Aus dem Institut für Agrar- und Ernährungswissenschaften

der Naturwissenschaftlichen Fakultät III

der Martin-Luther-Universität Halle-Wittenberg

A genome-wide association study to genetically dissect yield related traits in a diverse collection of spring barley landraces

Dissertation
zur Erlangung des Doktorgrades
der Agrarwissenschaften (Dr. agr.)

vorgelegt von
Herrn Celestine Wabila
Geb. am 08.08.1981 in Bamenda-Cameroon

Gutachter:
Prof. Dr. Andreas Graner (IPK Gatersleben)

Prof. Dr. Volker Mohler (Lfl Freising-Weihenstephan)

Verteidigung am: 07/05/2018

Contents

List of figures

List of Tables

Supplementary figure

Supplementary tables

Extra supplementary files (see attached compact disc)

File 1: BLUE values of all 14 traits measured in 2013 and 2014 for GWAS

File 2: Complete GWAS results of all SNPs associated with respective traits. Allele frequencies of all significant SNPs together with respective SNP effects are given

File 3: Percentage relatedness of each genotype to each subpopulation for all 16 subgroups at *K*=16

File 4: Population structure of all 261 accessions based on STRUCTURE at *K*=16

List of abbreviations

AM	Association mapping
Anova	Analysis of variance
AwnL	Awn length
AwnR	Awn roughness
BCC	Barley core collection
BLUE	Best linear unbiased estimate
cm	Centimeter
cM	Centi Morgan
CMLM	Compressed Mixed linear model
CV	Coefficient of variation
DArT	Diversity array technology
EL	Ear length
EST	Expressed sequence tag
FDR	False discovery rate
Gbp	Giga base pair
G x E	Genotype by environment
GLM	General linear model
GPS	Grains per spike
GWAS	Genome-wide association scan
HD	Heading date
HI	Harvest index
Kbp	Kilo base pair
LD	Linkage disequilibrium
LRC	Landrace collection
M X B	Morex X Barke
MAF	Minor allele frequency
MAS	Marker assisted selection
MLM	Mixed linear model
MTA	Marker trait association
N_C	Naked caryopsis
NJ	Neighbor joining
PCA	Principal component analysis

PCR	Polymerase chain reaction
Pht	Plant height
PIC	Polymorphic information content
QTL	Quantitative trait locus
RAPD	Random amplified polymorphic DNA
REML	Restriction estimate of maximum likelihood
RFLP	Restriction fragment length polymorphism
RILs	Recombinant inbred lines
SDA	Seed Area
SDL	Seed length
SDW	Seed width
SNP	Single nucleotide polymorphism
SPD	Spike density
SSR	Simple sequence repeat
Stigma	Stigma hairiness
TKW	Thousand kernel weight

1. General introduction

1.1 Barley taxonomy and geographical distribution

Hordeum vulgare L., the domesticated form of *Hordeum spontaneum* C. Koch is one of the founder crops of old world agriculture (Diamond 1993). Barley belongs to the grass family *Poaceae*, the tribe *Triticeae* and the genus *Hordeum* which is comprised of about 33 species and 45 taxa separated into four sections (Bothmer 1992), although as many as six sections have been suggested (Reid 1968). The division of the genus into four sections puts plants into groups that have similar morphological characteristics, life forms, similarities in ecology, and geographical area of origin (von Bothmer et al. 2003). Members of the *Triticeae* have a complex mode of speciation including polyploidy, interspecific and intergeneric hybridizations, which have resulted in a reticulate pattern of relationships and a wide geographical distribution including all major temperate areas. *Hordeum* is even present in the subtropics (von Bothmer et al. 2003). All *Hordeum*-species share similar diagnostic and morphological characters such as spike, glumes, lemmas, anthers, and leaves. Despite the homogenous structure in basic morphology and speciation, *Hordeum* shows a high degree of biological diversity with some species being annual with more or less strict inbreeding like *H. marinum* Huds, *H. murinum* L. and *H. vulgare* L. Some species are perennials with self-incompatibility like *H. bulbosum* L. while the majority of species are perennial with a versatile reproductive system (Bothmer et al. 2003b). Nearly half of the *Hordeum* species are polyploids (tetra- and hexaploids) including allo- and autopolyploids, hence the genus *Hordeum* is a good model to study speciation through polyploidization.

Hordeum spontaneum C. Koch, the immediate ancestor of cultivated barley is still abundant in nature and was first discovered in Turkey by the German botanist Carl Koch (Bothmer et al. 1995). It was described as separate species. However, based on several criteria, the progenitor form is nowadays regarded as a subspecies (*ssp. sponteneum* (C.Koch) Thell.) of *H. vulgare* L., and the cultivated form of barley (*ssp. vulgare*) being another subspecies.

Barley as a whole is well-adapted to marginal and stress-prone environments such as high soil salinity, and a more reliable crop than wheat or rice in regions which are colder or higher in altitude. Due to its wide adaption to marginal and extreme conditions, barley is found in a wide range of geographical distributions throughout the world. Today, barley is grown in fertile as well as in marginal areas under extreme conditions, including altitudes of up to

5,500 m in the Himalayas, in seasonal flooded areas in south east Asia, and in arid regions of the Mediterranean (Angessa and Li 2015; Bothmer et al. 2003b). Different morphological forms of barley include two-rowed, six-rowed, hull/hull-less, and hooded barley. These different forms can be differentiated into spring or winter type based on the requirement of vernalisation by the winter type as opposed to the spring type which requires no vernalisation treatment. This difference in both morphological and physiological characteristics is a reflection of the underlying genetic diversity which eases the adaptation of barley to different environments. Wild barley is predominantly a winter-type, the domesticated form is either spring or winter type. Studies of the wild subspecies (ssp. *spontaneum*) have mainly focused on taxonomy, distribution, morphological variation patterns and species relationships (Jakob et al. 2014; von Bothmer R. et al. 1995). Wild barley covers the whole natural distribution area from the Mediterranean to Middle Asia, the eastern Mediterranean with eastern Greece and Turkey, the Cyrenaica area of Libya and Egypt extending eastwards to Afghanistan, Turkmenia and Baluchistan in west Pakistan (Badr et al. 2000; Bothmer et al. 1995). Recent studies by Bayesian assignment analyses of multi-locus sequence data and paleo distribution modelling have further revealed three population clusters of wild barley, in the Levant, Turkey, and east of Turkey, respectively (Jakob et al. 2014).

1.2 Barley domestication

Barley was one of the first domesticated grain cereals in the Near East, along with einkorn and emmer wheat (Pourkheirandish et al. 2015). Archaeological evidence indicates that the movement of mankind from hunter-gathering to cultivation and agriculture occurred around the Fertile Crescent around 12,000 to 9,500 years ago with the domestication process lasting several centuries (Tanno and Willcox 2006; Weiss et al. 2006; Willcox et al. 2008). Allelic frequencies at 400 AFLP polymorphic loci studied in 317 wild and 57 cultivated barleys revealed the wild populations from Israel-Jordan to be molecularly more similar than to the cultivated gene pool leading to the early hypothesis of Israel-Jordan to be the main region of domestication (Badr et al. 2000). This hypothesis was later supported by the diagnostic allele I of the homeobox gene *BKn-3* (a gene involved in awn suppression in barley), rarely but exclusively found in Israel *ssp. spontaneum*. However, recent studies of wild and landrace (primitive domesticates) barley collections (Morrell and Clegg 2007a; Saisho and Purugganan 2007) and evidence of independent origins of the important domestication-related trait such as the brittle rachis (controlled by *btr1* and *btr2*) (Pourkheirandish et al.

2015) support the hypothesis of at least two independent domestication events followed by some degree of admixture amongst domesticates from distinct portions of the geographic range of wild barley (Fuller et al. 2011; Kilian et al. 2007). Early studies (Morrell and Clegg 2007a) on tough rachis showed two tough rachis mutants were associated with genetically distinct groups of domesticated barley, one associated with West and the other with the East, suggesting that a core centre of origin does not apply to barley. The two genes, non-brittle rachis 1 (*btr1*) and non-brittle rachis 2 (*btr2*), control spike disarticulation in barley with a mutation in either of the two genes converting the brittle rachis (wild) to a non-brittle (domesticated) type (Pourkheirandish et al. 2015). Pourkheirandish et al. (2015) reported two distinct regions where early farmers must have independently selected for mutations of the barley brittle rachis phenotype. Additional evidence to the presence of up to four domestication events of wild barley in the region of the Fertile Crescent has been gained from analysis of European barley, where at least two different wild barleys were adopted into central and northern Europe thus indicating the possibility of at least two domestication events (Jones et al. 2013). Further evidence of more than a centric origin of domestication has been reported by Poets and colleagues (Poets et al. 2015a) while examining 6152 single nucleotide polymorphisms (SNPs). Here, several regional groups of domesticated barley were discerned relating to Central Europe, Coastal Mediterranean, Asia and East Africa. Data from chloroplast DNA clearly indicate that wild barley as it is found today in the Fertile Crescent might not be the progenitor of barley cultivated in Eritrea/Ethiopian, indicating that an independent domestication might have taken place at the Horn of Africa (Ethiopia) (Orabi et al. 2007). Barley exist as hulled or hull-less, with early reports pointing to a single gene *(Nud)* controlling the hulled phenotype (Taketa et al. 2008) which also led to earlier suggestion of a single origin of domesticated hull-less barley. A recent study has reported more than one origin of domesticated hull-less barley, with a suggestion of Tibetan hull-less barley having an independent origin of domestication (Yu et al. 2016).

1.3 Early and modern use of barley and its economic importance

Barley together with other cereal grains such as emmer, einkorn and later modern wheat and rice were the staple food and probably the most important products of the world in earlier times. Both wheat and barley were grown in Turkestan in the third millennium B.C and were also a basic food source for the Sumerian diet. Despite the primary use of barley as a staple food, probably as porridge or bread, it was used in making beer or alcoholic beverages called

"barley wine" (Bishop 1936; Harlan 1978). In the earlier times, recipes from barley *puls*, an oily seasoned paste mixture was a popular food in Greece (Tannahill 1988). Barley was a common constituent of unleavened bread and porridge eaten by ancient Greeks and was also used as an energy food and a preferred diet by the Roman gladiators who were called *Hordearii* or "barley men" (Ceccarelli et al. 2010; Percival 1921).

Although barley was classified as an important food grain in ancient times, its use as a staple food source declined as other food grains such as wheat, rye and oats became abundant. Barley became relegated to the status of "poor man`s bread" (Zohary 1988). However, consumer's interest in nutrition and the health benefits of barley has helped in the restoration of barley's status in the human diet. Barley currently ranks fourth after maize (*Zea mays*), rice (*Oryza sativa*) and wheat (*Triticum aestivum*) in terms of worldwide acreage cereal production (FAO 2015). Today, barley is mostly used in the brewing and malting industry and also as an important source of animal feed in the developed world (Friedt and Ordon 2013). Considered a staple food source in several regions of the world particularly in Asia and North Africa, including Ethiopia where barley is still used for bread porridge (Baik and Ullrich 2008; von Bothmer R. et al. 1995), barley is rich in nutrients with its nutritional components generally reported as averages though the chemical composition may differ greatly due to genotype, cultural practices and growing conditions. Starch, fiber, and proteins make up the largest portion of the kernel with a variation in one of the components directly influencing the amounts of the other two (Table 1). Barley constitutes a rich source of dietary fiber with β-glucan being the most important in terms of human diet and health benefits (Van Hung 2016). The high amount of β-glucan helps lowering cholesterol and blood glucose levels.

Table 1. Typical composition (g/kg) of hulled and hull-less isotype barleys on dry matter basis.

	Hulled		Hull-less	
Item	Mean[a]	Range	Mean[a]	Range
Protein[b]	13.7	12.5 – 15.4	14.1	12.1 – 16.6
Starch	58.2	57.1 – 59.5	63.4	60.5 – 65.2
Sugars[c]	3.0	2.8 – 3.3	2.9	2.0 – 4.2
Lipids	2.2	1.9 – 2.4	3.1	2.7 – 3.9
Fiber	20.2	18.8 – 22.6	13.8	12.6 – 15.6
Ash	2.7	2.3 – 3.0	2.8	2.3 – 3.5

Source: Adapted from Äman and Newman (1986)
[a]*n*=3 (n= number of samples)
[b]*N* x 6.25 (*N* =measured value)
[c]Glucose, fructose, sucrose, and fructans.

Worldwide barley production increased from 133.584 million metric tons (MMT) in 2011 to 145.164 metric tons (MMT) in 2015 (http://www.fas.usda.gov/data/grain-world-markets-and-trade) with the European Union being the leading producer of barley (41.9%), followed by the former Soviet Union (21.94%), North America (9.28%), and the Middle East (8.51%). Barley production in Europe is mostly used in the malting and brewing industry and also as animal feed. Domestic consumption of barley in East Asia stood at 10.000 MMT despite the fact that East Asia only produced 8.658 MMT of barley in 2015 (http://www.fao.org/worldfoodsituation/csdb/en/).

1.4 The genepool of barley and its genetic diversity

Barley (H. *vulgare* L.) is a diploid predominantly self-pollinating grain crop with a basic chromosome number of n=7 (2n=14) and a large genome size of 5.1 Gb with approximately 84% of its genome mostly comprising of mobile elements or other repetitive structures (Dolezel et al. 1998; Mayer et al. 2012; Wicker et al. 2008). Barley and its related species have been classified into genepools as primary, secondary, and tertiary (Fig. 1.1) based on cross-ability, hybrid viability, and meiotic chromosome pairing (Harlan and Wet 1971). The primary genepool of barley includes domesticated barley (*H. vulgare ssp. vulgare*) and the wild form (*H. vulgare ssp. spontaneum*) which is both annual, diploid and are predominantly inbreeding. Both members of the primary genepool are cross-compatible producing vigorous,

viable, and fertile hybrids, facilitating the transfer of desired traits from the wild subspecies to cultivated barley. Within the primary genepool, wild and cultivated barley from several geographic areas are genetically highly diverse (Petersen et al. 1994). The secondary genepool includes only a single species *H. bulbosum* L. that shares the H genome with the primary genepool and consists of both diploid and tetraploid individuals which are mainly self-incompatible perennials. However, crosses between the two genepools are difficult. Diploid and tetraploid *H. bulbosum* from the secondary genepool is widely used to produce haploid barley through chromosome elimination (Bothmer et al. 1999; Kasha and Kao 1970; Pickering 1984). Several genes from *H. bulbosum* have been transferred to cultivated barley, providing a new source for breeding (Pickering 2000). The tertiary genepool is the largest and includes all the remaining species of *Hordeum*. Members belonging to the tertiary genepool are either diploid, tetraploid or hexaploid and consist of both annuals and perennials (Bothmer et al. 2003b; Bothmer et al. 1995).

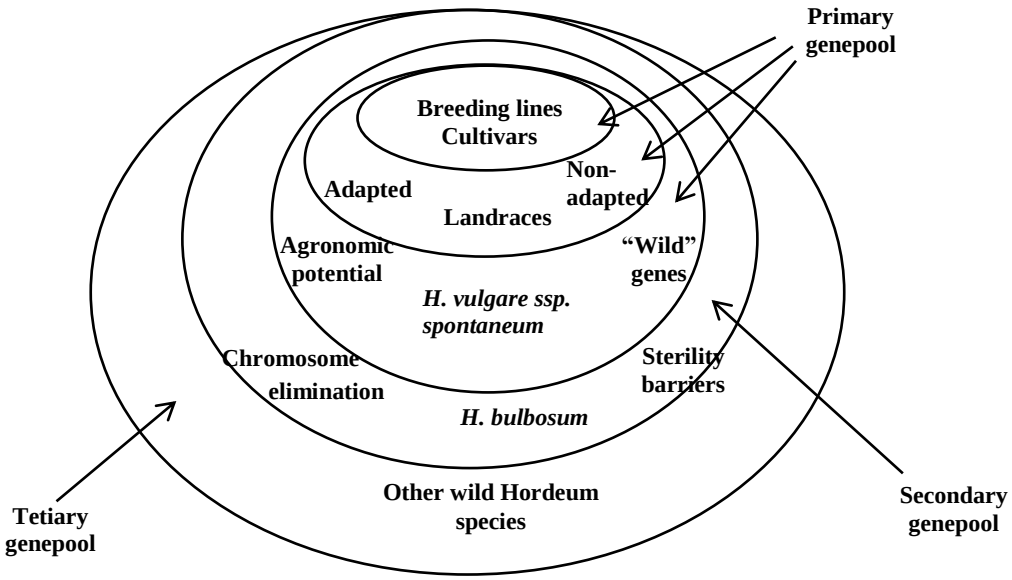


Figure 1.1. Genepools in cultivated barley by Harlan (*Hordeum vulgare*) (Bothmer et al. 1995)

Barley has a long history as a domesticated crop and the conscious selection of desired genotypes by farmers at early stages, together with natural selection created the rich source of variation found today in local varieties (Bothmer et al. 2003a). With the development of molecular markers, it has been easy to study genetic diversity in crops at the DNA level (Reif et al. 2003). There have been many studies on the genetic diversity of barley providing

evidence of a large variability present in the primary gene pool of barley which is surprising given the strong inbreeding nature of this species (Graner et al. 1994; Nandha and Singh 2014; Wang et al. 2010). Estimates of genetic diversity in barley mostly depend on the type of marker system used and the size and nature of the samples being investigates, such as RAPD (Fernandez et al. 2002; Meszaros et al. 2007), AFLP (Zhang and Ding 2007), ISSR (Fernandez et al. 2002), STS (Meszaros et al. 2007), and SSR (Pasam et al. 2014). Historically, the analysis of RFLPs was the first technique to study genetic diversity at the DNA level in barley. PCR-based markers such as SSRs and SNPs have increasingly replaced RFLPs due to their inherent technical advantages with SSR markers widely used since they are codominant, abundant and informative with their detection very simple (Matus and Hayes 2002; Nandha and Singh 2014). In general, SSR markers show the highest level of polymorphism in comparison to other genotypic markers (Russell et al. 1997) with the common pattern of genetic diversity in barley decreasing in the order of wild barley > landraces > cultivars.

1.5 Barley landraces and their importance

The use of crop wild relatives to improve crop performance has been well established for a very long time with examples dating back to more than 60 years. Crop wild relatives which includes the progenitor of crops such as *H. vulgare spp. sponta*neum as well as other species closely related to them have been very beneficial to modern agriculture, providing breeders with a broad pool of potentially useful genetic resources. Wild barley and modern landraces have been proven to be a useful source of genes (Ellis et al. 2000; Steffenson et al. 2007; Steffenson et al. 2016).

Barley landraces are heterogeneous populations developed by natural or farmer directed selection with a high local adaptation (Poets et al. 2015b). There has been much effort expended in the conservation of crop genetic resources in *ex situ* gene banks. A large proportion of gene bank materials comprises of accessions of traditional landraces of cultivated species. In barley *ssp. vulgare*, approximately one-half of the existing accessions in *ex situ* genebanks take the form of landraces according to information from the IPK Genebank Information Gystem (GBIS) http://gbis.ipk-gatersleben.de/ and the Biotechnology and Biological Sciences Research Council (BBSRC collection) (BBSRC 1999). Barley landrace populations are comprised of inbreeding lines and hybrid segregates generated by a low level of outcrossing (Nevo and Shewry 1992). Landraces harbour a rich source of genetic

diversity which has been exploited by The International Centre for Agriculture Research in Dry Areas (ICARDA) to improve yield and yield stability in dry areas. Similar to wild barley, landraces have high within-population diversity. In selected gene bank collections, at least 50-60% of the total genetic variation captured resides within populations, the remainder being accounted for by differences between landraces (Endresen et al. 2011; Poets et al. 2015a). Genetic diversity of landrace populations collected from Sardinia revealed that only 11% of the diversity detected by RAPD markers occurred between populations (Papa et al. 1998; Russell et al. 1997). Also, high level of genetic diversity was reported in a diverse collection of 1485 barley landraces originating from 41 countries genotyped with a set of 42 SSR markers (Pasam et al. 2014). The utilization of landrace genetic diversity as a source of crop improvement has been successful in many areas with examples including the introgression of dwarfing alleles (*Rht1 and Rht2*) derived from the Japanese wheat landrace "Shiro Daruma"(Kihara 1983), powdery mildew resistance allele *mlo11* derived from an Ethiopian barley landrace (Piffanelli et al. 2004), the barley yellow mosaic resistance gene, *rym4* (Graner and Bauer 1993), the boron-toxicity tolerances in barley obtained from the Algerian landrace "Sahara" (Sutton et al. 2007) and improvement of yield and abiotic stress adaptation (Dwivedi et al. 2016). Exploiting genetic diversity in the wild form and in landraces for crop improvement will help improve barley production in the future.

1.6 Barley breeding

Modern barley breeding started at the end of the last century aiming at improving yield along with the improvement of malting quality and yield stability. Through biotechnology-based and marker-assisted selection (MAS) approaches, different resistance genes have been combined or novel resistance genes have been introgressed from non-adapted germplasm into adapted cultivar's background. Other methods of plant biotechnology like anther and microspore culture for the rapid production of homozygous doubled haploids (DH) lines and cultivars have been implanted into barley breeding schemes. In *Hordeum bulbosum* method, anther or microspore culture has become very routine and focused on accelerated development of homozygous lines from segregating populations (Kang and Priyadarshan 2008). Examples of cultivars produced through anther culture include the spring barley cultivar "Henni" (D, 1995), the two-rowed winter barley "Anthere" (D, 1995); the six-rowed cultivars "Uschi" (D, 1997), "Sarah" (D, 1997) (Ullrich 2010). In addition to anther culture techniques, molecular marker technique which allows the transfer of selection steps from the

phenotypic to the genotypic level has offered new opportunities for more efficient barley breeding aiming at desired combinations of resistance, yield and quality (Weiskorn and Ordon 2003). Examples for the use of MAS in practical barley breeding include the pyramiding of resistance genes against barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2) by using markers tightly linked to *rym5* and *rym4* loci (Werner et al. 2005), molecular mapping of *Rph7.g* leaf rust resistance gene in barley (Brunner et al. 2000), stripe rust (Toojinda et al. 2000), cereal cyst nematode resistance (Kretschmer et al. 1997). Through MAS, resistance against the barley yellow dwarf virus (BYDV) was enhanced. By using DH lines and molecular markers, Habekuß et al. (2009) combined resistance genes *Ryd2* and *Ryd3* together with a QTL from cv. "Post" on barley chromosome 2H.

1.7 Barley as a model crop for genetic research

Barley with both emmer and einkorn wheat were domesticated simultaneously in the same area with the migration of all three crops and their subsequent adaptation to new areas following similar routes. Both crops have similar breeding prerequisites though wheat is a polyploid. Barley has an advantage to serve as a model crop due to its diploid genome and the easiness to create mutants and carry out genetic analysis which is difficult to achieve in wheat due to its polyploid nature (Bothmer et al. 2003a). The high level of chromosomal synteny between barley and other members of the *Triticeae* makes barley an ideal model crop for the whole *Triticeae* tribe (Bothmer et al. 2003b). The inbreeding nature coupled with the diploid habit of barley also makes inheritance studies easy to perform. Initially, the relatively large genome size has complicated molecular studies. More recently, the low cost of genotyping and the development of appropriate sequencing approaches (e.g. exome capture) have greatly reduced the complexity of the barley genome. There is a large number of closely and distantly related species in the *Triticeae* making barley a central organism suitable for the studies of evolution and relationships. A combination of crops like wheat, rye, barley and forage grasses makes the entire tribe a gigantic genepool for crop improvement. The importance of barley in research studies has promoted the screening of mutants, with more than 10.000 mutants already documented. Molecular genetics has also produced a significant impact on research into *Arabidopsis*, but there are more characterized mutants in barley than in *Arabidopsis* (Kalantidis et al. 2000; Weigel 2012). Amongst the characterized barley mutants, chlorophyll defects are the most common mutations and the first indication of

treatment success. In addition, barley has been widely used in cytogenetic research due its diploid nature and large chromosomes (Kunzel et al. 2000; Malysheva et al. 2003).

1.8 Barley genomic resources

Over the past two decades, there has been a great improvement in the development of molecular markers and genomic resources in barley. Several linkage maps have been generated since the construction of the first molecular linkage maps using RFLP markers (Graner et al. 1991; Heun et al. 1991). High density linkage maps were constructed using SSR (single sequence repeats) (Langridge et al. 1995), DArT (diversity array technology) and EST markers (Varshney et al. 2007; Wenzl et al. 2006). Recent development of new sequencing technologies (high-throughput genotyping platforms) and the low cost of genotyping has greatly advanced the analysis of SNP markers for association studies and fine mapping of genes (Close et al. 2009; Close et al. 2004; Davey et al. 2011).

A high density genetic map was constructed based on array genotyping by Comadran et al. (2012). Recently an improved high density genetic map (POPSEQ) was constructed by using a whole genome survey sequencing of genetic populations (Mascher et al. 2013) leading to additional SNPs being anchored to the genetic map generated by Comadran et al. (2012). By using a Genotyping by sequencing **(**GBS) protocol, Poland et al. (2012) were able to map over 34,000 SNPs and 240,000 tags on the barley Oregon Wolfe reference map. Furthermore, a physical map of 4.98 Gb has been developed with 3.90 Gb anchored to a high resolution genetic map (Mayer et al. 2012). Many efforts have been made in developing double haploid population (DH) lines which are a useful resources for research and breeding (Bjørnstad et al. 1992; Dwivedi et al. 2015; Powell et al. 1986) and TILLING mutant populations which can be used in screening candidate genes (Caldwell et al. 2004; Gottwald et al. 2009). By using these available genomic resources together with the genetic resources provided by the vast number of diverse landrace collections stored in different gene banks, mapping of candidate genes for yield improvement and other agronomic traits including disease resistance has been greatly advanced.

1.9 Quantitative trait loci (QTL) analysis

QTL detection is a tool for studying the underlying genetic architecture of complex traits and was developed more than 90 years ago (Sax 1923). The development of DNA markers in the 1980s created novel opportunities to identify QTLs for the characterization of quantitative

traits (Collard et al. 2005). Two approaches have been widely used in QTL analysis (linkage mapping and linkage disequilibrium mapping or association mapping) with both mapping strategies trying to exploit the fact that recombination breaks the genome into small fragments which can be correlated to the phenotype (Myles et al. 2009).

QTL analysis aims at linking phenotypic data and genotypic data (usually molecular markers) in an attempt to explain the basis of variation in complex traits (Falconer and Mackay 1996; Kearsey 1998; Lynch and Walsh 1998). The consistent trend in looking at QTLs is that phenotypes are frequently affected by a variety of interactions (genotype-by-environment, dominance, and epistatic interactions between QTLs). It appears that a substantial proportion of the phenotypic variation in many quantitative traits can be explained with few loci of large effects, with the remainder due to numerous loci of small effects (Mackay 2001; Remington and Purugganan 2003). For example, QTL mapping of flowering time in domesticated rice (*Oryza sativa*) identified six QTLs with the sum of the effects of the top five explaining 84% of the variation in the trait (Yamamoto et al. 1998; Yamamoto et al. 2000; Yano et al. 1997). After detecting QTLs on chromosomes, DNA markers tightly linked to the gene or QTL of interest can be used as a tool for MAS in breeding.

1.9.1 Linkage mapping vs linkage disequilibrium mapping

Linkage and linkage disequilibrium (LD) mapping both identify genotype-phenotype associations by identifying polymorphisms that are linked to functional alleles. Despite these similarities, linkage mapping is a highly controlled approach; individuals are crossed to generate a mapping population in which relatedness is known while linkage disequilibrium mapping is not controlled, but rather a natural experiment in which genotypic and phenotypic data are collected from a population in which relatedness is not controlled and correlations between genetic markers and phenotypes are sought within this population. Linkage disequilibrium mapping provides higher resolution compared to linkage mapping. Since linkage mapping relies on mapping populations developed from crosses between two contrasting parents, the experimenter can only exploit the recombination events that have occurred during the establishment of the mapping population which is a disadvantage since only a few recombination events have taken place and QTL are generally localized within large chromosomal regions (10 to 20 cM) due to a longer extent of LD. Furthermore, the QTL may be only segregating within one cross and therefore lack consistency across mapping populations and it remains unclear which QTL are representative in broader

germplasm (Holland 2007). On the other hand, linkage disequilibrium (LD) mapping offers a much higher mapping resolution than linkage mapping since all the recombination events that have accumulated in the evolutionary history of the sample are exploited and the detected QTL are representative for broad germplasm (Risch and Merikangas 1996; Spielman et al. 1993). In LD mapping, the number of QTLs one can map for a given phenotype is not limited to what segregates between two parents of a cross but rather by the number of real QTL underlying the trait and the degree to which the mapping population captures the total genetic diversity in nature (Zhu et al. 2008).

1.9.2 Genome-wide association mapping and linkage disequilibrium decay (LD)

Genome-wide association studies (GWAS) have revolutionized the genetic mapping of QTLs and were first established in human genetics where it is not possible to generate bi-parental mapping populations (Altshuler et al. 2008; Donnelly 2008). GWAS has become a powerful tool in detecting natural variation underlying complex traits in crops with the rapid development of sequencing technologies and computational methods (Nordborg and Weigel 2008; Rafalski 2002; Rafalski 2010; Waugh et al. 2009). Genome-wide association mapping or LD mapping is an association study that surveys the entire genome for genetic variants by exploiting the strength of linkage disequilibrium (LD) between markers and the association of these markers to traits under investigation. GWAS takes full advantage of ancestral recombination events to identify genetic loci underlying traits at a relatively high resolution. The term LD, defined as the non-random association of alleles at different loci was first coined out in 1960 (Lewontin and Kojima 1960). For LD mapping to be useful, markers tested for association must either be the causal allele or closely linked (in LD) with the causal allele (Jorde 2000; Kruglyak 1999). Thus, the genetic markers become proxies for the functional gene variant due to the high correlation of their genotypes with the genotype of the functional variant. In general, the strength of linkage between two markers is a function of the distance between them. The closer the markers are, the stronger they are in LD. Most of the genome falls into segments of strong LD, within which variants are strongly correlated with each other. The resolution of mapping a QTL is a function of how fast LD decays over distance and differs dramatically between species due to the differences in breeding systems (Flint-Garcia et al. 2003). Inbreeders such as rice (*Oryza sativa*) show a high degree of LD which can extend well beyond 100 kb as a result of selfing which reduces the opportunities for recombination (Garris et al. 2005; Nordborg 2000). In maize (*Zea mays*), LD decays

within 1 kb in landraces (Tenaillon et al. 2001), within 2 kb in diverse inbred lines (Wu et al. 2016), and may amount to 500 kb in commercial elite inbred lines (Remington et al. 2001). Barley, though a selfing crop, shows a varying extent of LD decay around 1 cM in wild and 3 - 4 cM in landraces (Munoz-Amatriain et al. 2014; Rodriguez et al. 2012). In modern cultivars, LD extends to around 9 - 15 cM (Comadran et al. 2009; Hamblin et al. 2010). LD decay also varies along chromosomal regions which is sometimes the result of selection which causes extended LD (Whitt et al. 2002). Since the resolution of QTL detection is a function of LD decay, the use of a diverse set of germplasm that exploits all recombination events that have occurred throughout the population history warrants high accuracy.

The strategy of GWAS is to place enough markers, usually SNP markers, across the genome so that functional alleles of the candidate gene will likely be in LD with at least one of the markers (Munoz-Amatriain et al. 2014; Rodriguez et al. 2012). Since the large genome size of most plant species makes it difficult to sufficiently capture all of the existing genetic variants with molecular markers, using a portion of the available genetic variants to represent the variation in a species can only be successful as a result of the extent of LD (Mangin et al. 2012). Thus the genome size of an organism and the extent of LD decay also define the estimated number of markers necessary to cover the whole genome (Kim et al. 2007). Previous reports have revealed that while 140,000 markers will provide a reasonable coverage for the 125 Mb of *Arabidopsis* genome (Kim et al. 2007), around 2 million markers will be required to cover the 457 Mb genome of grapevine, and between 10 to 15 million maybe necessary for the 2500 Mb of diverse maize varieties (Myles et al. 2009).

Conducted in barley, GWAS has already proven useful in studying multiple traits (Cockram et al. 2010; Comadran et al. 2011; Massman et al. 2011; Wang et al. 2012; Zhou and Steffenson 2013). Only a few GWAS have been carried on barley landraces, as most GWAS on barley has mainly focused on cultivated barley with landraces at most included as a sub-population as much effort is generally needed in order to generate single seed descents of collection of landraces and also to eliminate heterogeneity and to minimize heterozygosity. Nevertheless, the huge diversity within landrace material stored in Gene banks together with the available genomic resources can be exploited to fine map new variants for barley crop improvement. Since no specific number of markers has been reported in the case of barley, efficient GWAS in barley mainly relies on the number of accessions used and the extent of LD decay in the association panel under study.

1.9.3 Practical approaches for genome-wide association analysis

Complex patterns of population structure and genetic relatedness in wild and crop plants have been generated due to non-random mating, selection pressure, and other bottlenecks (Flint-Garcia et al. 2005; Nordborg et al. 2005). These are problematic when mapping a phenotype whose variation is correlated with genetic relatedness. Population structure and genetic relatedness may cause spurious associations, when markers only capture the genetic relatedness among individuals. Spurious associations as a result of population structure and genetic relatedness have long been described (Lander and Schork 1994), and a high number of false positive associations as a result of population structure has been reported in GWAS for flowering time in *Arabidopsis* (Aranzana et al. 2005).

Several methods have been developed to correct for genetic relatedness and population structure in GWAS. The first method developed to correct for genetic relatedness has been implemented in the software STRUCTURE (Evanno et al. 2005b; Pritchard et al. 2000a). It estimates the proportion of each individual's variation that comes from a particular sub-population. These estimates called "Q" are used as covariates to correct for population structure while calculating the relationship between one or more predictors (markers) and a continuous response variable (phenotype) in a General-linear model (GLM). An alternative method to the Q-matrix has been the use of principal component analysis (PCA) to reduce the high-dimensional genotypic data to a small number of dimensions. The axes of variation from these dimensions are used to calculate ancestry-adjusted genotypes and phenotypes (Price et al. 2006). The PCA-based approach is fast and performs similarly or better than STRUCTURE (Q-matrix) (Zhao et al. 2007). While the GLM was based mostly on pedigree information, random genetic markers from individuals within the association panel are often used to generate pairwise genetic relatedness matrix called kinship (K) which are used in a multiple regression model called the mixed-linear model (MLM) in GWAS. The K-matrix controls for population structure by correcting for background association resulting from genetic relatedness amongst various individuals (Hoffman 2013). This method has been widely used to predict breeding values in animals and plants (Hayes and Goddard 2001; Heffner et al. 2009; Schaeffer 2006) and was reported to out-perform the Q-matrix in correcting for population structure in GWAS in both animals and plants (Yu et al. 2006). Applying the MLM while correcting for genetic relatedness (K) in maize, humans, mouse, *Arabidopsis,* and potato has demonstrated that additional correction for pairwise relatedness

significantly leads to a reduction in type I error (false positives) and type II error (false negatives) compared to corrections involving only the Q-matrix in a GLM (Kang et al. 2008; Malosetti et al. 2007; Yu et al. 2006; Zhao et al. 2007). In general, the MLM approach with K is far superior to the GLM with Q or PCA since the K matrix captures relatedness between each possible pair of individuals in the sample. By contrast, Q or PCA captures only a few axes of variation. An extension of the MLM includes combining Q + K or PCA + K and appears to be more powerful and yield results comparable to the MLM with K alone (Flint-Garcia et al. 2005; Yu et al. 2006) though some authors have reported that the use of both Q + K or PCA + K may lead to over-correction and increased type II error (false negatives) (Kang et al. 2008; Stich and Melchinger 2009; Yu et al. 2006).

1.9.4 Genome-wide association scans and challenges

The development of numerous molecular markers (especially SNPs) for various plant species and the development of methods to control for confounding effects from population structure have led to the identification of marker-trait associations for diverse traits such as; flowering time, kernel composition, and kernel colour in maize (Palaisa et al. 2004; Thornsberry et al. 2001; Wilson et al. 2004), developmental and flowering-related traits in *Arabidopsis* (Atwell et al. 2010), flowering time in ryegrass (Skot et al. 2007) as well as disease resistance and several agronomic traits in barley (Haseneyer et al. 2010; Massman et al. 2011; Munoz-Amatriain et al. 2014; Pasam et al. 2012; Wehner et al. 2015). Although the MLM has provided a robust method to correct for relatedness in GWAS, attempts to map phenotypes that are strongly correlated with relatedness remain problematic. There is no simple way to statistically determine whether a genetic variant is a true QTL if the phenotype is so strongly correlated with relatedness that random genetic variants throughout the genome associate equally well with the trait. In this case, linkage mapping can come to rescue when encountering confounding effects of relatedness as in the case of detecting low frequency functional variants (Balasubramanian et al. 2006; Manenti et al. 2009). In such cases, controlled crosses will break up the genotype-phenotype covariance while enhancing the power for QTL detection.

1.9.5 Sample size and statistical power for GWAS

Determination of genomic regions associated with a phenotype of interest requires a sample size with sufficient statistical power. Genome-wide association studies require an effective

sample size in order to achieve an adequate statistical power since it involves the evaluation of thousands or hundreds of thousands of SNP markers (Klein 2007; Park et al. 2010). In GWAS, an effective sample size is defined as the minimum sample size needed to achieve an adequate statistical power, with the ability to predict true marker-trait associations increasing with an increased sample size. Many studies have already reported a minimum sample size of 100 to be adequate for GWAS in crops, with statistical power highly reduced when sample size is below 100 (Hintsanen et al. 2006; Pfeiffer and Gail 2003).

In genome-wide association studies, marker-trait associations are referred to be statistically significant when the p-value is less than a pre-set threshold value (α) of 0.05 above which the null hypothesis of no marker effect is rejected. By testing a large number of SNP markers in a genome-wide association study, multiple comparisons are made thus causing a multiple testing problem which incurs an increase in false positives (Gao 2011). The Bonferroni-corrected p-value has widely been used to determine the threshold of significant marker-trait associations and is mostly set to α=0.05 which is calculated by diving 0.05 by the total number of SNP markers analysed in a GWAS. Though the Bonferroni method of correcting for false positives (type I error) is very efficient, it is frequently too strict to consider the correlations among SNP markers and might fail for some traits since a fixed threshold is calculated and considered for all investigated traits (Spencer et al. 2009; Wu and Zhao 2009). An alternative method of correcting for false positive associations is the false discovery rate (FDR) approach (Storey and Tibshirani 2003). In the FDR approach, corrected p-values (q-values) for individual markers are calculated from the p-value distribution of all markers considered being significant according to the p-value cut-off, and are computed for each trait separately. The q-value is an extension of FDR (Benjamini and Hochberg 1995) and significant FDR values are mostly set at α=0.05 (5%) indicating that among all features called significant, 5% of these are truly null on average. FDR methods have been used in the detection of differentially expressed genes across two or more biological conditions (Tzeng et al. 2003), in genetic dissection of transcriptional regulation in yeast (Brem et al. 2002) and determination of true positive associations in several GWAS (Alqudah et al. 2014; Pasam et al. 2014).

1.10 Research objectives

Given the large potential of landraces as sources for new and useful genetic diversity, *'A Genome-wide association analysis was carried out on a diverse collection of 2-rowed spring barley landraces*' with the following main objectives:

1. to study the population structure in the panel

2. to analyse the pattern of diversity and to estimate the extent of linkage disequilibrium decay in the panel of barley landraces,

3. to evaluate field trials and genetically dissect QTL for 14 agronomic traits using a GWA approach.

2 Materials and methods

2.1 Plant material

The association panel consisted of 261 two-rowed spring barley landraces. In addition, three additional two-rowed spring barley accessions were included as controls. This collection is a subset of a diverse core reference set (LRC648) developed at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). The reference core set of 648 accessions (LRC648) are from a collection of 1493 accessions which were previously genotyped with 42 SSR markers to study population structure and genetic diversity in order to build up a diverse reference core set harbouring most of the genetic diversity of all spring barley landrace accessions maintained at the IPK Gene bank (Pasam et al. 2014). The GWA analysis was restricted to two-rowed spring barley of the LRC648 core set to minimize the effect of population structure which exists between two- and six-rowed barley. The selected 261 accessions originate from 28 countries covering three continents with 53 accessions coming from Africa, 32 from Asia and 176 from Europe (Fig. 2.1, Table S1).

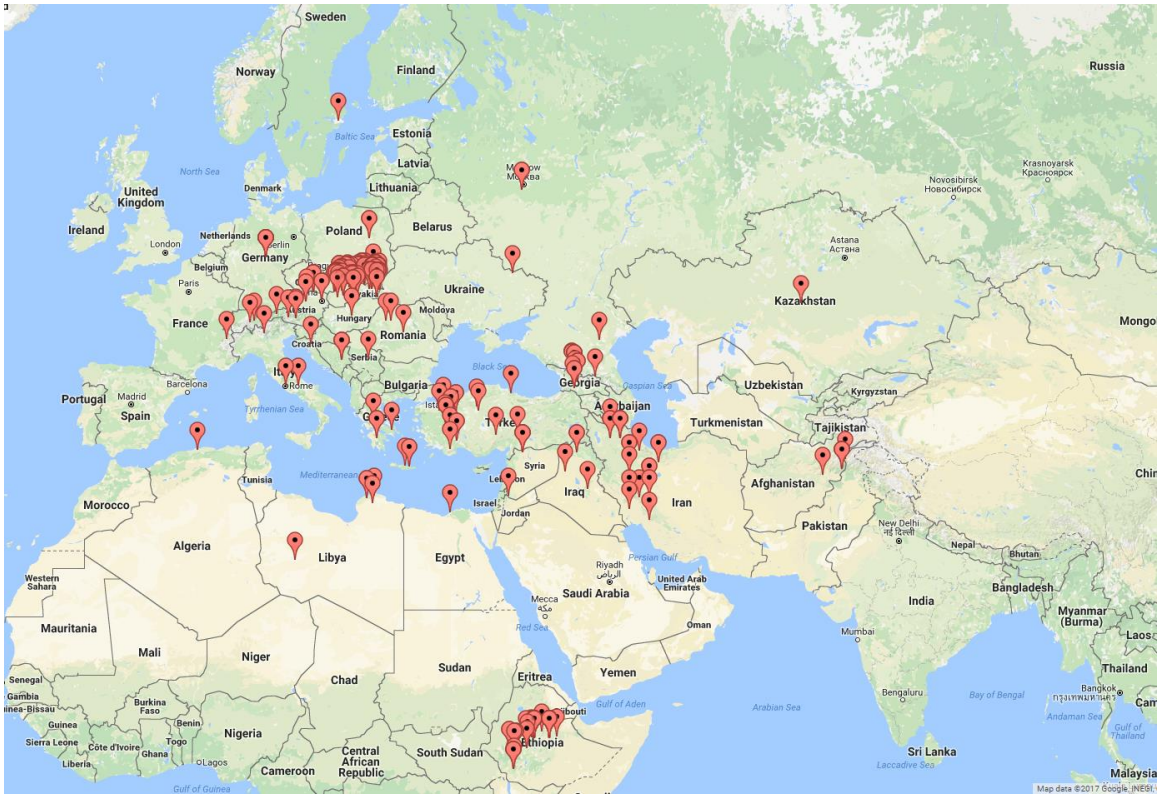


Figure 2.1. World map showing the geographic origin of all 261 accessions. Collection sites are based on latitudes and longitudes.

2.2 Sowing

For the field evaluation, 600 seeds were sown at 1.5 x 2m plots resulting in a sowing density of 250 seeds/m^2. The field trial was performed in a randomized plot design in two replications during the years 2013 and 2014 at IPK. The soil type was black soil with a pH range of 6.3 - 7.2. No growth regulators were applied during the entire experiment. To prevent cross contamination between plots during pollination and lodging, each barley plot was separated from neighbouring plots on either side by spring wheat following a chequerboard pattern (Fig. 2.2). In 2013, accessions were sown on the 16th of April while in 2014 the accessions were sown on the 20th of March. The two years showed differences in terms of temperature and precipitation (Table 2.1). At about four weeks after sowing, the fungicide Fandango (0.65/ha) + Aviator Xpro (0.65 l/ha) was applied against powdery mildew, Fastac SC (01.25 l/ha) against aphids and Pronto Plus (1.5 l/ha) against spike diseases.

Table 2.1. Average monthly weather conditions in IPK-Gatersleben during the field experiments of 2013 and 2014.

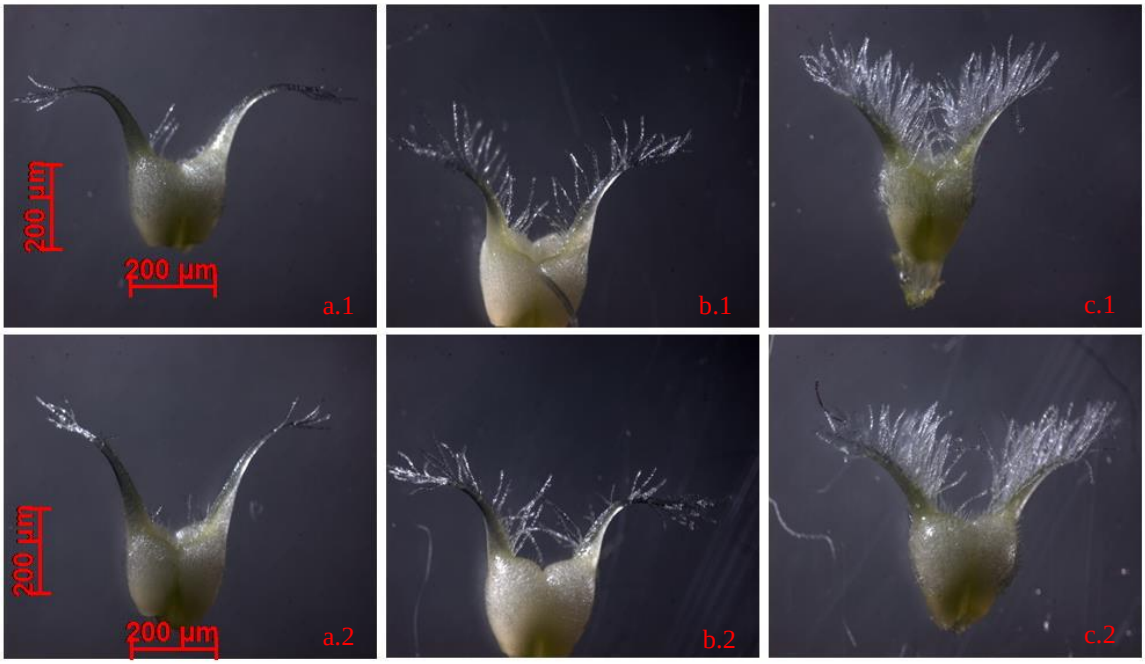
2013				2014			
Months	Temp (°c)	Precipitation (mm)	Humidity (%)	Months	Temp (°c)	Precipitation (mm)	Humidity (%)
April	8.6	22.7	74.8	March	6.8	3.9	78.4
May	12.74	103.1	81.7	April	11.4	29	77.9
June	16.8	14.4	74.1	May	12.9	99.6	77
July	20.1	33.1	72.2	June	16.3	70.2	76.4
August	18.6	26	71.2	July	20.3	101.4	77.3
Mean	15.4	39.9	74.8		13.5	60.82	77.4

Figure 2.2. Barley field trial at four months after sowing in 2013. Barley plots were separated on either side by spring wheat (red rectangles) as shown by few examples on the figure.

2.3 Phenotyping

In total, 14 agronomic traits were scored in both years (2013 and 2014). This included nine yield related and four morphological traits.

Heading date was scored as days after sowing (DAS) at GS 53 (when one-third of the spike was above the flag leaf) (Lancashire et al. 1991). Stigma hairiness was scored by dissecting three florets (middle and base of spike) taken from detached two main spikes of each plot (Fig. 2.3a and b). In order to get the exact stage in which the hairs on the stigma are clearly visible, spikes at different stages (close to anthesis and at anthesis) were collected for microscopic dissection and dissected florets were imaged with a binocular microscope (Zeiss, AxioCam 1Cc1) at x25 and further categorized as hairy (3), less hairy (2), no hairs (1).



[1]images taken from florets detached from the middle of spikes

[2]images taken from florets detached from base of spikes

Figure 2.3. Images of stigma taken under a binocular microscope at magnification x25. a represents barley stigma with no hairs (score=1), b represents barley stigma with few hairs (score =2), and c represents barley stigmas with many hairs (score = 3). a.1- c.2 indicates the position of florets in the spike: 1= taken from florets detached from the middle of spikes, 2 = taken from florets detached from the base of spikes.

Awn roughness

Awns were detached at maturity from two florets (at the middle and base) of two randomly selected spikes and images were taken with a binocular microscope (Zeiss, AxioCam 1Cc1) at x25. Saved images were later given numerical scores on a scale of 1 - 3 depending on the degree of barbs present on the awns. Barley awns with no barbs were given a score of 1; awns with few barbs a score of 2 and awns with many barbs were given a score of 3, respectively (Fig. 2.4 a, b and c).

Figure 2.4. Images of barley awns showing different densities of barbs. Images were taken at a magnification of x25. Examples from a) Barley awn with barbs (score 3). b) Barley awn with fewer barbs (score 2). c) Barley awn with no barbs (score 1).

Plant height

Plant height was recorded by measuring the height of plants from above ground at the centre of individual plots with a scaler in centimetres till the tip of the ear (without awns).

Other agronomic traits

The remaining traits were measured on three main tillers selected randomly from each plot:
Ear length (length of the spike from base of spike to tip of spike excluding the awns) and awn length (length from tip of ear to tip of awns) were measured during the harvest of those three tillers and the mean of all three measurements was recorded. The three main tillers were threshed and seeds were further used to determine the number of grains per spike (where the number of grains per spike was calculated by dividing total grain number by three). Spike density was scored as the ratio of the number of grains per spike to the ear length.

Spike density (length of internode) = number of grains per spike/ ear length

Thousand kernel weight of all seeds from the three main tillers along with seed length, seed width and seed area were measured using the grain analyser "Marvin" (GTA Sensorik GmbH, Neubrandenburg, Germany).
Harvest index was calculated on the three main tillers as the ratio of total grain yield and total above ground biomass at maturity (Huehn 1993).

Harvest index (Z) = X/Y = X/(X+S)
Where X= total grain weight, S = straw biomass, Y = total biomass

To score the presence or absence of hulls around the caryopsis (naked/covered trait), threshed seeds were visualized under a binocular microscope at x25 and scored as hulled (1) or hulless (2).

Figure 2.5. Images of barley seeds. Top panel: dorsal side of hulled seeds (left) and naked seeds (right); lower panel; ventral side of hulled seeds (left) and naked seeds (right).

2.4 Evaluation of phenotypic data

Phenotypic data were analysed by REML (Residual Maximum Likelihood) implemented in GenStat version 16.0 (Payne 2014).

To evaluate the quality of phenotypic data, the coefficient of determination r2 for each trait was calculated between the two replicates of each year and between the two different years while correcting for environmental effects between the two different years in a mixed-linear model (MLM) implemented in GenStat 16.0. For this, genotype x year (G x Y) was taken as a fixed effect. Trait means for both seasons were further displayed on a histogram by using the software package PAST 03.5 (Øyvind et al. 2015). By considering replicate effects as random effects and year effects as fixed effects, variance contribution by genotype (G), genotype by year (G x E or G x Y) and year (Y) was estimated by ANOVA (Analysis of Variance) in the software package GenStat 16th edition. Percentage variances were then further plotted on a histogram to evaluate the main components responsible for phenotypic variation within the association panel for each phenotypic trait.

To compute an adjusted phenotypic mean for each trait across both seasons, BLUES (Best Linear Unbiased estimates) were computed for each genotype while considering genotype x year (GxY) and genotype x replicate (G x Rep) interactions.

2.5 Broad sense heritability

Broad sense heritabilities (H^2) for all 14 traits were computed in GenStat 16.0 by incorporating genotypic variance over the total phenotypic variance, considering genotype by environment variance, the number of environments or seasons and number of replicates according to Nyquist (1991).

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{g\times e}^2}{e} + \frac{\sigma_e^2}{re}}$$

σ^2_g represents the genotypic variance, $\sigma^2_{g\,x\,e}$ the genotype × environmental variance, *e* is the number of environments (or seasons in this study), σ^2_e the environment variance, and *r* the number of replicates per environment.

To compute an adjusted phenotypic mean for each trait across both seasons, BLUES (Best Linear Unbiased estimator) were computed for each genotype while considering genotype x year (GxY) and genotype x replicate (G x Rep) interactions.

2.6 Genotyping

Illumina Infinium assay (9K iSelect chip)

All accessions used in the current study were genotyped using the iSelect 9K chip (Illumina, San Diego, USA) with SNP content and selection criteria described in Comadran et al. (2012). For genomic DNA extraction, about 5g of fresh leaf material from two weeks old seedlings was harvested, shock-frozen in liquid nitrogen and stored at -80 °C for DNA extraction. DNA was extracted and concentration measured according to the Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep protocol (Clarke 2009), and sent for genotyping to Trait Genetics GmbH (Gatersleben, Germany). SNP markers that failed in more than 5% of the genotypes were excluded. Individual SNPs which passed the above mentioned criteria were assigned to genetic positions by using the POPSEQ genetic map (Mascher et al. 2013) and the Morex x Barke genetic map (Comadran et al. 2012). In total,

5711 informative SNPs passed the filtering criteria, with 4801 markers having mapping positions leaving a total number of 910 unmapped SNPs. Only mapped SNPs were used for further analysis. In GWAS, a minor allelic frequency (MAF) threshold, MAF < 0.05 was set for which SNPs that did not meet this criteria were further excluded. This resulted to 4790 SNPs considered in GWA analysis.

2.7 Comparison of SSR and SNP makers in genetic relatedness study

The association panel consisted of 261 accessions which were a subset of a diverse core reference set (LRC648) earlier genotyped with 42 SSR markers (Pasam et al. 2014). A kinship matrix was first generated in the software package GenStat 16th edition (Payne 2014) using sequence information from 42 SSR and 4801 SNP markers, respectively. GenStat computes pair-wise genotypic comparison for all individuals at each marker locus returning the result as numeric values in a matrix table for each pair of genotypes. A score of 1 is given to accession pairs which are 100% identical at all marker loci and a score of zero is given to pairs which show no identity at any of the marker loci. From the matrix table, individuals which are 100% identical (having a score of 1) can be considered as potential duplicates. To compare genetic relatedness between accessions based on SSR and SNP markers, separate neighbour joining tree net-works were computed in the software package PAST3.0 using SSR and SNP marker information, respectively. The neighbour net computes a phylogeny tree which shows clusters of individuals as a result of how closely related they are based on the sequence information provided. Potential duplicate accessions will cluster on same spots on a neighbour-net tree.

2.8 Polymorphic information content

Polymorphic information content (PIC), major allele frequency, minor allele frequency (MAF) and gene diversity were calculated for each SNP (4801 SNPs), for and among all accessions using the software package Powermarker V 3.25. according to Liu and Muse (2005).

$$PIC_i = 1 - \sum_{J=1}^{n} P_{ij}^2$$

In this formula, P_{ij} is the frequency of the *j*th allele for marker *i*.

2.9 Population structure

In order to determine the number of subgroups within the association panel (261 accessions), three different approaches were implemented; i) principal component analysis was computed based on SNP marker information of all 4801 SNPs; ii) a neighbour-net and a neighbour joining tree were computed in the software package Splits Tree4 (Huson and Bryant 2006); ii) population structure was calculated with the software STRUCTURE (Pritchard et al. 2000b).

Principal component analysis

At first, a PCA that uses orthogonal transformation to convert a set of correlated variables into a set of linearly uncorrelated variables called principal components was performed by linear transformation of SNP data (all 4801 SNPs) into a new coordinate system in the software package PAST 3.05. The greatest variance lies on the first coordinate (first principal component), the second greatest variance on the second coordinate and so on.
In PCA, new variables Y_i that are a set of linear combination of the original variables X_i are calculated according to a statistical technique (Hotelling 1901).

$$Y_i = a_{i1} x_1 + a_{i2} x_2 + \ldots a_{ip} x_p \; ; \qquad i = 1 \ldots\ldots\ldots p$$

$x_1, x_{2 \ldots . Xp}$ can be viewed as the explanatory variables for a dependent variable Y_i, with loadings a_{i1} and a_{i2} as the regression coeficients.
The new variables *Yi* are derived in decreasing order of importance and are called principal components. The new variables or PCs have a variance equal to their corresponding eigenvalues.

$$Var(Y)_i = \lambda_i \quad \text{for all } i = 1 \ldots p$$

Small $\lambda_i \Leftrightarrow$ small variance $\Leftrightarrow$ data change little in the direction of component Y_i
The relative variance explained by each given PC is given by

$$\lambda_i \; / \Sigma \lambda_i$$

New variables Y_i are then presented in cluster format corresponding to their relative variance (Enki et al. 2013).

Neighbor-net and neighbor joining tree

A neighbor-net network was computed based on all 4801 SNP markers in the software package Splits Tree4 (Huson and Bryant 2006) for all 261 accessions. The split tree computes a phylogenetic tree network (neighbour-net) which represents differences within and between data sets. To study the number of sub-groups within the association panel and also to establish the relationship between the accessions, additional information (country of origin, presence or absence of caryopsis) was inferred to the neighbour-net diagram. Accessions were determined to clusters as a result of stratification if they share one or both of the additional information inferred onto the neighbor-net (country, and/or presence or absence of caryopsis).

Population structure by Q-matrix

The program STRUCTURE uses an alternative model to account for population structure developed by (Pritchard et al. 2000b) by assigning individuals into different groups known as Q-groups. STRUCTURE aims at delineating clusters of individuals on the basis of their genotypes at multiple loci using a Bayesian approach by assigning accessions to an assumed number *(K)* of different sub-groups minimizing LD and maximizing gametic-phase equilibrium within the population. The program was run for all 4810 markers to test for sub-group numbers from K=1 to 20 with 10 individual replications per K. For this, the admixture model with uncorrelated allele frequency settings was implemented with a burn-in-length of 20,000 iterations followed by 10.000 Markov Chain Monte Carlo iterations. At first, different burn-in iterations (20.000, 15.000, 10.000, and 5.000) were tested to determine the number of iterations which was sufficient to reach the maximum likelihood curve produced by the STRUCTURE software. For each K run, STRUCTURE produces a Q-matrix (Q_{ST}) which lists the membership coefficients of each accession in each group. The model choice criterion in STRUCTURE to determine the true K is an estimate of the posterior probability of the data for a given K, $\Pr(X \mid K)$ (Pritchard et al. 2000b), and is called LnP(D) or L(K).

The most likely number of sub-groups K was identified by applying the approaches of (Evanno et al. 2005a) and (Rosenberg et al. 2005). (Evanno et al. 2005a) proposed an ad hoc

statistic ΔK which determines the break in the slope of the LnP(D) probability function provided by STRUCTURE.

$$\Delta K = m\,(|L\,(K + 1)-2L\,(K) + L\,(K-1)|)/S\,[L\,(K)],$$

K denotes the assumed number of subgroups, L refers to the average of LnP(D) for the ten replications of the K^{th} STRUCTURE run, and *S* denotes the average standard deviation of the ten replications of the K^{th} STRUCTURE run. In order to estimate the average clusteredness of an individual or the extent to which an individual belonged to a single cluster rather than to a group of clusters, each STRUCTURE run computed a quantity according to the equation summarized by (Rosenberg et al. 2005);

$$G = \frac{1}{I}\sum_{i=1}^{I}\sqrt{\frac{K}{K-1}\sum_{k=1}^{K}(q_{ik} - 1/K)^2},$$

q_{ik} denotes the estimated membership coefficient for the *i*th individual in the *k*th cluster, *I* denotes the total number of individuals (261), and *K* denotes the total number of clusters (1 to 20). The factor $K/(K-1)$ was included so that a change in *K* would not produce a systematic change in clusteredness.

2.10 Linkage disequilibrium (LD) decay

A heatmap was generated for pairwise correlation r^2 between all mapped SNPs in the association panel by using TASSEL 3.0 (Trait Analysis by Association, Evolution and Linkage) (Bradbury et al. 2007). Pairwise SNP correlations r^2 were obtained in TASSEL from permutation tests at each marker locus. This was calculated by using the default settings of TASSEL 3.0 and 1000 iterations. From the heatmap, regions of high and low LD could be detected along various chromosomes (pair-wise correlation r^2 determined to be statistically significant at $P<0.001$). Intra-chromosomal LD decay was calculated for individual chromosomes separately in GenStat 16th edition (Payne 2013) by using squared allele frequency correlations r^2 between pairs of loci and the genetic distance between pair-wise markers (Rohlfs and Weir 2008) to generate LD plots. To investigate average LD decay across the entire genome in the association panel, pairwise correlation r^2 for all chromosomes generated in TASSEL 3.0 were plotted against the genetic distances in centi-Morgans (cM)

between the markers in GenStat 16th edition (Payne 2013). A second degree smoothing curve (Loess curve) was fitted to the plot and a critical r^2 value was derived from the distribution of r^2 values of unlinked markers (markers which were >50 cM apart) by square root transformation of those r^2 values to obtain a normally distributed random variable. The parametric 95th percentile of that distribution was then taken as the population-specific critical value of r^2, beyond which LD was likely to be as a result of genetic linkage. The point where the intersection of the loess curve fit to the syntenic r^2 was then considered as an estimate of average LD decay (Breseghello and Sorrells 2006).

2.11 Genome-wide association analysis

Corrected phenotypic trait means (BLUEs values) were computed for individual traits and used for GWAS analysis. All analyses were performed in TASSEL 2.1 (Bradbury et al. 2007) and results were further confirmed by re-analyzing the data in GenStat 16th edition (Payne 2013) to check for any discrepancies. In order to control for confounding effects which could yield false positive associations, two independent statistical approaches were tested. In the first approach, principal component analysis and the kinship matrix (PK) was used to control for confounding effect in a MLM while in the second approach, only genetic relatedness (kinship-matrix) was used to control for spurious association in a MLM. To determine best model performance, results from both approaches were compared.

The mixed linear model is an extension of the general linear model and follows the following equation (Yu et al. 2006),

$$y = X\tau + Zu + Z_g gj + e,$$

y denotes the data collected across *s* environments, *X*, *Zu* , and *Zg* are design matrices with *X* being of full rank, τ and *u* are vectors of fixed and random effects that model the experimental design and non-genetic effects, and *e* is a vector of residuals. Genetic relatedness (K) was generated in TASSEL 2.1 from SNP data and then fitted into the PK model.

Marker trait associations were plotted on Manhattan plots with P-values transformed to –logP values. For each trait, corrected P-values (q-values) were computed from the calculated GWAS P-values in the software R 2.15.3 (R Core Team 2013). From the calculated q-values, an FDR of 5% was implemented as a threshold in which marker-trait associations above this

threshold were considered to be significant. Nevertheless, not all traits yielded associations surpassing the calculated FDR threshold. In this case a traditional cut off P-value of $P \leq 0.001$ (-logP=3) was used.

3 Results

A collection of 261 two-rowed spring barley landraces was evaluated in the field in 2013 and 2014 in two replicates at IPK. Fourteen agronomic traits were scored with the purpose of identifying new candidate loci by GWAS. Due to the identification of potential duplicates, only 199 accessions were considered for phenotypic analysis.

The landrace collection exhibited a large phenotypic variation for all agronomic and morphological traits. Though differences were seen across both years for some measured agronomic traits, medium to high heritability values were recorded for all traits indicating the suitability of the data for performing GWAS.

3.1 Heading date

The landraces exhibited large variation in heading date in both years (Table 3.1). On average, heading started six days later in 2014 despite the earlier sowing. Nevertheless, maximum heading date for both seasons remained the same.

By constructing a neighbor joining tree with values of heading data scored across both years, accessions could be assigned to two groups (Fig. 3.1). One group which headed early was dominated mainly by accessions from Africa and Asia with an average heading date of 65 days after sowing (DAS) while the second group containing late heading accessions was mainly of European origin with an average heading of 71 DAS. The grouping of accessions is likely due to the difference in sensitivity at the pseudo-response regulator *Ppd-H1* locus which provides adaptation to photoperiod in barley (Turner et al. 2005). The causal SNP is included in the marker set used for GWAS. Spring barley lines from Western Europe show a reduced response to photoperiod thereby extending the period of vegetative growth to accumulate higher biomass which leads to higher yield. On the other hand, most spring accessions from Africa and Asia headed very early due to higher response to photoperiod as a result of increased sensitivity of the pseudo-response regulator *Ppd-H1*.

Although there was a big difference in heading date across the two years, both data were highly correlated indicating that accessions which headed early in 2013 also headed early in 2014 (Table 3.1).

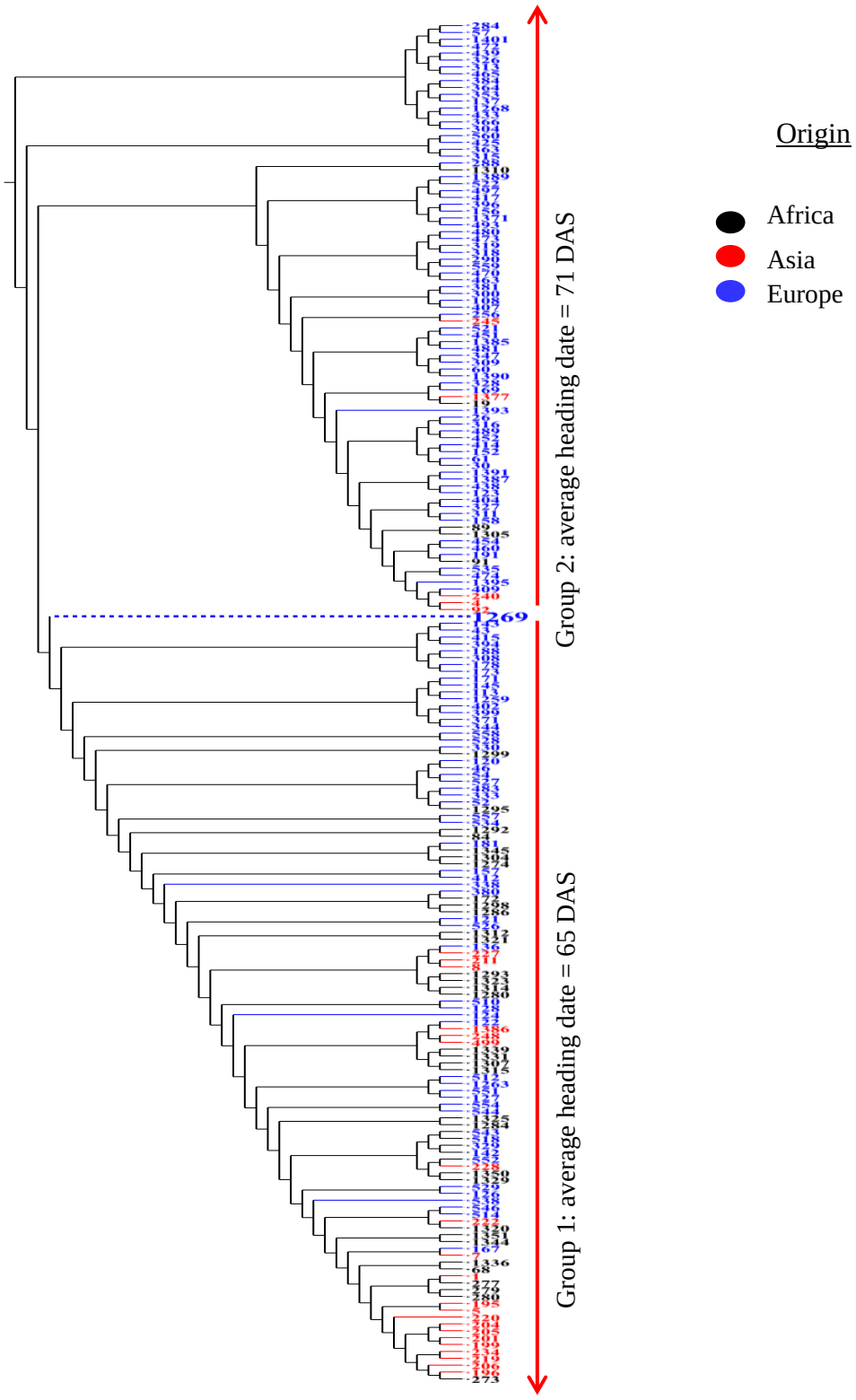


Figure 3.1. Neighbour joining tree of all 199 accessions generated based on heading date data from 2013 and 2014 field trials. The longest branch splits the accessions into two groups of early and late heading genotypes.

3.1.1 Other agronomic traits

A summary of descriptive statistics across both years is presented in Table 3.1. Correlations between the two years for all traits were calculated using Best Linear Unbiased Estimate values (BLUEs). Highly significant correlations (P<0.001) were recorded for all agronomic traits ranging from 0.59 for seed width to 0.95 for heading date (Table 3.1). Grains harvested and analyzed in 2013 had smaller seed width and seed area as compared to grains harvested in 2014. Broad sense heritability estimates were estimated for all traits and ranged from 0.79 - 0.95 indicating a high robustness of the phenotypic data and the suitability of the data for GWAS (Table 3.1).

Table 3.1. Descriptive statistics of eleven agronomic traits scored across two years. Minimum, maximum, mean, standard deviation (Sd) and coefficient of variation (Cv), correlation of both years (r^2) and broad sense heritability (h^2) across both years. Stigma hairiness, awn roughness and naked/hulled traits are not included in the table since they were scored as qualitative traits during both years.

	2013				2014				2013/2014		
Traits	Min	Max	Mean	Sd	Min	Max	Mean	Sd	r^2	h^2	*Cv*
HD (DAS)	50	83	66.5	4	63	82	72.5	4.7	0.85	0.77	0.07
Pht (cm)	55	106	80.5	10.7	52.5	112.5	82.5	11.5	0.8	0.89	0.15
GPS	10	30.8	20.4	4.2	7	31.5	19.3	4	0.81	0.91	0.18
HI	0.26	0.66	0.46	0.06	0.24	0.57	0.41	0.07	0.65	0.87	0.14
SPD	0.27	0.63	0.41	0.06	0.25	1	0.38	0.08	0.65	0.74	0.17
TKW (g)	35.9	76.8	56.4	6.2	23.4	64.6	44	5.1	0.66	0.83	0.1
AwnL (cm)	0	17.5	8.75	3.1	0	16.6	8.3	1.2	0.62	0.76	0.2
EL (cm)	5	12.8	8.9	1.6	3.58	12	7.8	1.2	0.73	0.88	0.15
SDA (mm^2)	24.3	48.6	36.4	3.7	17.8	31.7	24.8	2.6	0.84	0.92	0.11
SDL (mm)	7.5	16.9	12.2	1.2	6.8	12.9	9.9	1.1	0.91	0.94	0.11
SDW (mm)	3.2	5	4.1	0.16	3.8	4	3.9	0.14	0.59	0.65	0.04

*HD "heading date", Pht "plant height", GPS "grains per spike", HI "harvest index", SPD "spike density", TKW "thousand kernel weight", AwnL "awn length", EL "ear length", SDA "seed area", SDL "seed length", SDW "seed width".

Pairwise trait correlations are shown in Table 3.2. Grains per spike was highly correlated to ear length and spike density while thousand grain weight was highly correlated to seed area,

seed length and seed width. A high correlation r^2=0.92 was also observed between seed area and seed length. Plant height was moderately correlated to grains per spike and ear length but negatively correlated to harvest index. Significant correlation was also observed between grains per spike and awn roughness. Significant negative correlations were observed between harvest index and plant height and also harvest index and presence/absence of hulls which is likely as a result of the absence of hulls in hull-less barley which leads to a reduction in seed weight compared to hulled barley.

Table 3.2. Pairwise trait correlations r^2 between phenotypic data of 14 agronomic traits recorded across two years. Correlations r^2 were computed with BLUE values of respective traits.

	Correlation													
HD	-													
Pht	0.21	-												
GPS	0.72*	0.42*	-											
HI	0.07	-0.40*	0.33*	-										
SPD	0.23	0.06	0.51*	0.42*	-									
TKW	0.00	0.04	-0.13*	0.21*	-0.11	-								
AWNL	0.30*	0.06	0.17	-0.15	0.03	0.11	-							
EL	0.63*	0.43*	0.72*	-0.04	-0.22	0.00	0.20	-						
SDA	-0.32*	-0.10	-0.42*	0.15	-0.20	0.71*	-0.08	-0.33*	-					
SDL	-0.34*	-0.22	-0.41*	0.22	-0.23	0.52*	-0.17	-0.41*	0.9*	-				
SDW	0.21	0.12	0.13	0.05	0.03	0.60*	0.20	0.15	0.37	0.04	-			
AwnR	0.16	0.06	0.22*	-0.11	0.18	-0.12	0.36*	0.13	-0.30*	-0.33*	0.00	-		
Stig.H	0.24	0.11	0.31	-0.05	0.21	-0.21	0.20	0.20	-0.27	-0.32	0.05	0.32	-	
N-C	-0.09	0.10	-0.23	-0.71	-0.32*	-0.22	0.17	0.04	-0.31*	-0.51*	0.02	0.11	0.22	-
	HD	Pht	GPS	HI	SPD	TKW	AwnL	EL	SDA	SDL	SDW	AwnR	Stig.H	N-C

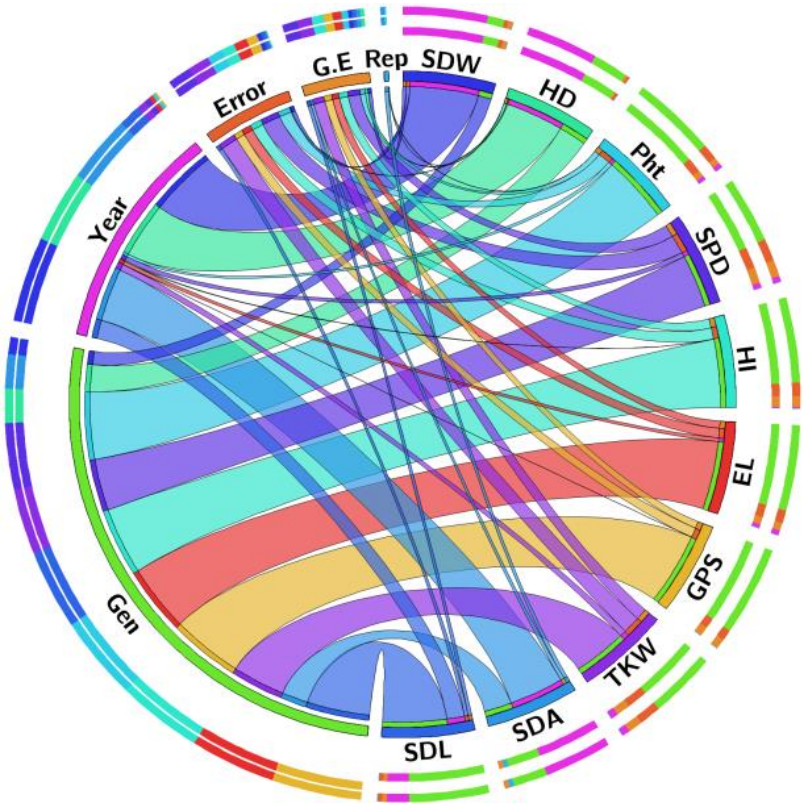
*Significant at P ≥ 0.001

HD “heading date”, Pht “plant height”, GPS “grains per spike”, HI “harvest index”, SPD “spike density”, TKW “thousand kernel weight”, AwnL “awn length”, EL “ear length”, SDA “seed area”, SDL “seed length”, SDW “seed width”,Stig.H “stigma hairiness”, N-C “ naked caryopsis”.

3.2 Variance components explained by Genotype, Genotype x year

With a significant difference between trait data calculated in 2013 and 2014 for most of the quantitative traits, variance contributions of genotype, genotype x environment and replicate effects were calculated using ANOVA (Fig. 3.2). Unless for heading date, seed area (SDA) and seed width (SDW), most of the variation across both years was explained by the genotypic difference. The effect of the difference in weather conditions between both seasons (Y) was highly significant for heading date (HD), seed area (SDA) and seed width (SDW), and contributed to 64%, 61%, and 77% of the difference in measured data across both years, respectively.

Nevertheless, the high heritability estimates observed for heading date, seed area and seed width implies that the environmental effect was equally distributed across genotypes in both years and hence the high correlation also observed between data for both years.



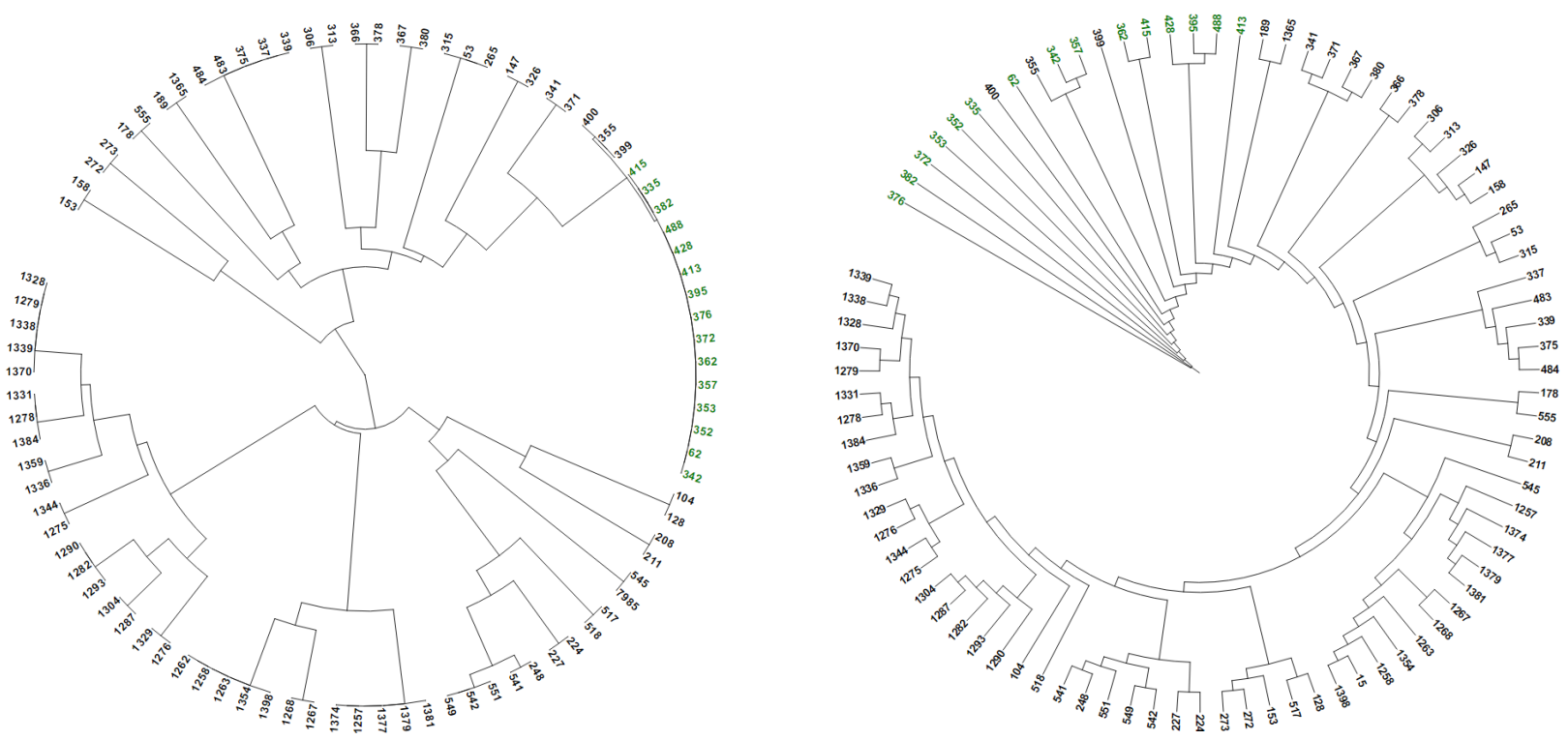
*GPS= grains per spike, EL= ear length, HD= heading date, HI= harvest index, Pht= plant height, SDA= seed area, SDL= seed length, SDW= seed width, TKW= thousand grain weight, SPD= spike density.

Figure 3.2. Interactive circos plot showing the influence of five factors, Genotype (G), Environment (year), Genotype x Environment (G.E) and Replicate (rep) on ten quantitative traits. Innermost circles represent the traits (right half) and different factors (left half). The size of interactive ribbons indicates the effect contribution of each of the five factors on respective traits. The length of each color on the

two outermost circles (right of figure above each respective trait) represents the size interaction between respective factors and various traits.

3.3 Comparison of SSR and SNP makers in genetic relatedness study

The present collection was previously genotyped with 42 SSR markers evenly distributed across the seven chromosomes (Pasam et al. 2014). A kinship matrix was generated based on all 5711 SNPs (without filtering for MAF) to study the genetic relatedness amongst all 261 accessions. In total, 30 groups of two or more genetically identical accessions involving a total of 88 accessions were identified (Fig. 3.3a, Table S1, and S2). These accessions were considered to be potential duplicates, with the smallest potential duplicate groups consisting of two accessions while the largest group contained 12 accessions. A closer look at all fourteen measured traits revealed that, members belonging to same cluster of potential duplicates also showed very high similarities (Table S2). According to the SSR markers, those accessions were not completely identical. A dissimilarity matrix was calculated for cluster development using a neighbor joining (NJ) method. An unrooted tree (Fig. 3.3.a,b) and a phylogram were generated to investigate and visualize the genetic relationships amongst all 88 potential duplicates based on either the 42 SSR or the 5711 SNPs (without filtering for MAF. Although the NJ tree generated with SSR markers did not show any potential duplicates, the corresponding accessions grouped close to each other. Inspection of passport data (GBIS/I) (http://gbis.ipk-gatersleben.de/) also revealed that accessions from the same cluster of potential duplicates were mostly of same geographic origin. For example, the largest group of potential duplicates consisted of 12 accessions (Fig. 3.3.a, Table S1). A majority of these accessions (11) came from Slovakia with a single accession originating from the Czech Republic. Moreover, passport data revealed that most of the accessions from Slovakia were collected in nearby locations during collecting trips in 1974 and 1977. However, some potential group of duplicates comprised of accessions from different geographical origin. Examples are duplicate groups 9 and 24 (Table S2). Accessions from group 9 originated from Austria, Ethiopia, CSFR and Russia. Group 24 on the other hand comprised of two accessions originating from Greece and Ethiopia which however, were hulled and hull-less, respectively. In this case, both accessions were dropped from the panel in the final GWAS, because of a likely mix-up of DNA used for finger-printing.



(a) (b)

Figure 3.3. Neighbor joining tree of 88 barley accessions representing 30 potential duplicate groups. Each accession is represented by the entry number with accession names presented in Table S2. a) Accessions were grouped into 30 duplicate groups based on 5711 SNPs and b) 42 SSR markers. Accessions in green belong to the largest group of potential duplicates comprising 12 accessions.

For each of the 30 groups of potential duplicates, one accession was selected to represent the whole group for further GWA analysis with the selected accession originating from same country as the majority of the accessions within the group. In case of only two potential duplicates originating from two different countries, an accession was selected at random to represent the whole group. In case of group 24 both accessions (one hulled, one hull-less) were removed from the panel. Subsequently, 199 accessions were finally considered for GWA analysis.

3.4 Marker distribution and polymorphic information content

Of all 4801 informative mapped SNPs considered for the 199 accessions, 11 had MAF < 0.05 and where excluded from further GWA analysis. A high percentage of the remaining 4790 SNP markers (>25% of total markers) had MAF within the range of 0.4 - 0.5 (Fig. 3.4). The total length of individual chromosomes, number of markers, average PIC and mean genetic diversity values are shown in Table 3.3. All seven chromosomes had mean PIC values ≥ 0.29. This is an indication that the markers were highly diverse and hence very informative. Genome-wide distribution of SNP markers across the seven chromosomes of barley was not uniform. Chromosome 1H had the minimum number of markers (453 SNPs) with an average of 3.41 markers per cM while chromosome 5H had the highest number of markers with an average of 5.6 markers per cM. The shortest chromosome was chromosome 4H with a total length of 114.3 cM while chromosome 5H was the longest having a genetic length of 169.5 cM.

Table 3.3. Summary table of marker coverage, mean PIC values of 4790 SNPs markers across the seven chromosomes of barley.

Chromosome	Length cM	No of Markers	Marker %	Coverage/ cM	Mean PIC
1H	132.8	453	9.46	3.41	0.30
2H	149.4	816	17.04	5.46	0.31
3H	154.9	758	15.82	4.89	0.31
4H	114.3	495	10.33	4.33	0.30
5H	169.5	965	20.15	5.69	0.29
6H	126.6	658	13.74	5.20	0.31
7H	140.7	645	13.47	4.58	0.30

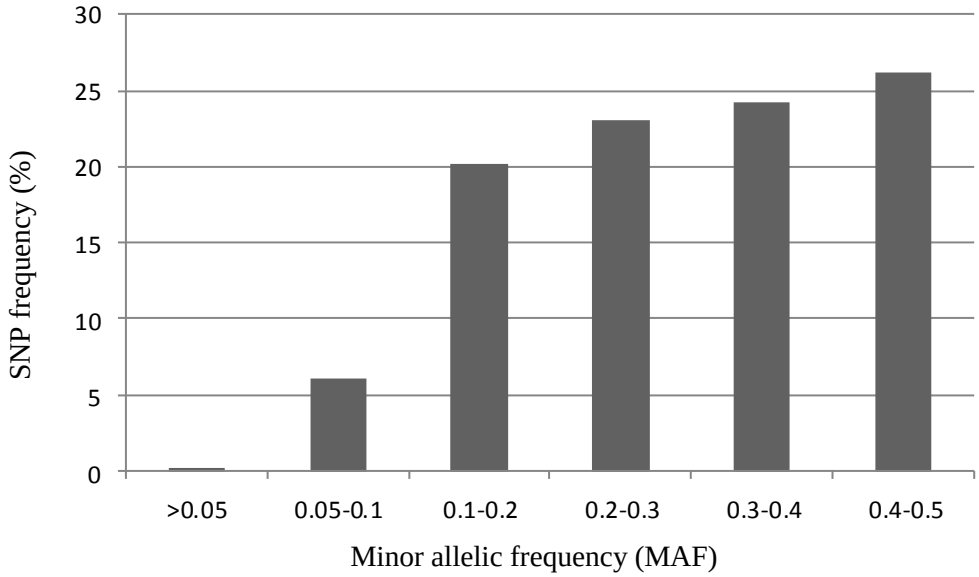


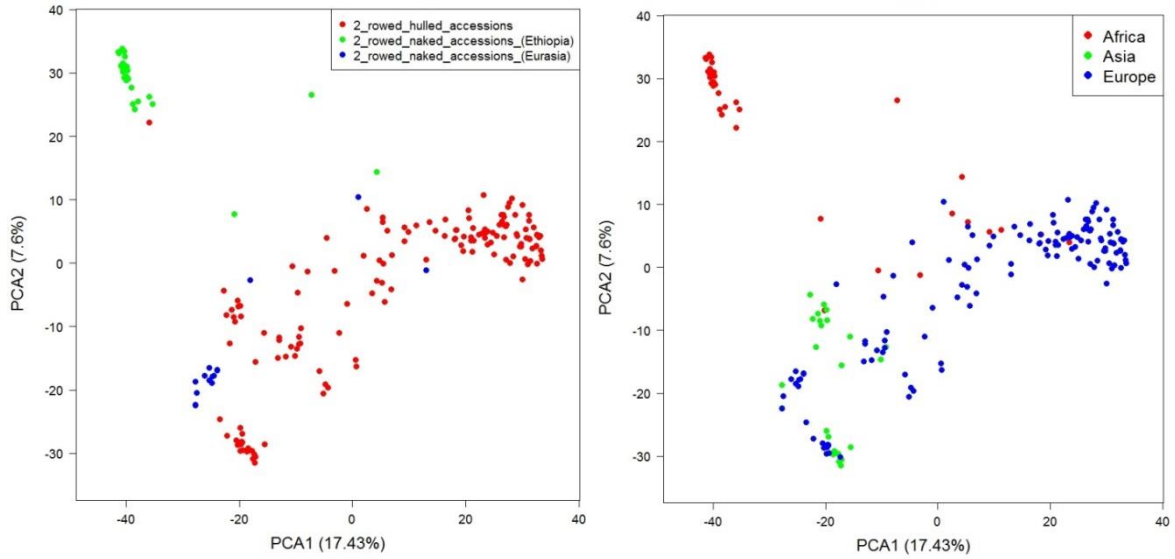
Figure 3.4. Distribution of SNP frequencies within the population of 199 spring barley landraces considered for GWAS.

3.5 Population structure

Three different methods were used to determine the number of sub groups within the association panel.

Cluster analysis using PCA approach

A principal component analysis (PCA) was computed for all 199 genotypes based on 4790 SNPs (SNPs below the MAF threshold excluded) to study population structure. Of the 199 genotypes considered for association analysis, 156 were hulled with most of the accessions originating from Slovakia. Only 43 accessions were naked with most of the naked accessions originating from Ethiopia. The first two principal components explained about 24% of the total variation with the first PC explaining 17.43% of the variation and the second PC explaining 7.6% of the total genetic variation (Fig. 3.5). In order to capture 50% of the total genetic variation within the genotypes, 18 PCAs' would have to be considered. In the PCA plot, a continuous distribution was observed within the hulled accessions with genotypes belonging to same country of origin close to each other. However, hull-less barley lines from Ethiopia showed a clear differentiation from other hulled and hull-less barley from Europe and Asia by forming a unique cluster far from all accessions on the PCA plot. This unique clustering of Ethiopian naked barley far from other accessions accounted for much of the variation observed in PC1.



(a) (b)

Figure 3.5. Population structure analysis of 199 spring barley landraces based on principal component analysis (PCA) shows clustering of accessions based on (a) presence or absence of caryopsis, (b) geographical origin. Eurasia = Accessions from Europe and Asia.

Neighbor-joining net

Results from the neighbor net were similar to PCA results. After inferring country of origin and the phenotypic trait presence/absence of caryopsis to the neighbor net diagram, accessions were seen to cluster predominantly according to their country of origin. A small group of naked barleys comprising mainly of genotypes from Europe and Asia formed a distinct cluster far from naked accessions from Ethiopia but were close to the hulled accessions (Fig. 3.6). This result was consistent with results from PCA analysis.

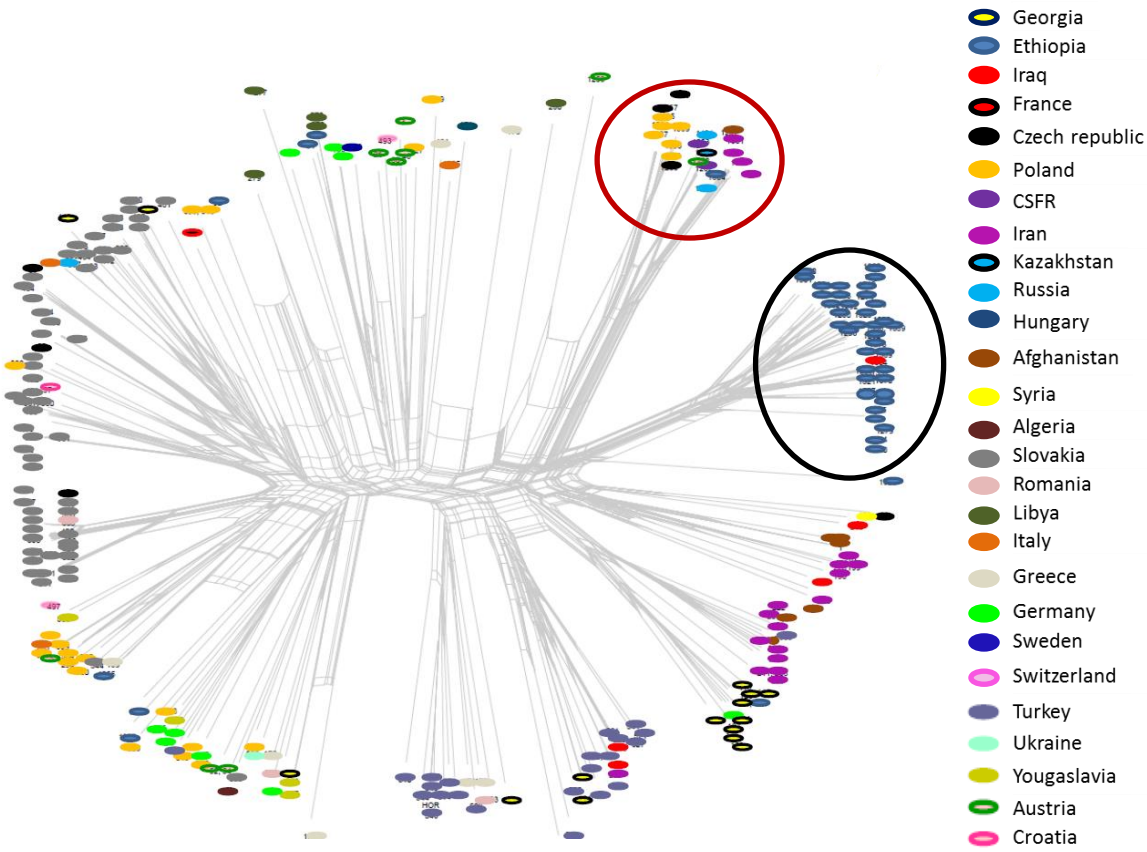


Figure 3.6. Neighbor joining net of all 199 spring barley accessions generated with splits tree software. Geographical origins of accessions are indicated by different colors. Two groups of naked accessions are indicated with black circle (naked accessions from Ethiopia) and red circle (naked accessions from Europe and Asia).

STRUCTURE (Q-matrix)

To determine the maximum number of sub groups within the association panel, population structure was analyzed on all 199 accessions using the software package STRUCTURE. The program was tested from *K*=1 to 20 with 20 iterations per *K*. An accession was assigned to a *K* group if ≥60% of the genomic information was estimated to belong to this group. Delta K (the mean log probability of the likelihood that an accession will belong to a group) was plotted against the number of subgroups. A maximum increase in Delta *K* was observed at *K*=2 (Fig. 2.12). At *K*=16, delta K reached its second maximum before flattening until *K*=20. This significant increase in delta K at *K*=2 indicates that the association panel could be separated into two major subgroups (Fig. 3.7). The two main groups could be further divided into 16 subgroups hence a second increase in delta *K* was observed at *K*=16.

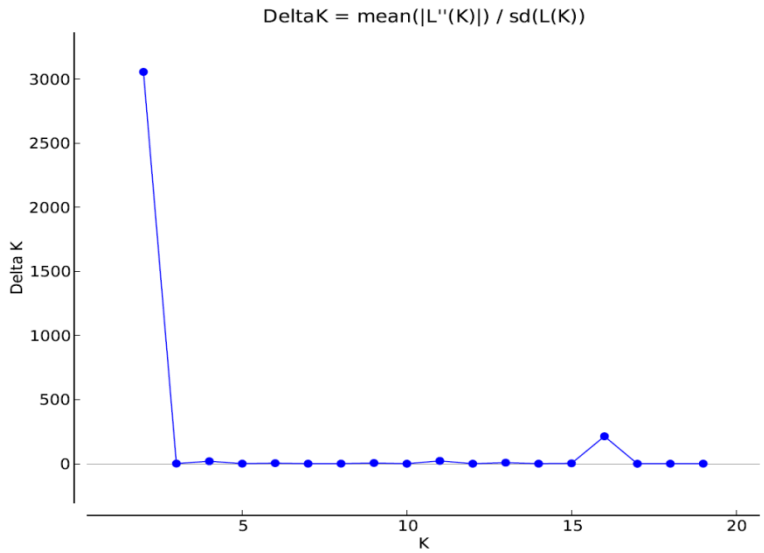


Figure 3.7. Plot of mean likelihood of delta *K* against the number of *K* groups. The highest peak observed at *K*=2 signifies the grouping of accessions into two groups while the small peak at *K*=16 signifies further grouping of accessions into 16 groups.

Q-matrix plot from STRUCTURE analysis at *K= 2* revealed the accessions to cluster mainly according to the caryopsis phenotype and geographic origin. The first group (Group 1) composed mainly of the hull-less accessions while the second group is composed mainly of hulled accessions (Fig. 3.8). Of all 199 accessions, 43 genotypes were naked with 27 of the 43 genotypes originating from Ethiopia. For all naked accessions from Ethiopia, a high percentage of the genomic information belonged to group 1, followed by naked accessions from Europe and Asia also showing a high percentage of genomic information belonging to group 1 as well. Group 2 was dominated by hulled accessions with only two naked accessions also grouping along the hulled accessions in this group. Of the hulled accessions in group 2, 39 were from Slovakia and seemed to be very unique since a majority had estimated membership coefficient of 100%. The clustering of accessions at *K*=2 was consistent with the result obtained from PCA analysis and split tree (neighbor joining net) confirming the reproducibility of these methods to detect the number of sub-groups within the association panel. Subgrouping of accessions at *K*=16 was mainly based on country of origin (extra supplementary) and was in concordance with the results obtained with split tree analysis.

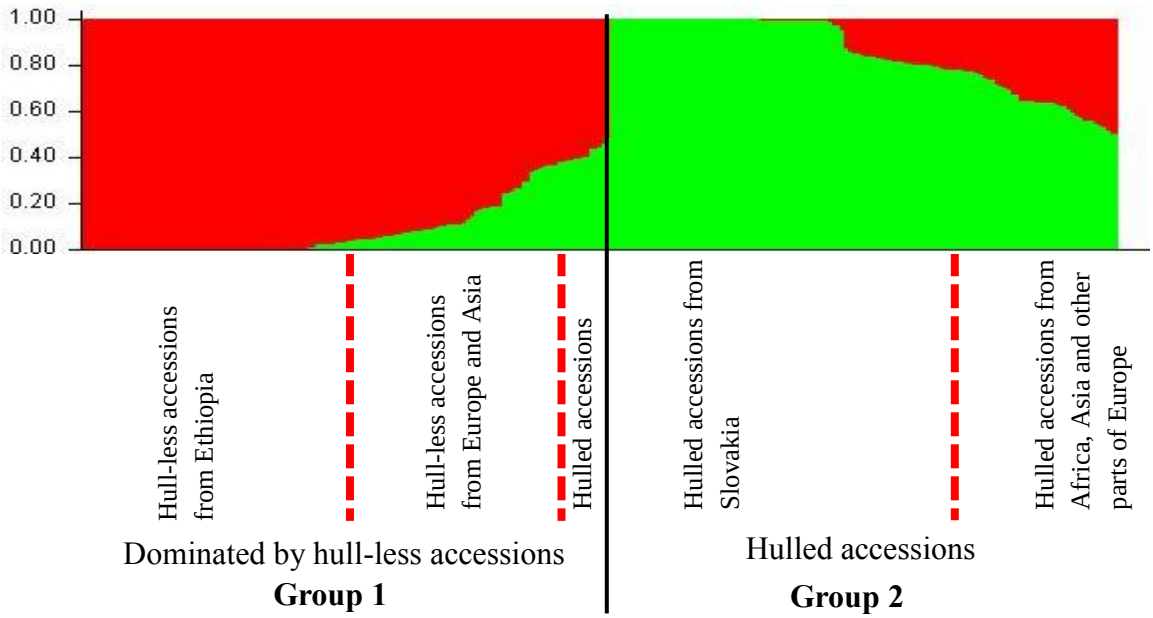


Figure 3.8. Q-matrix plot of STRUCTURE analysis at *K*=2. All 199 accessions are represented by the two colors. The separation of accessions into subgroup 1 and subgroup 2 was based on membership coefficient ≥ 0.6. Accessions are sorted first by membership coefficient.

3.6 Linkage disequilibrium (LD) decay

To study the extent of LD decay within the association panel of 199 landraces, pair-wise correlations between all 4790 markers were generated for each chromosome separately and results plotted against the genetic distance between marker pairs. The option Eigenanalysis was selected when computing LD to avoid background LD which could arise from population sub-stratification. LD varied along each chromosome with regions of high LD interspersed with regions of low LD. Pair-wise correlation r^2 was found to decay rapidly with genetic distance along each chromosome. To estimate the average genome-wide LD decay, a critical r^2 value (0.15) was calculated for all unlinked loci pairs (> 50cM). Beyond this threshold value, LD was assumed to be caused by genetic linkage. For all chromosomes, average intra-chromosomal LD decay ranged between 1 - 4.5 cM with high r^2 values observed for marker pairs within 6 cM distance and dropping drastically beyond 6 cM. Marker pairs beyond 6 cM were considered not to be in LD anymore and hence high correlation (r^2) beyond 6 cM was not considered the result of genetic linkage. Average genome-wide LD decay was determined by fitting a *Loess* curve and the point of interception between the *Loess* curve and the critical r^2 value (average genome-wide LD decay) was estimated to be around 2.5 cM (Fig. 3.9).

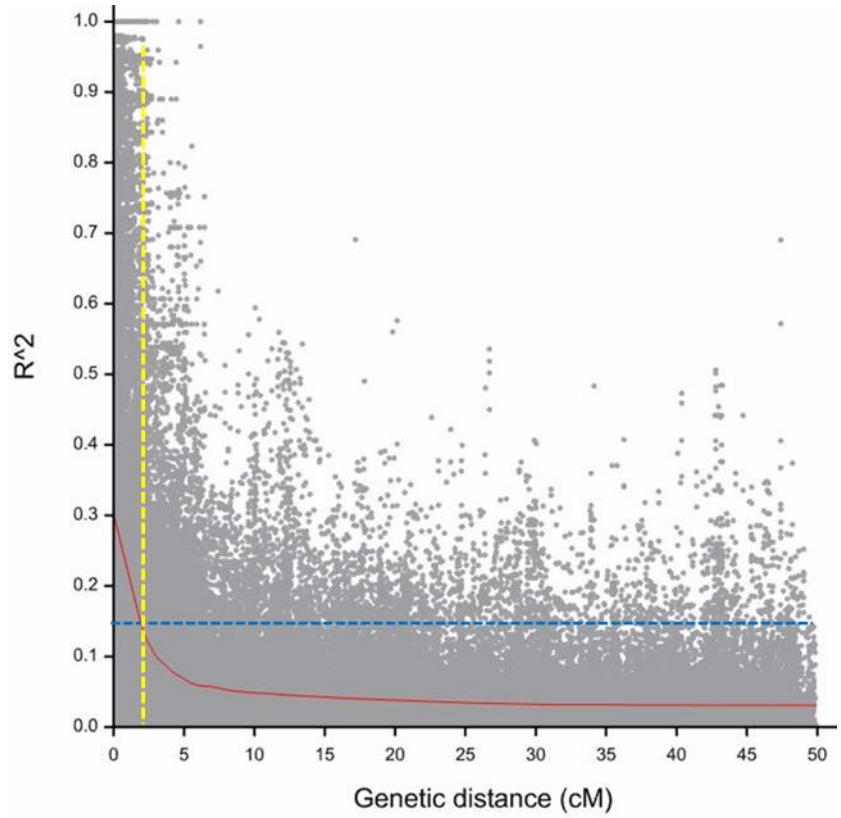
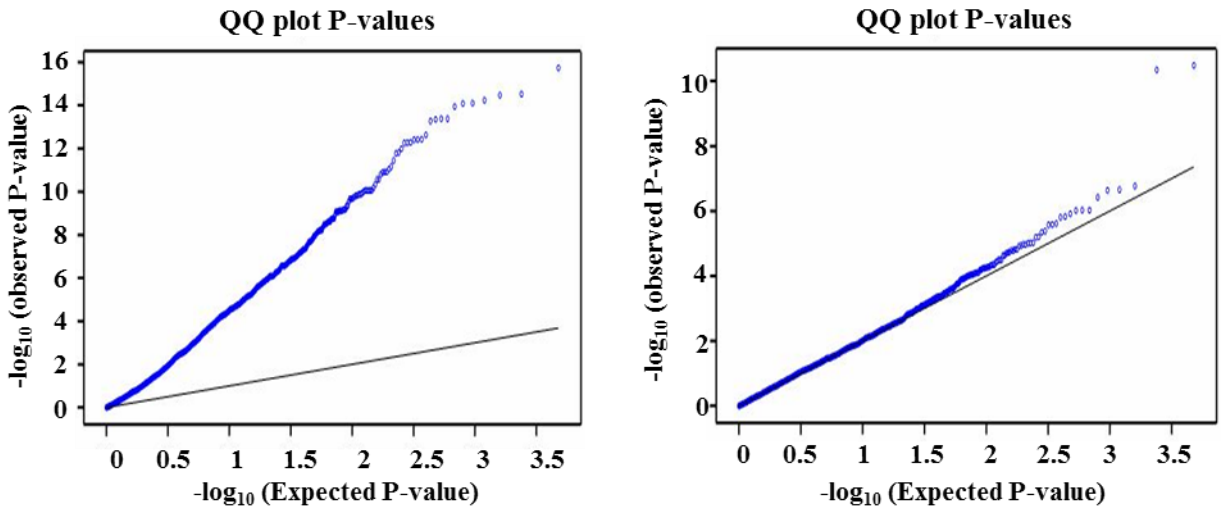


Figure 3.9. Average genome wide linkage disequilibrium (r^2) decay over genetic distance (cM) for the seven barley chromosomes. The *red line* shows the non-linear regression line of the LD decay. *Blue line* shows the intersection of an r^2 value of 0.15 with the regression line. The *yellow line* shows the average LD decay as a function of genetic distance in cM.

3.7 Genome-wide association analyses – model selection

GWAS were performed with the calculated BLUE values for the 14 traits. Different approaches were tested to detect associations between individual SNPs and measured traits. In the first approach, a mixed-linear model was implemented in TASSEL 2.0 (Bradbury et al. 2007) by using either the first two PCs or the results from Q-matrix at *K=2* together with kinship to correct for population structure. In the second approach, a mixed-linear model with kinship alone was tested in GenStat 16th edition and TASSEL 2.0. Implementing the QK, PC + K and K alone in the MLM resulted in a reduced P-value inflation (Fig. 3.10). Since similar results were observed for both mixed-linear model approaches, the MLM with K alone was chosen as the best model since it required less computational time.



(a) (b)

Figure 3.10. Quantile-Quantile plots showing of GWAS for harvest index, a) high p-value inflation for GWAS as a result of no structure correction, b) reduced p-value inflation for GWAS as a result of structure correction.

3.8 GWAS results

Based on the results described in previous chapters, GWAS was performed using GenStat 16th edition while controlling for genetic relatedness (K) in a mixed-linear model (MLM). True significant marker-trait associations (MTAs) were determined by calculating the false discovery rate (FDR) for each trait separately (P >0.05) to define a new threshold for –logP above which associations were considered as true positives. For all traits analyzed, the lowest FDR threshold was observed at a –logP= 2.6 for plant height and the highest at –logP= 7 for hull adherence (covered/naked). In total, 278 significant SNPs were detected for all 14 traits spanning across all the seven chromosomes of barley. Significant MTAs were not uniformly distributed across all chromosomes. The highest number of MTA was identified on chromosome 2H with 104 SNPs showing associations above the FDR threshold. Only eight SNPs were significantly associated on chromosome 6H with significant associations identified for some traits on all seven chromosomes (e.g, spike density and awn roughness). Heading date showed the highest number of MTAs (51 SNPs) with the least number of MTAs identified for seed width (12 SNPs). A summary of agronomic traits, chromosomes and number of MTAs on individual chromosomes is presented in Table 3.4. Localization of genomic regions significantly associated with all 14 traits investigated in current study is presented in Figure S1 of the supplementary section.

Table 3.4. Summary of marker trait associations exceeding the FDR threshold.

	Chromosome							
Trait	1H	2H	3H	4H	5H	6H	7H	Total
HD	3	36	3	1	2	1	5	51
Pht	1	4			23		1	29
GPS		30		4			1	39
HI	2	1	3	3	2		3	14
AwnR	5	3	8	4	3	3	2	28
N-C			4				4	8
TKW		2		2	4		7	15
EL		1	2				16	19
SDA	4	3	1	1	1		6	16
SDL	1	3	2	1	3		7	17
SDW	1	3	3	4	1			12
SPD	1	2	1	1	4	1	18	28
Stig.H	4	11	5	4	5	3	6	38
AwnL	3		3	5	3		3	17

*HD “heading date”, Pht “plant height”, GPS “grains per spike”, HI “harvest index”, SPD “spike density”, TKW “thousand kernel weight”, AwnL “awn length”, EL “ear length”, SDA “seed area”, SDL “seed length”, SDW “seed width”, N-C “ naked caryopsis” and stigma “stigma hairiness”.

In order to group significant marker-trait associations into loci, significant MTAs located within an interval of 5 cM were assigned to a single QTL. By grouping traits based on pair-wise phenotypic correlation (Table 3.2), circular Manhattan plots were generated in R 2.15.3. For visualization purpose, a threshold of –logP= 3 was set for all traits compared in each circular Manhattan plot. All SNPs significantly associated to individual traits are presented in Table S3 with all associated SNPs and their corresponding SNP effects presented as extra supplementary file (see attached compact disc).

3.8.1 Heading date, grains per spike, plant height and harvest index

A significant correlation was observed between heading date and grains per spike. Similarly, a weak correlation was observed between plant height and grains per spike, while no correlation was observed between plant height and heading date. Harvest index was not correlated with any of the traits.

Fifty-one SNPs were significantly associated with heading date at a threshold of –logP ≥3 (FDR) with most associations identified on chromosome 2H (Table S3). These SNPs were grouped into 12 QTL distributed over all seven chromosomes of barley except for chromosome 4H. The most significant associations were detected on chromosome 2H within the genomic region of two important flowering time loci, *Ppd-H1* and *HvCEN* (18.9 and 58 cM), respectively.

For grains per spike, thirty-nine MTAs above the FDR threshold of –log P= 2.9 were detected on chromosomes 2H, 4H, 5H, 6H and 7H. The majority of the MTAs co-located to genomic regions of flowering time genes or QTL pathways; *Ppd-H, HvCEN*, *HvFT4* and *HvCO4.* All thirty nine MTAs were grouped into nine potential QTL (Table S3). Although *Ppd-H1* derived SNPs, BK_13, BK-14 and BK_15 were also included in the SNP data set, the highest associations within the *Ppd-H1* locus were markers SCRI_RS_233272 and SCRI_RS_210172 for grains per spike (-logP=5.1) and heading date (–logP=8) respectively. These markers are in LD with the *Ppd-H1* SNPs and were all considered as a single QTL.

In total, twenty nine SNPs were significantly associated with plant height with the FDR threshold set at –logP ≥ 2.5. Except for chromosomes 3H and 7H with no MTAs, significant associations were identified on all barley chromosomes and were grouped into ten QTL. Most of the significant associations were identified on 5H (Table S3).

Fourteen SNPs exceeded the FDR threshold (–logP ≥ 2.9) for harvest index and were grouped into eleven QTL. Genomic regions of significant MTAs were located on all barley chromosomes except for chromosome 6H (Table S3). Significant associations were also observed on chromosome 1H within the genomic region of *HvCMF10*, a flowering time QTL of barley. Circular Manhattan plots of all four traits are presented in Figure 3.11. Out of the twelve QTL identified for heading date (HD), five were consistent with grains per spike indicating the pleiotropic nature of flowering time genes to other important yield traits.

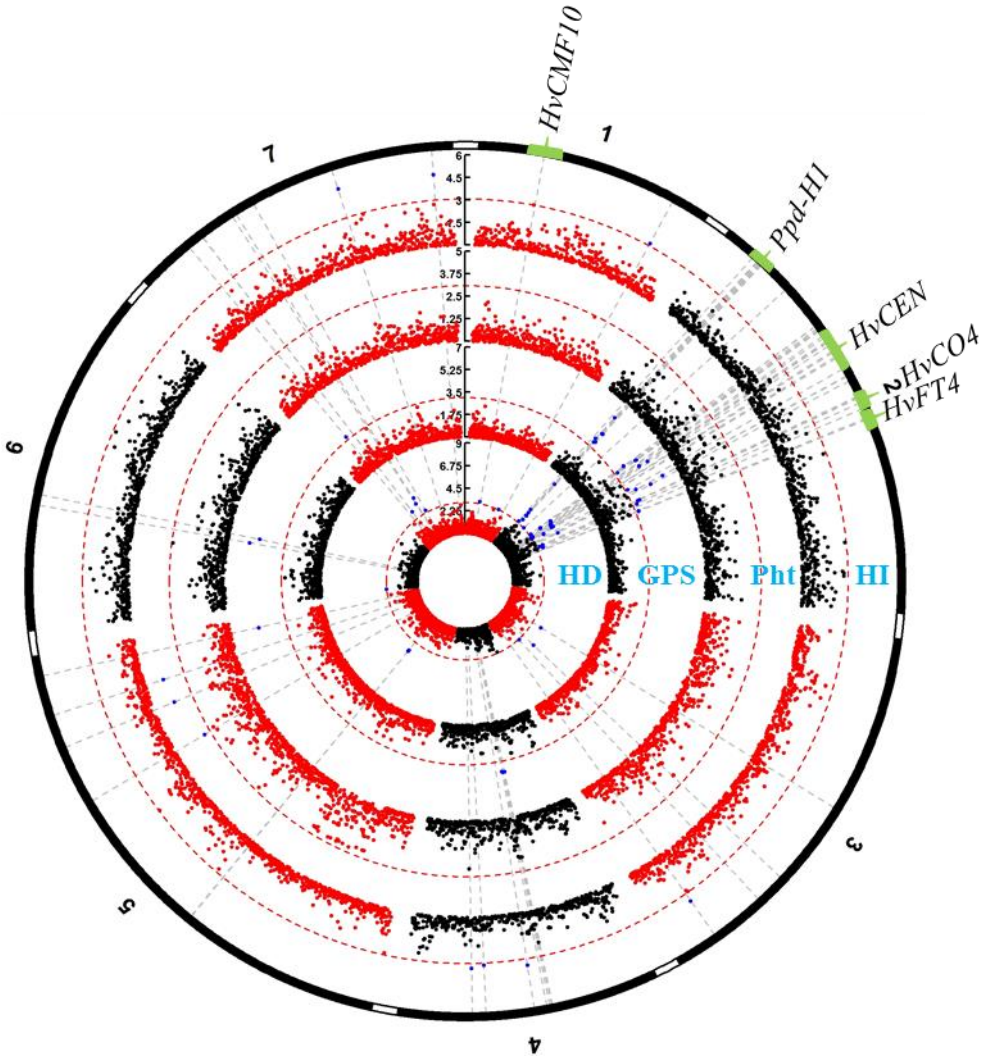


Figure 3.11. GWAS results for heading date (Hd), Grains per spike (GPS), plant height (Pht) and harvest index (HI). Blue dots indicate SNPs with –logP ≥ 3. Grey dashed lines indicate genomic regions associated with at least one of the investigated traits. Known genes within genomic region of some QTL are indicated outside the circle (green). All associated SNPs for each trait above the respective FDR threshold are presented in Table S3. The seven chromosomes of barley are numbered on the outer black circle separated by white borders.

3.8.2 Thousand kernel weight, Seed Area, seed width and seed length

Thousand kernel weight was significantly correlated with seed length, seed width and seed area. Similarly, significant correlation was observed between seed length and seed area. However, a weak correlation was observed between seed width and seed area.

In total, fifteen SNPs were associated with thousand kernel weight (-logP= 3). These MTAs were grouped into ten QTL located on chromosomes 2H, 4H, 5H and 7H (Table S3). The

highest association was detected on chromosome 4H. Significant associations were also detected within the vicinity of two flowering time loci *Ppd-H1* and *HvCEN* on 2H (Fig. 3.12). For seed area, (FDR threshold –logP= 2.9) sixteen SNPs were detected distributed on all barley chromosomes except 6H. Associated SNPs grouped into nine QTL. Some of the QTL corresponded to genomic regions of known flowering time pathways, e.g, one SNP on chromosome 1H at 46.5 cM was within the genetic interval of *HvCMF10*, a Constans-like flowering time QTL. Significant associations were also identified on 2HS at 57 cM and co-located to the genomic region of *HvCEN* (58.9cM).

Twelve SNPs were associated with seed width (FDR of –logP= 2.7), with associations detected on all barley chromosomes except 6H and 7H. All twelve associations were grouped into nine QTL. Two SNPs mapped to the genomic region of a flowering time QTL *HvFT2*.

For seed length, an arbitrary threshold of –logP= 3 was set for marker-trait associations since no marker-trait association was detected above the FDR threshold (-logP= 6). In total, fifteen SNPs were significantly associated with seed length with most of the associated SNPs on chromosome 7H. Marker-trait associations were all grouped into ten QTL.

For both seed length and seed area, significant associations were detected on the long arm of chromosome 7H (86 cM). Five SNPs within this genomic region were co-associated with both traits. Circular Manhattan plots of SNPs significantly associated with all four traits are presented in Figure 3.12**.**

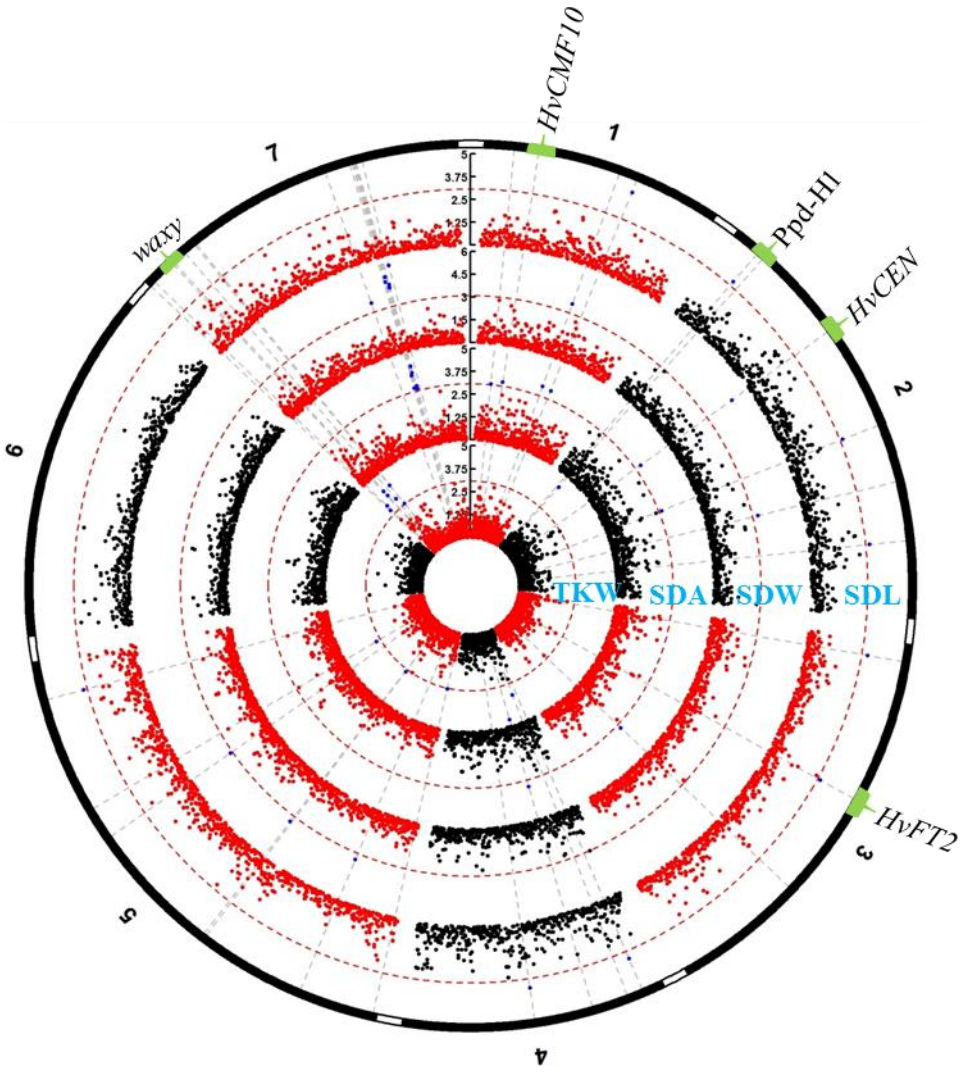


Figure 3.12. GWAS result for thousand kernel weight (TKW), seed area (SDA), seed length (SDL) and seed width (SDW). Blue dots indicate SNPs with –logP ≥ 3. Grey dashed lines indicate genomic regions associated with at least one of the investigated traits. Known gene(s) within genomic region of some QTL are indicated outside the circle (green). All associated SNPs for each trait above the respective FDR threshold are presented in Table S3. The seven chromosomes of barley are numbered on the outer black circle separated by white borders.

3.8.3 Spike density and ear length

Based on phenotypic data, no correlation was observed between spike density and ear length. For both traits, no SNP reached the FDR threshold and hence, an arbitrary threshold of –logP= 3 was set for marker-trait associations. The highest number of associations detected for both traits co-located to the centromeric region of 7H.

For spike density, 28 SNPs were detected at –logP ≥ 3 distributed across all seven chromosomes with 18 SNPs located near the centromere of chromosome 7H (70.2 – 70.8 cM). All MTAs were grouped into nine QTL with some of the QTL colocating to genomic regions of important flowering time genes (*Ppd-H1* and *HFT4*).

Nineteen SNPs were associated with ear length with MTAs identified on chromosomes 2H, 3H and 7H forming five QTL. Of these associations sixteen SNPs were detected on 7H with 15 of the 16 SNPs located close to the centromere of chromosome 7H (70.2 75.1 cM). The detected QTL for both traits within the centromeric region of chromosome 7H co-locates to the genomic region of the *dense spike 1* (*dsp1*). Circular Manhattan plots for both spike density and ear length showing co-localization of significant genomic regions are presented in Fig. 3.13.

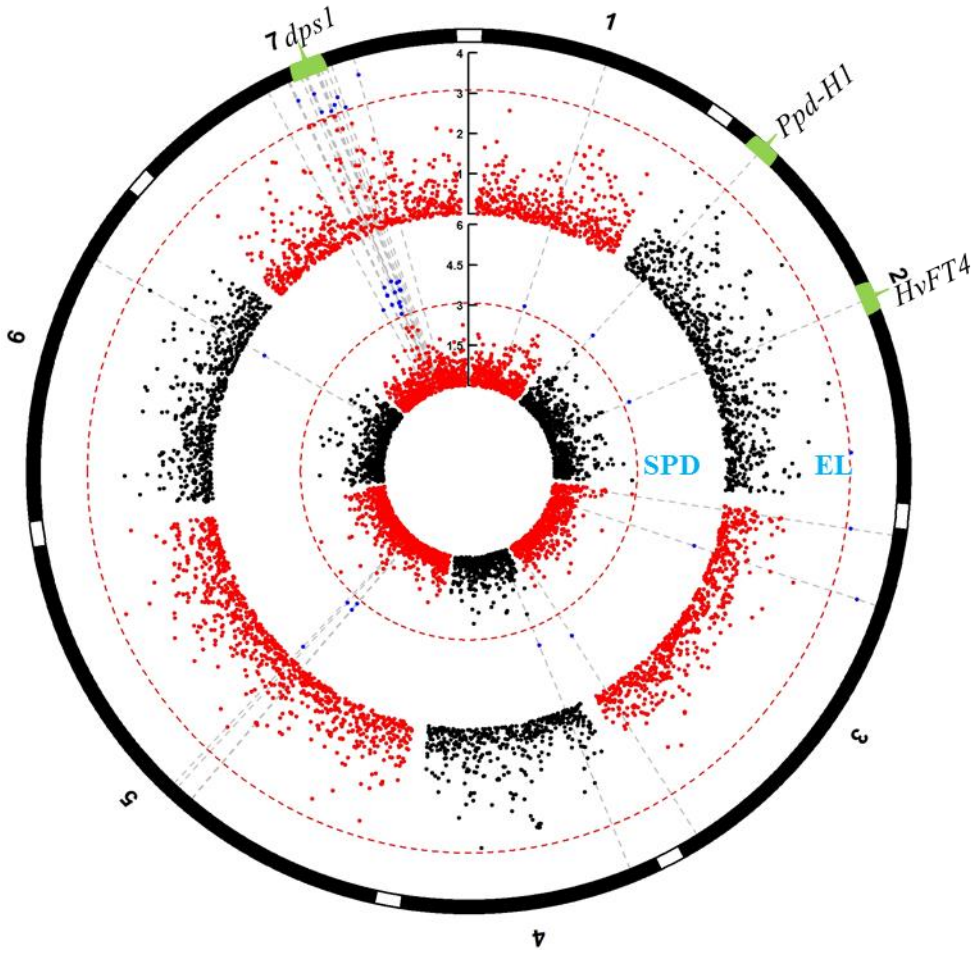


Figure 3.13. GWAS result for spike density (SPD), and earl length (EL). Blue dots indicate SNPs with –logP ≥ 3. Grey dashed lines indicate genomic regions associated with at least one of the investigated traits. Known gene(s) within genomic region of some QTL are indicated outside the circle (green). All associated SNPs for each trait above the respective FDR threshold are presented in

Table S3. The seven chromosomes of barley are numbered on the outer black circle separated by white borders.

3.8.4 Awn length, Awn roughness and stigma hairiness

As shown in Table 3.2, awn roughness was significantly correlated to awn length. On the other hand, weak correlations were observed between awn roughness and stigma hairiness and between stigma hairiness and awn length. For all three traits, significant associations were detected close to the centromere of chromosome 4H. For both awn length and awn roughness, no SNP reached the FDR threshold and hence an arbitrary threshold of –logP= 3 was used to determine significant associations.

Seventeen SNPs were associated with awn length with some of the identified MTAs within genetic regions of some flowering time QTL. Two SNPs on chromosome 1H at 46.5 cM co-located to the genomic region of *HvCMF10*, a Constance-like gene involved in flowering time in barley. All marker-trait associations grouped into ten QTL.

For awn roughness, twenty-eight SNPs were detected spanning across all chromosomes except on chromosome 6H. All MTAs were grouped into nine QTL.

For stigma hairiness, thirty eight MTAs were detected (FDR threshold –logP= 3.2) on all seven chromosomes and grouped into twenty-six QTL. Most of the SNPs associated with stigma hairiness were on chromosome 2H (11 SNPs). Some of the detected MTAs were within the vicinity of the flowering time loci *HvCEN*, *HvCO4* and *HvFT4*.

For all three traits, associations were detected within the genomic region of the awn roughness QTL (*raw1*) on chromosome 5H.

In general, most of the QTL for awn roughness, co-localized with QTL detected for awn length. Circular Manhattan plot of all three traits is presented in Fig. 3.14.

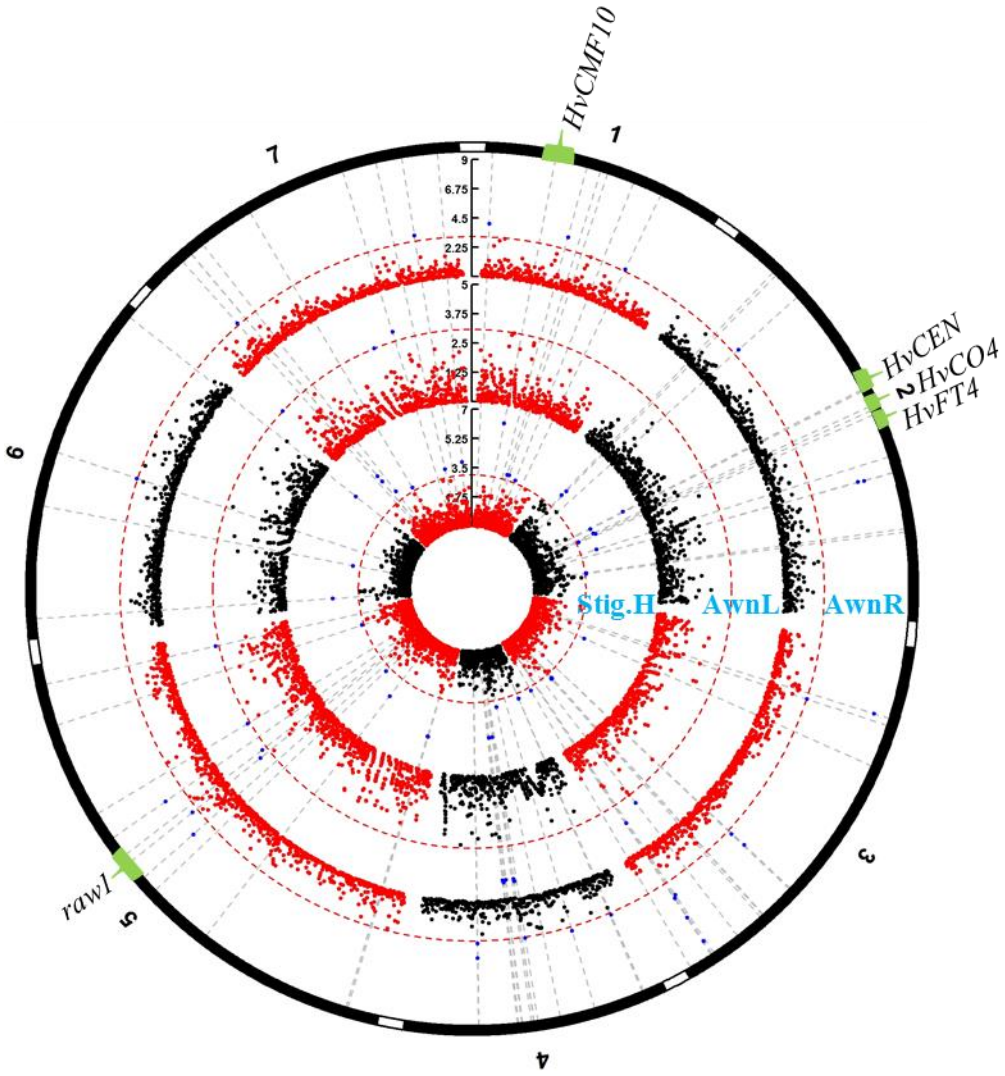


Figure 3.14. GWAS result across stigma hairiness (stigma), awn length (AwnL) and awn roughness (AwnR). Blue dots indicate SNPs with –logP ≥ 3. Grey dashed lines indicate genomic regions associated with at least one of the investigated traits. Known gene(s) within genomic region of some already reported QTL are indicated outside the circle (green). All associated SNPs for each trait above the respective FDR threshold are presented in Table S3. The seven chromosomes of barley are numbered on the outer black circle separated by white borders

3.8.5 Presence or absence of hulls

By considering all 199 accessions (43 naked accessions), eight SNPs were significantly associated with the barley hull/hulless phenotype above an FDR threshold of –logP=7. All eight SNPs were grouped into 2 QTL on chromosome 3H and 7H, respectively. Four SNPs on 7H spanned the genomic region of the *Nud* locus, with the highest SNP within this region

detected at 84.3 cM(-logP= 15.6). The 3H QTL is novel and was detected by 4 SNPs within the genetic interval of 45.4 – 46.03 cM. Within this genetic interval, the highest associations were detected at 46 cM (–logP=15.13). By further separating the naked accessions in the association panel according to geographical origin "naked accessions from Ethiopia vs naked associations from Europe and Asia", GWAS was further conducted to identify subpopulation specific QTL. For the association panel including 156 hulled accessions and 27 naked accessions from Ethiopia, only the 7H locus was detected. By considering all hulled barleys together with naked barleys only from Europe and Asia (15 accessions), two QTL were detected on chromosome 3H and 7H respectively. The 3H QTL was marked by 5 SNPs within the genetic interval of 45 – 51 cM with the highest associations detected at 46 cM (-logP=18) while the 7H QTL was detected in this panel by a single SNP at 70.8 cM with a –logP= 16. Circular Manhattan plots of all three association panels are presented in Fig. 3.15.

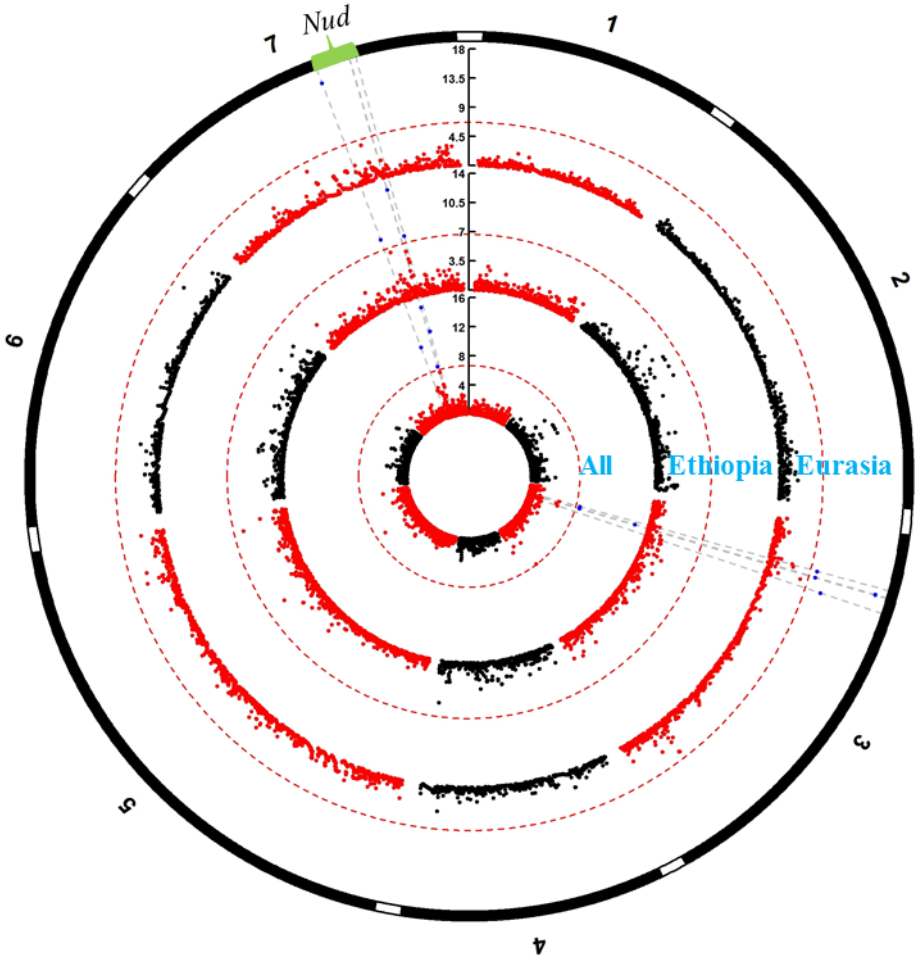


Figure 3.15. Maker-trait associations for hull-adherence across three different panels, a) the entire association panel (All), b) all hulled barley and naked barley of Ethiopian origin (Ethiopia) and c) all

hulled barley with naked barley from Europe and Asia (Eurasia). Blue dots indicate SNPs with –logP ≥ 7. Grey dashed lines indicate genomic regions detected in at least one of the investigated panel. Known gene(s) within the genomic region of detected QTL are indicated outside the circle (green). All associated SNPs above the respective FDR threshold are presented in Table S3. The seven chromosomes of barley are numbered on the outer black circle separated by white borders.

4 Discussion

Genome-wide association studies have been increasingly used in the identification of genomic regions associated with phenotypic traits. Barley landraces serve as a suitable material for uncovering new QTLs which can be introgressed into elite materials due to their historical origin with high local adaptation associated with traditional farming systems (Villa et al. 2005). In this study, GWAS was conducted for agronomic and morphologic traits in a diverse collection of 199 two-rowed spring barley landraces. Compared to previous population structure analysis with 42 SSR markers on this collection (Pasam et al. 2014), the present study recruited additional markers, reported associations with 14 agronomic traits, and conducted a comprehensive survey of population structure and LD decay using SNPs from an improved genotyping technique. By using different methods, the extent of population structure was investigated based on all informative SNPs (4790). The present analysis also revealed the presence of potential duplicate accessions within barley landraces stored in the IPK Gene bank. By modelling confounding effects from population structure and genetic relatedness, shared common associations were detected between a number of agronomic traits with many of these co-locating with flowering time genes or QTL, pointing to their importance for adaptation. Novel QTL were also identified for most of the traits which could be potentially exploited for future yield improvement.

4.1 Phenotypic variation and trait heritability

In the present study, BLUEs were calculated for all traits to adjust for environment and replicate effects. A broad phenotypic variation was observed for all 14 traits. For instance, a difference of 22.5 days was observed between the minimum and maximum heading date, and likewise TKW varied between 35.6 g and 70.5g. While working with a collection of barley breeding lines, cultivars and landraces, Munoz-Amatriain et al. (2014) reported a high level of phenotypic diversity for agronomic traits within a sub-collection of landraces originating from Europe and Asia. In the current study, a significant coefficient of variation was observed for all traits ranging from 4% for seed width to 18% for grains per spike. Compared with other previous studies involving 106 spring barley landraces (Žáková and Benková 2006), the coefficients of variation observed in the current study for thousand grain weight, plant height and grains per spike are two times higher likely caused by covering a higher level of diversity with a larger sample size. The high variation in phenotypic data is as a result of the genetic variation in adaptive traits and indicates the suitability of the present

collection in uncovering novel QTL for agronomic traits through genome-wide association studies. Detailed analysis of heading date divided the collection into two categories; one category consisted of early heading genotypes and was dominated by African and Asian genotypes while the second of late heading genotypes mainly dominated by genotypes from Europe. Both *Ppd-H1* and *HvCEN* play major roles in early heading in spring barley. Alqudah et al. (2014) and Turner et al. (2005) reported a reduced photoperiod sensitivity at the *Ppd-H1* locus in European spring barley accessions. It is likely that, most of the accessions from Europe showing early heading, carry the early allele at the *HvCEN* locus and accessions will show early heading irrespective of day length (Comadran et al. 2012). Early heading is an advantage for plants in regions of hot and dry summers (like in most parts of Africa and Asia) because plants can complete their life cycle before exposure to severe drought and heat. On the other hand, for accessions from East Europe were the summers are comparatively cool and humid, late heading is an advantage since the longer growing period is allowing crops to produce higher yields (Hershey 2005).

The year of 2013 was characterised by warm weather with low humidity and precipitation as compared to 2014 which was characterized by cooler temperatures with much rain. This difference in weather conditions resulted in a 13 days difference observed for heading date between early heading genotypes in 2013 and 2014. Accordingly, Karsai et al. (2008) reported the effects of temperature and light intensity on flowering time of barley as important factors influencing the number of days to heading. By investigating the effect of changing weather on cereals, Ingver et al. (2010) also reported a significant increase in days to heading in oat, wheat and barley in cooler and rainy periods during summer with days to heading reduced when the weather was warmer and sunny. However, the maximum heading date remained the same in both years. This can be attributed to the fact that late heading genotypes might not benefit from higher temperatures if a certain minimum temperature threshold is met or the plant has reached a critical physiological state. Heading date is one of the most heritable traits in barley as demonstrated in many studies (Maurer et al. 2016; Pasam et al. 2012), and accordingly, heading date was the trait with the highest broad sense heritability estimate (H^2 =0.95).

4.2 Population structure

The investigation of subgroups within a population is a prerequisite for LD mapping since this can lead to spurious associations. Different methods (PCA, STRUCTURE and splits tree)

were used to exploit population structure within the association panel with all methods yielding similar results. All three approaches revealed that the current collection of 199 spring barley landraces could be grouped into two major subgroups with the division mainly based on the presence/absence of hulls on caryopsis followed by geographical origin. Interestingly, in the PC analysis, naked barley from Asia and Europe clustered closely to the hulled barley but far from the Ethiopian naked barleys. This was consistent with results from the neighbour joining tree. The unique clustering of Ethiopian naked barley far from naked barley from Asia and Europe and also far from hulled accessions is concordant to the finding of Pasam et al. (2014) while analysing population structure with 42 SSR markers in the full core collection of both two- and six-rowed and is also concordant with reports from other population structure studies in barley landraces (Bellucci et al. 2013; Saisho and Purugganan 2007). Landraces are diverse and are structured mainly according to geographical origin according to studies using SSRs and chloroplast SSR data (Pasam et al. 2014; Russell et al. 2004; Russell et al. 2011). Based on the Q-matrix approach, the method of Evanno used in determining the number of subgroups in this study is not robust since it uses the variance of each STRUCTURE run as a divisor and this variance estimate may vary between runs for each *K* within a subgroup. Although the STRUCTURE approach (Fig. 3.8) revealed the maximum increase at K=2, indicating the presence of two major subgroups, the first two variance components in the PCA approach explained about 23% of the variation within the association panel. In general the main causes of population structure in barley are growth habit, row type and geographical origin (Hamblin et al. 2010; Malysheva-Otto et al. 2006). As the landrace subset contains only spring barley of the two-rowed type, sub-stratification within the panel was mainly based on geographical origin and the difference in morphotypes as a result of presence or absence of hulls around the caryopsis.

4.3 Linkage disequilibrium decay

The extent of LD decay plays a central role in association mapping. The number and density of markers needed for LD mapping is determined by the distance at which LD will persist in a chosen collection. In the present study, LD was calculated as the pairwise correlation r^2 between polymorphic markers on each chromosome. Average genome-wide LD decay was estimated at 2.5 cM. In cultivated barley, LD decay has been reported to decay between 10-15 cM when evaluated with SSR or SNP markers (Malysheva-Otto et al. 2006; Pasam et al. 2012). The extent of LD decay depends on effective population size, demographic history,

mating system, admixture, recombination rate and selection effects (Gaut and Long 2003; Nordborg et al. 2002). Consistent with the present study on barley landraces, Munoz-Amatriain et al. (2014) and Rodriguez et al. (2012) reported a rapid decay of LD at 3 cM while working with a population of barley landraces from Sardinia. A different pattern of LD decay has been reported in wild barley. Here LD decayed within 1 cM despite the high rate of selfing (Morrell et al. 2005). The fast decay in LD reported in wild barley as compared to the cultivated forms is likely a result of the higher number of historical recombination events. However, the reasons for the rapid LD decay in barley landraces than might be expect is unclear (Caldwell et al. 2006; Comadran et al. 2009; Zhang et al. 2009). It might be the consequence of "unique human-induced pseudo-outbreeding" coupled with "strong selection for advantageous alleles" in agriculture (Rostoks et al. 2006).

Given the total genetic map of 988.2 cM and the average LD decay of 2.5 cM in the present landrace collection, 395.2 (988.2/2.5) polymorphic SNP makers evenly distributed on each chromosome are required to cover the entire genome. Thus the 4970 SNPs used in this study provide sufficient marker coverage for association mapping. In a similar study with a collection of modern cultivars (Zhou et al. 2014), a lower number of polymorphic markers (113 SNPs) evenly distributed across the entire genome was required to cover the entire genome due to the large extent of LD decay in modern cultivars compared to landraces.

4.4 Comparison of SSR and SNP markers in genetic relatedness study

The maintenance of unknown duplicate samples in Gene bank collections is inefficient and costly. In most of the cases, redundant accessions have the same origin/passport information. In some cases identical accessions may have different country of origin, different phenotypic data, and different names. The identification of potential duplicates has neither been reliable nor cost effective until the era of high-throughput genotyping and sequencing technologies.

In the present study, a substantial number of duplicate accessions were detected, which accounted for 33% of the initial population selected for the field trial. Of all accessions initially selected, eighty-eight were redundant based on 5711 SNPs and formed 30 groups of potential duplicates. While this genetic identity was not observed with 42 SSR markers, the corresponding pattern on NJ-tree was similar to that observed with the SNPs as potential duplicates within each group based on SNP information were also closely related in the SSR-NJ tree diagram. Most of the duplicate accessions have similar passport data and also showed similarities at the trait (Table S2). For example, the largest group of potential duplicates with

12 accessions are from Slovakia (one genotype from Czech Republic) and were collected in 1974 and 1977 in a very small geographical region. However other potential duplicates originating from different geographical locations or differing in hull adherence were also detected. In the case of duplicate group 24, DNA samples might have been confused or one of the accessions may have been misclassified during the handling of seeds.

It is estimated that only one third of the total number of accessions conserved in *ex situ* Gene banks are distinct (FAO 2010) and duplications occur within and between Gene banks for the same crop. Different methods are implemented to handle duplicated accessions: i) keeping one accession and eliminating the rest from the collection, ii) combining the seeds of duplicated accessions, ii) remove identical accessions only from the "active" core (McCouch et al. 2012). Munoz-Amatriain et al. (2014) also identified duplicate accessions in a core collection of 2,417 barley accessions in which 14% were genetically identical. In the present collection, 5711 SNPs were considered for genetic relatedness analysis compared to 6,224 informative SNPs used by Munoz-Amatriain et al. (2014).

Mean genetic diversity and mean PIC values were also compared for both 42 SSR and 4790 SNPs (only informative SNPs above MAF of 5%) across 261 accessions. The mean genetic diversity and mean PIC values were higher with SSR (0.58 and 0.48 respectively) than with SNPs (0.38 and 0.3, respectively) indicating that both SSR and SNP markers were informative, although at different extent. The polymorphisms of both marker sets are generated through different mechanisms (replication slippage for SSR vs point mutation for SNPs) which can therefore provide different views of diversity study. Despite lower PIC value for the SNP marker set, SNPs are more reliable in diversity studies with the primary advantage that they occur in genomes at a much higher frequency than SSRs. By comparing genetic diversity and PIC in a collection of Indian rice varieties, Singh et al. (2013) reported an average PIC value of 0.23 with SNP markers lower than the 0.25 observed with SSR markers. Yu et al. (2009) suggested that over a 10 fold more SNPs than SSRs should be used in order to capture an equal amount of genetic diversity, while Van Inghelandt et al. (2010) suggested an excess of seven to eleven fold SNPs. Similar discriminant observation between SNPs and SSRs was reported in a GWAS in rice (Courtois et al. 2012) and sunflower (Filippi et al. 2015). The average PIC values of 0.48 and 0.3 for both SSR and SNP, respectively, suggest that both panels of markers have a high discriminant capacity for the present germplasm collection. Due to bi-allelic nature of SNPs, PIC values can range from 0 to 0.5 whereas for SSR markers which are multi-allelic, PIC values goes above 0.5 and can even go up to 1.

The discrepancy of SSR and SNP data for identifying potential duplicates in diverse germplasm arises from the different scoring methods. One of the problems encountered with SSR markers is the difficulty in scoring as seen by re-inspecting the raw data of Pasam et al. (2014) used in current study. In some cases, it was hard to read the peak precisely, which may lead to a faulty scoring at a locus for different genotypes.

In most cases, phenotypic data of accessions belonging to the same group of potential duplicates were also very similar (Table S2) underlining their status as true duplicates. The current finding reveals that based on phenotype alone, it is not easy to identify potential duplicates. In conclusion, though SSRs and SNPs are multi-allelic and bi-allelic in nature with different distribution pattern across the whole genome, the unique features of SNP markers such as, abundance in the genome, ability to generate polymorphism due to variation at single base level and their development from genic regions enabled these markers to present different spectrum of diversity. For this reason, the SNP markers were more preferable for the detection of potential duplicates and can be considered to be equally more useful in assessing genetic diversity since they reveal functional variations which can be potentially exploited for marker-trait association studies.

4.5 GWA analysis

The robustness of the phenotypic data, the model choice for GWA analysis and the method of handling false positive associations determines the reproducibility of GWA results. For GWA analysis, "BLUEs" were calculated for all traits to avoid any discrepancies in phenotypic data which could be as a result of environmental influence. Yearly effects where treated as fixed effects while replications were treated as random effects (Hess et al. 2006). As indicated in previous chapter 3.7 (Fig. 3.10), a strong reduction in p-value inflation was observed by correcting for population structure. No significant difference was observed for results from all three different mixed-linear model approaches (PCA + K, QK and K), probably due to the low level of population structure within the association panel (results not shown). However, the MLM with K proved to be more efficient in terms of computational time. Results from other studies have demonstrated that the MLM with K is more efficient in the control of population structure compared to PCK and QK when calculating GWA in a less structured panel since the inclusion of either PC or Q together with K might lead to over correction of population structure and hence resulting in false negatives or type II error (Kang et al. 2008; Stich and Melchinger 2009; Yu et al. 2006). By implementing the FDR (0.05) approach for

all traits, 339 significant marker-trait associations were detected across all 14 agronomic traits investigated in current study and were grouped into 79 potential QTL. Out of the 79 identified QTL, 45 were associated with two or more traits, with some of the QTL already reported in previous studies for the same related traits (Table S3).

Heading date and grains per spike

Flowering time is an important trait in barley as it is highly involved in adaptation. Three main classes of genes control flowering time in barley: those responsive to photoperiod (day length), those responsive to vernalisation, and *earliness per se*. All three classes influence the transition from vegetative to reproductive phase (Laurie 1997; Takahashi and Yasuda 1956). In this study, two of the three classes (photoperiod responsive and earliness per se) were identified for both heading date and number of grains per spike. For both traits, the highest associations were observed at the *Ppd-H1* locus (HD; -logP= 8.4 and GPS; -logP= 6). The functional SNP of *Ppd-H1* (BK_15) is included in the SNP marker set used in current study and was significant at –logP=4.2. Interestingly, within the genomic region of the *Ppd-H1* locus (19.9 cM), the most significant marker for both traits was SCRI_RS_233272 (18.9 cM). This SNP was still significant when BK_15 was used as a co-factor in the GWA analysis, indicating the association of this SNP with both traits (heading date and grains per spike) to be independent of *Ppd-H1*. However, SCRI_RS_233272 is surely in tight linkage with the *Ppd-H1* locus since both loci are less than 1cM apart. One of the main reasons for SCRI_RS_233272 to be highly associated to heading date instead of the *Ppd-H1* specific SNP (BK_15) could be due to the number of genotypes with missing genotypic information. From the genotypic data, all accessions were genotyped at the SCRI_RS_233272 locus whereas four accessions had missing information at the BK_15 locus. This difference in the number of accessions with SNP information at both loci could influence significance level at the different loci. However, the possibility of another major gene within the vicinity of *Ppd-H1* also involved in heading time cannot be ruled out. At BK_15, the minor allele "G" occurred with a frequency of 0.28 with an estimated effect of -2. This implies that replacing the reference allele "A" by "G" will lead to a decrease in heading date by 2 units and a decrease in number of grains per spike by -1.5 units. For SNP marker SCRI_RS_233272, the minor allele "G" occurred with a frequency of 0.25., replacing the reference allele "G" by "T" will lead to a decrease in heading date and grains per spike by a factor of 2.8 and 1.6 units, respectively. Many studies have already reported the importance of photoperiod

genes/QTL pathways in promoting flowering in many crops including barley (Turner et al. 2005) and wheat (Shaw et al. 2013). *Ppd-H1* encodes a pseudo-response regulator (*PPR*) which promotes flowering under long days in barley (Turner et al. 2005).

The circadian clock is an intrinsic regulator of biological processes oscillating within an ~24-h interval (Pittendrigh 1993) and is considered a mechanism by which plants recognize the optimal photoperiod for seasonal flowering (Imaizumi 2010). Gawroński (2013) reported a disruption of the circadian clock function by the *HvCEN* (*earliness per se* locus) of barley. In this study, the second highest associated SNPs were detected within the genomic region of *HvCEN* (2H at 56 - 60 cM) for both heading date and grains per spike. *HvCEN* affects flowering independent of day length and vernalisation requirement. Under 10 h, 13 h and 16 h photoperiod treatments after vernalisation and autumn sowing, Laurie et al. (1995) were able to detect *HvCEN* in a collection of winter x spring barley implying that *HvCEN* is photoperiod neutral.

Other candidate genes involved in flowering time identified for both traits included *HvCO4* (60 - 54 cM on 2H), *HvCO1* (51 - 55 cM on 7H) and *HvFT4* (67 cM on 2H). Both *HvCO1* and *HvCO4* belong to the CONSTANS-like gene family *(CO)* with members of this family known to regulate flowering time through the photoperiod pathway in *Arabidopsis* (long day plant) and rice (short day plant). Numerous homologs of *CO*-like genes (*HvCO1* - *HvCO18*) have been reported in barley but their roles in heading time pathway are still unclear (Cockram et al. 2012; Griffiths et al. 2003). An overexpression of *HvCO1* upregulates *HvFT1* (an *FT*-like gene of barley) thereby accelerating time to flowering in long-and short-day conditions (Campoli et al. 2012). Campoli et al. (2012) reported the overexpression of *HvCO1* in spring barley with a variation at *Ppd-H1* not affecting the diurnal expression of *HvCO*1. The involvement of *HvCO1* and other *CO*-like genes in inflorescence development of barley was reported by Alqudah et al. (2014). In the current study, significant associations for heading date were also detected within genomic regions of other *CO*-like genes for heading date (*HvCO12* on chromosome 7H and *HvCO18* on chromosome 2H) and within the vicinity of genomic regions harbouring *HvFT1* (34 cM on 7H), *HvFT2* (59 cM on 3H) and *HvFT3* (92 cM on 1H). Overexpression of *HvFT1*, *HvFT2* and *HvFT3* has been reported for early flowering in rice (Kikuchi et al. 2009), indicating these *FT-like* genes act as promoters of floral transition. Amongst these *FT-like* genes, *HvFT1* has been reported to be the key gene responsible for flowering in the barley *FT-like* gene family. *Ppd-H2*, a photoperiod sensitive gene, encodes *HvFT3* which promotes flowering of barley under short day lengths

(Faure et al. 2007; Kikuchi and Handa 2009). Kikuchi and Handa (2009) reported *HvFT3* expression in Morex (a spring barley variety carrying *Ppd-H2*) and not in Steptoe a (spring variety carrying *ppd-H2*) under long-and short day conditions with the expression increased under short day condition. Furthermore, significant associations were identified for heading date at 14.4 cM on 4H (-logP=3). Maurer et al. (2015) reported a *CCT* (*CONSTANS, CO-like*, and a *TOC1*) domain gene within this genomic region, with the *LOG (LONELY GUY)* gene also located within this genomic region. The *LOG* gene encodes a cytokinin-activating enzyme required for meristem activity. Loss of function of the *LOG* gene causes pre-mature termination of the shoot meristem in rice development (Kuroha et al. 2009).

No candidate genes have been reported in previous studies for grain number QTL identified on chromosomes 2H (74 cM) and 4H (54.4 cM) respectively. However, Ingvordsen et al. (2015) reported significant associations on 4H (50 cM) for grain number while Locatelli et al. (2013) detected significant associations on 4H (54 cM) for thousand grain weight on 4H (56 cM) for harvest index. These associations are within the vicinity of the identified QTL at 54 cM on 4H. Grain number is directly related to thousand grain weight and harvest index. Increase in grain number leads to a decrease in thousand grain weight due to decrease in seed size as a result of carbohydrate partitioning. For the 2H QTL (74 cM), 6 UDP-glycosyltransferase superfamily proteins are mapped within this genomic region (74 cM – 74.16 cM). Li et al. (2013b) reported the enhancement of grain production in rice by a rice zinc finger protein DST through controlling *Gn1a/OSCKX2 (grain number 1a/cytokinin oxidase 2)* expression. Reduced expression of *OsCKX2* causes cytokinin accumulation in inflorescence meristems and increases the number of reproductive organs, resulting in enhanced grain yield (Ashikari et al. 2005). Li et al. (2013b) reported *LOC_Os04g25440/Os04g0320700*, a rice gene containing a glucosyltransferase protein domain as a potential candidate of *Gn1a*. The genomic region of the *UDP-glycosyltransferase superfamily proteins* identified in this study (2H, 74 cM) is syntenic to the genomic region of *Os04g0320700*; hence the annotated *UDP-glycosyltransferases* are potential candidates. The corresponding alleles of both SNPs at 74 cM (T or A) reduced number of grains by 2 units when replaced with the reference alleles (C or G, respectively).

Other associations detected for both heading date and grains per spike corresponded to genomic regions of *HvCMF10* (1H at 47.8 cM) and *HvCMF13* (5H at 50.4 cM). Both *HvCMF10* and *HvCMF13* interact with other heading time genes to influence flowering time in barley. The identification of significant associations within the regions of many major

flowering time genes *Vrn-H3, HvCO1, Ppd-H1, HvFT4, Ppd-H2* in the present association panel, clearly demonstrate the power of the present collection and also the power of the mapping approach.

Plant height

Plant height is influenced by many quantitative genes/QTL (Ji-hua et al. 2007) and is mainly controlled by dwarfing, semi-dwarfing, and other plant height genes. Amongst the QTL identified for plant height in the current study, significant associations were detected in the genomic region of *HvCEN* on chromosome 2H (58 cM). Consistent with this observation, Tondelli et al. (2013) also reported significant associations for plant height at close vicinity of *HvCEN* in a collection of European spring barley cultivars. The identification of QTLs at *HvCEN* for plant height, heading date and grains per spike indicates that the inheritance of these traits is either linked functionally (pleiotropy) or physically (linkage disequilibrium). *HvCEN* is located in the centromere of 2H and is in LD with many loci within this region due to supressed recombination, hence it is likely possible that the identification of this region is not as a result of the involvement of *HvCEN* in height regulation but rather as a result of physical linkage (linkage disequilibrium). All SNPs associated with plant height within the genomic region of *HvCEN* showed a decrease in plant height by 4 units when substituted with the reference allele. Further on 2H, a QTL was detected at 8.6 cM. Wehner et al. (2015) reported a biomass QTL within this genomic region in a collection of winter barley. Three QTL were detected on chromosome 5H at 43.8cM, 143.7 cM and 152.4 cM. These QTL have been identified in other GWA analysis for biomass (Wehner et al. 2015). Ingvordsen et al. (2015) reported a QTL for the number of productive ears at 144.5 cM on chromosome 5H, while Tondelli et al. (2013) reported significant associations for necking at 145 cM on chromosome 5H.

Barley cultivars in North-western Europe mostly contain either of two dwarfing genes; *Denso* on chromosome 3H, presumed to be an ortholog of the rice green revolution gene *OsSd1*, or *Breviaristatum-e (ari-e)* on chromosome 5H. Based on genotyping by sequencing, Liu et al. (2014) was able to map the *Breviaristatum-e (ari-e)* between morex_contig_335403 (49 cM) and morex_contig_137133 (unmapped). In this study 18 SNPs were detected within the centromeric region (43 – 45 cM) of 5H. This QTL is 4 cM downstream of morex_contig_335403 (49 cM) and might be in LD with *ari-e* due to suppressed recombination around the centromere of 5H. Mutations in *Ari-e* cause semi-dwarfing and have been widely used in barley cultivar development to shorten straw length and reduced

severity to lodging. Alleles of all SNPs detected within the vicinity of *ari-e* showed a decrease in plant height by 4 units when substituted with the respective reference alleles. Furthermore, significant associations on 5H (8 cM) and 6H (53 cM) coincided to genomic regions of necking and lodging QTL reported in other GWAS (Tondelli et al. 2013). The co-localization of plant height QTL with reported necking and lodging QTL can be expected as these traits are cross-related by common factors of straw strength and potential for wind damage through height exposure. Three SNPs were significantly associated to plant height at 130 – 131.7 cM on 5HL. Amongst the list of genes within this genomic region are *AP2-like* ethylene-responsive transcription factors and Gibberellin receptors. Among the phytohormones, gibberellin (GA) is the most well-known involved in controlling stem elongation and a deficiency or insensitivity to GA could easily result in severe dwarfism as already reported in many different kinds of plant species such as; rice mutants independently mutated in any of the six GA biosynthetic enzymes (copalyl diphosphate synthase *(CPS)*, ent-kaurene synthase (KS), ent-kaurene oxidase *(KO)*, ent-kaurenoic acid oxidase *(KAO)*, GA 20 oxidase (*GA20ox)*, and GA 3-oxidase *(GA3ox)*) (Sakamoto et al. 2004), the *Arabidopsis* GA-insensitive *short internodes* (Fridborg et al. 1999) and the barley GA-deficient *grd2c* mutant (Wolbang et al. 2007). Gibberellin deficiency has been reported to pleiotropically induce culm bending in sorghum (Ordonio et al. 2014). On the other hand, *AP2/Ethylene-Responsive* Element Binding Factor (*ERF)* family of transcription factors are present only in the plant kingdom and is characterized by the presence of a highly conserved DNA-binding domain (Riechmann and Meyerowitz 1998). Members of the *AP2/ERF* family have various developmental and physiological processes such as internode elongation restriction in rice by down-regulation of gibberellin biosynthetic gene (Qi et al. 2011).

Harvest index

Harvest index as the ratio of grain to total biomass is highly influenced by environmental factors such as soil condition and temperature (Li et al. 2012; Shrotria and Singh 1988). However, genetic control of harvest index plays an important role in crop production. The intrinsic regulation of harvest index is controlled by many genes (Laza et al. 2004). In the current study, the genomic region of *HvCMF10* (1H, 47.8 cM) was significantly associated with harvest index. *HvCMF10* interacts with other flowering time genes in barley to promote early flowering (Alqudah et al. 2014). In the same region (47.5 cM), Wehner et al. (2015) reported a QTL for biomass yield, which is one component of harvest index. Also, significant

associations were detected at 135.5 cM (3HL) and 83.6 cM (4HL). Both genomic regions have been reported to be involved in grain yield and grain number in barley (Ingvordsen et al. 2015). Furthermore, SCRI_RS_235762 on the long arm of chromosome 4H (112 cM) and BOPA2_12_30590 (5H, 118.9 cM) were significantly associated with harvest index and co-locate with genomic regions containing other biomass yield QTL in barley (Wehner et al. 2015). Both QTL at 4H (112 cM) and 5H (118.9 cM) are at close proximity to *Vrn-H2* and *Vrn-H1*, 114.94 and 125.7 cM respectively. *Vrn-H1* and *Vrn-H2* play major role in determination of vernalisation requirement and hence are key players of flowering time. Genotypes carrying allelic combination of *Vrn-H1/Vrn-H2* (dominant at both vernalisation loci) show a spring growth habit (Mohammadi et al. 2013). Resequencing of *Vrn-H1*/*Vrn-H2* will be necessary to further analyse the allelic state of both loci in this population. Co-association of flowering time, biomass and grain yield QTLs with harvest index has also been reported in rice (Li et al. 2012) and is not surprising as these traits are co-related.

Thousand grain weight, seed width, seed length and seed area

Thousand grain weight is an important grain yield trait co-related with seed length, seed width and seed area. Significant associations were identified for TKW, SDL and SDA around the genomic region of *HvCEN* on chromosome 2H (58.6 cM). As discussed earlier, many flowering time genes or QTL might be involved in grain traits. But also, other genes in the same region might cause this association. Recently, a major QTL, *GL7*, encoding the plant-specific transcription factor *OsSPL13*, which positively regulates cell size in grain hull, resulting in enhanced rice grain length and yield was reported in rice (Si et al. 2016). Eleven predicted genes are allocated within 18.5 - 21 Mb on rice chromosome 7 with *Os07g0505200 (*also predicted as *LOC_Os07g32170* and referred to as *OsSPL13)* reported as the most likely candidate. *OsSPL13* belongs to the *SQUAMOSA PROMOTER BINDING PROTEIN (SBP)* family of transcription factors and is an orthologue of Arabidopsis *AT1G53160* and *AT3G15270* (squamosa promoter binding protein-like 4 and squamosa promoter binding protein-like 5) respectively. *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes have numerous important roles during plant growth and development (Imbe et al. 2004; Qi et al. 2012). Though *Os07g0505200* could not be located on the genome zipper (Mayer et al. 2012), one of the closest predicted genes *Os07g0506000* is in synteny with the centromeric region of barley chromosome 2H. A *SQUAMOSA PROMOTER BINDING PROTEIN* (MLOC_13032.1) maps on chromosome 2H at 58.8 cM. The closest SNP

(SCRI_RS_240011) to MLOC_13032.1 detected for TKW maps at 58.6 cM and has an estimated effect of 2.7, thereby indicating an increase in TKW by close to 3 units when the reference allele "C" is replaced with the allele "A".

The short arms of chromosome 5H (1.6 cM) and 7H (13.6 cM) were associated with TKW. Both regions have been reported to harbor QTL conferring positive effects on both grain number and number of ears with grains (Ingvordsen et al. 2015). The detected QTL at 13.6 cM on 7H is within the genomic region of the *waxy* gene. The *waxy* gene encodes a *granule-bound starch synthase I (GBSSI*) in barley. Variations in the *waxy* gene have a great effect in amylose synthesis and starch properties of barley (Li et al. 2014). Consistent with a previous study (Locatelli et al. 2013), significant MTAs were detected on the short arm of chromosome 7H for thousand grain weight (0.3-0.9 cM). Eleven *kelch repeat-containing proteins* and a *kinesin-like protein* are annotated within the genomic interval of 0 - 4 cM on 7H and could be considered as potential candidates. A sucrose fructan 6-fructosyltransferase also maps at 0.57 cM on chromosome 7H. Fructans have been reported to play an important role in assimilation partitioning in the vacuole of barley and wheat (Sprenger et al. 1995).

Furthermore, five markers were consistently associated with seed length and seed area within 84 - 86.6 cM on chromosome 7H. Several *Zinc finger-like proteins*, *cytochrome P450s* and a *serine carboxypeptidase* are annotated within this genetic region and are potential candidate genes. Gao et al. (2015) reported an additive effect of *GS3* and *qGL3* on rice grain length. *GL3.1/qGL3* is a major grain length locus and encodes a putative protein phosphate *(OsPPKL1)* containing Kelch domains (Qi et al. 2012; Zhang et al. 2012). Amongst eleven *GS3* candidate genes investigated in rice, *OsSCP16*, a putative *serine carboxypeptidase* homologue was highly regulated and playing a major role in seed length (Gao et al. 2015). Transgenic study in rice showed that Kelch domains functions as a negative regulators and are essential for the biological function of *OsPPKL1* (Gao et al. 2015). At the cellular level, *qGL3* functions by negatively modulating the longitudinal cell number in grain glumes (Gao et al. 2015).

Spike density, ear length

Five QTL were identified for ear length with two of these QTL overlapping with spike density QTL. For both spike density and ear length, 11 SNPs were detected at the centromeric region of chromosome 7H (70 cM) which is within the vicinity of dense spike (*dps1*) of barley (Taketa et al. 2011). However, the dense spike of barley is under the control

of several major genes. By high resolution mapping mutation, Taketa et al. *(2011)* reported the *DENSE SPIKE (dsp1)* as one of the major determinants of ear length. Consistent with this result, Gawenda et al. (2015) also reported significant MTAs on 7H (70 cM) for grain number per ear in a collection of German winter barley. Number of grains per ear, ear length and spike density are all interrelated, hence the detection of QTL overlapping for all three traits is not surprising. Significant associations were also detected within the vicinity of *Ppd-H1* (18.9 cM on 2H) and *HvFT4* (68.5 cM on 2H) indicating once more the importance of flowering time on yield components in barley landraces. Within the centromeric region of chromosome 3H, significant associations were detected for both spike density and ear length (3H, 51 cM). Of all the SNPs significantly associated within this region, BOPA2_12_30467 was the highest significant SNP for spike density (–logP=5.6). Amongst the genes located within this genomic region, a *cytokinin oxidase/dehydrogenase* (MLOC_81291.1) maps at 46 cM. Although the gene is 5 cM away from BOPA2_12_3067, it could still be in LD with MLOC_81291.1 since recombination is suppressed in the centromeric region. Searches revealed close synteny between MLOC_81291.1 and *LOC_Os01g10110* (Mayer et al. 2012). *LOC_Os01g10110* is located on rice chromosome 1 and encodes the zinc finger transcription factor "*DST"* which directly regulates *OsCKX2* expression in rice. *OsCKX2* is a rice *cytokinin oxidase/dehydrogenase* and its expression in the apical meristem leads to *OSCKX2* regulated *CK* accumulation in the shoot apical meristem (SAM) and, therefore, controls the number of reproductive organs in *dst* mutants leading to a lower plant height and longer rice panicle (Li et al. 2013a).

Wehner et al. (2015) reported significant associations at 25.7 cM for grain number on chromosome 4HS coinciding to the genomic region of a spike density QTL identified in this study (26.8 cM). Both spike density and number of grains are related as spike density was calculated as the ratio of grain number to ear length.

Awn length and awn roughness

The presence of awns is associated with, or tightly linked to numerous beneficial traits (Bariana et al. 2006), as barley awns play a significant role in seed dispersal, burial and photosynthesis (Yuo et al. 2012a). Till date, only few genes controlling awning in grass species have been map-based cloned; *Awn-1* and *Awn-2* in rice (Gu et al. 2015; Luo et al. 2013) and *SHORT INTERNODES (SHI)(short awn 2)* in barley (Yuo et al. 2012b). *Awn-1 (An-1)* regulates awn length in rice by encoding a *BHLH* protein which is intensely expressed

at the apex of lemma primordia, specifically causing continuous cell division to form long awns (Luo et al. 2013). The *short awn 2 (lks2)* which produces about 50% shorter than normal awns in barley is a natural variant restricted to Eastern Asia. *Lks2* encodes a *SHI-family protein transcription factor*. In the current study, a significant correlation was observed between awn length and awn roughness. Three QTL were consistently detected for both awn roughness and awn length with associations detected within the vicinity of the flowering time QTL *HCMF10* on 1H centromeric region (46.5 cM). The highest significant associations were detected on 3H (142 cM) and 4H (60 cM) for awn roughness and awn length, respectively. Within a small interval on chromosome 3H (142.2 – 142.6 cM), 4 SNPs were highly associated with awn roughness with a single SNP within this region (BOPA2_12_20198) also associated with awn length. On the other hand, 5 SNPs within the interval of 59.5 cM – 60.3 cM on 4H were highly associated with awn length. Close to this genomic region, marker SCRI_RS_202326 (4H, 57 cM) was associated with awn roughness. Though the corresponding barley homologue of both candidate genes (*Os04g0350700 and Os04g0351333)* reported for rice *Awn-1* could not be located within any of the detected QTL, three *BHLH DNA-binding superfamily proteins* and a *Zinc finger protein* maps within 141 – 142 cM on 3H. On the other hand four *basic helix-loop-helix transcription factors* and a *MAD-box transcription factor* maps within the interval of 57- 60 cM on 4H. These genes could be potential candidates since several genes controlling awn length in rice are reported to belong to either of these gene families. BHLH and *BHLH* plays a role in cell division especially on lemma primordia (Luo et al. 2013). The *short awn 2 (lks2)* is located on the long arm of chromosome 7H and could be in LD with the awn length QTL detected at 76.7 cM on chromosome 7H. Consistent with this study, the smooth awn *(raw1)* has been mapped on the long arm of chromosome 5H (116 cM) (Franckowiak 2008) which is closed to the genomic region of significant associations detected on 5H (114.8 – 120 cM) for awn roughness. A single SNP was associated to awn length at 0.3 cM on chromosome 7H. Ingvordsen et al. (2015) also reported significant MTAs on the short arm of chromosome 7H (1.9 cM) for number of grains per ear with the associated SNPs having a positive effect on number of grains per ear. The identification of significant associations consistent for both awn length and awn roughness indicates that both traits are cross-related and hence are either controlled in most cases by same genes or same genetic pathways as a result of pleiotropy or are controlled by genes which are in linkage disequilibrium.

Stigma hairiness

Stigma hairs are the receptive tissue on which pollen germinates during flowering time. Scarcity of hairs has been reported in some cases and may prevent the normal reception and germination of pollen (Harvey et al. 1968). Till date, there are only few reports on the genetic basis of stigma hairiness with a close correlation reported between stigma hairiness and awn roughness (Harvey et al. 1968). SCHOLZ (1963) and WOODWARD (1949) reported a type of female sterility associated to barley with smooth awns and previously attributed sterility in smooth-awn barley as a result of poor pollen reception due to lack of stigma hairs. Though no significant phenotypic correlation was observed between awn roughness and stigma hairiness in this study, significant associations were detected for both traits within the genomic region of *HCMF10* on 1H (46.5 cM). Within this genomic region, BOPA2_12_11301 (-logP= 6.3) was highly associated with stigma hairiness. Though the direct relationship between flowering time and stigma hairiness remains unclear, most flowering time genes or QTL are also involved in other traits. However the centromeric region of 3H contains many more genes that could be the causal gene for the identified QTL. Four QTL detected for awn length (3H at 51.1 cM, 4H at 20.9 and 57.5 cM, 7H at 21 cM) coincided with QTL for stigma hairiness supporting early observations on the interrelation of both traits (Harvey et al. 1968) and the pleiotropic nature of the corresponding QTL. The 4H QTL (57. 5 cM) also collocates to genomic region of a QTL detected for awn roughness in current study, further illustrating all three traits to be cross-related. As already mentioned above, *BHLH* transcription factors and *MAD-box transcription factors* are annotated within the genomic region of this QTL and will be potential candidates. A single marker was highly associated to stigma hairiness at 48.4 cM on chromosome 5H which close proximity to the genomic region of *Breviaristatum-d (ari-d). Breviaristatum-d* is a short awn gene which produces stigma with sparse hairs (Yuo et al. 2012a).

Presence or absence of hulls (hulled/hull-less)

The presence or absence of caryopsis in barley is an important agronomic trait as it is directly linked to the dietary use. The trait has been reported to be controlled by the *Nud* gene which maps to the long arm of chromosome 7H (Franckowiack and Konishi 1997). Till date, the *Nud* locus is the only cloned gene reported to control the barley caryopsis phenotype (Taketa et al. 2008). The *Nud* gene encodes an ethylene *response factor (ERF) family transcription factor (TF)* that is involved in the lipid biosynthesis pathway (Taketa et al. 2008). A complete

deletion of the *nud* locus leads to the naked caryopsis phenotype and has been reported to be a monophyletic natural mutation of domesticated barley following the analysis of a wide range of worldwide naked barley (Taketa et al. 2008). Surprisingly, in the current study, two QTL were detected for hull adherence on 3H and 7H, respectively. The highest association was on 7H (84.3 cM) in the region of *Nud*. Since *Nud*-specific SNPs are not included on the SNP array employed in this study or on the POPSEQ linkage map, the position of the gene was estimated to be at 78.3 cM by reference of the SNP data from this study to the most recent reference sequence of the barley genome (Ariyadasa et al. 2014). The position of the 7H QTL which spanned a 13.5 cM interval (70.8 – 84.3 cM) also spans the approximate location of *Nud* (75-80 cM) (Russell et al. 2016). Also, Muñoz-Amatriaín et al. (2014) reported significant associations at 84.3 cM on 7H in a collection of barley landraces and suggested these associations to be in LD with the *Nud*. Yu et al. (2016) recently reported the *nud1.g* distinctly found in naked barley collected from Tibet. The *nud1.g* contains a non-synonymous SNP T634A when compared with the functional *Nud* gene and co-segregates with the naked phenotype. Though Taketa et al. (2004) proposed a single monophyletic origin (probably in south western Iran) of naked barley, several authors have suggested more than one origin (including east of the Fertile Crescent) (Morrell and Clegg 2007b; Saisho and Purugganan 2007). Dai et al. (2014) reported great differences between hull-less barley from Tibet and modern cultivated barley at the genomic level and suggested Tibetan hull-less barley to have existed in an early stage of domestication which lends support to the identification of the *nud1.g,* found exclusively in Tibetan naked barley (Yu et al. 2016). In order to screen for subpopulation specific QTL, naked barley within the current panel were subdivided based on the geographical origin into Ethiopian naked barley (27 accessions) and naked barley originating from Europe and Asia (15 accessions). For the association panel comprising of hulled accessions and naked accessions from Ethiopia (183 accessions), significant associations were detected only on chromosome 7H. For the subpanel including all hulled accessions together with naked barley from Europe and Asia (172 accessions), significant associations were detected on chromosome 3H spanning from 45 cM - 51 cM and for a single marker on 7H (70.8 cM). The 3H QTL is novel and subpopulation specific. In a recent association analysis of a large germplasm set comprising 2417 accessions fingerprinted with the same iSelect array, significant association were detected with SNPs located around the *Nud* locus as well as with unmapped SNPs (Munoz-Amatriain et al. 2014). It cannot be ruled out that some of these unmapped SNPs may be located in chromosomal regions corresponding to the loci on 3H detected in this study. All 4 SNPs detected on

chromosome 3H were in strong LD (Fig. 21). For the QTL on 7H spanning the *Nud*, all 4 SNPs were in weak to moderate LD (r^2= 0.12 – r^2= 0.5), which is not surprising since the detected SNPs were within a large genetic interval of more than 13 cM (Fig. 2.22). A complex LD pattern was observed between all SNPs on 3H and the SNPs on 7H as only BOPA2_12_20685 at 70.8 cM on 7H was in weak LD with the 3H SNPs (r^2=0.24 - 0.29). The LD between the highest associated SNP on 7H (SCRI_RS_4562, 84.4 cM) and the 3H SNPs was very weak (r^2=0.08). It is still unclear to whether the 3H QTL is independent of *Nud* since the single association on 7H detected in the subpanel comprising of naked barley from Europe and Asia is around 8 cM away from *Nud* and other SNPs detected on 7H in the whole panel were below the MAF of 5% in the subpanel comprising of only naked barleys from Europe and Asia, probably as a result of the small number of naked barley (15 accessions).

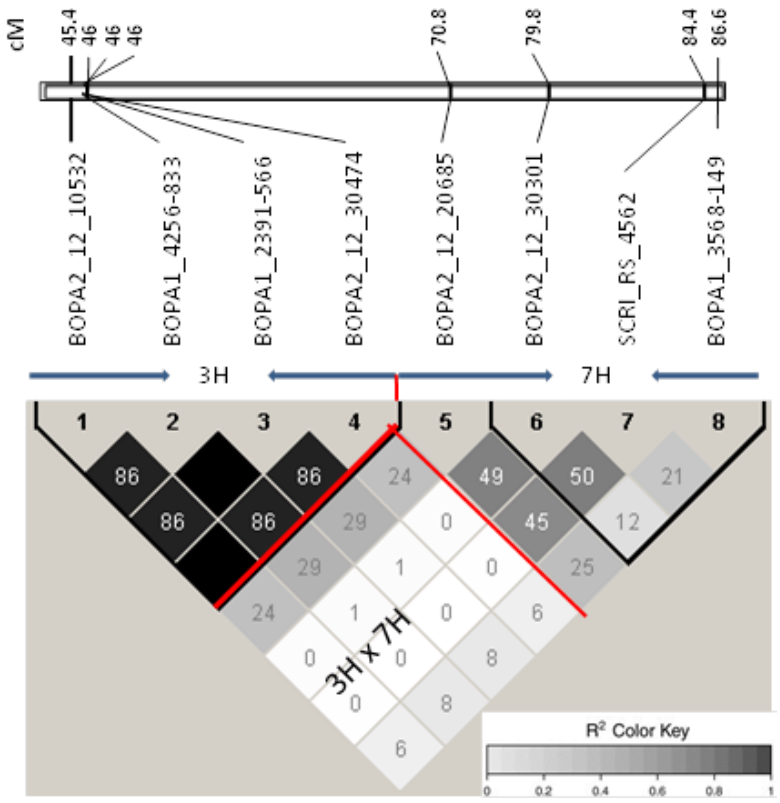


Figure 4.1. Heatplot of LD between all SNPs associated with naked caryopsis on chromosome 3H and 7H in the whole association panel. Pair-wise R^2 values are indicated in boxes (86= 0.86, 6=0.06). All SNPs were grouped into two LD blocks (Gabriel et al. 2002) indicated on the figure by dark borders (SNPs 1- 4 and 6 - 8), except for BOPA2_12_20685 at 70.8 cM on 7H which does not fall into any of the detected LD blocks.

Duan et al. (2015) suggested the regulation of one transcription factor gene *(SHN1)*, six cutin biosynthesis genes (*ATT1*, *LCR*, *GPAT6*, *GPAT8*, *LACS2*, and *HTH*), four cutin related genes (*FDH1*, *W AX2*, *CER9*, and *ACC1*), and two cutin transporter genes (*Eibi1* and *WBC11*) by the *Nud*. Interestingly, *Eibi1* is a full ABC-G transporter identified in barley (MLOC_62487.1) for cutin deposition (Chen et al. 2011) and is located at 45.9 cM on chromosome 3H (Mascher et al. 2013) within the QTL interval (45.4 cM – 51 cM). According to (Duan et al. 2015), *Nud* transcription factor regulates the cutin biosynthesis pathway, leading to a thick and loose cutin, which forms a high permeable cuticle on the caryopsis of covered barley and is responsible for the hull-caryopsis fusion. A deletion or low expression of the *Nud* leads to a functional cuticle covering the pericarp of naked caryopsis which defines a perfect boundary of the pericarp to separate the caryopsis from the hulls in hull-less barley. Functional characterization of spontaneous recessive *eibi1* mutants in wild barley (*Hordeum vulgare ssp. spontaneum*) has revealed a defective leaf cuticle which leads to an increase water loss rate and hyper-susceptibility to drought (Chen et al. 2004). Also, *Epb1*, a major Cysteine proteinase responsible for the degradation of endosperm storage proteins in barley is located within the genomic region of the 3H novel QTL. However, the direct involvement of *Epb1*and hull adherence has not been reported. Owing to the fact that there has been no evidence that any of the described mutations in *Eibi1*or a mutation in *Epb1* results in a reduction in hull adherence or the formation of a naked caryopsis, more information about the structure and function of *Eibi1*or *Epb1* in relation to the barley caryopsis is required to vet the potential effects on hull adherence. It is also possible that the QTL on 3H may not be directly linked to hull adherence, but maybe have been unconsciously selected for a targeted end use of naked barley in Europe and Asia.

5 Summary and outlook

Genome wide association studies (GWAS) have been widely used to detect genomic regions significantly associated with phenotypic traits of interest. Due to the use of natural populations, the resolution of mapping a QTL is highly increased in comparison to bi-parental mapping. However, the power to capture a common variant associated to a phenotype depends on the extent of linkage disequilibrium (LD) in the population under investigation.

In the current study, a collection of 261 2-rowed spring barley landraces genotyped by using the iSelect 9K chip was used for GWAS. This collection is a subset of a core collection (LRC 648) which was previously genotyped with 42 SSR markers in a population structure study (Pasam et al. 2014). Due to the identification of 88 potential duplicates based on iSelect SNPs, only 199 accessions were used for GWAS analysis. The identification of these potential duplicates was important as duplicate accessions can alter GWAS results as overrepresentation of a genotype will increase the MAF at a genotyped locus. This finding also indicates the inherent inadequacies of SSR markers in diversity studies.

Average genome wide LD decay for the whole panel was estimated at 2.5 cM with no strong sub-stratification observed which is in concordance to other studies on 2-rowed spring barley landraces (Muñoz-Amatriaín et al. 2014; Rodriguez et al. 2012).

In total, 14 agronomic traits were scored in 2013 and 2014 at IPK-Gatersleben. For all traits, a high phenotypic variance was observed across both years with heritability estimates ranging between 0.74 – 0.95. From a total of 79 QTL identified in the current study, significant associations for many yield traits co-located with genomic regions of the flowering time genes *Ppd-H1* and *HvCEN* lending strength to the hypothesis of their pleiotropic nature. Within the genomic region of *Ppd-H1*, the most significant SNP was SCRI_RS_233272 for heading and grains per spike respectively. This SNP is 1 cM away from the *Ppd-H1* specific SNPs. Though just 1 cM away from *Ppd-H1* specific SNPs, the possibility of another flowering time gene closed to *Ppd-H1* cannot be ruled out. Also, important agronomic traits such as plant height and thousand grain weight were associated at close proximity to already reported loci , *arie-e* and *waxy*, respectively.

Apart from already significant associations at already reported loci, synteny searches between barley and rice revealed two potential candidates MLOC_13032.1 for thousand grain weight,

seed area and MLOC_16287.1 for grains per spike on chromosome 2H at 58.6 and 74 cM respectively. MLOC_13032.1 belongs to the *SQUAMOSA PROMOTER BINDING PROTEIN (SBP)* family of transcription factors and is an orthologue of Arabidopsis *AT1G53160* and *AT3G15270*. The rice homolog of MLOC_13032.1; *Os07g0505200* has been reported to increase grain size in rice (Imbe et al. 2004; Qi et al. 2012). On the other hand MLOC_16287.1 belong the UDP-glycosyltransferase superfamily proteins. Reduced expression of OsCKX2, the rice homolog of MLOC_16287.1 causes cytokinin accumulation in inflorescence meristems leading to increase number of grains.

Another interesting discovery is the identification of a second QTL for barley hull adherence on chromosome 3H which was specific for naked barley originating from Europe and Asia. The observation that this association is specific to naked barley from Eurasia reflects the allelic diversity at this genomic region in the whole population. Till date, only the *Nud* gene on chromosome 7H has been identified, cloned and functionally characterized with regards to barley hull adherence. Within the genomic region of the 3H QTL is *Eibi1*, an ABC transporter involved in cutin biosynthesis which has been reported to be regulated by the *Nud* and *Epb1*, a major Cysteine proteinase responsible for the degradation of endosperm storage proteins in barley. However, the direct involvement of *Eibi1* in hull adherence cannot be elucidated. Though there has been no report on the involvement of both *Eibi1* and *Epb1* in hull adherence, these two genes are potential candidates and need to be investigated in more detail.

Though the low LD in the current panel can be exploited for candidate gene discovery, increasing the number of SNPs by using other sequencing platforms such as GBS might help in detecting further QTL and also in narrowing down the genetic interval of detected QTL which could serve for marker assisted selection. In addition, resequencing of *nud* in all naked landraces used in current study will shed a light on whether there is a complete deletion of *nud* in all naked barley as already reported or whether there exist naked barleys with no or partial deletion. Furthermore, expression analysis of *Eibi1* and *Epb1* will be necessary in determining the different expression level of these candidates between the two groups of naked barley (Ethiopia vs Eurasia).

Also, increasing the number of trial sites will further improve the phenotypic data quality especially in the case of traits like heading date, seed width and seed length which was greatly affected by poor weather in 2014. However, this improvement might be very minimal as high heritability estimates were observed for these traits across the two years of field trial.

Field trials with barley landraces are also difficult to handle due to the high degree of lodging and pre-harvest sprouting.

As observed in this study, the current subpopulation of 199 genotypes is very useful for association study. Narrowing down the genetic interval of large effect QTL by increasing the number of markers will improve the resolution of these QTL which could be applied for marker assisted selection and further integrated into breeding schemes for yield enhancement.

6 Zusammenfassung

Genomweite Assoziationsstudien (GWAS) finden weithin Verwendung, um genomische Regionen zu detektieren, die signifikant mit phänotypischen Merkmalen von Interesse assoziiert sind. Aufgrund der Verwendung natürlicher Populationen ist die Auflösung der Kartierung eines QTL im Vergleich zu einer bi-parentalen Population deutlich erhöht. Die Fähigkeit, eine mit einem Phänotyp assoziierte gemeinsame Variante zu erfassen, hängt jedoch vom Ausmaß des Kopplungsungleichgewichts (LD) in der untersuchten Population ab.

In der aktuellen Studie wurde eine Sammlung von 261 zwei-reihigen Sommergerste-Landsorten, welche unter Verwendung des iSelect 9K-Chips genotypisiert wurden, mittels GWAS analysiert. Diese Sammlung ist eine Untergruppe einer Kernsammlung (LRC 648), die zuvor in einer Populationsstrukturstudie mit 42 SSR-Markern genotypisiert wurde (Pasam *et al*. 2014). Aufgrund der Identifizierung von 88 möglichen Duplikaten auf Basis von iSelect-SNPs wurden nur 199 Akzessionen für die GWAS-Analyse verwendet. Die Identifizierung dieser potentiellen Duplikate war wichtig, da die doppelten Akzessionen die GWAS-Ergebnisse verändern können, da eine Überrepräsentation eines Genotyps die MAF an einem genotypisierten Locus erhöht. Dieser Befund weist auch auf die inhärenten Unzulänglichkeiten von SSR-Markern in Diversitätsstudien hin.

Der durchschnittliche genomweite LD-Abfall für die gesamte Kollektion wurde auf 2,5 cM geschätzt, wobei keine starke Subschichtung beobachtet wurde, was in Übereinstimmung mit anderen Studien zu zwei-reihigen Landsorten von Sommergerste steht (Muñoz-Amatriaín *et al*. 2014; Rodriguez *et al*. 2012).

Insgesamt wurden in den Jahren 2013 und 2014 im IPK-Gatersleben 14 agronomische Merkmale bonitiert. Für all Merkmale wurde über beide Jahren eine hohe phänotypische Varianz beobachtet, wobei die Heritabilitätsschätzungen zwischen 0,74 und 0,95 lagen. Von insgesamt 79 in der aktuellen Studie identifizierten QTL, co-lokalisieren signifikante Assoziationen für viele Ertragsmerkmale mit genomischen Regionen welche die Blühzeitgene *Ppd-H1* und *HvCEN* beinhalten, was die Hypothese ihrer pleiotropen Effekte stärkt. Innerhalb der genomischen Region von *Ppd-H1* war SCRI_RS_233272 der signifikanteste SNP sowohl für das Ährenschieben, als auch für die Anzahl der Körner pro Ähre. Obwohl dieser SNP nur 1 cM von *Ppd-H1*-spezifischen SNPs entfernt ist, kann die Möglichkeit eines weiteren Blühzeitgens, das sich in unmittelbarer Nähe zu *Ppd-H1* befindet,

nicht ausgeschlossen werden. Auch andere wichtige agronomische Merkmale wie die Pflanzenhöhe und das Tausendkorngewicht waren mit Loci assoziiert, welche sich in enger Nachbarschaft zu den bereits beschriebenen Loci *arie-e* und *waxy* befinden.

Abgesehen von signifikanten Assoziationen an bereits bekannten Loci, ergab die Syntenie-Suche zwischen Gerste und Reis zwei potentielle Kandidaten auf Chromosom 2H: *MLOC_13032.1* für Tausendkorngewicht und Samenfläche bei 58,6 cM, und *MLOC_16287.1* für die Anzahl der Körner pro Ähre bei 74 cM. *MLOC_13032.1* gehört zur SQUAMOSA PROMOTER BINDING PROTEIN (SBP) Familie von Transkriptionsfaktoren und ist ein Ortholog der Arabidopsis Gene *AT1G53160* und *AT3G15270*. Für das Reis-Homolog von *MLOC_13032.1*, Os07g0505200, wurde beschrieben, dass es die Korngröße in Reis erhöht (Imbe *et al*. 2004; Qi *et al*. 2012). *MLOC_16287.1* gehört zu den UDP-Glycosyltransferase-Superfamilie-Proteinen. Eine reduzierte Expression von *OsCKX2*, dem Reis-Homolog von *MLOC_16287.1*, verursacht eine Cytokinin-Akkumulation in Infloreszenz-Meristemen, was zu einer Erhöhung der Anzahl der Körner führt.

Eine weitere interessante Entdeckung ist die Identifizierung eines zweiten QTL für die Spelzen-Adhäsion auf Chromosom 3H. Die Beobachtung, dass diese Assoziation spezifisch für „Nacktgerste“ aus Eurasien ist, spiegelt die Alleldiversität in dieser genomischen Region in der gesamten Population wider. Bislang wurde nur das *Nud* Gen auf Chromosom 7H identifiziert, kloniert und funktionell in Bezug auf Spelzen-Adhäsion in Gerste charakterisiert. Innerhalb der genomischen Region des 3H QTL befindet sich *Eibi1*, ein ABC-Transporter, der an der Cutin-Biosynthese beteiligt ist und als von *Nud* und *Epb1* reguliert beschrieben wurde. *Epb1* kodiert eine wichtigen Cystein-Proteinase, die für den Abbau von Endosperm-Speicherproteinen in Gerste verantwortlich ist. Die direkte Beteiligung von *Eibi1* an der Spelzen-Adhäsion konnte jedoch nicht aufgeklärt werden. Obwohl es keine Berichte über die Beteiligung von *Eibi1* und *Epb1* an der Spelzen-Adhäsion gibt, sind diese beiden Gene potentielle Kandidaten und sollten näher untersucht werden.

Obwohl die niedrige LD in der aktuellen Kollektion für die Entdeckung von Kandidatengenen ausgenutzt werden kann, könnte einer Erhöhung der Anzahl der SNPs durch Verwendung anderer Sequenzierungsplattformen wie GBS bei der Detektion weiterer QTL helfen und auch die genetischen Intervalle der detektierten QTL eingrenzen, um markergestützte Selektion zu ermöglichen. Darüber hinaus wird die Resequenzierung des *Nud* Gens in allen „nackten“ Landsorten, die in der aktuellen Studie verwendet werden,

Aufschluss darüber geben, ob, wie bereits berichtet, eine vollständige Deletion von *Nud* bei allen „Nacktgersten“ vorliegt oder ob „nackte“ Gerstenakzessionen ohne oder mit nur einer teilweisen Deletion existieren. Darüber hinaus wird eine Expressionsanalyse von *Eibi1* und *Epb1* notwendig sein, um mögliche unterschiedliche Expressionsniveaus dieser Kandidaten zwischen den zwei Gruppen von „Nacktgerste“ (Äthiopien vs. Eurasien) zu bestimmen.

Auch die Erhöhung der Anzahl der Versuchsstandorte wird die phänotypische Datenqualität weiter verbessern, insbesondere im Hinblick auf Merkmale wie Ährenschieben, Samenbreite und Samenlänge, welche 2014 durch schlechtes Wetter stark beeinflusst wurde. Diese Verbesserung könnte jedoch sehr gering ausfallen, da hohe Heritabilitätsschätzungen für diese Merkmale während der zwei Jahre des Feldversuchs beobachtet wurden. Feldversuche mit Gerstenlandrassen sind aufgrund der niedrigen Standfestigkeit und des hohen Grades an Keimung vor der Ernte ebenfalls schwierig zu handhaben.

Wie in dieser Studie beobachtet wurde, ist die aktuelle Subpopulation von 199 Genotypen sehr hilfreich für Assoziationsstudien. Eine Einengung der genetischen Intervalle von QTL mit großem Effekt durch Erhöhung der Anzahl von Markern wird die Auflösung dieser QTL verbessern, die für eine markergestützte Selektion verwendet und weiter in Züchtungsschemata zur Ertragssteigerung integriert werden könnten.

7 Supplementary Tables

Table S1. All two-rowed spring barley landrace accessions investigated in current study. Potential duplicate accessions are given the same group number

Entry No	Accession name	Row type	Phenotype	Potential duplicate group	Origin
1279	HOR2566	2	naked	1	Ethiopia
1328	HOR5917	2	naked	1	Ethiopia
1338	HOR5964	2	naked	1	Ethiopia
1339	HOR5965	2	naked	1	Ethiopia
1370	HOR14033	2	naked	1	Ethiopia
1278	HOR2551	2	naked	2	Ethiopia
1331	HOR5921	2	naked	2	Ethiopia
1384	HOR10484	2	naked	2	Iraq
1336	HOR5929	2	naked	3	Ethiopia
1359	HOR6966	2	naked	3	Ethiopia
1275	HOR1729	2	naked	4	Ethiopia
1344	HOR6478	2	naked	4	Ethiopia
1282	HOR3290	2	naked	5	Ethiopia
1290	HOR3592	2	naked	5	Ethiopia
1293	HOR3595	2	naked	5	Ethiopia
1287	HOR3536	2	naked	6	Ethiopia
1304	HOR4288	2	naked	6	Ethiopia
1276	HOR2546	2	naked	7	Ethiopia
1329	HOR5918	2	naked	7	Ethiopia
1257	HOR1431	2	naked	8	Afghanistan
1374	HOR2774	2	naked	8	Iran
1377	HOR2777	2	naked	8	Iran
1379	HOR2779	2	naked	8	Iran
1381	HOR2781	2	naked	8	Iran
1258	HOR3777	2	naked	9	Austria
1262	HOR9579	2	naked	9	CSFR
1263	HOR14336	2	naked	9	CSFR
1354	HOR6905	2	naked	9	Ethiopia
1398	BCC1489	2	naked	9	Russia
1267	HOR7385	2	naked	10	Czech Republic
1268	HOR7386	2	naked	10	Czech Republic
211	HOR2819	2	hulled	11	Iran
208	HOR2815	2	hulled	11	Iran

549	HOR8099	2	hulled	12	Turkey
551	HOR8113	2	hulled	12	Turkey
542	HOR7970	2	hulled	12	Turkey
541	HOR7969	2	hulled	13	Turkey
248	HOR10651	2	hulled	13	Iraq
224	HOR2849	2	hulled	14	Iran
227	HOR2853	2	hulled	14	Iran
517	HOR634	2	hulled	15	Turkey
518	HOR655	2	hulled	15	Turkey
115	HOR9612	2	hulled	16	Georgia
7895	HOR7985	2	hulled	16	Turkey
272	HOR9880	2	hulled	17	Libya
273	HOR9883	2	hulled	17	Libya
153	HOR2195	2	hulled	18	Germany
158	HOR3757	2	hulled	18	Germany
178	HOR937	2	hulled	19	Greece
555	HOR203	2	hulled	19	Ukraine
367	HOR7365	2	hulled	20	Slovakia
380	HOR7389	2	hulled	20	Slovakia
366	HOR7364	2	hulled	21	Slovakia
378	HOR7377	2	hulled	21	Slovakia
306	HOR7540	2	hulled	22	Poland
313	HOR8817	2	hulled	22	Poland
375	HOR7373	2	hulled	23	Slovakia
337	BCC1481	2	hulled	23	Russia
339	HOR7335	2	hulled	23	Slovakia
483	HOR9695	2	hulled	23	Slovakia
484	HOR9696	2	hulled	23	Slovakia
1365	HOR11435	2	naked	24	Ethiopia
189	HOR2674	2	hulled	24	Greece
53	HOR9855	2	hulled	25	Austria
265	HOR11123	2	hulled	25	Italy
315	HOR8825	2	hulled	25	Poland
147	HOR354	2	hulled	26	Germany
326	HOR8854	2	hulled	26	Poland
341	HOR7337	2	hulled	27	Slovakia
371	HOR7369	2	hulled	27	Slovakia
355	HOR7353	2	hulled	28	Slovakia
399	HOR8633	2	hulled	28	Slovakia
400	HOR8634	2	hulled	28	Slovakia

335	HOR9453	2	hulled	29	Romania
382	HOR7391	2	hulled	29	Slovakia
415	HOR8649	2	hulled	29	Slovakia
353	HOR7351	2	hulled	30	Slovakia
62	HOR7344	2	hulled	30	Czech Republic
342	HOR7338	2	hulled	30	Slovakia
352	HOR7350	2	hulled	30	Slovakia
357	HOR7355	2	hulled	30	Slovakia
362	HOR7360	2	hulled	30	Slovakia
372	HOR7370	2	hulled	30	Slovakia
376	HOR7374	2	hulled	30	Slovakia
395	HOR8629	2	hulled	30	Slovakia
413	HOR8647	2	hulled	30	Slovakia
428	HOR8679	2	hulled	30	Slovakia
488	HOR9700	2	hulled	30	Slovakia
26	HOR108	2	hulled		Austria
89	HOR7956	2	hulled		Ethiopia
92	HOR10259	2	hulled		Ethiopia
104	HOR10761	2	hulled		Ethiopia
136	HOR10741	2	hulled		Georgia
256	HOR10782	2	hulled		Italy
290	HOR7519	2	hulled		Poland
316	HOR8828	2	hulled		Poland
327	HOR10410	2	hulled		Poland
363	HOR7361	2	hulled		Slovakia
439	HOR9642	2	hulled		Slovakia
489	HOR1732	2	hulled		Sweden
493	HOR12166	2	hulled		Switzerland
497	HOR12206	2	hulled		Switzerland
1259	HOR3965	2	naked		Austria
1305	HOR4450	2	naked		Ethiopia
1371	HOR337	2	naked		Germany
1393	HOR8827	2	naked		Poland
5	HOR1707	2	hulled		Afghanistan
11	BCC2	2	hulled		Afghanistan
195	HOR2785	2	hulled		Iran
196	HOR2792	2	hulled		Iran
201	HOR2797	2	hulled		Iran
204	HOR2807	2	hulled		Iran
205	HOR2808	2	hulled		Iran

206	HOR2809	2	hulled		Iran
219	HOR2833	2	hulled		Iran
228	HOR2857	2	hulled		Iran
234	HOR2872	2	hulled		Iran
277	HOR10162	2	hulled		Libya
279	HOR10169	2	hulled		Libya
280	HOR10280	2	hulled		Libya
1307	HOR4453	2	naked		Ethiopia
1325	HOR5827	2	naked		Ethiopia
1	BCC3	2	hulled		Afghanistan
4	HOR1675	2	hulled		Afghanistan
7	HOR1751	2	hulled		Afghanistan
8	HOR1794	2	hulled		Afghanistan
19	HOR10092	2	hulled		Algeria
30	HOR828	2	hulled		Austria
43	HOR3955	2	hulled		Austria
46	HOR3958	2	hulled		Austria
52	HOR9853	2	hulled		Austria
54	HOR9856	2	hulled		Austria
57	HOR182	2	hulled		Croatia
60	BCC1436	2	hulled		Czech Republic
61	BCC1437	2	hulled		Czech Republic
68	HOR8658	2	hulled		Egypt
84	HOR7939	2	hulled		Ethiopia
91	HOR10258	2	hulled		Ethiopia
108	HOR9919	2	hulled		France
113	HOR9607	2	hulled		Georgia
120	HOR9619	2	hulled		Georgia
121	HOR9626	2	hulled		Georgia
122	HOR9630	2	hulled		Georgia
123	HOR9816	2	hulled		Georgia
124	HOR9869	2	hulled		Georgia
126	HOR9871	2	hulled		Georgia
127	HOR10357	2	hulled		Georgia
128	HOR10463	2	hulled		Georgia
137	HOR10749	2	hulled		Georgia
142	HOR10972	2	hulled		Georgia
143	HOR10973	2	hulled		Georgia
145	HOR11320	2	hulled		Georgia
152	HOR2072	2	hulled		Germany

156	HOR2212	2	hulled		Germany
157	HOR2962	2	hulled		Germany
167	HOR14364	2	hulled		Germany
169	HOR15961	2	hulled		Germany
171	HOR725	2	hulled		Greece
172	HOR753	2	hulled		Greece
173	HOR869	2	hulled		Greece
181	HOR1122	2	hulled		Greece
188	HOR1943	2	hulled		Greece
191	HOR3759	2	hulled		Hungary
199	HOR2795	2	hulled		Iran
220	HOR2838	2	hulled		Iran
222	HOR2844	2	hulled		Iran
240	HOR10610	2	hulled		Iraq
245	HOR10620	2	hulled		Iraq
284	HOR7484	2	hulled		Poland
288	HOR7515	2	hulled		Poland
300	HOR7531	2	hulled		Poland
304	HOR7537	2	hulled		Poland
308	HOR7542	2	hulled		Poland
309	HOR8808	2	hulled		Poland
311	HOR8812	2	hulled		Poland
318	HOR8832	2	hulled		Poland
319	HOR8833	2	hulled		Poland
328	HOR10600	2	hulled		Poland
329	HOR10635	2	hulled		Poland
330	HOR1391	2	hulled		Romania
333	HOR9450	2	hulled		Romania
338	HOR7334	2	hulled		Slovakia
344	HOR7340	2	hulled		Slovakia
347	HOR7343	2	hulled		Slovakia
364	HOR7362	2	hulled		Slovakia
381	HOR7390	2	hulled		Slovakia
384	HOR7393	2	hulled		Slovakia
394	HOR8627	2	hulled		Slovakia
396	HOR8630	2	hulled		Slovakia
402	HOR8636	2	hulled		Slovakia
404	HOR8638	2	hulled		Slovakia
407	HOR8641	2	hulled		Slovakia
409	HOR8643	2	hulled		Slovakia

412	HOR8646	2	hulled		Slovakia
414	HOR8648	2	hulled		Slovakia
417	HOR8651	2	hulled		Slovakia
425	HOR8675	2	hulled		Slovakia
433	HOR9636	2	hulled		Slovakia
438	HOR9641	2	hulled		Slovakia
451	HOR9658	2	hulled		Slovakia
452	HOR9660	2	hulled		Slovakia
454	HOR9662	2	hulled		Slovakia
460	HOR9670	2	hulled		Slovakia
463	HOR9673	2	hulled		Slovakia
465	HOR9675	2	hulled		Slovakia
470	HOR9680	2	hulled		Slovakia
472	HOR9682	2	hulled		Slovakia
473	HOR9684	2	hulled		Slovakia
474	HOR9685	2	hulled		Slovakia
480	HOR9692	2	hulled		Slovakia
481	HOR9693	2	hulled		Slovakia
499	HOR4469	2	hulled		Syria
510	HOR509	2	hulled		Turkey
512	HOR527	2	hulled		Turkey
514	HOR572	2	hulled		Turkey
521	HOR902	2	hulled		Turkey
522	HOR1178	2	hulled		Turkey
526	HOR1260	2	hulled		Turkey
527	HOR1408	2	hulled		Turkey
528	HOR1415	2	hulled		Turkey
529	HOR1740	2	hulled		Turkey
534	HOR4078	2	hulled		Turkey
535	HOR4093	2	hulled		Turkey
538	HOR7964	2	hulled		Turkey
543	HOR7977	2	hulled		Turkey
544	HOR7978	2	hulled		Turkey
545	HOR7985	2	hulled		Turkey
546	HOR8006	2	hulled		Turkey
552	HOR8158	2	hulled		Turkey
554	HOR10567	2	hulled		Turkey
557	HOR757	2	hulled		Yugoslavia
558	BCC1434	2	hulled		Yugoslavia
559	HOR4016	2	hulled		Yugoslavia

560	HOR4017	2	hulled		Yugoslavia
1269	HOR7387	2	naked		Czech Republic
1274	HOR1726	2	naked		Ethiopia
1280	HOR2720	2	naked		Ethiopia
1284	HOR3292	2	naked		Ethiopia
1286	HOR3294	2	naked		Ethiopia
1292	HOR3594	2	naked		Ethiopia
1295	HOR3597	2	naked		Ethiopia
1298	HOR3600	2	naked		Ethiopia
1299	HOR3601	2	naked		Ethiopia
1310	HOR5199	2	naked		Ethiopia
1312	HOR5417	2	naked		Ethiopia
1314	HOR5595	2	naked		Ethiopia
1315	HOR5689	2	naked		Ethiopia
1320	HOR5702	2	naked		Ethiopia
1321	HOR5712	2	naked		Ethiopia
1323	HOR5719	2	naked		Ethiopia
1345	HOR6765	2	naked		Ethiopia
1350	HOR6891	2	naked		Ethiopia
1351	HOR6892	2	naked		Ethiopia
1385	HOR949	2	naked		Italy
1386	BCC1496	2	naked		Kazakhstan
1387	HOR7533	2	naked		Poland
1389	HOR8809	2	naked		Poland
1390	HOR8810	2	naked		Poland
1391	HOR8813	2	naked		Poland
1395	HOR8830	2	naked		Poland
1401	HOR4463	2	naked		Russia

Table S2. Passport data of potential duplicate accessions along with BLUES from field experiments of all fourteen traits investigated in current study. Each potential duplicate group is indicated by number in the column "Group".

Entry	Group	ACQDATEE	Origin	HI	GPS	EL	AWNL	HD	Pht	SDA	SDL	SDW	TKW	SPD	Stig.H	RS	N-C
1279	1	--	Ethiopia	0.32	22.83	9.50	5.92	63.50	81.00	22.98	7.53	3.93	43.25	2.40	2	2	naked
1370	1		Ethiopia	0.32	19.75	8.93	6.12	62.25	82.50	21.68	7.25	3.75	39.42	2.21	2	2	naked
1328	1	--	Ethiopia	0.33	20.67	8.85	5.92	61.50	83.25	22.75	7.45	3.83	41.13	2.34	2	2	naked
1339	1	--	Ethiopia	0.35	22.58	8.60	7.11	62.50	83.25	23.03	7.53	3.85	42.50	2.63	2	2	naked
1338	1	--	Ethiopia	0.31	22.67	8.97	6.22	61.75	89.25	23.13	7.63	3.85	40.49	2.53	2	2	naked
1278	2	--	Ethiopia	0.29	21.42	10.88	12.08	63.25	97.50	28.90	9.30	4.08	49.22	1.97	2	2	naked
1331	2	--	Ethiopia	0.32	21.83	10.38	11.38	62.50	93.50	27.98	9.13	3.90	45.24	2.10	2	2	naked
1384	2	1985	Iraq	0.36	20.75	9.75	11.88	62.50	94.25	28.60	9.10	4.03	48.77	2.13	2	2	naked
1336	3	--	Ethiopia	0.32	20.17	8.19	0.00	60.00	85.75	21.30	7.23	3.70	38.59	2.46	2	0	naked
1359	3	--	Ethiopia	0.37	19.92	8.50	0.00	61.00	83.00	21.95	7.38	3.75	39.08	2.34	2	0	naked
1275	4	--	Ethiopia	0.33	19.42	8.22	12.90	59.50	84.25	23.38	8.35	3.63	43.17	2.36	2	2	naked
1344	4	--	Ethiopia	0.34	18.92	9.67	12.67	61.25	77.75	25.90	8.98	3.70	47.33	1.96	2	2	naked
1290	5	1963	Ethiopia	0.35	21.25	8.93	13.31	64.25	77.50	26.80	8.70	4.08	50.66	2.38	2	2	naked
1293	5	1959	Ethiopia	0.33	19.83	9.50	12.95	64.00	78.50	28.05	9.10	4.10	51.77	2.09	2	2	naked
1282	5	1959	Ethiopia	0.34	19.92	8.97	12.81	63.00	79.25	27.58	9.05	4.00	49.36	2.22	2	2	naked
1287	6	--	Ethiopia	0.34	18.58	8.42	12.58	66.25	75.00	27.80	8.83	4.00	51.12	2.21	2	2	naked
1304	6	--	Ethiopia	0.37	17.58	8.20	12.14	65.75	78.25	27.15	8.83	4.03	48.28	2.14	2	2	naked
1276	7	--	Ethiopia	0.35	19.83	6.37	12.32	63.50	79.25	30.25	9.20	4.28	57.80	3.11	2	2	naked
1329	7	--	Ethiopia	0.34	20.33	5.88	12.28	62.00	84.25	28.03	9.00	4.03	52.17	3.46	2	2	naked
1257	8	1935	Afghanistan	0.41	20.50	8.92	11.46	60.75	76.25	27.25	8.73	4.00	55.66	2.30	2	2	naked

1381	8	1954	Iran	0.40	18.50	8.92	10.21	58.75	75.75	26.25	8.73	3.93	53.59	2.07	2	2	naked
1379	8	1954	Iran	0.42	19.00	8.34	11.15	59.25	69.50	28.18	9.00	4.08	56.83	2.28	2	2	naked
1377	8	1954	Iran	0.38	20.08	9.78	12.10	70.50	65.00	26.00	9.08	3.70	43.95	2.05	2	2	naked
1374	8	1954	Iran	0.44	19.25	7.35	10.98	59.00	72.50	26.88	8.73	3.95	56.72	2.62	2	2	naked
1258	9	1964	Austria	0.40	19.83	9.00	12.21	62.00	74.25	26.78	8.78	3.93	57.61	2.20	2	2	naked
1354	9	1968	Ethiopia	0.41	20.17	8.81	11.55	62.00	70.00	27.78	8.98	4.05	58.31	2.29	2	2	naked
1263	9		CSFR	0.41	21.83	8.92	12.77	63.00	75.50	27.43	9.08	3.93	58.79	2.45	2	2	naked
1398	9	1938	Russia	0.41	21.00	9.25	13.08	64.25	67.50	28.80	9.40	4.05	59.19	2.27	2	2	naked
1262	9	--	CSFR	0.42	18.67	8.53	12.97	66.00	71.25	25.85	8.78	3.90	54.12	2.19	2	2	naked
1267	10	1974	Czech Rep.	0.40	24.67	9.22	13.63	69.50	80.00	24.93	8.25	3.80	46.52	2.68	2	2	naked
1268	10	1974	Czech Rep.	0.37	22.83	10.13	14.25	69.25	78.75	24.88	8.28	3.80	50.49	2.25	2	2	naked
208	11	1954	Iran	0.43	18.50	7.67	9.68	61.50	63.75	32.28	11.88	3.85	51.11	2.41	1	2	hulled
211	11	1954	Iran	0.46	17.00	7.68	9.26	63.75	63.25	30.35	11.13	3.83	49.14	2.21	1	2	hulled
549	12	1976	Turkey	0.42	15.67	6.49	14.86	62.50	81.50	35.15	11.23	4.08	67.35	2.41	2	2	hulled
551	12	1976	Turkey	0.48	18.75	7.58	12.87	63.00	80.75	33.20	10.78	4.05	65.78	2.47	2	2	hulled
542	12	1976	Turkey	0.45	17.92	7.72	13.48	64.25	83.75	32.73	10.75	4.00	63.21	2.32	2	2	hulled
248	13	1987	Iraq	0.45	20.17	7.67	12.25	62.75	82.50	33.25	10.83	4.08	63.87	2.63	2	2	hulled
541	13	1976	Turkey	0.44	19.67	7.93	11.98	63.00	84.25	33.98	11.03	4.10	64.67	2.48	2	2	hulled
224	14	1954	Iran	0.44	18.83	6.69	11.90	62.50	71.25	35.58	11.30	4.36	60.01	2.81	2	1	hulled
227	14	1954	Iran	0.57	19.33	7.03	11.61	63.50	73.25	36.83	11.63	4.30	57.88	2.75	2	1	hulled
517	15	1928	Turkey	0.45	23.42	8.71	9.73	62.75	84.50	29.15	10.00	3.93	48.81	2.69	1	2	hulled
518	15	--	Turkey	0.48	18.75	8.32	12.43	62.00	71.50	34.70	12.08	3.95	62.45	2.25	1	2	hulled
115	16	1981	Georgia	0.48	21.92	8.62	10.28	64.00	75.50	35.15	11.30	4.18	66.27	2.54	1	1	hulled

7985	16	1976	Turkey	0.54	19.33	8.47	10.10	62.25	60.75	30.15	10.55	3.83	53.74	2.28	1	1	hulled
272	17	1981	Libya	0.39	17.17	4.92	14.38	56.75	65.25	29.18	10.15	3.88	45.96	3.49	2	2	hulled
273	17	1981	Libya	0.38	16.17	5.17	14.14	56.75	68.75	29.45	10.33	3.85	45.36	3.13	2	2	hulled
153	18	--	Germany	0.39	25.83	10.78	15.11	74.25	91.75	27.48	9.28	3.90	52.42	2.40	2	2	hulled
158	18	1964	Germany	0.40	27.58	11.04	14.43	72.25	93.50	27.03	9.33	3.93	52.21	2.50	2	2	hulled
178	19	1941	Greece	0.42	24.67	10.46	12.46	68.00	87.00	27.28	9.80	3.72	47.42	2.36	2	2	hulled
555	19	--	Ukraine	0.41	23.08	7.78	8.77	67.50	90.85	26.75	9.40	3.81	52.19	2.97	2	2	hulled
367	20	1974	Slovakia	0.49	26.00	9.35	11.33	66.25	84.50	26.00	9.03	3.85	49.12	2.78	2	2	hulled
380	20	1974	Slovakia	0.52	25.33	9.61	12.23	65.25	78.75	26.88	9.15	3.98	50.72	2.64	2	2	hulled
366	21	1974	Slovakia	0.48	26.50	9.79	13.04	69.25	77.25	26.13	9.00	3.90	50.58	2.71	2	2	hulled
378	21	1974	Slovakia	0.43	25.25	10.13	12.79	67.00	86.25	26.90	9.20	3.93	50.01	2.49	2	2	hulled
306	22	1976	Poland	0.49	26.33	9.84	12.62	69.00	74.50	26.33	8.93	3.83	49.70	2.68	2	2	hulled
313	22	1978	Poland	0.47	27.42	10.49	12.54	68.75	82.50	28.50	9.53	3.98	52.13	2.61	2	2	hulled
375	23	1974	Slovakia	0.54	27.92	9.73	13.48	70.75	62.25	27.15	9.10	4.00	49.59	2.87	2	2	hulled
484	23	1981	Slovakia	0.47	25.17	9.94	11.73	66.00	77.00	26.23	9.08	3.88	50.01	2.53	2	2	hulled
337	23	1935	Russia	0.44	22.42	9.59	12.43	68.25	79.50	27.75	9.68	3.83	49.72	2.34	2	2	hulled
483	23	1981	Slovakia	0.49	26.67	9.93	11.94	67.25	75.50	25.25	8.73	3.88	50.18	2.69	2	2	hulled
339	23	1974	Slovakia	0.47	25.33	9.69	11.58	66.50	76.25	25.78	9.00	3.83	47.79	2.61	2	2	hulled
189	24	1942	Greece	0.48	26.08	10.08	12.17	70.00	83.75	26.10	9.35	3.75	46.05	2.59	2	2	hulled
1365	24	1994	Ethiopia	0.36	27.75	10.42	11.26	73.25	77.50	22.10	7.75	3.70	42.72	2.66	2	2	naked
53	25	1982	Austria	0.50	24.42	9.87	13.75	70.25	64.75	26.08	9.10	3.88	51.31	2.47	2	2	hulled
265	25	1993	Italy	0.50	25.00	9.98	13.58	69.00	66.75	27.38	9.20	3.98	54.67	2.51	2	2	hulled
315	25	1978	Poland	0.52	24.83	8.89	13.07	68.50	62.25	27.15	9.10	3.95	52.92	2.79	2	2	hulled

147	26	--	Germany	0.40	24.17	10.74	12.83	68.75	92.00	29.13	9.68	3.98	52.12	2.25	2	2	hulled
326	26	1978	Poland	0.41	25.50	9.84	12.39	68.75	90.50	27.93	9.48	3.90	52.07	2.59	2	2	hulled
341	27	1974	Slovakia	0.44	27.17	10.29	10.75	70.00	81.25	26.73	9.40	3.75	48.81	2.64	2	2	hulled
371	27	1974	Slovakia	0.45	28.08	10.71	11.02	68.25	83.75	25.50	9.03	3.73	48.65	2.62	2	2	hulled
355	28	1974	Slovakia	0.46	28.83	10.21	11.29	69.25	78.50	27.08	9.33	3.88	51.09	2.82	2	2	hulled
399	28	1977	Slovakia	0.48	27.17	9.52	10.43	68.25	80.25	25.13	8.88	3.75	46.04	2.85	2	2	hulled
400	28	1977	Slovakia	0.49	27.33	9.42	11.17	69.00	74.00	25.50	8.93	3.83	45.75	2.90	2	2	hulled
335	29	--	Romania	0.45	28.25	10.28	10.51	68.25	81.25	25.88	9.08	3.85	47.90	2.75	2	2	hulled
382	29	1974	Slovakia	0.48	26.92	9.54	11.63	68.25	79.50	26.18	9.20	3.83	47.14	2.82	2	2	hulled
415	29	1977	Slovakia	0.46	26.17	9.97	10.23	67.75	82.00	25.88	9.25	3.75	47.33	2.62	2	2	hulled
376	30	1974	Slovakia	0.44	26.25	10.69	10.68	68.00	85.25	27.70	9.48	3.90	52.47	2.46	2	2	hulled
352	30	1974	Slovakia	0.50	27.42	10.00	10.79	68.25	81.00	26.78	9.23	3.85	49.73	2.74	2	2	hulled
488	30	1981	Slovakia	0.49	27.08	10.34	11.17	70.25	74.25	25.85	9.03	3.85	48.68	2.62	2	2	hulled
362	30	1974	Slovakia	0.46	27.75	10.16	10.68	68.50	76.25	26.25	9.18	3.83	49.50	2.73	2	2	hulled
353	30	1974	Slovakia	0.48	27.25	9.65	10.77	69.00	73.75	25.85	9.08	3.78	47.95	2.82	2	2	hulled
372	30	1974	Slovakia	0.46	26.67	9.29	10.75	68.50	77.00	26.63	9.25	3.90	49.84	2.87	2	2	hulled
395	30	1977	Slovakia	0.50	28.67	10.89	10.36	70.50	75.25	26.55	9.35	3.85	48.34	2.63	2	2	hulled
342	30	1974	Slovakia	0.44	26.08	10.02	11.72	68.75	78.75	26.05	9.13	3.80	48.15	2.60	2	2	hulled
62	30	1974	Czech Rep.	0.44	26.83	9.96	11.26	69.25	74.25	26.60	9.13	3.85	49.32	2.69	2	2	hulled
357	30	1974	Slovakia	0.45	27.17	9.56	11.17	69.00	75.75	27.65	9.65	3.88	49.13	2.84	2	2	hulled
413	30	1977	Slovakia	0.48	27.25	9.41	11.13	68.75	78.25	25.75	9.05	3.80	47.23	2.90	2	2	hulled
428	30	1977	Slovakia	0.50	26.58	10.15	10.93	68.00	79.50	26.38	9.20	3.82	48.42	2.62	2	2	hulled

*HD “heading date”, Pht “plant height”, GPS “grains per spike”, HI “harvest index”, SPD “spike density”, TKW “thousand kernel weight”, AwnL “awn length”, EL “ear length”, SDA “seed area”, SDL “seed length”, SDW “seed width” and Stig.H “stigma hairiness”, RS “awn roughness”,N-C “naked/covered”.

Table S3. SNPs significantly associated with at least one or more traits above the respective FDR threshold. Second column (from right to left) show the grouping of SNPs into QTL based on estimated LD decay. Already reported genes or QTL corresponding to different traits at close vicinity to detected QTL in current study are also listed in column 2 (from right to left) with corresponding literature in the first right column. From left to right are SNPs, chromosome and SNP position on respective chromosome. –logP value of each associated SNP is presented under the corresponding trait.

			-logP values of SNPs significantly associated with each trait														**Candidate(s)/QTL corresponding to different traits**	**Literature**
SNPs	**Chr.**	**Pos. (cM)**	**HD**	**GPS**	**Pht**	**HI**	**TKW**	**SDA**	**SDL**	**SDW**	**EL**	**SPD**	**AwnL**	**AwnR**	**Stig.H**	**N-C**		
BOPA2_12_30945	1	3.6												4.1			*Number of ears with grains*	(Ingvordsen et al. 2015)
BOPA2_12_30876	1	22.3												3				
SCRI_RS_14227	1	27.3						3.1										
SCRI_RS_151764	1	46.5						3.3					3				*HvCMF10*	(Alqudah et al. 2014)
BOPA2_12_30683	1	46.6											3.1					
BOPA1_ABC10636-1-4-285	1	47.8				3.1												
BOPA1_8613-278	1	47.8																
BOPA1_3217-929	1	47.8	3.3															
BOPA2_12_11301	1	47.9													6.3			
BOPA1_ABC13652-1-2-156	1	49.6												4				
SCRI_RS_229636	1	52.7													3.4			
SCRI_RS_213455	1	54.5													3.4			

BOPA1_1190-86	1	55.9										3.3					
SCRI_RS_120605	1	61.5												3.1			
BOPA1_10070-1435	1	61.5															
BOPA1_ABC16273-1-1-48	1	61.8						3.6	3.7								
BOPA1_5772-1176	1	72.4	2.9														
SCRI_RS_181353	1	76.8													3.2		
BOPA1_4691-721	1	83.6						3.6		4.6						*Biomass yield*	(Wehner et al. 2015)
BOPA2_12_31319	1	92.4	3.1														
BOPA1_2711-234	1	95.9												3.2		*Ppd-H2/HvFT3*	
SCRI_RS_232650	1	97.6			3												
BOPA2_12_21522	1	117.5				3.2											
SCRI_RS_141771	2	8.6			2.7												
BOPA1_12224-363	2	12.1								3.3							
SCRI_RS_154030	2	13.5	3.3													*Biomass yield*	(Wehner et al. 2015)
SCRI_RS_149462	2	14.4					4.6										
SCRI_RS_159228	2	17.6							3								
SCRI_RS_205712	2	18.9		3.8								3.7					
SCRI_RS_210172	2	18.9	4.9	6													
SCRI_RS_233272	2	18.9	8.4	5.1													

SCRI_RS_153798	2	18.9	4.5	3.5														
BK_12	2	19.9	4.2	4.4													*Ppd-H1*	(Comadran et al. 2012; Turner et al. 2005)
BK_14	2	19.9	4.2	4.4														
BK_16	2	19.9	4.2	4.4														
BOPA2_12_30871	2	19.9	4.2	4.4														
BOPA2_12_30872	2	19.9	4.2	4.4														
BK_15	2	19.9	3.6	3.6														
BOPA1_5880-2547	2	23.2													4			(Ingvordsen et al. 2015)
SCRI_RS_182270	2	24.5													4.4		*Grain yield*	
SCRI_RS_141655 SCRI_RS_229103	2 2	41.9 43.7	3.1												3.1		*HvCO18*	(Alqudah et al. 2014)
SCRI_RS_152206	2	48.4												3.5				
BOPA1_4049-233	2	56.5		4														
SCRI_RS_194640	2	56.5		4														
SCRI_RS_177375	2	57	3.3															
SCRI_RS_100054	2	57	4.4	3.3														
BOPA1_ABC04893-1-1-248	2	57		5				3.4	5									
BOPA1_4659-1261	2	57	5.6	5.1														
SCRI_RS_237688	2	57.1	4.4															

SCRI_RS_167594	2	57.4	5.6														
SCRI_RS_196026	2	57.7	5.6														
SCRI_RS_4969	2	58	3.3														
SCRI_RS_155067	2	58	3.1	5													
SCRI_RS_157207	2	58	3.1	5												*HvCEN*, MLOC_13032.1	(Alqudah et al. 2014) (Si et al. 2016)
SCRI_RS_134925	2	58	5.4	5.7													
SCRI_RS_222769	2	58.1	3.3														
SCRI_RS_127347	2	58.1	3														
BOPA1_1917-848	2	58.1	5.4														
SCRI_RS_208320	2	58.1	5.6														
SCRI_RS_240011	2	58.6					3.1										
SCRI_RS_116920	2	58.6		2.9													
SCRI_RS_235063	2	58.6	4.3														
BOPA2_12_30265	2	58.6	4.3														
BOPA1_7489-442	2	58.8			2.6												
BOPA1_4717-386	2	58.8		3.3													
BOPA1_9191-263	2	58.8	3.1														
SCRI_RS_231725	2	58.8	3.1														
SCRI_RS_144776	2	58.9			2.6												
SCRI_RS_165574	2	58.9		2.8													

SCRI_RS_135468	2	58.9		3.8														
BOPA1_6911-866	2	58.9		6.2														
SCRI_RS_1502	2	59.1	5.6	5.1														
BOPA2_12_11121	2	59.1		2.9	2.9													
SCRI_RS_125516	2	59.1	3.1												6.4			
BOPA2_12_30275	2	59.1													4.2			
BOPA2_12_10545	2	62	3.8															
BOPA1_6804-1197	2	62.5		3.8														
SCRI_RS_162413	2	62.5	3.4	3.6														
SCRI_RS_198848	2	64.6													4.2		*HvCO4*	(Alqudah et al. 2014)
BOPA2_12_31021	2	64.6		2.9														
SCRI_RS_156871	2	67.3		3.4														
SCRI_RS_9469	2	67.4													4.3			
BOPA1_6280-1098	2	67.4	3															
SCRI_RS_159024	2	67.9	3														*HvFT4*	(Alqudah et al. 2014)
SCRI_RS_73	2	67.9	3.7															
SCRI_RS_221992	2	67.9													3.2			
BOPA1_6852-506	2	68.5										3.3						
SCRI_RS_158285	2	74.2		3.3														
BOPA2_12_10859	2	74.4		3.3														

SCRI_RS_166540	2	75.6						3.2		3.2								
SCRI_RS_221795	2	75.6						2.9										
SCRI_RS_116694	2	86.8													4.1			
SCRI_RS_157097	2	94.4												6.8				
SCRI_RS_147985	2	94.4												7.2				
SCRI_RS_227965	2	106.9							3.4									
BOPA1_4218-1230	2	114.4													3.2		*Grain yield*	(Ingvordsen et al. 2015)
BOPA1_9701-925	2	114.9													3.2			
BOPA1_2464-1228	2	120								3.3								
SCRI_RS_119513	2	120				3												
SCRI_RS_226193	2	133.3									3.1							
SCRI_RS_119379	3	2.4									3.2							
SCRI_RS_129198	3	8.8								3.4								
BOPA2_12_10532	3	45.4														9.3		(Li et al. 2013a)
BOPA1_2391-566	3	46														15.1		
BOPA1_4256-833	3	46														15.1		
BOPA2_12_30474	3	46														9.3	*(Eibi1)*	
SCRI_RS_205711	3	46.3												8.4			*MLOC_81291.1*	
BOPA2_12_30467	3	51.1									3.8	5.7		3.1				
BOPA2_12_30002	3	51.8													3.6			

SCRI_RS_128810	3	54.5								2.9							*HvFT2*	(Alqudah et al. 2014)
SCRI_RS_12836	3	54.7								3.2								
SCRI_RS_127140	3	59.1	4.1															
SCRI_RS_147950	3	83.6							3									
SCRI_RS_138193	3	88	4.5											4.7				
BOPA1_3787-1223	3	88.2						3.3						3.4				
SCRI_RS_234342	3	100.3	3.1														*HvCMF1*	(Maurer et al. 2015)
BOPA2_12_31220	3	104.3													3.4			
SCRI_RS_151711	3	104.3													3.5			
SCRI_RS_153915	3	104.5													3.4			
SCRI_RS_209963	3	124.5														6		
SCRI_RS_177313	3	126.7											3.5					
SCRI_RS_175038	3	126.7											3.1					
SCRI_RS_168360	3	133				3											*Biomass yield*	(Wehner et al. 2015)
SCRI_RS_168977	3	133				2.9												
SCRI_RS_184593	3	135.3				3.6												
SCRI_RS_164726	3	142.2												4.2				
BOPA2_12_20198	3	142.6											4	8.4				
BOPA1_4403-885	3	142.6												4.4				
SCRI_RS_206483	3	142.6												6.2				

SCRI_RS_237864	3	144.4										4					
BOPA2_12_30055	3	151													3.3		
BOPA2_12_30150	4	14.4							3.1								
BOPA2_12_11300	4	14.4	3														
SCRI_RS_162743	4	20.9												3.3			
BOPA1_1593-1597	4	21.2								3.5							
BOPA2_12_30394	4	21.2													3.4		
BOPA1_4616-503	4	26.8										3.8					
SCRI_RS_75805	4	41.9					3.7										
SCRI_RS_2937	4	51												3.4			
BOPA1_ABC24906-1-1-279	4	51					5	3.4									
SCRI_RS_66562	4	54.3													3.5		
BOPA1_245-433	4	54.3		3.9													
BOPA2_12_10088	4	54.3		3.9													
BOPA2_12_30060	4	54.5		3.9													
BOPA2_12_31297	4	54.6		3.9													
SCRI_RS_235738	4	57.3				3.5											
SCRI_RS_202326	4	57.5												3.2			
BOPA1_4276-1082	4	58.1								3.6							
SCRI_RS_206179	4	59.5											4.6				

SCRI_RS_218193	4	59.5											4.5					
SCRI_RS_189861	4	59.5													5.2			
BOPA2_12_30620	4	59.8											4.5					
SCRI_RS_132067	4	59.8													5.2			
SCRI_RS_165912	4	60.1											4.6					
SCRI_RS_229658	4	60.1																
BOPA2_12_30455	4	60.3											4.5					
SCRI_RS_146941	4	61.5								2.9								
SCRI_RS_8100	4	61.5								2.9								
BOPA1_9149-1316	4	73.6				3.2												
BOPA2_12_31231	4	81.6				3.4											*Grain yield*	(Ingvordsen et al. 2015)
BOPA2_12_20648	4	81.6												3.3				
BOPA2_12_31523	4	81.6												4.4				
SCRI_RS_235762	4	112				3.5											*Vrn-H2*	(Comadran et al. 2012; Maurer et al. 2015; Wehner et al. 2015)
BOPA1_3417-1451	5	0.1				2.9											*Number of ears with grains*	(Ingvordsen et al. 2015)
BOPA2_12_31023	5	1.6					3.1											
SCRI_RS_220165	5	31.9													5.5			
BOPA2_12_30410	5	32.4													5.5			

SCRI_RS_145275	5	43			2.6												*Biomass yield, TKW,*	(Liu et al. 2014; Pasam et al. 2012; Wehner et al. 2015)
SCRI_RS_172679	5	43.5			2.6													
SCRI_RS_91468	5	43.6			2.6													
SCRI_RS_149232	5	43.7			2.6												*ari-e*	
BOPA2_12_10899	5	43.8							3.5									
SCRI_RS_63610	5	43.8			2.8													
SCRI_RS_145116	5	44			2.7													
SCRI_RS_6422	5	44			2.7													
SCRI_RS_167575	5	44			2.6													
SCRI_RS_200057	5	44			2.8													
SCRI_RS_137196	5	44.1			2.7													
SCRI_RS_218911	5	44.1			2.8													
BOPA2_12_30575	5	44.2			2.7													
SCRI_RS_156086	5	44.2			2.7													
BOPA1_5565-1908	5	44.2			2.6													
SCRI_RS_149440	5	44.2			2.6													
SCRI_RS_161655	5	44.2			2.6													
SCRI_RS_208177	5	44.2			2.6													
SCRI_RS_165878	5	46.5					3.1											
SCRI_RS_153476	5	48.2					3.2											

BOPA2_12_11512	5	48.4							3.5									
BOPA2_12_10953	5	48.4													4.4			
BOPA1_ABC07010-1-2-150	5	50.4	4.2														*HvCMF13*	
BOPA1_2482-126	5	50.4	4.1															
BOPA2_12_20297	5	55.6										3.5						
SCRI_RS_205235	5	55.7										3.3						
SCRI_RS_204275	5	80.2										3.5					*Biomass yield*	(Ingvordsen et al. 2015)
BOPA1_11944-542	5	80.3										5.8						
BOPA2_12_30962	5	87.4			2.9													
BOPA1_6170-304	5	95													3.4			
BOPA2_12_30619	5	98.9												4.6				
SCRI_RS_135254	5	114.8											3.6	3.5			*raw1*,	(Franckowiak 2008)
SCRI_RS_7191	5	119											3.3					
BOPA2_12_30590	5	119				3												
SCRI_RS_202774	5	121.2							3.2								*Vrn-H1*	(Maurer et al. 2015; Mohammadi et al. 2013)
SCRI_RS_174710	5	121.7												4.7				
SCRI_RS_188141	5	130.7			2.7								3.2				*Plant height*	(Pillen et al. 2003)
SCRI_RS_157935	5	131.2			3.5		3.3	2.9										
SCRI_RS_185613	5	131.7			2.8													

SCRI_RS_167426	5	143.7			4.1												*Wricke's ecovalence (W^2) for ear stability*	(Ingvordsen et al. 2015)
BOPA2_12_30666	5	151.9													3.2		*Biomass yield*	(Wehner et al. 2015)
BOPA2_12_30003	5	152.4			4.3													
SCRI_RS_150686	5	159.5								3.3								
SCRI_RS_4753	5	166.8		5.2														
SCRI_RS_203132	5	167.6											3.2					
BOPA1_5159-579	6	1.8	3.1															
SCRI_RS_129888	6	15.2													4.6			
BOPA2_12_30032	6	53.6			3												*Biomass yield*	(Wehner et al. 2015)
SCRI_RS_72672	6	54.9		5.8														
SCRI_RS_140091	6	54.9													3.5			
SCRI_RS_233266	6	54.9		5.1														
BOPA2_12_11386	6	58.5												3.1				
SCRI_RS_137215	6	71.7												3				
SCRI_RS_175709	6	86.8										5.6						
SCRI_RS_17542	6	118.4													5.2			
SCRI_RS_161101	7	0.3					3.4						3.1				*Grain number*	(Ingvordsen et al. 2015)
SCRI_RS_101902	7	0.9					3.7											
SCRI_RS_235422	7	13.6					4.4										*waxy*	(Li et al. 2014)
SCRI_RS_153202	7	13.9					4.1											

SCRI_RS_127224	7	21												3.3			*Grain number*	(Ingvordsen et al. 2015)
SCRI_RS_200391	7	23.7					3.6											
SCRI_RS_140096	7	24.2													5.1			
SCRI_RS_171103	7	24.2													4.6			
SCRI_RS_237670	7	26.7		3.3														
BOPA2_12_10878	7	27.3					4.4											
SCRI_RS_150053	7	27.6					3.1											
SCRI_RS_236580	7	30													3.7			
BOPA2_12_30894	7	34.2	3.9														*Vrn-H3/HvFT1*	(Casas et al. 2011; Cockram et al. 2007; Kikuchi and Handa 2009; Maurer et al. 2015)
BOPA2_12_30895	7	34.3	3.9															
SCRI_RS_178619	7	51.3		2.9													*HvCO1, HvSS1, grain number*	(Alqudah et al. 2014; Ingvordsen et al. 2015)
BOPA2_12_30880	7	54.4	4.4															
BOPA1_2669-1012	7	61.5	5														*HvFT2*	(Alqudah et al. 2014)
BOPA1_3186-1560	7	65.4													3.3		*HvCO12*	(Alqudah et al. 2014)
BOPA2_12_30125	7	67.9	3.3															
SCRI_RS_149647	7	70.2										3.6					*Biomass yield*	(Wehner et al.

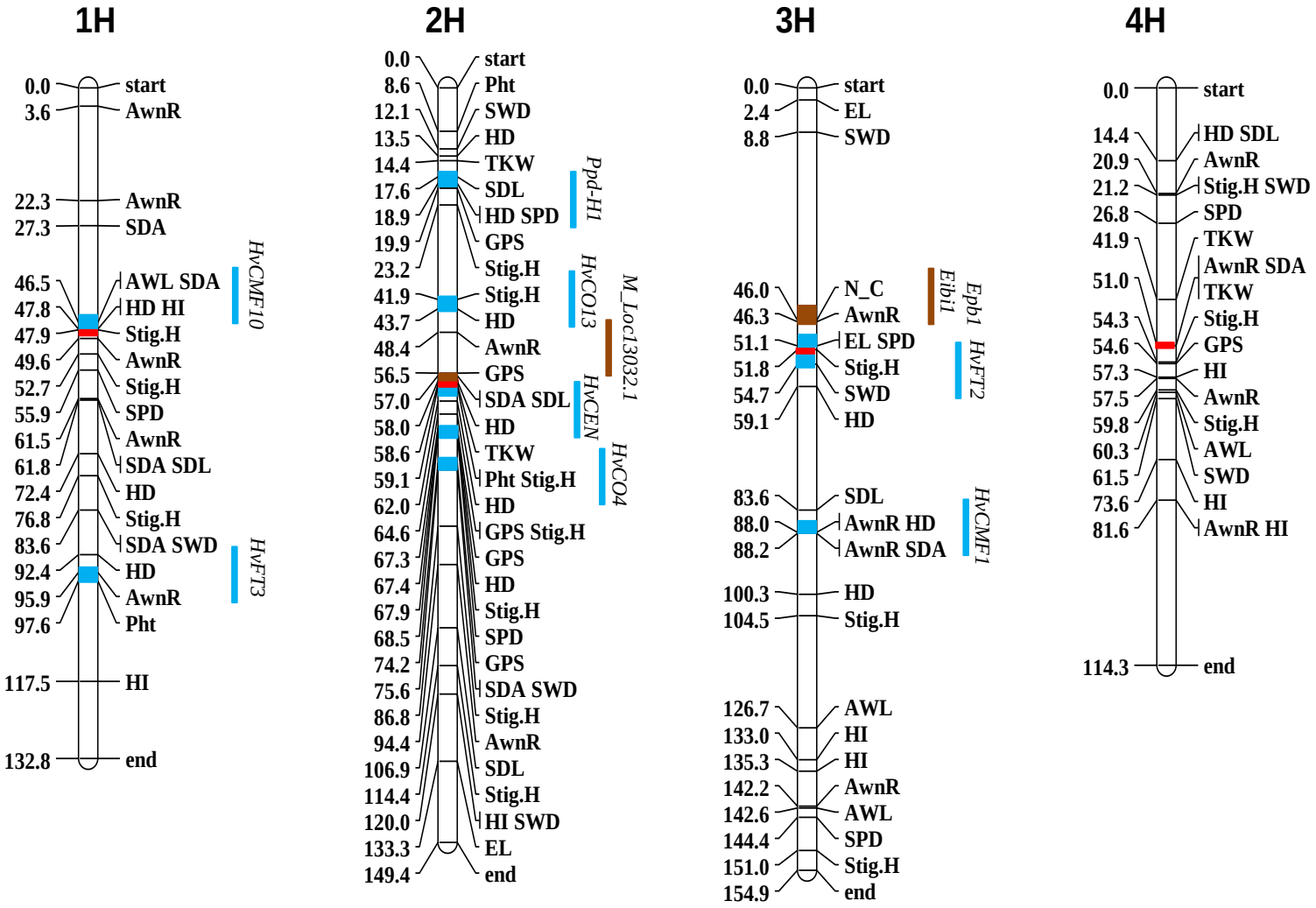
SCRI_RS_132879	7	70.5									3	4					2015)
SCRI_RS_141164	7	70.5									3	4					*Dsp1*
SCRI_RS_164251	7	70.5										3.6					(Taketa et al. 2011)
BOPA2_12_30486	7	70.5									3.7	4.3					
SCRI_RS_237881	7	70.6									3						
BOPA2_12_10698	7	70.6										3.2					
BOPA1_7712-674	7	70.6										3.4					
BOPA2_12_30760	7	70.6										3.6					
SCRI_RS_228070	7	70.6									3	4					
SCRI_RS_134797	7	70.6									3.7	4.4					
BOPA1_5138-265	7	70.6									3.2	3.6					
SCRI_RS_157035	7	70.7											3.1				
SCRI_RS_13927	7	70.7									3.2	4.3					
BOPA1_2924-1189	7	70.7									3	4					
BOPA2_12_30544	7	70.7									3	4					
SCRI_RS_148407	7	70.7									3	4					
SCRI_RS_185707	7	70.7									3	4					
BOPA2_12_20685	7	70.8														16.1	
BOPA2_12_10222	7	70.8									3.3	4.4					
BOPA2_12_30998	7	73.2							3.7		3.5						

BOPA1_3232-201	7	75.1									3.2						*Nud*	(Munoz-Amatriain et al. 2014; Taketa et al. 2008)
BOPA2_12_30565	7	76.5				5.1												
SCRI_RS_194291	7	76.7											3.5					
SCRI_RS_4562	7	84.3						3.3								15.6		
SCRI_RS_146640	7	86						4	4.7									
BOPA1_6468-770	7	86				3					3.8							
SCRI_RS_148722	7	86						4.1	5.1									
BOPA2_12_31395	7	86						4.5	4.4									
SCRI_RS_146157	7	86						3.4	4.3									
BOPA1_3568-149	7	86.6						3.2	4.5							22.3		
SCRI_RS_4556	7	91.9							5.7									
BOPA2_12_20684	7	97.3												3.6				
SCRI_RS_195908	7	118.4												3.5				
BOPA1_1590-544	7	131.4												3.9				
BOPA2_12_20949	7	131.4				4.8												

HD “heading date”, Pht “plant height”, GPS “grains per spike”, HI “harvest index”, SPD “spike density”, TKW “thousand kernel weight”, AwnL “awn length”, EL “ear length”, SDA “seed area”, SDL “seed length”, SDW “seed width”,Stig.H “stigma hairiness”, N-C “ naked caryopsis”

7.1 Supplementary Figure:

(a)



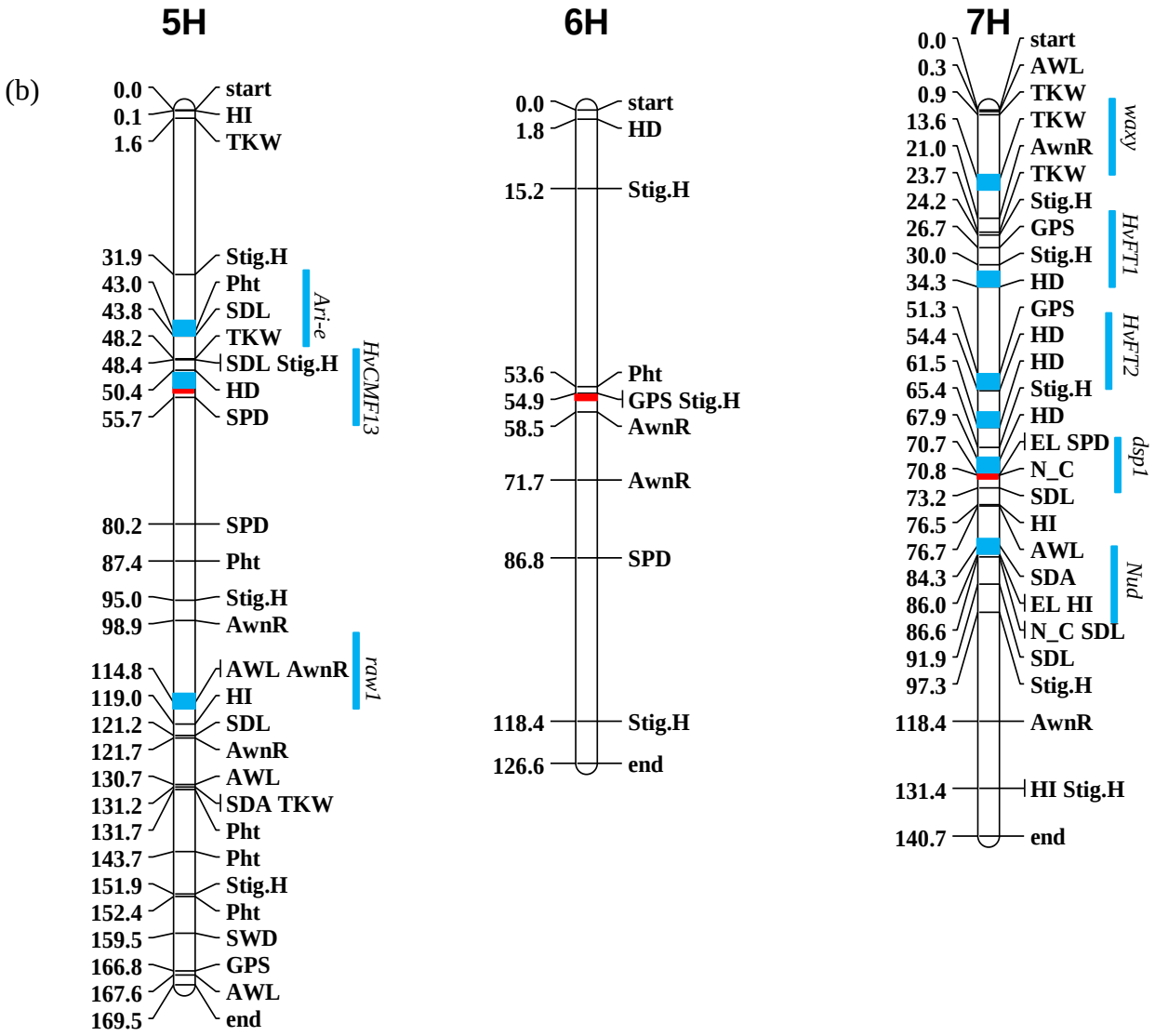


Figure S1, a and b. GWAS for all traits. Localization of genomic regions containing already reported genes/QTL are indicated in blue. Brown segments represent chromosomal section for candidate genes. A red segment on each chromosome represents the centromere. Only the genetic position of the highest associated SNP within a QTL interval was considered for respective traits.

8 References

Alqudah AM, Sharma R, Pasam RK, Graner A, Kilian B, Schnurbusch T (2014) Genetic Dissection of Photoperiod Response Based on GWAS of Pre-Anthesis Phase Duration in Spring Barley. Plos One 9

Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. Science 322:881-888

Angessa TT, Li C (2015) Exploration and Utilization of Genetic Diversity Exotic Germplasm for Barley Improvement. Exploration, Identification and Utilization of Barley Germplasm:223

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. Plos Genet 1:e60

Ariyadasa R, Mascher M, Nussbaumer T, Schulte D, Frenkel Z, Poursarebani N, Zhou RN, Steuernagel B, Gundlach H, Taudien S, Felder M, Platzer M, Himmelbach A, Schmutzer T, Hedley PE, Muehlbauer GJ, Scholz U, Korol A, Mayer KFX, Waugh R, Langridge P, Graner A, Stein N (2014) A Sequence-Ready Physical Map of Barley Anchored Genetically by Two Million Single-Nucleotide Polymorphisms. Plant Physiol 164:412-423

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741-745

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465:627-631

Badr A, Sch R, El Rabey H, Effgen S, Ibrahim H, Pozzi C, Rohde W, Salamini F (2000) On the origin and domestication history of barley (Hordeum vulgare). Mol Biol Evol 17:499-510

Baik BK, Ullrich SE (2008) Barley for food: Characteristics, improvement, and renewed interest. J Cereal Sci 48:233-242

Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, Weigel D (2006) The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of Arabidopsis thaliana. Nat Genet 38:711-715

Bariana H, Parry N, Barclay I, Loughman R, McLean R, Shankar M, Wilson R, Willey N, Francki M (2006) Identification and characterization of stripe rust resistance gene Yr34 in common wheat. Theor Appl Genet 112:1143-1148

BBSRC (1999) http://www.jic.bbsrc.ac.uk/welcome.htm

Bellucci E, Bitocchi E, Rau D, Nanni L, Ferradini N, Giardini A, Rodriguez M, Attene G, Papa R (2013) Population structure of barley landrace populations and gene-flow with modern varieties. Plos One 8:e83891

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 57:289-300

Bishop CW (1936) Origin and early diffusion of traction-plough. *Antiquity* 10: 261-278

Bjørnstad Å, Skinnes H, Thoresen K (1992) Comparisons between doubled haploid lines produced by anther culture, the Hordeum bulbosum-method and lines produced by single seed descent in barley crosses. Euphytica 66:135-144

Bothmer Rv (1992) The wild species of Hordeum: relationships and potential use for improvement of cultivated barley. Barley: genetics, biochemistry, molecular biology and biotechnology: 3-18

Bothmer RV, K. Sato, H. Knüpffer, Th. van Hintum (2003a) Barley diversity—an introduction. In: R. von Bothmer, Th. van Hintum, H. Knüpffer and K. Sato (eds), Diversity in Barley (*Hordeum vulgare*), pp. 3–8. Elsevier Science B.V., Amsterdam, The Netherlands.

Bothmer Rv, K. Sato, T. Komatsuda, S. Yasuda, G. Fischbeck (2003b) The domestication of cultivated barley. In: R. von Bothmer, Th. van Hintum, H. Knüpffer and K. Sato (eds), Diversity in Barley (Hordeum vulgare), pp. 9–27. Elsevier Science B.V., Amsterdam, The Netherlands

Bothmer RV, N. Jacobsen, C. Baden, C. Jørgensen, I. Linde-Laursen (1995) An ecogeographical study of the genus Hordeum. Systematic and ecogeographic studies on crop genepools, 7. IPGRI, Rome, 2nd ed 129p

Bothmer RV, Salomon B, Linde-Laursen I (1999) Chromosome pairing patterns in interspecific Hordeum lechleri x H-vulgare (cultivated barley) hybrids with 2n=21-29. Hereditas 131:109-120

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752-755

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165-1177

Brunner S, Keller B, Feuillet C (2000) Molecular mapping of the Rph7.g leaf rust resistance gene in barley (Hordeum vulgare L.). Theor Appl Genet 101:783-788

Caldwell DG, McCallum N, Shaw P, Muehlbauer GJ, Marshall DF, Waugh R (2004) A structured mutant population for forward and reverse genetics in barley (Hordeum vulgare L.). Plant J 40

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

Campoli C, Drosse B, Searle I, Coupland G, von Korff M (2012) Functional characterisation of HvCO1, the barley (Hordeum vulgare) flowering time ortholog of CONSTANS. Plant J 69:868-880

Casas AM, Djemel A, Ciudad FJ, Yahiaoui S, Ponce LJ, Contreras-Moreira B, Gracia MP, Lasa JM, Igartua E (2011) HvFT1 (VrnH3) drives latitudinal adaptation in Spanish barleys. Theor Appl Genet 122:1293-1304

Ceccarelli S, Grando S, Maatougui M, Michael M, Slash M, Haghparast R, Rahmanian M, Taheri A, Al-Yassin A, Benbelkacem A, Labdi M, Mimoun H, Nachit M (2010) Plant breeding and climate changes. J Agr Sci 148:627-637

Chen G, Komatsuda T, Ma JF, Nawrath C, Pourkheirandish M, Tagiri A, Hu Y-G, Sameri M, Li X, Zhao X, Liu Y, Li C, Ma X, Wang A, Nair S, Wang N, Miyao A, Sakuma S, Yamaji N, Zheng X, Nevo E (2011) An ATP-binding cassette subfamily G full transporter is essential for the retention of leaf water in both wild barley and rice. Proceedings of the National Academy of Sciences 108:12354-12359

Chen G, Sagi M, Weining S, Krugman T, Fahima T, Korol AB, Nevo E (2004) Wild barley eibi1 mutation identifies a gene essential for leaf water conservation. Planta 219:684-693

Clarke JD (2009) Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. Cold Spring Harbor protocols 2009:pdb.prot5177

Close TJ, Bhat PR, Lonardi S, Wu YH, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao SAM, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. Bmc Genomics 10

Close TJ, Wanamaker SI, Caldo RA, Turner SM, Ashlock DA, Dickerson JA, Wing RA, Muehlbauer GJ, Kleinhofs A, Wise RP (2004) A new resource for cereal genomics: 22K barley GeneChip comes of age. Plant Physiol 134:960-968

Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J Exp Bot 58:1231-1244

Cockram J, Thiel T, Steuernagel B, Stein N, Taudien S, Bailey PC, O'Sullivan DM (2012) Genome Dynamics Explain the Evolution of Flowering Time CCT Domain Gene Families in the Poaceae. Plos One 7

Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, Luo ZW, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WTB, Ramsay L, Mackay I, Balding DJ, Waugh R, O'Sullivan DM, Consortium A (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. P Natl Acad Sci USA 107:21611-21616

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

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

Comadran J, Russell JR, Booth A, Pswarayi A, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, van Eeuwijk FA, Thomas WTB, Romagosa I (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

Comadran J, Thomas WTB, van Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, Romagosa I, Hackett CA, Russell JR (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured Hordeum vulgare association-mapping population for the Mediterranean basin. Theor Appl Genet 119:175-187

Courtois B, Frouin J, Greco R, Bruschi G, Droc G, Hamelin C, Ruiz M, Clement G, Evrard JC, van Coppenole S, Katsantonis D, Oliveira M, Negrao S, Matos C, Cavigiolo S, Lupotto E, Piffanelli P, Ahmadi N (2012) Genetic Diversity and Population Structure in a European Collection of Rice. Crop Sci 52:1663-1675

Dai F, Chen Z-H, Wang X, Li Z, Jin G, Wu D, Cai S, Wang N, Wu F, Nevo E (2014) Transcriptome profiling reveals mosaic genomic origins of modern cultivated barley. Proceedings of the National Academy of Sciences 111:13403-13408

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499-510

Diamond JM (1993) Domestication of Plants in the Old-World, 2nd Edition - Zohary,D, Hopf,M. Nature 366:523-523

Dolezel J, Greilhuber J, Lucretti S, Meister A, Lysak MA, Nardi L, Obermayer R (1998) Plant genome size estimation by flow cytometry: Inter-laboratory comparison. Ann Bot-London 82:17-26

Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. Nature 456:728-731

Duan R, Xiong H, Wang A, Chen G (2015) Molecular Mechanisms Underlying Hull-Caryopsis Adhesion/Separation Revealed by Comparative Transcriptomic Analysis of Covered/Naked Barley (Hordeum vulgare L.). International Journal of Molecular Sciences 16:14181-14193

Dwivedi SL, Britt AB, Tripathi L, Sharma S, Upadhyaya HD, Ortiz R (2015) Haploids: constraints and opportunities in plant breeding. Biotechnology advances 33:812-829

Dwivedi SL, Ceccarelli S, Blair MW, Upadhyaya HD, Are AK, Ortiz R (2016) Landrace Germplasm for Improving Yield and Abiotic Stress Adaptation. Trends Plant Sci 21:31-42

Ellis R, Forster B, Robinson D, Handley L, Gordon D, Russell J, Powell W (2000) Wild barley: a source of genes for crop improvement in the 21st century? J Exp Bot 51:9-17

Endresen DTF, Street K, Mackay M, Bari A, De Pauw E (2011) Predictive Association between Biotic Stress Traits and Eco-Geographic Data for Wheat and Barley Landraces. Crop Sci 51:2036-2055

Enki DG, Trendafilov NT, Jolliffe IT (2013) A clustering approach to interpretable principal components. J Appl Stat 40:583-599

Evanno G, Regnaut S, Goudet J (2005a) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620

Evanno G, Regnaut S, Goudet J (2005b) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, 4th ed. (London, Prentice Hall, 1996)

FAO (2010) The second report on the state of the word's plant genetic resources for food and agriculture. http://wwwfaoorg/docrep/013/i1500e/i1500e00htm

FAO (2015) http://www.fao.org/worldfoodsituation/csdb/en/

Faure S, Higgins J, Turner A, Laurie DA (2007) The FLOWERING LOCUS T-like gene family in barley (Hordeum vulgare). Genetics 176:599-609

Fernandez ME, Figueiras AM, Benito C (2002) The use of ISSR and RAPD markers for detecting DNA polymorphism, genotype identification and genetic diversity among barley cultivars with known origin. Theor Appl Genet 104:845-851

Filippi CV, Aguirre N, Rivas JG, Zubrzycki J, Puebla A, Cordes D, Moreno MV, Fusari CM, Alvarez D, Heinz RA, Hopp HE, Paniego NB, Lia VV (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. Bmc Plant Biol 15

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374

Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054-1064

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. The Plant Journal 44:1054-1064

Franckowiack J, Konishi T (1997) Naked caryopsis. Barley Genet Newsletter 26:51–52

Franckowiak JD (2008) Tables of Barley Genetic Stock Descriptions Barley Genetics Newsletter 38:134-164

Fridborg I, Kuusk S, Moritz T, Sundberg E (1999) The Arabidopsis dwarf mutant shi exhibits reduced gibberellin responses conferred by overexpression of a new putative zinc finger protein. Plant Cell 11:1019-1031

Friedt W, Ordon F (2013) Barley Production and Breeding in Europe: Modern Cultivars Combine Disease Resistance, Malting Quality and High Yield. In: Zhang G, Li C, Liu X (eds) Advance in Barley Sciences: Proceedings of 11th International Barley Genetics Symposium. Springer Netherlands, Dordrecht, pp 389-400

Fuller DQ, Willcox G, Allaby RG (2011) Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. World Archaeology 43:628-652

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

Gao X (2011) Multiple testing corrections for imputed SNPs. Genet Epidemiol 35:154-158

Gao X, Zhang X, Lan H, Huang J, Wang J, Zhang H (2015) The additive effects of GS3 and qGL3 on rice grain length regulation revealed by genetic and transcriptome comparisons. Bmc Plant Biol 15:156

Garris AJ, Tai TH, Coburn J, Kresovich S, McCOUCH S (2005) Genetic structure and diversity in Oryza sativa L. Genetics 169:1631-1638

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502-1506

Gawenda I, Thorwarth P, Gunther T, Ordon F, Schmid KJ (2015) Genome-wide association studies in elite varieties of German winter barley using single-marker and haplotype-based methods. Plant Breeding 134:28-39

Gawroński P (2013) Earliness per se 3 locus from wheat (Triticum L. sp.) and barley (Hordeum vulgare L.) disrupts circadian clock function. Thesis submitted to the university of Halle, Germany

Gottwald S, Bauer P, Komatsuda T, Lundqvist U, Stein N (2009) TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene HvHox1. BMC Research Notes 2:1-14

Graner A, Bauer E (1993) Rflp Mapping of the Ym4 Virus-Resistance Gene in Barley. Theor Appl Genet 86:689-693

Graner A, Jahoor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, Wenzel G, Herrmann RG (1991) Construction of an Rflp Map of Barley. Theor Appl Genet 83:250-256

Graner A, Ludwig WF, Melchinger AE (1994) Relationships among European Barley Germplasm .2. Comparison of Rflp and Pedigree Data. Crop Sci 34:1199-1205

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant Physiol 131:1855-1867

Gu BG, Zhou TY, Luo JH, Liu H, Wang YC, Shangguan YY, Zhu JJ, Li Y, Sang T, Wang ZX, Han B (2015) An-2 Encodes a Cytokinin Synthesis Enzyme that Regulates Awn Length and Grain Production in Rice. Mol Plant 8:1635-1650

Habekuß A, Riedel C, Schliephake E, Ordon F (2009) Breeding for resistance to insect-transmitted viruses in barley–an emerging challenge due to global warming. Journal für Kulturpflanzen 61:53-61

Hamblin MT, Close TJ, Bhat PR, Chao SM, Kling JG, Abraham KJ, Blake T, Brooks WS, Cooper B, Griffey CA, Hayes PM, Hole DJ, Horsley RD, Obert DE, Smith KP, Ullrich SE, Muehlbauer GJ, Jannink JL (2010) Population Structure and Linkage Disequilibrium in US Barley Germplasm: Implications for Association Mapping. Crop Sci 50:556-566

Harlan JR (1978) On the origin of barley. Pages 10 - 36 in : *Barley: Origin , Botany Culture, Winter Hardiness, Genetics, Utilization, pests*. Agriculture Handbook 338. US. Department of Agriculture, Washinton, DC.

Harlan JR, Wet JMJd (1971) Toward a rational classification of cultivated plants. *Taxon* 20:509-517

Harvey BL, Reinberg.E, Somaroo BH (1968) Inheritance of Female Sterility in Barley. Can J Plant Sci 48:417-&

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

Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1-12

Hershey CH (2005). Plant Breeding News 161 an electronic newsletter of applied plant breeding, Sponsored by FAO and Cornell University; http://wwwfaoorg/WAICENT/FAOINFO/AGRICULT/AGP/AGPC/doc/services/pbnhtml 128

Hess KT, Schmidt KD, Zocher M (2006) Multivariate loss prediction in the multivariate additive model. Insur Math Econ 39:185-191

Heun M, Kennedy AE, Anderson JA, Lapitan NLV, Sorrells ME, Tanksley SD (1991) Construction of a Restriction-Fragment-Length-Polymorphism Map for Barley (Hordeum-Vulgare). Genome 34:437-447

Hintsanen P, Sevon P, Onkamo P, Eronen L, Toivonen H (2006) An empirical comparison of case-control and trio based study designs in high throughput association mapping. Journal of medical genetics 43:617-624

Hoffman GE (2013) Correcting for population structure and kinship using the linear mixed model: theory and extensions. Plos One 8:e75707

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156-161

Hotelling Pa (1901) Principal component analysis

Huehn M (1993) Harvest Index Versus Grain Straw-Ratio - Theoretical Comments and Experimental Results on the Comparison of Variation. Euphytica 68:27-32

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254-267

Imaizumi T (2010) Arabidopsis circadian clock and photoperiodism: time to think about location. Curr Opin Plant Biol 13:83-89

Imbe T, Akama Y, Nakane A, Hata T, Ise K, Ando I, Uchiyamada H, Nakagawa N, Furutachi H, Horisue N (2004) Development of a multipurpose high-yielding rice variety Takanari. Bulletin of the National Institute of Crop Science (Japan)

Ingver, I. Tamm, Ü. Tamm, Koppel TKaR (2010) The characteristics of spring cereals in changing weather in Estonia Agronomy Research 8 (Special Issue III), 553–562, 2010

Ingvordsen CH, Backes G, Lyngkjaer MF, Peltonen-Sainio P, Jahoor A, Mikkelsen TN, Jorgensen RB (2015) Genome-wide association study of production and stability traits in barley cultivated under future climate scenarios. Mol Breeding 35

Jakob SS, Rödder D, Engler JO, Shaaf S, Özkan H, Blattner FR, Kilian B (2014) Evolutionary History of Wild Barley (Hordeum vulgare subsp. spontaneum) Analyzed Using Multilocus Sequence Data and Paleodistribution Modeling. Genome Biology and Evolution 6:685-702

Ji-hua T, Wen-tao T, Jian-bing Y, Xi-qing M, Yi-jiang M, Jin-rui D, Jian-Sheng L (2007) Genetic dissection of plant height by molecular markers using a population of recombinant inbred lines in maize. Euphytica 155:117-124

Jones G, Charles MP, Jones MK, Colledge S, Leigh FJ, Lister DA, Smith LMJ, Powell W, Brown TA, Jones H (2013) DNA evidence for multiple introductions of barley into Europe following dispersed domestications in Western Asia. Antiquity 87:701-713

Jorde L (2000) Linkage disequilibrium and the search for complex disease genes. Genome research 10:1435-1444

Kalantidis, Kriton;, Briarty, Greg; L, Wilson, Zoe A. (2000) Arabidopsis mutant characterization; microscopy, mapping, and gene expression analysis. Wilson, Zoe A. Practical Approach Series; Arabidopsis. 77-103

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709-1723

Kang M, Priyadarshan PM (2008) Breeding Major Food Staples. Wiley

Karsai I, Kőszegi B, Kovács G, Szűcs P, Mészáros K, Bedő Z, Veisz O (2008) Effects of temperature and light intensity on flowering of barley (Hordeum vulgare L.). Acta Biologica Hungarica 59:205-215

Kasha KJ, Kao KN (1970) High Frequency Haploid Production in Barley (Hordeum vulgare L.). Nature 225:874-876

Kearsey MJ (1998) The principles of QTL analysis (a minimal mathematics approach). J Exp Bot 49:1619-1623

Kihara H (1983) in Proceedings of the 6th International Wheat Genetic Symposium, ed Sakamoto S(Plant Germplasm Institute, University of Kyoto, Kyoto, Japan)

Kikuchi R, Handa H (2009) Photoperiodic control of flowering in barley. Breeding Sci 59:546-552

Kikuchi R, Kawahigashi H, Ando T, Tonooka T, Handa H (2009) Molecular and Functional Characterization of PEBP Genes in Barley Reveal the Diversification of Their Roles in Flowering. Plant Physiol 149:1341-1353

Kilian B, Ozkan H, Walther A, Kohl J, Dagan T, Salamini F, Martin W (2007) Molecular diversity at 18 loci in 321 wild and 92 domesticate lines reveal no reduction of nucleotide diversity during Triticum monococcum (einkorn) domestication: Implications for the origin of agriculture. Mol Biol Evol 24:2657-2668

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat Genet 39:1151-1155

Klein RJ (2007) Power analysis for genome-wide association studies. Bmc Genet 8:58

Kretschmer J, Chalmers K, Manning S, Karakousis A, Barr A, Islam A, Logue S, Choe Y, Barker S, Lance R (1997) RFLP mapping of the Ha 2 cereal cyst nematode resistance gene in barley. Theor Appl Genet 94:1060-1064

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139-144

Kunzel G, Korzun L, Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154

Kuroha T, Tokunaga H, Kojima M, Ueda N, Ishida T, Nagawa S, Fukuda H, Sugimoto K, Sakakibara H (2009) Functional Analyses of LONELY GUY Cytokinin-Activating Enzymes Reveal the Importance of the Direct Activation Pathway in Arabidopsis. The Plant Cell 21:3152-3169

Lancashire PD, Bleiholder H, Vandenboom T, Langeluddeke P, Stauss R, Weber E, Witzenberger A (1991) A Uniform Decimal Code for Growth-Stages of Crops and Weeds. Ann Appl Biol 119:561-601

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037-2048

Langridge P, Karakousis A, Collins N, Kretschmer J, Manning S (1995) A Consensus Linkage Map of Barley. Mol Breeding 1:389-395

Laurie DA (1997) Comparative genetics of flowering time. Plant Mol Biol 35:167-177

Laurie DA, Pratchett N, Bezant JH, Snape JW (1995) Rflp Mapping of 5 Major Genes and 8 Quantitative Trait Loci Controlling Flowering Time in a Winterxspring Barley (Hordeum-Vulgare L) Cross. Genome 38:575-585

Laza M, Peng S, Huang J, Sheehy J, Visperas R, Zhong X, Centeno G (2004) Rice yields decline with higher nighttime temperature from global warming. Philippine Journal of Crop Science (Philippines)

Lewontin R, Kojima K-i (1960) The evolutionary dynamics of complex polymorphisms. Evolution:458-472

Li Q, Pan ZF, Deng GB, Long H, Li ZY, Deng XQ, Liang JJ, Tang YW, Zeng XQ, Tashi N, Yu MQ (2014) Effect of Wide Variation of the Waxy Gene on Starch Properties in Hull-less Barley from Qinghai-Tibet Plateau in China. J Agr Food Chem 62:11369-11385

Li S, Zhao B, Yuan D, Duan M, Qian Q, Tang L, Wang B, Liu X, Zhang J, Wang J, Sun J, Liu Z, Feng Y-Q, Yuan L, Li C (2013a) Rice zinc finger protein DST enhances grain production through controlling Gn1a/OsCKX2 expression. Proceedings of the National Academy of Sciences 110:3167-3172

Li S, Zhao B, Yuan D, Duan M, Qian Q, Tang L, Wang B, Liu X, Zhang J, Wang J, Sun J, Liu Z, Feng Y-Q, Yuan L, Li C (2013b) Rice zinc finger protein DST enhances grain production through controlling Gn1a/OsCKX2 expression. P Natl Acad Sci USA 110:3167-3172

Li XB, Yan WG, Agrama H, Jia LM, Jackson A, Moldenhauer K, Yeater K, McClung A, Wu DX (2012) Unraveling the Complex Trait of Harvest Index with Association Mapping in Rice (Oryza sativa L.). Plos One 7

Liu H, Bayer M, Druka A, Russell JR, Hackett CA, Poland J, Ramsay L, Hedley PE, Waugh R (2014) An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. Bmc Genomics 15

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129

Locatelli A, Cuesta-Marcos A, Gutiérrez L, Hayes PM, Smith KP, Castro AJ (2013) Genome-wide association mapping of agronomic traits in relevant barley germplasm in Uruguay. Mol Breeding 31:631-654

Luo J, Liu H, Zhou T, Gu B, Huang X, Shangguan Y, Zhu J, Li Y, Zhao Y, Wang Y, Zhao Q, Wang A, Wang Z, Sang T, Wang Z, Han B (2013) An-1 Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. The Plant Cell 25:3360-3376

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits (Sunderland, MA, Sinauer, 1998)

Mackay TFC (2001) Quantitative trait loci in Drosophila. Nat Rev Genet 2:11-20

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to Phytophthora infestans in potato. Genetics 175:879-889

Malysheva-Otto LV, Ganal MW, Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (Hordeum vulgare L.). Bmc Genet 7

Malysheva L, Sjakste T, Matzk F, Röder M, Ganal M (2003) Molecular cytogenetic analysis of wheat barley hybrids using genomic in situ hybridization and barley microsatellite markers. Genome 46:314-322

Manenti G, Galvan A, Pettinicchio A, Trincucci G, Spada E, Zolin A, Milani S, Gonzalez-Neira A, Dragani TA (2009) Mouse genome-wide association mapping needs linkage analysis to avoid false-positive loci. Plos Genet 5

Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity 108:285-291

Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Munoz-Amatriain M, Close TJ, Wise RP, Schulman AH, Himmelbach A, Mayer KFX, Scholz U, Poland JA, Stein N, Waugh R (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J 76:718-727

Massman J, Cooper B, Horsley R, Neate S, Dill-Macky R, Chao S, Dong Y, Schwarz P, Muehlbauer G, Smith K (2011) Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. Mol Breeding 27:439-454

Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome 45:1095-1106

Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. Bmc Genomics 16

Maurer A, Draba V, Pillen K (2016) Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. J Exp Bot

Mayer KFX, Waugh R, Langridge P, Close TJ, Wise RP, Graner A, Matsumoto T, Sato K, Schulman A, Muehlbauer GJ, Stein N, Ariyadasa R, Schulte D, Poursarebani N, Zhou RN, Steuernagel B, Mascher M, Scholz U, Shi BJ, Langridge P, Madishetty K, Svensson JT, Bhat P, Moscou M, Resnik J, Close TJ, Muehlbauer GJ, Hedley P, Liu H, Morris J, Waugh R, Frenkel Z, Korol A, Berges H, Graner A, Stein N, Steuernagel B, Taudien S, Groth M, Felder M, Platzer M, Brown JWS, Schulman A, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Taudien S, Sampath D, Swarbreck D, Scalabrin S, Zuccolo A, Vendramin V, Morgante M, Mayer KFX, Schulman A, Conso IBGS (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711-+

McCouch SR, McNally KL, Wang W, Hamilton RS (2012) Genomics of Gene Banks: A Case Study in Rice. Am J Bot 99:407-423

Meszaros K, Karsai I, Kuti C, Banyai J, Lang L, Bedo Z (2007) Efficiency of different marker systems for genotype fingerprinting and for genetic diversity studies in barley (Hordeum vulgare L.). S Afr J Bot 73:43-48

Mohammadi M, Torkamaneh D, Nikkhah H-R (2013) Correlation of Vernalization Loci VRN-H1 and VRN-H2 and Growth Habit in Barley Germplasm. International Journal of Plant Genomics 2013:9

Morrell PL, Clegg MT (2007a) Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. Proceedings of the National Academy of Sciences 104:3289-3294

Morrell PL, Clegg MT (2007b) Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. P Natl Acad Sci USA 104:3289-3294

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. P Natl Acad Sci USA 102:2442-2447

Muñoz-Amatriaín M, Cuesta-Marcos A, Endelman JB, Comadran J, Bonman JM, Bockelman HE (2014) The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. PLoS One 9

Munoz-Amatriain M, Cuesta-Marcos A, Endelman JB, Comadran J, Bonman JM, Bockelman HE, Chao SM, Russell J, Waugh R, Hayes PM, Muehlbauer GJ (2014) The USDA Barley Core Collection: Genetic Diversity, Population Structure, and Potential for Genome-Wide Association Studies. Plos One 9

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. The Plant Cell 21:2194-2202

Nandha PS, Singh J (2014) Comparative assessment of genetic diversity between wild and cultivated barley using gSSR and EST-SSR markers. Plant Breeding 133:28-35

Nevo E, Shewry P (1992) Origin, evolution, population genetics and resources for breeding of wild barley, Hordeum spontaneum, in the Fertile Crescent. Barley: genetics, biochemistry, molecular biology and biotechnology:19-43

Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics 154:923-929

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. Nat Genet 30:190-193

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R (2005) The pattern of polymorphism in Arabidopsis thaliana. PLoS biology 3:1289

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456:720-723

Nyquist WE (1991) Estimation of Heritability and Prediction of Selection Response in Plant-Populations. Crit Rev Plant Sci 10:235-322

Orabi J, Backes G, Wolday A, Yahyaoui A, Jahoor A (2007) The Horn of Africa as a centre of barley diversification and a potential domestication site. Theor Appl Genet 114:1117-1127

Ordonio RL, Ito Y, Hatakeyama A, Ohmae-Shinohara K, Kasuga S, Tokunaga T, Mizuno H, Kitano H, Matsuoka M, Sazuka T (2014) Gibberellin deficiency pleiotropically induces culm bending in sorghum: an insight into sorghum semi-dwarf breeding. Sci Rep-Uk 4

Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. P Natl Acad Sci USA 101:9885-9890

Papa R, Attene G, Barcaccia G, Ohgata A, Konishi T (1998) Genetic diversity in landrace populations of Hordeum vulgare L. from Sardinia, Italy, as revealed by RAPDs, isozymes and morphophenological traits. Plant Breeding 117:523-530

Park J-H, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42:570-575

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. Bmc Plant Biol 12

Pasam RK, Sharma R, Walther A, Ozkan H, Graner A, Kilian B (2014) Genetic Diversity and Population Structure in a Legacy Collection of Spring Barley Landraces Adapted to a Wide Range of Climates. Plos One 9

Payne RW (2014) Developments from analysis of variance through to generalized linear models and beyond. Ann Appl Biol 164:11-17

Payne RW, Murray, D.A., Harding, S.A., Baird, D.B. & Soutar, D.M. (2013) GenStat for Windows 16th Edition. VSN International (2013).

Percival J (1921) *The Wheat Plant*. Duckworth, London

Petersen L, Ostergard H, Giese H (1994) Genetic Diversity among Wild and Cultivated Barley as Revealed by Rflp. Theor Appl Genet 89:676-681

Pfeiffer RM, Gail MH (2003) Sample size calculations for population- and family-based case-control association studies on marker genotypes. Genet Epidemiol 25:136-148

Pickering R (2000) Do the wild relatives of cultivated barley have a place in barley improvement? In: Barley Genetics VIII. Proc. 8"Int. Barley Genet., Symp., Adelaide, Australia, pp. 223-230

Pickering RA (1984) The Influence of Genotype and Environment on Chromosome Elimination in Crosses between Hordeum-Vulgare-L X Hordeum-Bulbosum L. Plant Sci Lett 34:153-164

Piffanelli P, Ramsay L, Waugh R, Benabdelmouna A, D'Hont A, Hollricher K, Jørgensen JH, Schulze-Lefert P, Panstruga R (2004) A barley cultivation-associated polymorphism conveys resistance to powdery mildew. Nature 430:887-891

Pillen K, Zacharias A, Leon J (2003) Advanced backcross QTL analysis in barley (Hordeum vulgare L.). Theor Appl Genet 107:340-352

Pittendrigh CS (1993) Temporal Organization - Reflections of a Darwinian Clock-Watcher. Annu Rev Physiol 55:16-54

Poets AM, Fang Z, Clegg MT, Morrell PL (2015a) Barley landraces are characterized by geographically heterogeneous genomic origins. Genome Biol 16:1-11

Poets AM, Fang Z, Clegg MT, Morrell PL (2015b) Barley landraces are characterized by geographically heterogeneous genomic origins. Genome Biol 16:1

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. Plos One 7

Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, Azhaguvel P, Sakuma S, Dhanagond S, Sharma R, Mascher M, Himmelbach A, Gottwald S, Nair SK, Tagiri A, Yukuhiro F, Nagamura Y, Kanamori H, Matsumoto T, Willcox G, Middleton CP, Wicker T, Walther A, Waugh R, Fincher GB, Stein N, Kumlehn J, Sato K, Komatsuda T (2015) Evolution of the Grain Dispersal System in Barley. Cell 162:527-539

Powell W, Caligari PDS, Thomas WTB (1986) Comparison of Spring Barley Lines Produced by Single Seed Descent, Pedigree Inbreeding and Doubled Haploidy. Plant Breeding 97:138-146

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904-909

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Pritchard JK, Stephens M, Donnelly P (2000b) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Qi P, Lin Y-S, Song X-J, Shen J-B, Huang W, Shan J-X, Zhu M-Z, Jiang L, Gao J-P, Lin H-X (2012) The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. Cell Research 22:1666-1680

Qi WW, Sun F, Wang QJ, Chen ML, Huang YQ, Feng YQ, Luo XJ, Yang JS (2011) Rice Ethylene-Response AP2/ERF Factor OsEATB Restricts Internode Elongation by Down-Regulating a Gibberellin Biosynthetic Gene. Plant Physiol 157:216-228

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL
http://www.R-project.org/

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94-100

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174-180

Reid DAW, G. A.; USA, USDA (1968) Taxonomy, botany, classification, and world collection. Barley: Origin, Botany, Culture, Winter Hardiness, Genetics, Utilization & Pests ;Taxonomy and Evolution:1998 pp. 1961-1984

Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, Srinivasan G, Bohn M, Frisch M (2003) Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci 43:1275-1282

Remington DL, Purugganan MD (2003) Candidate genes, quantitative trait loci, and functional trait evolution in plants. Int J Plant Sci 164:S7-S20

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences 98:11479-11484

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biological chemistry 379:633-646

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516-1517

Rodriguez M, Rau D, O'Sullivan D, Brown AHD, Papa R, Attene G (2012) Genetic structure and linkage disequilibrium in landrace populations of barley in Sardinia. Theor Appl Genet 125:171-184

Rohlfs RV, Weir BS (2008) Distributions of Hardy-Weinberg Equilibrium Test Statistics. Genetics 180:1609-1616

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW (2005) Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet 1:e70

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Grainer A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. P Natl Acad Sci USA 103:18656-18661

Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, Powell W (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389-398

Russell J, Dawson IK, Flavell AJ, Steffenson B, Weltzien E, Booth A, Ceccarelli S, Grando S, Waugh R (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

Russell J, Mascher M, Dawson IK, Kyriakidis S, Calixto C, Freund F, Bayer M, Milne I, Marshall-Griffiths T, Heinen S, Hofstad A, Sharma R, Himmelbach A, Knauft M, van Zonneveld M,

Brown JWS, Schmid K, Kilian B, Muehlbauer GJ, Stein N, Waugh R (2016) Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. Nat Genet 48:1024-1030

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714-722

Saisho D, Purugganan MD (2007) Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World. Genetics 177:1765-1776

Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K, Miyao A, Hirochika H, Kitano H, Ashikari M, Matsuoka M (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiol 134:1642-1653

Sax K (1923) The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS. Genetics 8:552-560

Schaeffer L (2006) Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123:218-223

SCHOLZ F (1963) Induced smooth-awned mutants and the inheritance of stigma hairiness in barley. Barley Ncu'slcttcr 6,43-44 (Quoted with the author's permission)

Shaw LM, Turner AS, Herry L, Griffiths S, Laurie DA (2013) Mutant Alleles of Photoperiod-1 in Wheat (Triticum aestivum L.) That Confer a Late Flowering Phenotype in Long Days. Plos One 8

Shrotria P, Singh R (1988) Harvest index-A useful selection criteria in sorghum. Sorghum-Newsletter Utter Pardesh India 31

Si LZ, Chen JY, Huang XH, Gong H, Luo JH, Hou QQ, Zhou TY, Lu TT, Zhu JJ, Shangguan YY, Chen EW, Gong CX, Zhao Q, Jing YF, Zhao Y, Li Y, Cui LL, Fan DL, Lu YQ, Weng QJ, Wang YC, Zhan QL, Liu KY, Wei XH, An K, An G, Han B (2016) OsSPL13 controls grain size in cultivated rice. Nat Genet 48:447-+

Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Singh NK, Singh R (2013) Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. Plos One 8

Skot L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J, Sanderson R, Armstead IP, Thomas ID (2007) Association of candidate genes with flowering time and water-soluble carbohydrate content in Lolium perenne (L.). Genetics 177:535-547

Spencer CC, Su Z, Donnelly P, Marchini J (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. Plos Genet 5:e1000477

Spielman RS, Mcginnis RE, Ewens WJ (1993) Transmission Test for Linkage Disequilibrium - the Insulin Gene Region and Insulin-Dependent Diabetes-Mellitus (Iddm). Am J Hum Genet 52:506-516

Sprenger N, Bortlik K, Brandt A, Boller T, Wiemken A (1995) Purification, cloning, and functional expression of sucrose: fructan 6-fructosyltransferase, a key enzyme of fructan synthesis in barley. Proceedings of the National Academy of Sciences 92:11652-11656

Steffenson BJ, Olivera P, Roy JK, Jin Y, Smith KP, Muehlbauer GJ (2007) A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. Crop Pasture Sci 58

Steffenson BJ, Solanki S, Brueggeman RS (2016) Landraces from mountainous regions of Switzerland are sources of important genes for stem rust resistance in barley. Alpine Bot 126:23-33

Stich B, Melchinger AE (2009) Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. Bmc Genomics 10

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. P Natl Acad Sci USA 100:9440-9445

Sutton T, Baumann U, Hayes J, Collins NC, Shi B-J, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. Science 318:1446-1449

Takahashi R, Yasuda S (1956) GENETIC STUDIES OF SPRING AND WINTER HABIT OF GROWTH IN BARLEY*. Berichte des Ohara Instituts für landwirtschaftliche Biologie, Okayama Universität 10:245-308

Taketa S, Amano S, Tsujino Y, Sato T, Saisho D, Kakeda K, Nomura M, Suzuki T, Matsumoto T, Sato K, Kanamori H, Kawasaki S, Takeda K (2008) Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. P Natl Acad Sci USA 105:4062-4067

Taketa S, Kikuchi S, Awayama T, Yamamoto S, Ichii M, Kawasaki S (2004) Monophyletic origin of naked barley inferred from molecular analyses of a marker closely linked to the naked caryopsis gene (nud). Theor Appl Genet 108:1236-1242

Taketa S, Yuo T, Sakurai Y, Miyake S, Ichii M (2011) Molecular mapping of the short awn 2 (lks2) and dense spike 1 (dsp1) genes on barley chromosome 7H. Breeding Sci 61:80-85

Tannahill R (1988) *Food in History*, rev. ed. Penguin, London

Tanno K, Willcox G (2006) How fast was wild wheat domesticated? Science 311:1886-1886

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proceedings of the National Academy of Sciences 98:9161-9166

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286-289

Tondelli A, Xu X, Moragues M, Sharma R, Schnaithmann F, Ingvardsen C, Manninen O, Comadran J, Russell J, Waugh R (2013) Structural and temporal variation in genetic diversity of European spring two-row barley cultivars and association mapping of quantitative traits. The Plant Genome 6

Toojinda T, Broers L, Chen X, Hayes P, Kleinhofs A, Korte J, Kudrna D, Leung H, Line R, Powell W (2000) Mapping quantitative and qualitative disease resistance genes in a doubled haploid population of barley (Hordeum vulgare). Theor Appl Genet 101:580-589

Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science 310:1031-1034

Tzeng J-Y, Byerley W, Devlin B, Roeder K, Wasserman L (2003) Outlier detection and false discovery rates for whole-genome DNA matching. J Am Stat Assoc 98:236-246

Ullrich SE (2010) Barley: Production, Improvement, and Uses. Wiley

Van Hung P (2016) Phenolic compounds of cereals and their antioxidant capacity. Crit Rev Food Sci 56:25-35

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289-1299

Varshney RK, Marcel TC, Ramsay L, Russell J, Roder MS, Stein N, Waugh R, Langridge P, Niks RE, Graner A (2007) A high density barley microsatellite consensus map with 775 SSR loci. Theor Appl Genet 114:1091-1103

Villa TCC, Maxted N, Scholten M, Ford-Lloyd B (2005) Defining and identifying crop landraces. Plant Genetic Resources 3:373-384

von Bothmer R, van Hintum T, Knüpffer H, Sato K (2003) Diversity in Barley (Hordeum vulgare). Elsevier Science

von Bothmer R., Jacobsen N BC, Jørgensen RB., Baden C., I. L-L (1995) An Ecogeographical Study of the Genus Hordeum, . 2nd edn Systematic and Ecogeographical Studies on Crop Genepools 7:IBPGR, Rome

Wang J-m, Yang J-m, Zhu J-h, Jia Q-j, Tao Y-z (2010) Assessment of genetic diversity by simple sequence repeat markers among forty elite varieties in the germplasm for malting barley breeding. Journal of Zhejiang University Science B 11:792-800

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

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

Wehner GG, Balko CC, Enders MM, Humbeck KK, Ordon FF (2015) Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. Bmc Plant Biol 15

Weigel D (2012) Natural Variation in Arabidopsis: From Molecular Genetics to Ecological Genomics. Plant Physiol 158:2-22

Weiskorn D-HC, Ordon PDF (2003) Strategies of breeding for durable disease resistance in cereals. Progress in Botany. Springer, pp 138-167

Weiss E, Kislev ME, Hartmann A (2006) Autonomous cultivation before domestication. Science 312:1608-1610

Wenzl P, Li HB, Carling J, Zhou MX, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesna J, Cakir M, Poulsen D, Wang JP, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. Bmc Genomics 7

Werner K, Friedt W, Ordon F (2005) Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). Mol Breeding 16:45-55

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. Proceedings of the National Academy of Sciences 99:12959-12962

Wicker T, Narechania A, Sabot F, Stein J, Vu GT, Graner A, Ware D, Stein N (2008) Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. Bmc Genomics 9

Willcox G, Fornite S, Herveux L (2008) Early Holocene cultivation before domestication in northern Syria. Veg Hist Archaeobot 17:313-325

Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. The Plant Cell 16:2719-2733

Wolbang CM, Davies NW, Taylor SA, Ross JJ (2007) Gravistimulation leads to asymmetry of both auxin and gibberellin levels in barley pulvini. Physiologia plantarum 131:140-148

WOODWARD RW (1949) The inheritance of fertility in the lateral florets of the four barley group. Agron. J. 41, 317 - 322

Wu Y, San Vicente F, Huang K, Dhliwayo T, Costich DE, Semagn K, Sudha N, Olsen M, Prasanna BM, Zhang X (2016) Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theor Appl Genet 129:753-765

Wu Z, Zhao H (2009) Statistical power of model selection strategies for genome-wide association studies. Plos Genet 5:e1000582

Yamamoto T, Kuboki Y, Lin SY, Sasaki T, Yano M (1998) Fine mapping of quantitative trait loci Hd-1, Hd-2 and Hd-3, controlling heading date of rice, as single Mendelian factors. Theor Appl Genet 97:37-44

Yamamoto T, Lin HX, Sasaki T, Yano M (2000) Identification of heading date quantitative trait locus Hd6 and characterization of its epistatic interactions with Hd2 in rice using advanced backcross progeny. Genetics 154:885-891

Yano M, Harushima Y, Nagamura Y, Kurata N, Minobe Y, Sasaki T (1997) Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. Theor Appl Genet 95:1025-1032

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203-208

Yu JM, Zhang ZW, Zhu CS, Tabanao DA, Pressoir G, Tuinstra MR, Kresovich S, Todhunter RJ, Buckler ES (2009) Simulation Appraisal of the Adequacy of Number of Background Markers for Relationship Estimation in Association Mapping. Plant Genome-Us 2:63-77

Yu SY, Long H, Deng GB, Pan ZF, Liang JJ, Zeng XQ, Tang YW, Tashi N, Yu MQ (2016) A Single Nucleotide Polymorphism of Nud Converts the Caryopsis Type of Barley (Hordeum Vulgare L.). Plant Molecular Biology Reporter 34:242-248

Yuo T, Yamashita Y, Kanamori H, Matsumoto T, Lundqvist U, Sato K, Ichii M, Jobling SA, Taketa S (2012a) A SHORT INTERNODES (SHI) family transcription factor gene regulates awn elongation and pistil morphology in barley. J Exp Bot 63:5223-5232

Yuo T, Yamashita Y, Kanamori H, Matsumoto T, Lundqvist U, Sato K, Ichii M, Jobling SA, Taketa S (2012b) A SHORT INTERNODES (SHI) family transcription factor gene regulates awn elongation and pistil morphology in barley. J Exp Bot

Žáková M, Benková M (2006) Characterization of spring barley accessions based on multivariate analysis. Communications in Biometry and Crop Science 1:124-134

Zhang D, Ding Y (2007) [Genetic diversity of wild close relatives of barley in Tibet of China revealed by AFLP]. Yi Chuan 29:725-730

Zhang LY, Marchand S, Tinker NA, Belzile F (2009) Population structure and linkage disequilibrium in barley assessed by DArT markers. Theor Appl Genet 119:43-52

Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, Wu Y, Tang H, Qian Q, Li J, Zhang H (2012) Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. Proc Natl Acad Sci U S A 109:21534-21539

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P (2007) An Arabidopsis example of association mapping in structured samples. Plos Genet 3:e4

Zhou H, Steffenson BJ (2013) Association mapping of septoria speckled leaf blotch resistance in US barley breeding germplasm. Phytopathology 103:600-609

Zhou H, Steffenson BJ, Muehlbauer G, Wanyera R, Njau P, Ndeda S (2014) Association mapping of stem rust race TTKSK resistance in US barley breeding germplasm. Theor Appl Genet 127:1293-1304

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. The plant genome 1:5-20

Zohary D (1988) Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe, and the Nile Valley 3rd Edition

Acknowledgements

I would like to extend my sincere appreciation to Prof. Dr. Andreas Graner (director of IPk and head of the Gene Bank department) for accepting me as a PhD student in his group and also for his immense scientific contribution towards the success of my research work.

Special thanks to Dr. Benjamin Kilian and Dr. Kerstin Neumann not forgetting Dr. Rajiv Sharma for their support, encouragements and guindiance that shaped my scientific career. Thanks to Dr. Raj Pasam for developing the diverse landrace core collection (LRC 648) used in this research. I am also very thankful to the "Experimental Fields and Nurseries" groups at IPK, members of the Genomics of Genetic Resources, all members of the Genome diversity group for their support during the field trials.

I deeply appreciate Dr. Britt Leps for her assistanstance regarding office appointments and Miss Lynn Main for taking the images of barley seeds used in current research.

It is my pleasure to convey a debt of gratitude to Prof. Dr. Nicolaus von Wiren, chair of the Leibniz Graduate-School, Dr. Thorsten Schnurbusch for accepting to be my mentor and Prof. Dr. Volker Mohler for accepting to review my thesis.

I will also like to aknowledge my family members back home espcially my uncle "officier de police Tambah Godwin Ndungbowo" for his strong words of encouragement during the start of my PhD not forgetting other family members and friends residing abroad.

I dedicate this thesis to my grand father "Ba Tambah Sampson Fongoh" who raised me and has always been a positive figure in my life.

I am grateful for the financial support from the Leibniz Graduate-School "Yield Formation in cereals-overcoming yield-limiting factors" (FKZ: SAW-2012-IPK-4).

Celestine Wabila
Finkenweg 2, OT Gatersleben,
06466 seeland
✆: 015202662504
✉: wabila@ipk-gatersleben.de

Curriculum vitae

Personal Data

Geburtsdatum:	08.08.1981
Geburtsort:	Bamenda, Kamerun
Nationalität:	Kamerunisch, unbefristete Arbeitserlaubnis
Einreise in die BRD:	2005

Academic background

03/2017 - date	**Researcher at IPK Gatersleben**
Research topic	High-throughput phenotyping and GWAS in a diverse collection of spring Wheat (WHEALBI project)
09/2012 – 03/2017	**PhD student at the institute of plant genetics and crop plant research (IPK-Gatersleben)**
Research topic	Genome-wide association study to genetically dissect yield related traits in a diverse collection of spring barley landraces
10/2009 – 08/2012	**Masters of Science in Agrobiotechnology**
Name of institution	University of Giessen Institute of plant breeding
Master thesis	Expression analysis of candidate genes responsible for resistance in wheat against *F. graminearum* mycotoxin " DON"
09/2005 - 08/2009	**Bachelor of science in applied biology**
Name of institution	Fachochschule Bonn-Rhein-Sieg
Bachelor thesis	Determination of anti-oxidative properties of Biophenols using the stable free radical 2,2 diphenyl-1-picrylhydrazyl "DPPH"
01/2003 - 07/2003	**Zertifikat Deutsch Als Fremdsprache**
Name of institution	Goethe Institute Yaunde
09/1999 – 06/2001	**General Certificate of Education, A Levles (Arbitur)**
Name of Institution	Mervick Billingual Grammar School
09/1992 – 06/1999	**General Certificate of Education, O Levels**
Name of institution	Government high school Bali
09/1986 – 06/1992	**First school leaving certificate**
Name of institution	Government primary school Gungong

Publication:

Adel H. Abdel-Ghani, Kerstin Neumann, **Celestine Wabila**, Rajiv Sharma, Sidram Dhanagond, Saed J. Owais, Andreas Börner, Andreas Graner, Benjamin Kilian (2014) .

Diversity of germination and seedling traits in a spring barley *(Hordeum vulgare L.)* collection under drought simulated conditions. Gent. Resources and crop evolution 62:275-292.

Celestine Wabila[1], Kerstin Neumann[1], Benjamin Kilian[1,2], Andreas Graner[1] .

A tiered approach to genome-wide association analysis reveals novel QTL controlling the adherence of hulls to the caryopsis of barley seeds (**in preparation**)

Celestine Wabila[1]**,** Kerstin Neumann[1], Rajiv Sharma[1,2], Benjamin Kilian[1,3], Andreas Graner[1]

Genome-wide association study reveals novel QTL for yield and yield related traits in a diverse collection of 261 spring barley landraces (**in preparation**)

ERKLÄRUNG

Herr Celestine Wabila

Leibniz Institute of Plant Genetics
and Crop Plant Research (IPK)
Genebank/ Genome Diversity
Corrensstrasse 3, OT Gatersleben
06466 Seeland

Eidesstattliche Erklärung

Hiermit erkläre ich, dass mit dieser wissenschaftlichen Arbeit noch keine vergeblichen Promotionsversuche unternommen wurden. Die eingereichte Dissertation mit dem Thema: "*Genome-wide association study to genetically dissect yield related traits in a diverse collection of spring barley landraces*" habe ich selbständig und nur unter Verwendung der angegeben Literatur und Hilfsmittel angefertigt. Des Weiteren erkläre ich, dass keine Strafverfahren gegen mich anhängig sind.
Gatersleben, den ……
Celestine Wabila

Declaration concerning Criminal Record and Pending Investigations
I hereby declare that I have no criminal record and that no preliminary investigations are pending against me.

04.10.2017 — Celestine Wabila
Date — *Signature of the applicant*

Declaration under Oath

I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

04.10.2017 — Celestine Wabila
Date — *Signature of the applicant*