from the genome of Hordeum spontaneum into cultivated barley (Hordeum vulgare ssp vulgare). Euphytica 191:231–243. doi: 10.1007/s10681-012-0795-9

Larsson SJ, Lipka AE, Buckler ES (2013) Lessons from Dwarf8 on the Strengths and Weaknesses of Structured Association Mapping. PLoS Genet 9:e1003246. doi: 10.1371/journal.pgen.1003246

Laurie DA, Pratchett N, Bezant JH, Snape JW (1994) Genetic analysis of a photoperiod response gene on the short arm of chromosome 2(2H) of Hordeum vulgare (barley). Heredity 72:619–627. doi:10.1038/hdy.1994.85

Li JZ, Huang XQ, Heinrichs F, et al (2005) Analysis of QTLs for yield, yield components, and malting quality in a BC3-DH population of spring barley. Theor Appl Genet 110:356–363. doi: 10.1007/s00122-004-1847-x

Li JZ, Huang XQ, Heinrichs F, et al (2006) Analysis of QTLs for yield components, agronomic traits, and disease resistance in an advanced backcross population of spring barley. Genome 49:454–466. doi: 10.1139/g05-128

Lin C-S, Poushinsky G (1985) A modified augmented design (type 2) for rectangular plots. Can J Plant Sci 65:743–749. doi: 10.4141/cjps85-094

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi: 10.1093/bioinformatics/bti282

Liu W, Maurer HP, Reif JC, et al (2013) Optimum design of family structure and allocation of resources in association mapping with lines from multiple crosses. Heredity 110:71–79. doi: 10.1038/hdy.2012.63

Long AD, Macdonald SJ, King EG (2014) Dissecting complex traits using the Drosophila Synthetic Population Resource. Trends Genet 30:488–495. doi: 10.1016/j.tig.2014.07.009

Lundqvist U, Franckowiak JD (2003) Chapter 5 Diversity of barley mutants. In: Roland von Bothmer T van H, Helmut Knüpffer and Kazuhiro Sato (ed) Developments in Plant Genetics and Breeding. Elsevier, pp 77–96

Lundqvist U, Franckowiak JD (1997) BGS 315, Six-rowed spike 3, vrs3. Barley Genet Newsl 264–265.

Lundqvist U, Wettstein-Knowles PV (1982) Dominant mutations at Cer-yy change barley spike wax into leaf blade wax. Carlsberg Res Commun 47:29–43. doi: 10.1007/BF02907795

Mace ES, Hunt CH, Jordan DR (2013) Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. Theor Appl Genet 126:1377–1395. doi: 10.1007/s00122-013-2059-z

March TJ, Richter D, Colby T, et al (2012) Identification of proteins associated with malting quality in a subset of wild barley introgression lines. PROTEOMICS 12:2843–2851. doi: 10.1002/pmic.201200117

Marquez-Cedillo LA, Hayes PM, Jones BL, et al (2000) QTL analysis of malting quality in barley based on the doubled-haploid progeny of two elite North American varieties representing different germplasm groups. Theor Appl Genet 101:173–184. doi: 10.1007/s001220051466

Mascher M, Muehlbauer GJ, Rokhsar DS, et al (2013a) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J 76:718–727. doi: 10.1111/tpj.12319

Mascher M, Richmond TA, Gerhardt DJ, et al (2013b) Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. Plant J 76:494-505. doi: 10.1111/tpj.12294

Massman J, Cooper B, Horsley R, et al (2011) Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. Mol Breed 27:439–454. doi: 10.1007/s11032-010-9442-0

Mather DE, Tinker NA, LaBerge DE, et al (1997) Regions of the Genome That Affect Grain and Malt Quality in a North American Two-Row Barley Cross. Crop Sci 37:544-554. doi: 10.2135/cropsci1997.0011183X003700020039x

Matus I, Corey A, Filichkin T, et al (2003) Development and characterization of recombinant chromosome substitution lines (RCSLs) using Hordeum vulgare subsp. spontaneum as a source of donor alleles in a Hordeum vulgare subsp. vulgare background. Genome 46:1010–1023. doi: 10.1139/g03-080

Maurer A, Draba V, Jiang Y, et al (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics 16:290. doi: 10.1186/s12864-015-1459-7

McCouch S, Baute GJ, Bradeen J, et al (2013) Agriculture: Feeding the future. Nature 499:23–24. doi: 10.1038/499023a

McMullen MD, Kresovich S, Villeda HS, et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740. doi: 10.1126/science.1174320

Mohammadi M, Blake TK, Budde AD, et al (2015) A genome-wide association study of malting quality across eight U.S. barley breeding programs. Theor Appl Genet 128:705–721. doi: 10.1007/s00122-015-2465-5

Moragues M, Comadran J, Waugh R, et al (2010) Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. Theor Appl Genet 120:1525–1534. doi: 10.1007/s00122-010-1273-1

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. Nat Rev Genet 13:85–96. doi: 10.1038/nrg3097

Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. Proc Natl Acad Sci 104:3289–3294. doi: 10.1073/pnas.0611377104

Morrell PL, Gonzales AM, Meyer KKT, Clegg MT (2014) Resequencing Data Indicate a Modest Effect of Domestication on Diversity in Barley: A Cultigen With Multiple Origins. J Hered 105:253–264. doi: 10.1093/jhered/est083

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp. spontaneum) despite high rates of self-fertilization. Proc Natl Acad Sci 102:2442–2447. doi: 10.1073/pnas.0409804102

Muñoz-Amatriaín M, Cuesta-Marcos A, Endelman JB, et al (2014) The USDA Barley Core Collection: Genetic Diversity, Population Structure, and Potential for Genome-Wide Association Studies. PLoS ONE 9:e94688. doi: 10.1371/journal.pone.0094688

Muñoz-Amatriaín M, Moscou MJ, Bhat PR, et al (2011) An Improved Consensus Linkage Map of Barley Based on Flow-Sorted Chromosomes and Single Nucleotide Polymorphism Markers. Plant Genome J 4:238-249. doi: 10.3835/plantgenome2011.08.0023

Muñoz-Amatriaín M, Xiong Y, Schmitt MR, et al (2010) Transcriptome analysis of a barley breeding program examines gene expression diversity and reveals target genes for malting quality improvement. BMC Genomics 11:1–15. doi: 10.1186/1471-2164-11-653

Myles S, Peiffer J, Brown PJ, et al (2009) Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. Plant Cell 21:2194–2202. doi: 10.1105/tpc.109.068437

Naz AA, Arifuzzaman M, Muzammil S, et al (2014) Wild barley introgression lines revealed novel QTL alleles for root and related shoot traits in the cultivated barley (Hordeum vulgare L.). BMC Genet 15:107. doi: 10.1186/s12863-014-0107-6

Nielsen R (2000) Estimation of Population Parameters and Recombination Rates From Single Nucleotide Polymorphisms. Genetics 154:931–942.

Pasam RK, Sharma R, Malosetti M, et al (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol 12:16. doi: 10.1186/1471-2229-12-16

Pascual L, Desplat N, Huang BE, et al (2014) Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnol J 13:565-577. doi: 10.1111/pbi.12282

Peiffer JA, Flint-Garcia SA, De Leon N, et al (2013) The Genetic Architecture of Maize Stalk Strength. PLoS ONE 8:e67066. doi: 10.1371/journal.pone.0067066

Peiffer JA, Romay MC, Gore MA, et al (2014) The Genetic Architecture Of Maize Height. Genetics 196:1337–1356. doi: 10.1534/genetics.113.159152

Pillen K, Zacharias A, Léon J (2003) Advanced backcross QTL analysis in barley (Hordeum vulgare L.). Theor Appl Genet 107:340–352. doi: 10.1007/s00122-003-1253-9

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci 108:6893–6898. doi: 10.1073/pnas.1010894108

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLoS ONE 7:e32253. doi: 10.1371/journal.pone.0032253

Pourkheirandish M, Komatsuda T (2007) The Importance of Barley Genetics and Domestication in a Global Perspective. Ann Bot 100:999–1008. doi: 10.1093/aob/mcm139

Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. Nat Rev Genet 11:459–463. doi: 10.1038/nrg2813

Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. Nature 457:843–848. doi: 10.1038/nature07895

Rafalski A (2002a) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100. doi: 10.1016/S1369-5266(02)00240-6

Rafalski JA (2002b) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci 162:329–333. doi: 10.1016/S0168-9452(01)00587-8

Rasmusson DC, Phillips RL (1997) Plant Breeding Progress and Genetic Diversity from De Novo Variation and Elevated Epistasis. Crop Sci 37:303–310. doi: 10.2135/cropsci1997.0011183X003700020001x

Rebetzke GJ, Verbyla AP, Verbyla KL, et al (2014) Use of a large multiparent wheat mapping population in genomic dissection of coleoptile and seedling growth. Plant Biotechnol J 12:219–230. doi: 10.1111/pbi.12130

Rostoks N, Ramsay L, MacKenzie K, et al (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci 103:18656–18661. doi: 10.1073/pnas.0606133103

Roy JK, Smith KP, Muehlbauer GJ, et al (2010) Association mapping of spot blotch resistance in wild barley. Mol Breed 26:243–256. doi: 10.1007/s11032-010-9402-8

Russell J, Booth A, Fuller J, et al (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389–398. doi: 10.1139/g03-125

Russell J, Dawson IK, Flavell AJ, et al (2011) Analysis of >1000 single nucleotide polymorphisms in geographically matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes. New Phytol 191:564–578. doi: 10.1111/j.1469-8137.2011.03704.x

Saal B, Korff M von, Léon J, Pillen K (2011) Advanced-backcross QTL analysis in spring barley: IV. Localization of QTL × nitrogen interaction effects for yield-related traits. Euphytica 177:223–239. doi: 10.1007/s10681-010-0252-6

Sannemann W, Huang BE, Mathew B, Léon J (2015) Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Mol Breed 35:1–16. doi: 10.1007/s11032-015-0284-7

Sayed MA, Schumann H, Pillen K, et al (2012) AB-QTL analysis reveals new alleles associated to proline accumulation and leaf wilting under drought stress conditions in barley (Hordeum vulgare L.). BMC Genet 13:1–12. doi: 10.1186/1471-2156-13-61

Schmalenbach I, Körber N, Pillen K (2008) Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. Theor Appl Genet 117:1093–1106. doi: 10.1007/s00122-008-0847-7

Schmalenbach I, March TJ, Bringezu T, et al (2011) High-Resolution Genotyping of Wild Barley Introgression Lines and Fine-Mapping of the Threshability Locus thresh-1 Using the Illumina GoldenGate Assay. G3 Genes Genomes Genet 1:187–196. doi: 10.1534/g3.111.000182

Schmalenbach I, Pillen K (2009) Detection and verification of malting quality QTLs using wild barley introgression lines. Theor Appl Genet 118:1411–1427. doi: 10.1007/s00122-009-0991-8

Schnaithmann F, Kopahnke D, Pillen K (2014) A first step toward the development of a barley NAM population and its utilization to detect QTLs conferring leaf rust seedling resistance. Theor Appl Genet 127:1513-1525. doi: 10.1007/s00122-014-2315-x

Schnaithmann F, Pillen K (2013) Detection of exotic QTLs controlling nitrogen stress tolerance among wild barley introgression lines. Euphytica 189:67–88. doi: 10.1007/s10681-012-0711-3

Shannon LM, Yandell BS, Broman K (2013) Users Guide for New BCsFt Tools for R/qtl.

Slatkin M (2008) Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat Rev Genet 9:477–485. doi: 10.1038/nrg2361

Smith KP, Rasmusson DC, Schiefelbein E, et al (2010) Registration of "Rasmusson" Barley. J Plant Regist 4:167-170. doi: 10.3198/jpr2009.10.0622crc

Steffenson BJ, Olivera P, Roy JK, et al (2007) A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. Aust J Agric Res 58:532–544. doi: 10.1071/AR07123

Stracke S, Presterl T, Stein N, et al (2007) Effects of Introgression and Recombination on Haplotype Structure and Linkage Disequilibrium Surrounding a Locus Encoding Bymovirus Resistance in Barley. Genetics 175:805–817. doi: 10.1534/genetics.106.063800

Stupar RM, Specht JE (2013) Chapter Four - Insights from the Soybean (Glycine max and Glycine soja) Genome: Past, Present, and Future. In: Sparks DL (ed) Advances in Agronomy. Academic Press, pp 177–204

Tabangin ME, Woo JG, Martin LJ (2009) The effect of minor allele frequency on the likelihood of obtaining false positives. BMC Proc 3:S41. doi: 10.1186/1753-6561-3-S7-S41

Taketa S, Amano S, Tsujino Y, et al (2008) Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. Proc Natl Acad Sci 105:4062–4067. doi: 10.1073/pnas.0711034105

Tanksley SD, McCouch SR (1997) Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. Science 277:1063–1066. doi: 10.1126/science.277.5329.1063

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203. doi: 10.1007/BF00223376

Technow F (2011) R Package mvngGrAd: Moving Grid Adjustment In Plant Breeding Field Trials.

Tian F, Bradbury PJ, Brown PJ, et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162. doi: 10.1038/ng.746

Trevaskis B, Hemming MN, Peacock WJ, Dennis ES (2006) HvVRN2 Responds to Daylength, whereas HvVRN1 Is Regulated by Vernalization and Developmental Status. Plant Physiol 140:1397–1405. doi: 10.1104/pp.105.073486

Turner A, Beales J, Faure S, et al (2005) The Pseudo-Response Regulator Ppd-H1 Provides Adaptation to Photoperiod in Barley. Science 310:1031–1034. doi: 10.1126/science.1117619

Ullrich SE (2010) Significance, Adaptation, Production, and Trade of Barley. In: Ullrich SE (ed) Barley. Wiley-Blackwell, pp 3–13

USDA-NASS (2015) United States Department of Agriculture, National Agricultural Statistics Service. http://quickstats.nass.usda.gov/results/2DAB2E75-77F6-315F-87BB-66EC5DB562C5. Accessed 23 Jan 2015

Van Hintum T, Menting F (2003) Chapter 12 Diversity in ex situ genebank collections of barley. In: Roland von Bothmer T van H, Helmut Knüpffer and Kazuhiro Sato (ed) Developments in Plant Genetics and Breeding. Elsevier, pp 247–257

Verbyla AP, George AW, Cavanagh CR, Verbyla KL (2014) Whole-genome QTL analysis for MAGIC. Theor Appl Genet 127:1753–1770. doi: 10.1007/s00122-014-2337-4

Vilhjálmsson BJ, Nordborg M (2013) The nature of confounding in genome-wide association studies. Nat Rev Genet 14:1–2. doi: 10.1038/nrg3382

Von Korff M, Wang H, Léon J, Pillen K (2008) AB-QTL analysis in spring barley: III. Identification of exotic alleles for the improvement of malting quality in spring barley (H. vulgare ssp. spontaneum). Mol Breed 21:81–93. doi: 10.1007/s11032-007-9110-1

Von Korff M, Wang H, Léon J, Pillen K (2006) AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (H. vulgare ssp. spontaneum). Theor Appl Genet 112:1221–1231. doi: 10.1007/s00122-006-0223-4

Von Korff M, Wang H, Léon J, Pillen K (2005) AB-QTL analysis in spring barley. I. Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley. Theor Appl Genet 111:583–590. doi: 10.1007/s00122-005-2049-x

Von Korff M, Wang H, Léon J, Pillen K (2004) Development of candidate introgression lines using an exotic barley accession (Hordeum vulgare ssp. spontaneum) as donor. Theor Appl Genet 109:1736–1745. doi: 10.1007/s00122-004-1818-2

Von Korff M von, Léon J, Pillen K (2010) Detection of epistatic interactions between exotic alleles introgressed from wild barley (H. vulgare ssp. spontaneum). Theor Appl Genet 121:1455–1464. doi: 10.1007/s00122-010-1401-y

VSN International (2009) ASReml-R, version 3.0.

Wang B, Chee PW (2010) Application of advanced backcross quantitative trait locus (QTL) analysis in crop improvement. J Plant Breed Crop Sci 2:221–232.

Wang G, Schmalenbach I, von Korff M, et al (2010) Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines. TAG Theor Appl Genet Theor Angew Genet 120:1559–1574. doi: 10.1007/s00122-010-1276-y

Waugh R, Jannink J-L, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. Curr Opin Plant Biol 12:218–222. doi: 10.1016/j.pbi.2008.12.007

Wendler N, Mascher M, Nöh C, et al (2014) Unlocking the secondary gene-pool of barley with next-generation sequencing. Plant Biotechnol J 12:1122–1131. doi: 10.1111/pbi.12219

Wenzl P, Carling J, Kudrna D, et al (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci 101:9915–9920. doi: 10.1073/pnas.0401076101

Wettstein-Knowles P von (1990) New alleles of Cer-yy and cer-b. Barley Genet Newsl 20:66–68.

Windels CE (2000) Economic and Social Impacts of Fusarium Head Blight: Changing Farms and Rural Communities in the Northern Great Plains. Phytopathology 90:17–21. doi: 10.1094/PHYTO.2000.90.1.17

Xu Y, Zhu L, Xiao J, et al (1997) Chromosomal regions associated with segregation distortion of molecular markers in F2 , backcross, doubled haploid, and recombinant inbred populations in rice (Oryza sativa L.). Mol Gen Genet 253:535–545. doi: 10.1007/s004380050355

Yan L, Fu D, Li C, et al (2006) The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proc Natl Acad Sci 103:19581–19586. doi: 10.1073/pnas.0607142103

Yan L, Loukoianov A, Blechl A, et al (2004) The Wheat VRN2 Gene Is a Flowering Repressor Down-Regulated by Vernalization. Science 303:1640–1644. doi: 10.1126/science.1094305

Yan L, Loukoianov A, Tranquilli G, et al (2003) Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci 100:6263–6268. doi: 10.1073/pnas.0937399100

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic Design and Statistical Power of Nested Association Mapping in Maize. Genetics 178:539–551. doi: 10.1534/genetics.107.074245

Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. doi: 10.1038/ng1702

Yun SJ, Gyenis L, Bossolini E, et al (2006) Validation of Quantitative Trait Loci for Multiple Disease Resistance in Barley Using Advanced Backcross Lines Developed with a Wild Barley. Crop Sci 46:1179-1186. doi: 10.2135/cropsci2005.08-0293

Yun SJ, Gyenis L, Hayes PM, et al (2005) Quantitative Trait Loci for Multiple Disease Resistance in Wild Barley. Crop Sci 45:2563-2572. doi: 10.2135/cropsci2005.0236

Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360. doi: 10.1038/ng.546

Zhou H, Muehlbauer G, Steffenson B (2012) Population structure and linkage disequilibrium in elite barley breeding germplasm from the United States. J Zhejiang Univ Sci B 13:438–451. doi: 10.1631/jzus.B1200003

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. Plant Genome J 1:5-20. doi: 10.3835/plantgenome2008.02.0089

Zohary D, Hopf M, Weiss E (2012) Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin. Oxford University Press