Aus dem Institut für Agrar- und Ernährungswissenschaften

der Naturwissenschaftlichen Fakultät III

der

Martin-Luther-Universität Halle-Wittenberg

# GENOME-WIDE ASSOCIATION STUDIES IN DIVERSE GENE POOLS OF BARLEY

Dissertation

Zur Erlangung des akademischen Grades doctor agriculturarum (Dr. agr.)

von

Rajiv Sharma M.Sc.

geb. am 14.04.1982 in Dinanagar, India

**Gutachter:**

Prof. Dr. Andreas Graner (Gatersleben)

Prof. Dr. Klaus Pillen (Halle)

Prof. Søren K. Rasmussen (Copenhagen)

**Verteidigt am:** 07.01.2013

**Halle/Saale 2012**

**Table of Contents**

**Aerial view showing field plots (IPK year 2009)**

**Picture showing field plots (IPK year 2010)**

**List of Abbreviations**

| | |
|---|---|
| % | Percentage |
| 78P | Powdery mildew isolate name |
| 9k | Nine thousand |
| BCC | Barley core collection |
| Bgh | *Blumeria graminis* f sp. *hordei* |
| BLUE | Best Linear Unbiased Estimator |
| BOPA1 | Barley Oligo Pool Assay 1 |
| cm | Centimeter |
| cM | Centi Morgan (Genetic distances) |
| D12/12 | Powdery mildew isolate name |
| DArT | Diversity Array Technology |
| EST | Expressed Sequence Tagged |
| FDR | False Discovery Rate |
| Fr | Frost Tolerance |
| Garea | Grain area |
| Gbp | Giga base pair |
| GEar | Grains per ear |
| Genobar collection or HVCC_Genobar | 224 world-wide spring barley collection |
| Glength | Grain length |
| GLM | General Linear Model |
| GWAS | Genome-Wide Association Studies |
| Gwidth | Grain width |
| Hd | Heading date |
| HSC | *Hordeum spontaneum* collection |
| Ht | Height |
| HVCC | *Hordeum vulgare* cultivar collection (spring + winter barley) |
| HVCC_S | *Hordeum vulgare* cultivar collection spring barley |
| HVCC_W | *Hordeum vulgare* cultivar collection winter barley |
| IBSC | International Barley Sequencing Consortium |
| iSelect | ILLUMINA SNP platform with 7864 SNPs from barley |
| Kb | Kilo base pairs |
| LD | Linkage Disequilibrium |
| LRC | Landraces collection |
| LRC_J | Landraces from Jordan |
| LRC_S | Landraces from Syria |
| MAF | Minor Allele Frequency |
| MAS | Marker Assisted Selection |
| Mha | Million hectares |
| MLM | Mixed Linear Model |
| MxB | Morex x Barke RILs |
| NIRS | Near Infrared Reflectance Spectrometer |
| NJ | Neighbor Joining |
| PC | Protein content |
| PCA | Principal Component Analysis |
| PCR | Polymerase chain reaction |
| PIC | Polymorphic Information Content |
| PM | Powdery Mildew |
| PMdla | Powdery mildew resistance under detached leaf assay condition |
| PMfld | Powdery mildew resistance under field condition |

| | |
|---|---|
| QTL | Quantitative trait loci |
| REML | Restriction Estimate of Maximum Likelihood |
| RFLP | Restriction fragment length polymorphism |
| RIL | Recombinant Inbred lines |
| SC | Starch content |
| Sl | Spike length |
| SSR | Simple sequence repeats |
| TGW | Thousand grain weight |
| W_l | ratio of grain width by length |
| Yld | Grain yield |

## List of Tables



## List of Figures

## List of Supplementary Figures

**List of Supplementary Tables (Please see attached CD)**

| |
|---|
| **Table S1**: Genotypes, pedigree and origin of the germplasm sampled and genotyped. Serial number (SNO); Gene pool name (HVCC_W and HVCC_S gene pools); Acc no. (accession number); Entry name (name of the variety); Row type (two or six rowed type); year (year of the release); origin (country of origin); pedigree (pedigree information of the variety); Breeding institute (Institute or company who breeded the variety); marker type available (BOPA1 or iSelect) are shown in the table |
| **Table S2**: Genotypic descriptions of LRC gene pool. Genotype name, code of the region (based on the region from Russell et al. 2011), origin, longitude, latitude (East and North) and the region of collection are shown |
| **Table S3**: Genotypic descriptions of wild barley (HSC) gene pool. Genotype name, site code (based on the region from Hübner et al. (2009)), location, X and Y UTM coordinates from the respective locations |
| **Table S4**: Genotypic descriptions of Genobar collection (HVCC_Genobar). Genotype name, row type, region of origin [Europe (EU), West Asia and North Africa (WANA), 40 from East Asia (EA) and 30 from the Americas (AM)] and country of origin are shown |
| **Table S5**: GWAS results for the frost tolerance (Fr) in winter (HVCC_W) gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2). Marker name (SNP), chromosome (Chrom.), map position in centimorgan (cM), minor allele frequency (MAF), marker effect (Effect) and minus log P significance (-log10 (P)) of the marker are shown |
| **Table S6**: GWAS results for the heading date (Hd) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S7:** GWAS results for the height (Ht) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S8:** GWAS results for the spike length (Sl) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S9:** GWAS results for the row type (Rtype) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S10:** GWAS results for the grains per ear (GEar) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S11:** GWAS results for the thousand grain weight (TGW) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S12:** GWAS results for the grain area (Garea) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S13:** GWAS results for the grain length (Glength) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |
| **Table S14:** GWAS results for the grain width (Gwidth) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) $\geq$ 2) |

| |
|---|
| **Table S15:** GWAS results for the grain width / length (W_l) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) ≥ 2) |
| **Table S16:** GWAS results for the grain yield (Yld) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using BOPA1 SNP markers (-log10 (P) ≥ 2) |
| **Table S17:** GWAS results for the starch content (SC) in winter (HVCC_W), spring (HVCC_S) and landraces (LRC) barley gene pool using BOPA1 SNP markers (-log10 (P) ≥ 2) |
| **Table S18:** GWAS results for the protein content (PC) in winter (HVCC_W), spring (HVCC_S) and landraces (LRC) barley gene pool using BOPA1 SNP markers (-log10 (P) ≥ 2) |
| **Table S19:** GWAS results for the powdery mildew field infection (PMfld) in winter (HVCC_W), spring (HVCC_S) and powdery mildew detached leaf assay in winter (HVCC_W), spring (HVCC_S), landraces (LRC) and wild barley (HSC) barley gene pools using BOPA1 SNP markers (-log10 (P) ≥ 2) |
| **Table S20**: GWAS results for the frost tolerance (Fr) in winter (HVCC_W) gene pool using iSelect SNP markers (-log10 (P) ≥ 2). Marker name (SNP), chromosome (Chrom.), map position in centimorgan (cM), minor allele frequency (MAF), marker effect (Effect), minus log P significance (-log10 (P)) and mapping information (Map) [1 means recombinant inbred Morex x Barke genetic position and 2 means Close et al. (2009) and LD mapping information (Comadran et al. (unpublished))] are shown |
| **Table S21:** GWAS results for the heading date (Hd) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S22:** GWAS results for the height (Ht) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S23:** GWAS results for the spike length (Sl) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S24:** GWAS results for the row type (Rtype) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S25:** GWAS results for the grains per ear (GEar) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S26:** GWAS results for the thousand grain weight (TGW) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S27:** GWAS results for the grain area (Garea) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S28:** GWAS results for the grain length (Glength) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S29:** GWAS results for the grain width (Gwidth) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |

| |
|---|
| **Table S30:** GWAS results for the grain width / length (W_l) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S31:** GWAS results for the grain yield (Yld) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S32:** GWAS results for the starch content (SC) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S33:** GWAS results for the protein content (PC) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S34:** GWAS results for the powdery mildew field infection (PMfld) in winter (HVCC_W) and spring (HVCC_S) barley gene pool using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S35:** GWAS results for the powdery mildew detached leaf assay (PMdla) in winter (HVCC_W), spring (HVCC_S) and spring (HVCC_Genobar) barley gene pools using iSelect SNP markers (-log10 (P) ≥ 2) |
| **Table S36:** GWAS scan results for all traits in HVCC_W gene pools using BOPA1 SNP markers (-log10 (P) ≥ 2). Marker name (SNP), chromosome (Chrom.), map position in centimorgan (cM), minor allele frequency (MAF) and minus log P significance (-log10 (P)) of the marker are shown |
| **Table S37:** GWAS scan results for all traits in HVCC_S gene pools using BOPA1 SNP markers. Only significant associations (-log10 (P) ≥ 2) are shown |
| **Table S38:** GWAS scan results for all traits in HVCC_W gene pools using iSelect SNP markers. Only significant associations (-log10 (P) ≥ 2) are shown |
| **Table S39:** GWAS scan results for all traits in HVCC_S gene pools using iSelect SNP markers. Only significant associations (-log10 (P) ≥ 2) are shown |
| **Table S40:** Genotype names that contains *mlo* locus |

# 1. Introduction

## 1.1 General introduction

*Hordeum vulgare* L. (cultivated barley) (Linde-Laursen et al. 1997; von Bothmer et al. 1995) is one of the major crop and a model species in the tribe Triticeae. Barley is the fourth major cereal crop after maize (*Zea mays*), rice (*Oryza sativa*) and wheat (*Triticum aestivum*) in terms of the total food production (http://faostat.fao.org/site/291/default.aspx (June 20 2008)). Barley is a self pollinating diploid species with large genome size (> 5 Gbp). It's genome consists of highly repetitive elements and is almost twelve times the size of rice genome (Bennett and Smith 1976; Doležel et al. 1998; Wicker et al. 2008). Cultivated barley is diploid with 2n=14 chromosomes. Due to its diploid state and shared genome collinearity with other Triticeae species, barley became a model plant also for the hexaploid wheat. Barley is mainly used for feeding, malting, distilling and to less extent also as calorie food source for human consumption mainly in marginal areas with problematic soils and scant rainfall (Grando and Macpherson 2005). Barley has good adaptability to extreme and marginal conditions that has led to its widespread cultivation and distribution throughout the world (von Bothmer et al. 1995). It was a main staple crop since its domestication and was used as food throughout the old world. Later, wheat replaced major areas of barley cultivation. In the oldest Roman literature, barley was described as an award given to the champions of the games (Andrew 2008). The gladiators were called barley men or "hordearii" due to its rich dietary value. Barley with its good nutritional values is still used as a staple food in mountainous areas of Central Asia, in South-West Asia and Northern Africa including Ethiopia. The history of human civilization and barley domestication is closely interwoven. The earliest archeological remains of wild barley were found in the Fertile Crescent area, where the first signs of storing and use of wild cereals was reported (Diamond 1997; Kislev et al. 1992; Zohary and Hopf 2000). Due to its vast morphological and environmental adaptability, many types of barley are grown throughout the world (von Bothmer et al. 2003a) and these are viz., winter, spring, two-rowed, six-rowed, awned, awnless, hooded, covered, naked, hull less, malting, feed and food types.

Barley has been used for a variety of purposes and therefore has vast economic importance. Around 55-60% of barley is used for feed, 30-40% for malt, 2-3% for food and 5% for seed (Ullrich 2010). It is the major dietary source for ruminant and non ruminant livestock, poultry, and fish. Mostly feed barley varieties have higher yield than malting barley varieties. Yield

advantage of 10-20% is reported between the highest yielding feed and malting barley varieties (Blake et al. 2010). At Global level, an estimated 20 Mt of barley grain have been exported and imported annually in this century that accounts for about US$3 billion per year (Ullrich 2010). Significant (48%) increase in the value of malt export and imports was observed worldwide, from an average of US$1.35 billion in the year 2000 to about US$2.0 billion in 2005 (Ullrich 2010). Global trend shows that the area under barley production is declining from 80 million hectares in the 1970s to less than 60 million hectares in 2008 (Friedt et al. 2010). High yielding modern cultivars have ensured the continuous yield increase in almost all barley growing areas. However, abiotic and biotic stresses incur major losses to barley production and are the major focus areas in breeding programs in order to achieve further yield stability.

### 1.2 Barley classification and domestication

Barley belongs to the Poaceae family and Triticeae tribe along with wheat and rye. Barley is considered to be of monophyletic origin within the Pooideae subfamily of Poaceae (Blattner 2004; Devos 2005, 2010; Jakob and Blattner 2006). Present cultivated barley (*Hordeum vulgare* L.) was domesticated along with other cereals ca. 10,000 years ago from its wild relative (*Hordeum spontaneum* (C. Koch) Thell or *H. vulgare* ssp. spontaneum) (Badr et al. 2000; Bennett and Smith 1976; Kilian et al. 2009; Salamini et al. 2002).

The genus *Hordeum* comprises of 32 species and 45 taxa. All *Hordeum* species have similar three single flowered spikelets at each rachis node called triplet (von Bothmer et al. 1995). Although *Hordeum* species share the same basic morphology, high levels of genetic diversity are found among the species (von Bothmer et al. 2003b). Some *Hordeum* species are annual (*H. marinum* Huds., *H. murinum* L. and *H. pusillum* Nutt) and some are perennials (*H. bulbosum* L. and *H. brevisubulatum* (Trin.) Link), most of them are inbreeding but some species show self-incompatibility (eg. *H. bulbosum* L. and *H. brevisubulatum* (Trin.) Link) (von Bothmer et al. 2003b).

According to the gene pool concept of Harlan and de Wet (Harlan and de Wet 1971), *Hordeum* species have been grouped into three gene pools. The primary gene pool includes cultivated barley, and landraces along with their progenitor *H. vulgare* ssp. *spontaneum*. The secondary gene pool comprises of one species, *H. bulbosum*, and the rest of the species are in the tertiary gene pool that are mostly polyploids with tetraploid and hexaploid genome types (Harlan 1971; von Bothmer et al. 1995; von Bothmer et al. 2003b). Autoploidy is found in

two species, *H. bulbosum* and *H. brevisubulatum*. Most of the other polyploids are allopolyploids and majorly segmental allopolyploids (von Bothmer et al. 2003b).

In the primary gene pool, cultivated and wild barley are sexually compatible and occasionally some feral hybrid types were found in the areas where they are growing alongside each other (Harlan 1971). The secondary gene pool, *Hordeum bulbosum* harbors many agronomically interesting genes but has been introgressed in the cultivated gene pool only in few instances e.g. for disease resistance and insect resistance. *Hordeum bulbosum* technique for doubled haploid production is widely used and exploited for chromosomal engineering, but is being replaced increasingly by anther and microspore techniques (Kasha and Kao 1970; Pickering and Johnston 2005; Sanei et al. 2011; Zeller 1998).

Barley is one of the first domesticated crops in the Fertile Crescent (Badr et al. 2000; Kilian et al. 2009; Zohary and Hopf 2000). It was an important primary cereal in ancient Egypt and in Mesopotamia (Harlan 1995). In several pre-agricultural sites, wild barley grains have been found. At the shore of Sea of Galilee 21,000 years old wild barley remains were found (Kislev et al. 1992). Wild remains found at these sites were older than the domestication of barley and supports the theory that wild barley was collected by early humans even before the domestication. Earliest charred carbonized remains of domesticated barley are of two-rowed type (van Zeist 1970; Hillman et al. 1989), but six-rowed types appeared at Ain Ghazal at around 9000-8500 years ago (Rollefson et al. 1985; Willcox 1997). From the eastern Fertile Crescent sites of Ganj Dareh, Chogah Golan and Chia Sabz, mixtures of wild and domesticated barley were found. These samples were 10,500 calculated Before Present (B.P.) years old. In the location of Southern Levant, non-brittle barley with increased seeds size was found. It confirms the observation of wild barley presence along with the non-brittle types (Riehl et al. 2011; Tanno and Willcox 2011). All these findings conclude that domestication of barley was slow and involved several steps of selection.

Wild barley (*H. spontaneum*) is distributed throughout the Fertile Crescent area. In addition to the primary habitats, *H. spontaneum* stands occur in eastern Mediterranean, Western Asia, Turkmenia and Eastern Afghanistan (Harlan and Zohary 1966) and also in secondary habitats such as Morocco and Abyssinia (Badr et al. 2000; Molina-Cano et al. 1987).

Domesticated barley spread to Europe from the Mediterranean region, to Africa and eastwards through Iran and Afghanistan into India and China (Zohary and Hopf 2000). East Asian

barley differ from European and North African landraces as they have frequently naked caryopsis and are mainly six-rowed types (Knupffer et al. 2003; Vavilov 1926). In addition, some evidence from the distribution of brittle rachis loci in the barley germplasm highlighted the independent evolution and distribution of cultivated barley. Brittleness of rachis is controlled by two linked complementary genes *Btr1* and *Btr2* and is prevalent in all wild barley. However, the recessive allele *btr1* has been reported to be present in most occidental cultivars whereas the *btr2* allele is present in most oriental cultivars (Komatsuda et al. 2004; Takahashi 1955; Zohary 1999).

Different diversity studies have been performed to identify the region where barley was domesticated. Investigating 317 wild and 57 domesticated cultivars using 400 AFLP loci Badr et al. (2000) reported monophyletic origin of barley domestication in the Israel-Jordan area. From this region wild barley populations were more similar to the domesticated gene pool than any other wild barley population. Morrell and Clegg (2007) found two origins of domestications based on haplotype frequencies. One was within the Fertile Crescent that gave rise to the majority of European and American cultivars. The second was the western foothills of the Zagros Mountains 1,500-3,000 km east of the Fertile Crescent. Central; and Far-Eastern germplasm arises from the second domestication. However, different hypotheses on the origin and sites of domestication are subject of debate (Molina-Cano et al. 2005; Orabi et al. 2007; Saisho and Purugganan 2007; Zohary and Hopf 2000). Barley domestication involved the natural as well as un-intentional selection by early farmers that probably led to gradual changes in plant architecture. Six-rowed spike, non-brittle rachis, and naked caryopsis evolved in the cultivated germplasm (Kilian et al. 2009; Salamini et al. 2002). Barley is sensitive to photoperiod, but mutation in the photoperiod genes led to photoperiod insensitivity in barley and thus its distribution range extended to diverse geographical areas across the continents (Jones et al. 2008; Takahashi et al. 1963; von Bothmer et al. 2003b).

Spike morphology is important in barley as six-rowed barley has yield advantage over the two- rowed due to presence of three fertile spikelets at each rachis node (von Bothmer et al. 1995; von Bothmer et al. 2003b). In six-rowed cultivated barley all three spikelets are fertile and bear seeds whereas in wild and cultivated two-rowed barley the lateral spikelets are suppressed and do not produce any seed. Variation in spike morphology is observed in natural and induced mutants (Lundqvist and Lundqvist 1987; Lundqvist et al. 1996;). Six-rowed spike phenotype is reported to be under the control of five independent genes, *Vrs1* (2H), *Vrs2* (5H), *Vrs3* (1H), *Vrs4* (3H) and *Vrs5* (4H) (Lundqvist et al. 1997). The major spike

morphology locus (*Vrs1*) was recently cloned (*HvHox1*) using map based cloning approach. The dominant allele of *Vrs1* suppresses the lateral spikelets in six-rowed genotypes resulting in a two-rowed phenotype (Komatsuda et al. 2007). The dominant allele (*Vrs1*) was found to be widespread in wild barley and confirmed previous findings that two-rowed barley is the ancestral form and six-rowed evolved later from two-rowed barley. Another locus, *Vrs5* (*Int-C*), was reported to control male fertility and enlarged lateral spikelets in two-rowed types (Ramsay et al. 2011). A genome-wide association scan of 190 barley cultivars using SNP markers was used to identify and characterize this gene (Ramsay et al. 2011). This gene was significantly associated to the markers on chromosome 4H in the panel comprising of two-rowed and six-rowed types. Mutant screenings identified the barley ortholog (*HvTB1*), maize domestication gene - *Teosinte Branched 1* as candidate for the *Vrs5* (*Int-C*) gene. In future, cloning of other spike morphology genes will shed more light on the evolution and functionality of these loci.

Barley is grown under different seasons of the year. In some areas it requires cold winter period before flowering. This growth habit is mainly under the control of vernalization genes at the *Vrn-H1*, *Vrn-H2*, and *Vrn-H3* loci (Takahashi and Yasuda 1971). Epistatic interaction between *Vrn-H1* and *Vrn-H2* loci causes winter and spring growth habit (Szucs et al. 2007). Most variation for the vernalization is due to the *Vrn-H1* and *Vrn-H2* loci in European-cultivated barley, and winter alleles at *Vrn-H3* is considered fixed genotypes but can provide adaptive variation in flowering behavior in barley (Cockram et al. 2007; Yan et al. 2006). Spring allele at *Vrn-H3* locus promotes early flowering and is found mostly in exotic germplasm (Takahashi and Yasuda 1971; Wang et al. 2010). In addition, early flowering with and without vernalization can be due to the various alleles at flowering loci such as *Ppd-H1* affecting photoperiodic response (Turner et al. 2005).

### 1.3 Genetic bottleneck

Cultivated crops have several distinguishing characters that differentiate them from their wild progenitor and are collectively called ‘‘domestication syndrome’’ (Hammer 1984). Since domestication, crop plants have experienced selection forces that have changed their genetic base. This has resulted in the fixation and narrowing of the genetic base in cultivated barley in comparison to wild species (Tanksley and McCouch 1997). For increasing the crop productivity under the changing environmental conditions, wild gene pools and landraces can

be utilized in plant breeding for crop improvement (Bockelman and Valkoun 2010; Kilian et al. 2006; von Korff et al. 2004; Xu et al. 2012).

Molecular markers have paved the way to explore diversity at genomic level. In most of the crop plants as barley, molecular linkage maps are developed that can be used to access the genetic diversity at molecular level (Graner et al. 1991; Heun et al. 1991; Kleinhofs et al. 1993).

Diversity studies on wheat and barley cultivars have reported the change in diversity after domestication and during the extensive plant breeding practices. In a study on barley cultivars, allele frequency shifts at certain genomic regions were observed due to breeders selection (Russell et al. 2000). In another study Kilian et al. (2006), signs of genetic bottlenecks in a panel of cultivars from Turkey in comparison to wild barley have been observed at seven loci. In domesticated genotypes, two loci (*Amy1* and *GAPDH*) were monomorphic and 87% of the nucleotide diversity was lost in the remaining loci. More haplotypes were found in wild barley in comparison to domesticated genotypes. Several other studies found similar results which emphasized the presence of bottlenecks during barley domestication. (Christiansen et al. 2002; Donini et al. 2000; Khlestkina et al. 2004; Kilian et al. 2006; Koebner et al. 2003; Malysheva-Otto et al. 2006; Ordon et al. 2005; Roussel et al. 2004; Russell et al. 2000).

There is an urgent need to scrutinize and recruit new alleles that have been lost during the process of domestication and modern breeding processes (Kilian et al. 2006; Tanksley and McCouch 1997). Initially, the use of genetic resources was limited to resistance breeding in crop plants. Their potential for yield improvement is not appropriately investigated yet. Only few reports have demonstrated their utility for yield improvement (Cox and Frey 1984; Takeda and Frey 1987). Nevertheless, their practical value was limited for various reasons e.g. linkage drag often hampered their applicability in practical plant breeding. However, with the advent of molecular markers, usefulness of wild relatives for crop improvement became conceivable in practical terms in comparison to the conventional approaches (deVicente and Tanksley 1993; Frary et al. 2000; Tanksley and McCouch 1997; Xiao et al. 1996). Plant breeding efforts in the last century have improved the productivity of the crop plants. However, relying only on the elite germplasm has severely reduced the genetic base. As many alleles were lost during the process of domestication and breeding (Tanksley and McCouch 1997), there is great need to mine for useful alleles. In recent years, several approaches like

allele mining, advanced back crossed QTL analysis (AB-QTL), multi-parent advanced generation intercross (MAGIC) population, nested association mapping (NAM) population and association mapping approaches have been developed in crops that can be used to broaden the genetic basis of the domesticated germplasm by moving exotic and un-used alleles (Bhullar et al. 2009; Cavanagh et al. 2008; Pillen et al. 2003; von Korff et al. 2004).

### 1.4 Genetics and genomics resources in barley

Several genetic linkage maps were developed for barley using different markers (Haseneyer et al. 2010a). The first molecular linkage map using Restriction Fragment Length Polymorphism (RFLP) markers was developed in 1991 (Graner et al. 1991; Heun et al. 1991b). Genetic and genomics resources in barley have been evolving continuously from three-point linkage test derived classical map to the high density SSR (sequence tagged microsatellite), diversity array technology (DArT) and expressed sequence tagged (EST) based saturated linkage maps (Close et al. 2009; Lundqvist et al. 1996; Marcel et al. 2007; Rostoks et al. 2005; Sato et al. 2009; Stein et al. 2007; Varshney et al. 2007; Wenzl et al. 2006). Polymerase Chain Reaction (PCR) based genetic markers have revolutionized the development of genetic linkage maps and gene discovery.

From expressed sequence tags (ESTs) and pilot sequencing studies, numerous SNP (Single Nucleotide Polymorphism) markers were developed for barley (Close et al. 2009; Kota et al. 2003; Sato et al. 2009). Currently, high-throughput SNP genotyping platforms with a capacity to genotype several hundred genotypes with thousands of SNPs are available at low costs and have revolutionized the practical uses of genomics (Illumina$^{TM}$: Veracode, Goldengate and the iSelect chip assay, Affymetrix gene chip) (Atwell et al. 2010; Close et al. 2009; Close et al. 2004). High resolution genetic map together with the sequence information in barley has huge potential for candidate gene discovery using conservation of synteny based searches in sequenced grass genomes (Mayer et al. 2011).

Despite the complexity of the barley genome, genomic resources can efficiently be exploited to clone and to characterize candidate genes (Saisho and Takeda 2011). For instance, the major durable disease resistant gene *mlo* that is effective against all barley powdery mildew races and the *Vrs1* gene responsible for row type morphology were cloned (Buschges et al. 1997; Komatsuda et al. 2007). Mutant lines for barley have been extensively studied. Several

genes involved in plant development were reported and cloned from barley mutants and are listed in Table A.

Table A: List of genes cloned or mapped from barley mutants

| Genes | References |
|---|---|
| Nitrate reductase | (Somers et al. 1983) |
| Anthocyanin pathway genes | (von Wettstein 2007) |
| Floral bract phenotype *knox3* | (Muller et al. 1995) |
| Naked caryopsis phenotype *NUD* | (Taketa et al. 2008) |
| Plant height *Uzu* | (Chono et al. 2003) |
| Dominant dwarf mutant *SLENDER1* | (Chandler et al. 2002) |
| Endosperm Development genes | (Clarke et al. 2008; Felker et al. 1983; Morell et al. 2003; Roder et al. 2006) |
| Disease resistance genes *Necrotic1*, *Necrotic.S1* | (Rostoks et al. 2006; Zhang et al. 2009a) |
| Row type gene *Vrs1* and *int-C* | (Komatsuda et al. 2007; Ramsay et al. 2011) |

Despite these enlisted successess in using barley mutants for characterizing and studying various pathways in barley development, it has not been explored and exploited as *Arabidopsis* mutants due to it's large genome and not systematic avilability of the mutant populations. In recent past, Druka et al. (2011) have developed 881 backcross lines that segregate for morphological and developmental traits. These genotypes were characterized with more than 3,000 SNP markers and the genetic location of 426 mutant allele chromosomal segments with good resolution was revealed. Mutation frequency is very low and developing such populations takes years. However, introduction of TILLING (Targeting Local Lesions In Genomes) a reverse genetic approach that combines mutagenesis with high throughput genome scale screening of mutations in genes of interest is complimentary to the forward mutant screening approaches (McCallum et al. 2000). In barley efforts were made to generate such large TILLING mutant populations (Caldwell et al. 2004; Gottwald et al. 2009; Talamè et al. 2008). The main advantages of such resources are that they can be used to screen candidate genes mutations along with precise phenotype of limited mutant lines. Also barley physical map will be available in the coming years then these resources can be used to annotate the genes that are barley specific and cannot be annotated from the available

sequenced genomes such as *Arabidopsis* that is dicot and have different morphology then barley. These vast mutant resources will be instrumental in dissecting the pathways and in identifying the genes underlying the mutant phenotype as shown by Ramsay et al. (2011) for cloning *int-C* gene.

### 1.5 Genome-wide association studies and population structure

Association mapping or linkage disequilibrium (LD) mapping refers to the non-random association of alleles at different loci. The term LD was first used in 1960s by Lewontin and Kojima (Slatkin 2008). Loci that are near to each other tend to have strong LD (Slatkin 2008). The extent of LD differs in different crops and gene pools. The scale and the extent of LD determine the marker number required to cover the genome (Myles et al. 2009). The extent of LD determines the resolution of marker-trait association. High LD in a population indicates that fewer markers are required to detect the marker-trait associations, but with low resolution. (Myles et al. 2009). Contrastingly, low LD in a population indicates that dense marker coverage is needed, but resolution of mapping is higher (Waugh et al. 2009). Association mapping at genome-wide level is highly successful in human genetics (Syvanen 2005). However, genome wide association studies (GWAS) have recently started in plants (Waugh et al. 2009). Statistical framework that suits well to the plant genetics was developed and implemented in order to control the population structure of natural plant populations (Flint-Garcia et al. 2003; Rafalski 2002; Rafalski 2010).

In almost all natural populations, non-random distribution of genotypes within population exists due to complex relationship among them, and thus causes population structure. Population structure often results in spurious associations as allele frequencies are biased among the subpopulations. Different statistical methods were implemented to correct for population structure (Flint-Garcia et al. 2003; Mackay and Powell 2007). Among them, general linear model (GLM) based on structured association (SA), genomic control (GC) and family based tests were implemented to control type I error in association studies (Abecasis et al. 2000; Devlin and Roeder 1999; Pritchard et al. 2000). In case of SA, random markers are used to estimate population structure and then implemented in a statistical framework to control false positives. On the other hand, in GC random markers are used to control the false positives assuming structure has similar affect on all loci. Recently, mixed model analysis that accounts for multiple levels of relatedness was proposed and is also applied in many association genetics studies (Yu et al. 2006). In order to develop the mixed model analysis,

kinship or principal components from the markers are used to account for population structure (Price et al. 2006; Yu et al. 2006; Zhang et al. 2010). Although association mapping is a straight forward approach, the inherent population structure in natural populations can lead to false positives and strong spurious associations (Rafalski 2010). Hence, determining and implementing appropriate statistical models to control the false positives in genome-wide association studies is an important pre-requisite (Stich et al. 2008). In most cases, mixed model frameworks performs comparatively better and are extensively used in association studies (Atwell et al. 2010; Comadran et al. 2011b; Huang et al. 2010a).

The use of natural populations for association mapping has added advantage over the conventional QTL mapping, as there is no need to develop the mapping population (Rafalski 2010). In comparison to the bi-parental QTL mapping, association mapping captures multiple allele segregation in natural populations. LD in the association panels is due to several generations of recombination (Myles et al. 2009). In addition, varietal historical data can be used directly in genome-wide association approaches to characterize the morphological and phenotypic traits at genomic level (Cockram et al. 2010; Kraakman et al. 2004; Wang et al. 2011). Association mapping can be applied to elite germplasm using phenotypic and genotypic data generated in plant breeding programs (Jansen et al. 2003). Recently, method for adopting family-based association test is proposed that can be applied to plant breeding programs (Stich et al. 2006). Thus, association genetics potentially identify genetic variants which control complex traits and addresses several issues of QTL-MAS in breeding programmes (Mackay and Powell 2007).

Several complex disease loci were identified in human populations using GWAS approach (International HapMap Consortium 2007; The International Hapmap Consortium 2005; Wellcome Trust Case Control Consortium 2007). In plants, GWAS was reported initially on self pollinated model plant *Arabidopsis thaliana* and cross pollinated crop species maize. In *Arabidopsis*, population structure and pattern of genome wide LD was revealed using diverse genotypes. In these studies LD decay was found rapidly decaying within 50 kb region in *Arabidopsis* (Nordborg et al. 2005; Remington et al. 2001; Thornsberry et al. 2001). Further, genome-wide association studies have successfully identified flowering time and pathogen resistance related genes using structured populations of *Arabidopsis* (Aranzana et al. 2005). On the other hand, in the cross pollinated species maize, LD extent was < 1kb in landraces, 2kb in diverse inbred lines and 100 kb in elite inbred lines (Ching et al. 2002; Remington et

al. 2001; Tenaillon et al. 2001). The studies in maize reiterate the dependency of LD decay and LD patterns on the selected germplasm (Rafalski 2010; Waugh et al. 2009).

Even in self pollinating crop plant like barley, low LD has been observed in landraces and wild barley populations (Caldwell et al. 2006; Morrell et al. 2005). One of the first studies in barley using SNP markers revealed that few thousand SNP markers can cover the whole genome in elite barley as the LD is large (Rostoks et al. 2006). Since then, several association studies have been published in barley using candidate gene re-sequencing and different high-throughput marker sets (Illumina Goldengate assay, DArT assay) (Cockram et al. 2008; Cockram et al. 2010; Comadran et al. 2011b; Comadran et al. 2009; Haseneyer et al. 2010b; Pasam et al. 2012; Stracke et al. 2009). Majority of barley association genetic studies reported structured populations based on spike morphology and growth habits (Comadran et al. 2011b; Pasam et al. 2012; Wang et al. 2010; Wang et al. 2011; Waugh et al. 2010). Further, within the row type, population structure was observed in different studies due to geographical and breeding histories (Rodriguez et al. 2012; Zhang et al. 2009b).

Further, the captured allelic diversity in natural populations has revealed the evolutionary and adapted signatures of the species. Population structure and population genetics statistics in natural populations helps in understanding the species spread and distribution (Fournier-Level et al. 2011; Turner et al. 2005).

In barley, many candidate gene based association mapping studies have been published (Haseneyer et al. 2008; Stracke et al. 2009), and recently, GWAS approach has also been reported in many studies (Cockram et al. 2010; Comadran et al. 2011; Comadran et al. 2008; Kraakman et al. 2004; Kraakman et al. 2006; Pasam et al. 2012; Roy et al. 2010). In these studies despite the large barley genome, few thousand markers were used and significant marker-trait associations were observed. One obvious reason for the success was presence of larger LD (~10cM) extent in cultivars that requires only few thousand markers to find association at genome level, in barley (Waugh et al. 2009; Waugh et al. 2010). At the genome-wide level, studies have been successful in identifying major genes (Waugh et al. 2009). However, few publications have emphasized on complex trait identification that are mostly polygenic and control many agronomic traits (Waugh et al. 2009). Atwell et al. (2010) have demonstrated at genome-wide level, significant marker-trait association of 107 phenotypes in *Arabidopsis thaliana* using 250,000 SNP markers for different traits.

Nevertheless, in most cases allele effects in association mapping studies were low, as low as 2-3 percent of the total variation for some traits (Pasam et al. 2012).

GWAS has the potential to narrow down to candidate gene regions for each trait depending on the marker number, trait, size and resolution of the population. For instance, Huang et al. (2010) performed GWAS for 14 agronomic traits using ~3.6 million SNPs identified in genotyping by sequencing approach (GBS) in a collection of diverse rice landraces. Loci that were previously cloned were also detected in this study and the significantly associated markers were only 1-3 genes away from the known candidate genes. The level of significance was in most cases higher within the nearby flanking genes than in the functionally characterized known candidate genes. These results clearly indicate the complexity of multiple polymorphisms and the effect of population structure. In barley, *int-C* gene was characterized using genome-wide association followed by identification of candidate gene by searching the sequenced rice genome. Seventeen independent mutants were instrumental in revealing *Teosinte Branched 1* as the candidate gene. These results fueled research in the genome-wide association in crop plants. In addition, some successful examples for gene discovery using LD approaches have revolutionized the QTL discovery.

The fact that in some cases associated SNP can lead down to the candidate genes using syntenic information from already sequenced related species has further added value to the GWAS (Cockram et al. 2010; Houston et al. 2012).

### 1.6 Agronomic traits in barley

One of the most important objectives in plant breeding is to understand the genetic basis of complex agronomic and developmental traits. Those usually are under the control of major and minor quantitative trait loci (QTL). In barley several studies reported QTL for agronomic traits but often they are narrow to specific populations (Barley QTL workbook at Graingene http://wheat.pw.usda.gov/GG2/index.shtml). In several instances plant developmental genes like vernalization (*Vrn-H1* and *Vrn-H2*), flowering time (*Ppd-H1*) and height (*denso*, *uzu*) has been reported to have direct affect on the agronomic traits (Backes et al. 1995; Hackett et al. 1992; Li et al. 2005; Thomas et al. 1991; Thomas et al. 1995; Wang et al. 2010). Almost all agronomical important traits like yield and yield related component traits have strong environmental influence that also suggest identifying and dissecting these traits at molecular level is difficult. Thus needs good genotyping and robust phenotyping. With the availability of high throughput genotyping platforms genotyping is not difficult but phenotyping and

generation of high resolution mapping population is still cost intensive. Moreover most high resolution bi-parental mapping populations are segregating for few traits. On the other hand association mapping seems promising approach where several genotypes can be phenotyped and genotyped. The high genetic resolution based on LD decay has power to detect QTL from medium to high resolution depending upon extent of LD. Therefore it has wider applicability across genotypes. Recently association mapping has been successfully employed on agronomic traits in barley. Significant marker-trait association in several cases reported. In addition to major developmental genes population specific QTL were also located in these studies that needs further characterization to elucidate their role in plant development (Pasam et al. 2012; Ramsay et al. 2011; Stracke et al. 2009).

### 1.7 Importance of Powdery mildew in barley

More than 80 different diseases attack barley, but from them only a handful can cause economic losses (Mathre 1997). Among them, one of the major barley diseases is powdery mildew. It is caused by biotrophic pathogen *Blumeria graminis* f. sp. *hordei (Bgh).* It occurs in almost all barley production areas, but is most common and damaging in northern Europe, Japan and United States (Kiesling 1985; Mathre 1997). Its spread is favored in cool and humid weather conditions, but can also occur in warmer and semiarid environments (Mathre 1997). Development and deployment of resistant cultivars is one of the major objectives in barley breeding. Many different resistant alleles have been identified in cultivated and wild barley (Dreiseitl and Dinoor 2004; Jørgensen 1994). Thus, easily followed R gene resistance introgression is mostly followed in breeding. More than 30 alleles of the barley Mla complex that specify resistance are found in cultivars, worldwide. Many of them are molecularly characterized and belong to the coiled-coil, nucleotide binding site, leucine-rich repeat (CC-NBS-LRR) class of genes and they confer race specific resistance (Halterman et al. 2001; Jones 2001; Zhou et al. 2001). However, the effectiveness of resistance genes can be easily overcome due to the appearance of new virulence type races (Brown 1994; Wolfe and McDermott 1994). The other promising approach is to accumulate non-race specific (NR) durable resistant QTL but due to the complexity of QTL inheritance, it is difficult and less favored in practical breeding (Kou and Wang 2010). Moreover, in barley, interesting example is mlo11 allele that confers durable broad spectrum resistance against powdery mildew, for more than 30 years (Jørgensen 1992). However, the mechanism of quantitative resistance that usually is slow but effective against broad range of races is yet not discovered. Recently, a co-evolutionary model of race-specific resistance Vs non-race specific durable resistance was

proposed by Jones and Dangl (2006) and it has created interest among scientists to characterize NR resistance in crop plants. Also, powdery mildew infection in barley is one of the most studied host-pathogen infection systems. Some of the pioneering work on powdery mildew (*Bgh*) disease infection involved the identification of pathogen recognition factors, signal transduction and resistance responses (see review) (Schulze-Lefert and Panstruga 2003). In addition, the genome of Bgh was sequenced recently and the effector proteins were discovered (Bindschedler et al. 2009; Godfrey et al. 2010; Spanu et al. 2010). Due to the close evolutionary relevance, it is also a model species for the Erysiphales clade that include important pathogen like *Blumeria graminis* f. sp. *tritici* that cause powdery mildew in wheat, resulting in yield loss (Spanu et al. 2010). These developments together with large body of knowledge assist in understanding the basics of quantitative disease resistance and have relevance across the species (Jones and Dangl 2006). QTL discovery based on GWAS is quite promising and can unravel the complexity of durable non-race specific resistance mechanism.

## 1.8 Objectives of the study

In the light of the above mentioned successes and available genetic and genomic resources "*Genome-wide association studies in diverse gene pools of barley*" were carried out. The main objectives of this study were:

1. To determine the extent of LD in cultivated (winter, spring), landrace and wild barley populations.

As reported from different studies on barley, LD can vary and LD has to be determined before embarking on genome-wide association studies – mainly for two reasons. First the resolution of the GWAS is directly based on the LD decay and second, long range inter chromosomal LD between different chromosomes can lead to false positives in association studies. Also no extensive studies of LD comparing cultivated, landraces and wild barley have been published yet. In this study cultivated, landraces and wild barley were compared and their LD was determined.

2. To study genome-wide association for powdery mildew disease resistance in different gene pools of barley under field and controlled detached leaf assay conditions.

Among the yield reducing constrains on barley, powdery mildew disease caused by *Blumeria graminis* f. sp. *hordei* is considered one of the major pest that can cause heavy yield losses. In order to characterize the durable non-race specific quantitative resistance in barley, genome-wide association scan was performed using poly-virulent isolates of Bgh. In addition, powdery mildew resistance under field grown condition was also characterized in the cultivated barley as it gives us more inside view of learning race specific and non-race specific resistance at genome-wide level in the natural barley populations.

3. To study genome-wide association studies for agronomic and seed quality traits in cultivated winter and spring barley populations.

Barley is cultivated in winter and spring season and its growth is orchestrated by a series of vernalization and photoperiod genes. In addition different breeding history resulted in different sets of QTL selected and are segregated in the cultivated barley. In order to study genome-wide association in the cultivated gene pool, European winter and spring barleys

were selected and phenotyped extensively in multi-environmental field trials. GWAS across the winter and spring barley were compared and discussed in detail in this study.

# 2. Materials and Methods

## 2.1 Plant materials

Three diverse barley gene pools were investigated in the "Exploitation of Barley Diversity" (EXBARDIV) project in the framework of European Research Area in Plant Genetics (ERA-PG): cultivars (*Hordeum vulgare* L.), landraces (*Hordeum vulgare* L.) and wild barley accessions (*Hordeum spontaneum* (C. Koch.) Thell.). In addition, one association mapping panel was investigated from the GABI-GENOBAR project. These gene pools are described as following:

### 2.1.1 *Hordeum vulgare* cultivar collection (HVCC)

This HVCC gene pool consisted of 282 spring (HVCC_S) and 112 winter (HVCC_W) barley genotypes. These genotypes were selected from eighteen countries of Europe. Both two-rowed and six-rowed spike genotypes were included in the study (Sup. Table S1). All genotypes were provided by the project partners and were single seed descended in order to remove the heterogeneity and ensure the purity of seeds. Seeds were later multiplied at CRA (CRA – Fiorenzuola d'Arda, Italy (project partner)) for multi-location field trials. Genotype descriptions their pedigree, country of origin, rowed type, growth habit, year of release and the marker type used in this study are provided as supplementary information (Sup. Table S1). For some genotypes, varietal characteristics like rowed-type, winter, spring growth habits found in our field trials did not match with the varietal release specifications of seed board of the genotypes. Therefore, these genotypes were written as unknown in Sup. Table S1. Nevertheless, these genotypes did not affect our GWAS, as DNA was sampled from the same plants that were multiplied and used in the experiments.

### 2.1.2 Landraces gene pool (LRC)

Landraces collected by Dr. Eva Welzien from Syria and Jordan (Weltzien 1988) were used in this study. These landraces are maintained at the International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria. From this landrace collection 238 (LRC_S) genotypes from Syria and 79 genotypes (LRC_J) from Jordan were selected and used in the current study. They were selected from different parts of Syria and Jordan viz. South Jordan, North Jordan and South Syria, Central Syria, West Syria, Central Syria and

North east Syria. (modified from Russell et al. 2011 Sup. Table S2). These landraces were previously characterized using SSR and SNP markers by Russell et al. (2003, 2011). These were also multiplied in the field at CRA.

### 2.1.3 Wild barley gene pool (HSC)

*Hordeum spontaneum* the progenitor of cultivated barley was used in the study to represent the wild barley gene pool (referred as *Hordeum spontaneum collection*- HSC). HSC consists of 216 wild barley accessions sampled and purified in Israel. These wild barley accessions sampled across different Eco-geographic and soil regimes were genotyped with 45 SSR markers by Hübner et al. (2009). As these genotypes were well characterized and studied in different on-going projects, we considered the HSC gene pool for powdery mildew detached leaf assay screening. Genotype names, site code, their location and the regions coordinate in UTM (X and Y) are provided as Sup. Table S3 (modified from Hübner et al. (2009) Sup. Table S3).

### 2.1.4 Additional association panel

**The GENOBAR collection of worldwide origin (224 cultivars; HVCC_Genobar)**

In addition, a well characterized association mapping panel comprising of a diverse worldwide collection of 224 spring barley is available at IPK. This collection was mainly used for powdery mildew resistance screening in this study. These genotypes were extensively genotyped and phenotyped in various other studies (Haseneyer et al. 2010b; Pasam et al. 2012; Stracke et al. 2009). This collection represents a diverse set of spring barley genotypes and comprises of 127 two-rowed and 97 six-rowed barleys. These genotypes were from the different eco-geographical regions of the world, 109 accessions originated from Europe (EU), 45 from West Asia and North Africa (WANA), 40 from East Asia (EA) and 30 from the Americas (AM) (see modified from Pasam et al. (2012) Sup. Table S4).

## 2.2 Genotyping

### 2.2.1 SNP genotyping by Illumina GoldenGate genotyping Assay

Leaves from the single plants grown under greenhouse conditions were used for DNA extraction. Tepnel Nucleoplex plant DNA extraction kits (Tepnel Life Sciences PLC, Manchester, UK) were used to extract high quality DNA at James Hutton Institute (formerly SCRI, UK). Quality and quantity of DNA were checked and two micrograms were used for genotyping at the Southern California Genotyping Consortium facility at the University of

California, Los Angeles, USA. All gene pools were genotyped using the BOPA1 (Barley Oligo Pool Assay) that contained 1536 SNP markers (Close et al. 2009). Allele calls were made using the clustering algorithm GenTrain available in Bead-Studio 2.0 (Illumina Inc., http://www.illumina.com). The algorithm in BeadStudio was developed for outbreeders, where three clusters (two homozygotes and one heterozygote) are expected. However, in highly self pollinated barley we expected only two clusters therefore all SNPs were manually checked. Their relative performance was discussed in Moragues et al. (2010). The consensus map positions from Close et al. (2009) were used as genetic framework to place the SNP markers on seven barley chromosomes.

### 2.2.2 SNP genotyping by iSelect SNP Chip

The Illumina iSelect platform using 9k chip was used to genotype DNA sampled from the gene pools (HVCC_S, HVCC_W and HVCC_Genobar). This chip consists of SNPs detected by RNA sequencing of ten diverse barley cultivars, along with 1482 BOPA1, 1456 BOPA2 SNP markers (Comadran et al. unpublished). All genotypes were genotyped at TraitGenetics GmbH Gatersleben, Germany. Allele calls were made and from the successful 7842 SNPs. SNPs with $\geq 0.05$ minor allele frequency (MAF) (4923 SNPs in HVCC_W; 5753 SNPs in HVCC_S and 6461 SNPs in HVCCC_Genobar) were considered for further analysis. Primarily, the mapping information from Morex X Barke recombinant inbred lines (349 RIL) was used. Markers that were not placed on Morex X Barke RIL map were assigned genetic positions from BOPA1 and BOPA2 consensus map (Close et al. 2009) and in addition LD was used to place rest of the iSelect markers (Comadran et al. in preparation).

## 2.3 Population structure and linkage disequilibrium (LD)

BOPA1 SNP markers were used to determine the population structure and the extent of LD in all EXBARDIV gene pools. Principal component analysis (PCA) was used to plot the distribution of genotypes spatially and to estimate the population structure (Price et al. 2006). SNP markers (common 1356 SNP markers across gene pools) were used to perform PCA using TASSEL 2.1 software (http://www.maizegenetics.net/) and the significant component values were plotted using excel 2007 (Microsoft office 2007 package). Further separately for each gene pool Neighbour Joining (NJ) clustering using Euclidean distances and bootstrap N = 1000 was generated using PAST Software (http://folk.uio.no/ohammer/past/). LD between markers was calculated by the squared correlation coefficient ($r^2$) (Weir 1979) method between mapped SNPs using HAPLOVIEW version 4.2 software (Barrett et al. 2005). SNPs with minor allele frequencies (< 0.05%) were removed from the analysis. LD heatmaps

depicting high and low LD regions of seven chromosomes for HVCC_S, HVCC_W, LRC and HSC were generated. Significant LD ($r^2$ values, $P < 0.001$) were plotted against genetic distances (cM) in Excel 2007 (Microsoft Office package 2007). Unlinked $r^2$ (genetic distance $\geq$ 50cM) estimates were squared root transformed to approximate a normally distributed random variable and the parametric $95^{th}$ percentile of that distribution was taken as a critical value of $r^2$, beyond which LD is likely to be caused by genetic linkage (Breseghello and Sorrells 2006).

**2.4 Phenotypic analysis**

In order to generate robust phenotypic data, multi-environmental field trials were performed. The Randomized Complete Block (RCB) design was used to plant HVCC and LRC gene pools in two years (2008-09 and 2009-10) at four locations (CRA – Fiorenzuola d'Arda, Italy; JHI – Dundee, UK earlier SCRI; IPK – Gatersleben, Germany; UniHalle – Halle, Germany) under autumn and spring sown conditions. Seeds were sown in replicated trails with two replicates for each genotype with ~ 2 $m^2$ plots and grown according to the local agronomic practice for sowing rate and other inputs. The enlisted grain yield, yield components, plant phenological, seed quality and powdery mildew resistance traits were recorded at each trial (Table 1).

**Table 1**: Phenotypic traits scored during 2009-10 in field trials for HVCC_W, HVCC_S and LRC gene pools. CRA, HAL, IPK and SCR are the location codes for CRA – Fiorenzuola d'Arda, Italy (CRA); JHI – Dundee, UK earlier SCRI (SCR); IPK – Gatersleben, Germany (IPK); UniHalle – Halle, Germany (HAL); 09 and 10 are the respective years of the field trials. Traits abbreviations are given in brackets.

| | | CRA09 | CRA10 | HAL09 | HAL10 | IPK09 | IPK10 | SCRI09 | SCRI10 |
|---|---|---|---|---|---|---|---|---|---|
| **HVCC (Winter)** | | | | | | | | | |
| 1 | Frost (Fr) | X | X | X | X | X | X | X | X |
| 2 | Powdery Mildew (PMfld) | | | | | X | X | | |
| 3 | Heading date (Hd) | X | X | X | X | X | X | X | X |
| 4 | Height (Ht) | X | X | X | X | X | X | X | X |
| 5 | Spike length (Sl) | X | X | X | X | X | X | X | X |
| 6 | Grains / Ear (GEar) | X | X | X | X | X | X | X | X |
| 7 | Yield (Yld) | X | X | X | X | X | X | X | X |
| 8 | Thousand Grain Weight (TGW) | X | X | X | X | X | X | X | X |
| 9 | Grain Area (Garea) | X | X | X | X | X | X | X | X |
| 10 | Grain Length (Glength) | X | X | X | X | X | X | X | X |
| 11 | Grain Width (Gwidth) | X | X | X | X | X | X | X | X |
| 12 | Width / length (W_L) | X | X | X | X | X | X | X | X |
| 13 | Protein content (PC) | X | | X | | X | | X | |
| 14 | Starch content (SC) | X | | X | | X | | X | |
| **HVCC (Spring)** | | | | | | | | | |
| 1 | Powdery Mildew (PMfld) | | | | | X | X | | |
| 2 | Heading date (Hd) | X | X | X | X | X | X | X | X |
| 3 | Height (Ht) | X | X | X | X | X | X | X | X |
| 4 | Spike length (Sl) | X | X | X | X | X | X | X | X |
| 5 | Grains / Ear (GEar) | X | X | X | X | X | X | X | X |
| 6 | Yield (Yld) | X | X | X | X | X | X | X | X |
| 7 | Thousand Grain Weight (TGW) | X | X | X | X | X | X | X | X |
| 8 | Grain Area (Garea) | X | X | X | X | X | X | X | X |
| 9 | Grain Length (Glength) | X | X | X | X | X | X | X | X |
| 10 | Grain Width (Gwidth) | X | X | X | X | X | X | X | X |
| 11 | Width / length (W_L) | X | X | X | X | X | X | X | X |
| 12 | Protein content (PC) | X | | X | | X | | X | |
| 13 | Starch content (SC) | X | | X | | X | | X | |
| **LRC** | | | | | | | | | |
| 1 | Protein content (PC) | | | | | | | X | |
| 2 | Starch content (SC) | | | | | | | X | |

### 2.4.1 Phenological traits

***Frost damage (Fr)*** was scored for the winter (HVCC_W) barley gene pool in field trials. The percentage of leaf death due to frost damage was scored on a visual basis during and after melting of snow (Fig 1).

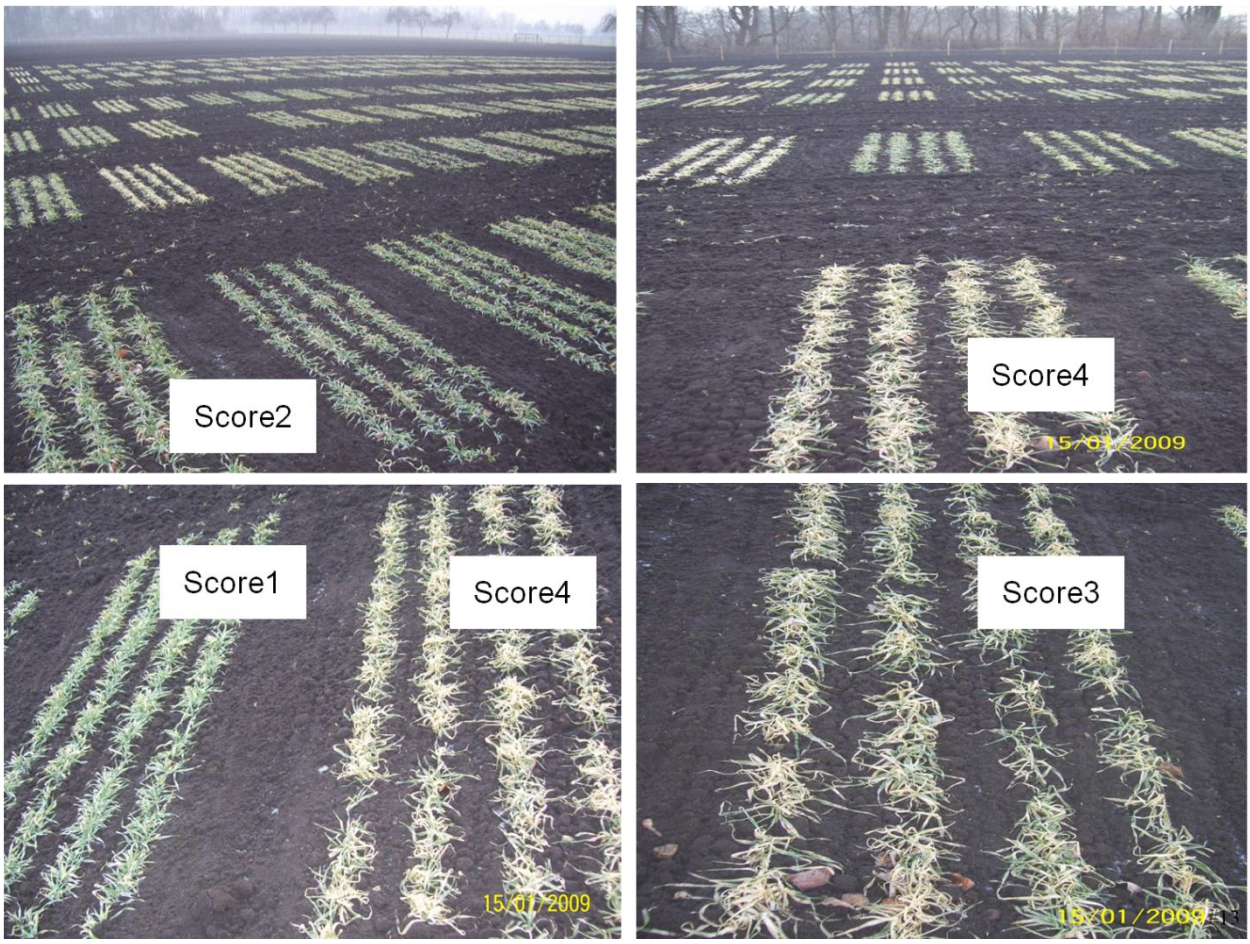



**Fig 1**: Showing frost damage in the winter fields 2008-09. Frost damage score 1-4 is shown wherein 1 means no damage and 4 means complete damage of the plants

***Heading date (Hd)*** was scored in days from the date of sowing until 50 percent of the plants in the plot had reached growth stage @ GS53 (Lancashire et al. 1991). The growth stage at which a third of the ear emerges from the plant is considered as @ GS53.

***Height (Ht)*** was measured in centimeters (cm) from the soil to the ear tip excluding awns.

***Spike length (Sl)*** was measured in cm from the spike tip to the base of the spike excluding awns.

### 2.4.2 Grain traits

After maturity plants were harvested with a combine harvester. Moisture content was measured and the harvested grains were dried to get the uniform moisture content. Then the grain yield (Yld) was measured in grams. Additionally, five spikes from randomly selected plants from the middle of each plot were harvested, dried, and manually threshed. Later, a seed analyzer "Marvin" (GTA Sensorik GmbH, Neubrandenburg, Germany) was used to measure grains per ear (GEar), thousand-grain weight (TGW), seed area (Garea), seed width (Gwidth), seed length (Glength) and Width / length (W_l).

### 2.4.3 Grain quality traits

Seed samples were ground and NIRS (Near Infrared Reflectance Spectrometer device (Multi Purpose Analyzer (MPA) from Bruker)) was used to estimate starch content (SC) and protein content (PC) of the samples. In order to calibrate and validate NIRS estimation, 160 samples were measured using Kjeldahl method and 200 samples were measured using polarimetric method for estimating starch and protein content. The estimation was highly correlated with the NIRS values from the same samples. Later, NIRS was used to estimate SC and PC for all samples. These traits were measured at the Julius Kühn-Institut (JKI) Bundesforschungsinstitut für Kulturpflanzen Institut für Resistenzforschung und Stresstoleranz OT Groß Lüsewitz Germany by Dr. Gisela Jansen and colleagues.

### 2.4.4 Powdery mildew (PM)

In order to characterize powdery mildew disease resistance two experiments were conducted under the field conditions (at location IPK) the percentage of leaf infection was scored for the HVCC_S and HVCC_W gene pool. In addition a PM detached leaf assay (PMdla) was conducted under controlled conditions for all gene pools (Table S1, S2,S3 and S4).

For the PM detached leaf assay (PMdla), two poly-virulent pathotypes of *Bgh* (78P and D12/12; for the virulence spectrum see Table 2) were used to screen all gene pools. As the powdery mildew pathogen is biotrophic, two isolates were maintained on the susceptible barley variety "Golden Promise". Scoring was performed using the resistant genotype (Ingrid Back-Crossed mlo5 containing genotype) and a susceptible control genotype (Morex) as controls. Seeds were sown in 96 well plates under disease free conditions in the green house. Two weeks old seedlings were used for the detached leaf assay. Second leaf from two weeks old plants was used for the detached leaf assay. Phytoagar gel plates were used in order to

maintain moisture and anti-senescent conditions for the detached leaves. The phytoagar plates were prepared by boiling 1% phytoagar in microwave oven, followed by adding 250µl of Benzimidazol (concentration 40mg Benzimidazole dissolved in 1ml Ethanol) after cooling to 50$^{o}$C. The solution was poured into the plates and left for some minutes for solidification. Leaf samples were cut into two halves and each half of the leaf was mounted on to two different plates using pins. One plate was inoculated with isolate 78P and the other plate with isolate D12/12. Two individuals of each genotype were used as biological replicate. Each plate was carefully inoculated using conidia (from isolates maintained) under aseptic conditions under a hood. Different lab coats and gloves were used for inoculation with each isolate to avoid cross contamination. After cleaning with disinfectants, a second inoculation was performed. Difference of twenty minutes maintained between two inoculations. In order to monitor inoculation density, one microscopic slide was placed inside the plate that was later checked for the number of conidia spores/mm$^2$ threshold of 5.0 spores / mm$^2$ was set per inoculation. Subsequently the plates were incubated at room temperature for one week under disease free conditions. One week old plates showed visual leaf infection symptoms. These were scored on a scale from 0 to 3 (Fig 2). The scores were later converted into percent leaf infection as described in Fig 2. To confirm the repeatability and to avoid escapes of resistant scores, screening for powdery mildew resistance by detached leaf assay was repeated for the genotypes with resistance scores 0 and 1. Average leaf infection across the isolates was calculated outliers were removed and the infection score was used for the powdery mildew detached leaf assay association analysis.

**Table 2**: Virulence spectrum of the powdery mildew isolates 78P and D12/12. Depending upon the percent leaf infection with PM conidia was scored on a 0-4 scale. Wherein 0 is resistant, 1 moderate resistant, 2 moderate susceptible, 3 susceptible and 4 highly susceptible used as reference. A set of differential cultivars with known R-genes and uncharacterized (U) genes was used as reference. Highlighted columns (pink) are resistant and (yellow) are susceptible scores. (Source: Kerstin Flath, JKI Klein-Machnow).

| Differentials hosts | Resistance gene(s) | 78P | D12/12 |
|---|---|---|---|
| Alexis | *mlo9* | 0 | 0 |
| Amazone | *Mlg, U* | 2 | 4 |
| Apex | *mlo11, Mlg, Ml(CP)* | 0 | 0 |
| Aura | *Mla6, Mla14, Mlg, Ml(CP)* | 3 | 0 |
| Banteng | *Ml(Ba)* | 4 | 4 |
| Borwina | *Ml(BW1,2)* | 2-3 | 2 |
| Camilla | *U* | 2 | 4 |
| Dura | *Ml(Dr)* | 4 | 3-4 |
| Goldi | *U, Mla12, MlLa* | 3-4 | 0 |
| Gunnar | *Mla3, Ml(Tu2)* | 0 | 2-3 |
| Hellas | *He* | 4 | 3-4 |
| *Hordeum spontaneum* | *Mla16* | 0 | 0 |
| *Hordeum spontaneum* | *Mla17* | 1-2 | 2 |
| *Hordeum spontaneum* | *Mla18* | 0 | 0 |
| *Hordeum spontaneum* | *Mla19* | 0 | 0 |
| *Hordeum spontaneum* | *Mla20* | 1-2 | 0 |
| *Hordeum spontaneum* | *Mla26* | 1-2 | 0 |
| *Hordeum spontaneum* | *Mla27* | 0 | 4 |
| *Hordeum spontaneum* | *Mla28* | 2-3 | 0 |
| *Hordeum spontaneum* | *Mlf* | 2 | 1 |
| *Hordeum spontaneum* | *Mlj* | 2-3 | 2-3 |
| Katharina | *U* | 3 | 4 |
| Kredit | *Ml(Kr)* | 4 | 2-3 |
| Lerche | *Mlg, Ml(CP), Mla7* | 3 | 2 |
| Lotta | *Ml(Ab)* | 3-4 | 3 |
| Maresi | *Mla12, Ml(AB), Mlg, Ml(CP)* | 2-3 | 3 |
| Marinka | *Mlg, Ml(CP), Mla7* | 4 | 2-3 |
| Meltan | *U, Mla13, Ml(Im9), Ml(Hu4)* | 3 | 0 |
| Olga | *U* | 4 | 0 |
| P01 | *Mla1* | 1-2 | 3 |
| P02 | *Mla3* | 1-2 | 3 |
| P03 | *Mla6, Mla14* | 4 | 0 |
| P04 B | *Mla7, U* | 4 | 4 |
| P06 | *Mla7, Ml(LG2)* | 4 | 0 |
| P08B | *Mla9* | 4 | 0 |

| | | | |
|---|---|---|---|
| P09 | *Mla10, Ml(Du2)* | 4 | 0 |
| P10 | *Mla12* | 4 | 4 |
| P11 | *Mla13, Ml(Ru3)* | 4 | 0 |
| P12 | *Mla22* | 0 | 3 |
| P13 | *Mla23* | 2 | 2 |
| P14 | *Mlra* | 4 | 3-4 |
| P15 | *Ml(Ru2)* | 2-3 | 3 |
| P17 | *Mlk* | 4 | 2-3 |
| P18 | *Mlnn* | 4 | 3 |
| P19 | *Mlp* | 2-3 | 3-4 |
| P20 | *Mlat* | 2 | 3-4 |
| P21 | *Mlg, Ml(CP)* | 3 | 3 |
| P22 | *mlo5* | 0 | 0 |
| P23 | *MI(La)* | 4 | 3 |
| P24 | *MlH* | 4 | 3-4 |
| Sissy | *Mla12, Mlg, Ml(CP)* | 4 | 3-4 |
| Steffi | *Ml(St1,2)* | 1-2 | 3-4 |
| Teo | *Mla7, Mlg, Ml(CP), Mlk, MLa* | 3 | 1 |
| Thuringia | *Ml(St1,2), U* | 2 | 3-4 |
| Trumpf | *Mla7, Ml(AB)* | 4 | 3-4 |

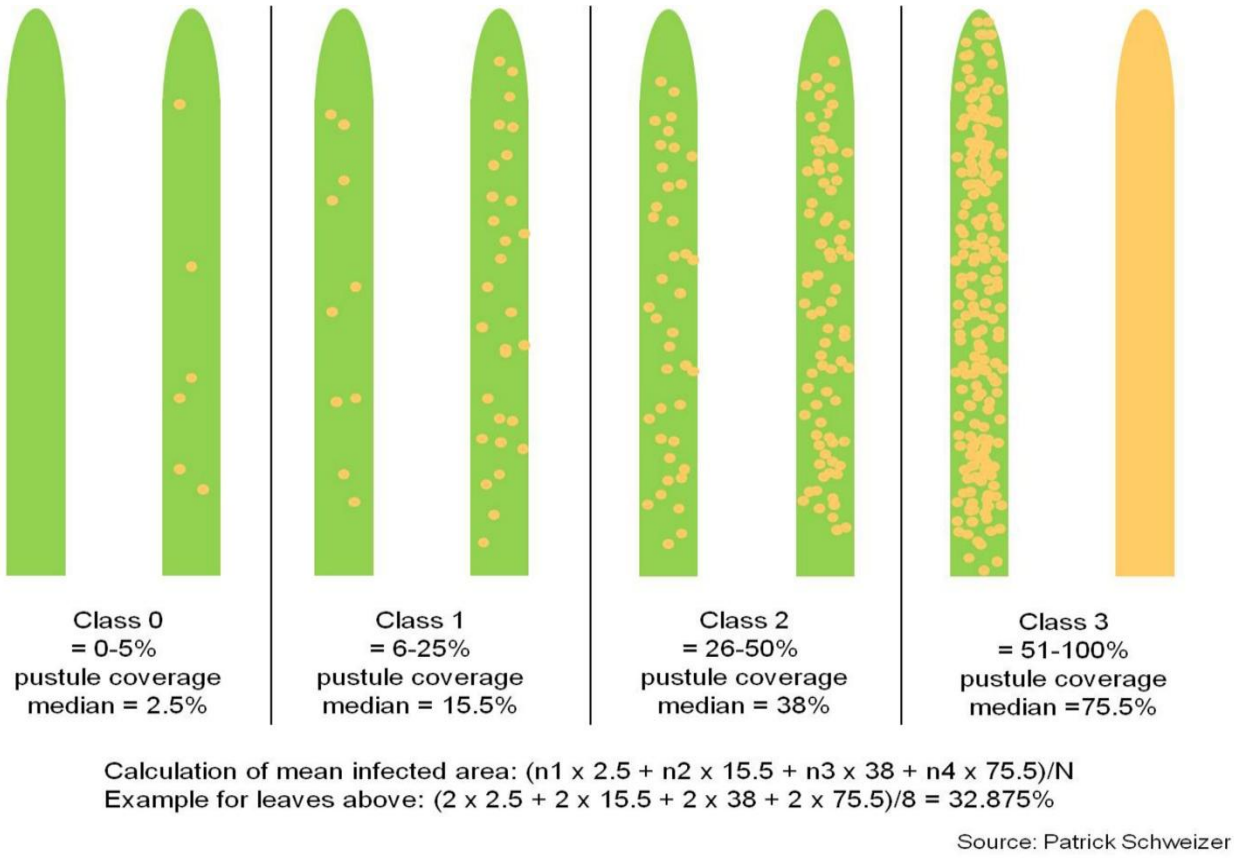



**Fig 2**: Graphical representation of the powdery mildew detached leaf assay infection showing the pustule growth (dots) on the leaf surface. According to the PM pustules covering the leaf area, three classes are defined in the detached leaf assay viz. Class 0 with 0-5% (median 2.5%); Class 1 with 6-25% (median 15.5%); Class 2 with 26-50% (median 38%) and Class 3 with 51-100% (median 75.5%) leaf covered with pustule. (Source: Dr. Patrick Schweizer, IPK Gatersleben Germany)

### 2.5 Phenotypic and association analysis

Phenotypic data was analyzed using Genstat 13$^{th}$ Edition (Payne 2009). Restriction Estimate of Maximum Likelihood (REML) was used to generate genotypic means (BLUE -Balanced Linear Unbiased Estimate) from each field trial after removing the outliers. Corresponding to the year and the location, separated BLUEs were generated considering the best spatial model for each year and each location. Later all errors and models were considered in a meta-analysis to generate BLUEs for each trait across year and location. Summary statistics of each trait viz. minimum, maximum, mean, standard deviation, variance, coefficient of variation and repeatability [genetic variance/ (genetic variance + residual variance)] range for each trait was generated. Narrow sense heritability is proportion of additive variation compared to the total genetic and error variation, whereas broad sense heritability is strictly proportion of total genetic effects compared to the total phenotypic variation. In this study broad sense heritability was used as traits were measured in muti-location trials and this reflects

repeatability of our data. As the data was analyzed using mixed model broad sense heritability was estimated using the formula (G + G.S + G.Y + G.Y.S)/(G + G.S + G.Y + G.Y.S + E) where G is genotypic, S is site, Y is year and E is error variance (. means interactions) estimated in REML using Genstat software. In order to estimate the relationship between traits, a correlation matrix was calculated between all traits. GWA could lead to spurious associations due to population structure. Different models were proposed to correct for population structure (Kang et al. 2008; Stich et al. 2008; Zhang et al. 2010; Zhu and Yu 2009). In this study different models were compared. In most cases, mixed model using kinship matrix was performing better than any other model. As an example, (Sup. Fig S1) cumulative observed p-value of GWAS powdery mildew resistance in percentages for naïve model (without correction for population), general linear model (GLM) using Q matrix (Q matrix from Structure software (Pritchard et al. 2000)), mixed model using principal components and kinship matrix was plotted. A model that efficiently controls false positives should approximate to a uniform distribution of the p-value. A mixed model based on a kinship matrix and principal components was performing best as it approximated a straight uniform distribution of the p-values. Hence, only the kinship model will be applied and discussed in whole study. In this model markers were treated as fixed effects and a k matrix (kinship) generated from 362 uniformly distributed SNP markers (i.e., 1 informative SNP per 3cM) across the seven chromosomes was fitted as a random effect in order to reduce false positives (Type I error). TASSEL version 2.1 was used to calculate the associations (http://www.maizegenetics.net/). In order to reduce further false positives, the conservative threshold of $-\log_{10}$ (P value) of 2 (P-value 0.01) was chosen in declaring significant associations. Additionally using more conservative approach, false discovery rate threshold based on multiple test hypothesis q values at $\alpha$ 0.05 significance was calculated for each trait and provided in the supplementary table (Storey and Tibshirani 2003) (Table S33, S34, S35, S36). The q value is similar to p value, but is measure of significance in terms of false discovery rate rather than false positive rate. The false discovery rate of a test is defined as the expected proportion of false positives among the declared significant results (Benjamini and Hochberg 1995). For an example 0.05 FDR means we expect 5 SNP are false positives from the 100 SNPs.

# 3 Results

## 3.1 Population structure

One of the pre-requisites for association mapping is determining the population structure of the collection used in GWAS. In order to estimate population structure principal component analysis was performed for all genotypes using BOPA1 SNP markers. Strong population structure was observed in the whole collection (Fig 3.1) with the first two principal components collectively explaining 30 percent of the variation. Principal component 1 (PC 1) explained 19.12 percent of the variation and separated the cultivated (HVCC) from the landraces (LRC) and wild barley (HSC) gene pools. Principal component 2 (PC 2) explained another 10.98 percent of the variation. Based on PC2, HVCC genotypes were majorly split into two-rowed and six-rowed types. Further spring and winter growth habit type genotypes clustered separately. Spring barley (HVCC_S_2) two rowed and winter barley (HVCC_W_2) two rowed were distinct from the spring (HVCC_S_6) six rowed and winter (W_6) six rowed genotypes. However, pronounced differences in structuring according to the row-type were observed in spring barley genotypes in comparison to the winter barley genotypes (Fig 3.1). This could be due to the lower number of winter (112) in comparison to spring barley (282) genotypes. Another probable reason might be that winter two rowed and six rowed were not bred separately. Landrace genotypes were more widely distributed across the axis in comparison to wild barley genotypes that look like a compact mass. Further, Jordanian and Syrian landraces (LRC_J and LRC_S) clustered separately. On Principal component 3 (PC 3) two-rowed and six-rowed types in spring and winter barley genotypes (HVCC_S_2, HVCC_S_6, HVCC_W_2 and HVCC_W_6) were showing split with wider split observed for spring two- and six-rowed genotypes (See Sup Fig S2 and S3). The wide distribution of cultivated (HVCC) barley gene pool on principal component axis reflects their diversity. On the other hand, low diversity was observed in wild (HSC) and landraces (LRC) gene pools although high diversity was reported in previous studies with the same LRC and HSC genotypes (Hübner et al. 2009; Russell et al. 2003). One of the major reasons of the differences in this study and the earlier study was the marker types and their informativeness. Most SNP markers in this study performed well in cultivated gene pool (Fig 3.2). On the other hand, around 19 percent of SNP markers in LRC and 12 percent in HSC gene pools were having minor allele frequency (MAF) $\leq 0.05$. In addition, the polymorphic information content (PIC) was high in cultivars (HVCC) in comparison to the wild (HSC) and landraces

(LRC) gene pools (Fig 3.2). It was quite clear from PCA that major structuring of the genotype was based on the population type viz. cultivars (spring: HVCC_S and winter: HVCC_W), landraces (LRC) and wild (HSC) barley, respectively. Thus, we considered them as separate populations and genome-wide association analysis was performed and discussed separately for each gene pool in this study.

Further to investigate population structure in the respective gene pools, STRUCTURE 2.3.3 software and Neighbour joining cluster analysis (based on Euclidean distances) was performed (Falush et al. 2003; Pritchard et al. 2000). For STRUCTURE 2.3.3, correlated allele frequency and admixture model was run assuming group k = 1 to 20 (hypothetical number of groups) and 5 run per k value was used. For robust analysis, initial burn-in period was set to 50000 followed by 1000000 Markov Chain Monte Carlo (MCMC) iterations. To determine the suitable k value "Ln *P*(*D*)" obtained from structure runs were plotted against the k and smallest k after which "Ln *P*(*D*)" reached plateau was considered as the best value to assign individuals to clusters (Pritchard et al. 2000). In this study, k plateaus after k=2 for winter and spring barley gene pools (Sup. Fig S4 and S5) although smaller increase in Ln *P*(*D*) value was observed. In addition, delta (k) method proposed by Evanno et al. (Evanno et al. 2005) was also used to confirm the observation. According to which, delta (k) is the mean of second order rate of change in Ln *P*(*D*) values of given k divided by the standard deviation and the curve shows upper delta (k) value at k= 2. In addition, Neighbour joining cluster analysis confirms existence of two major groups in cultivars (winter and spring types) (Sup Fig S6). Major clustering in these two groups within cultivars was due to the spike morphology type (Fig 3.3). Two-rowed, six-rowed genotypes were clustered and group completely separate in winter, and spring barley gene pools except few admixture genotypes (Fig 3.3). In landraces (LRC), according to the origin Syrian and Jordanian landraces were observed clustering separately with few admixtures (Sup. Fig S7). Wild barley genotypes were somewhat clustered according to the sites of collection (Sup. Fig S8).

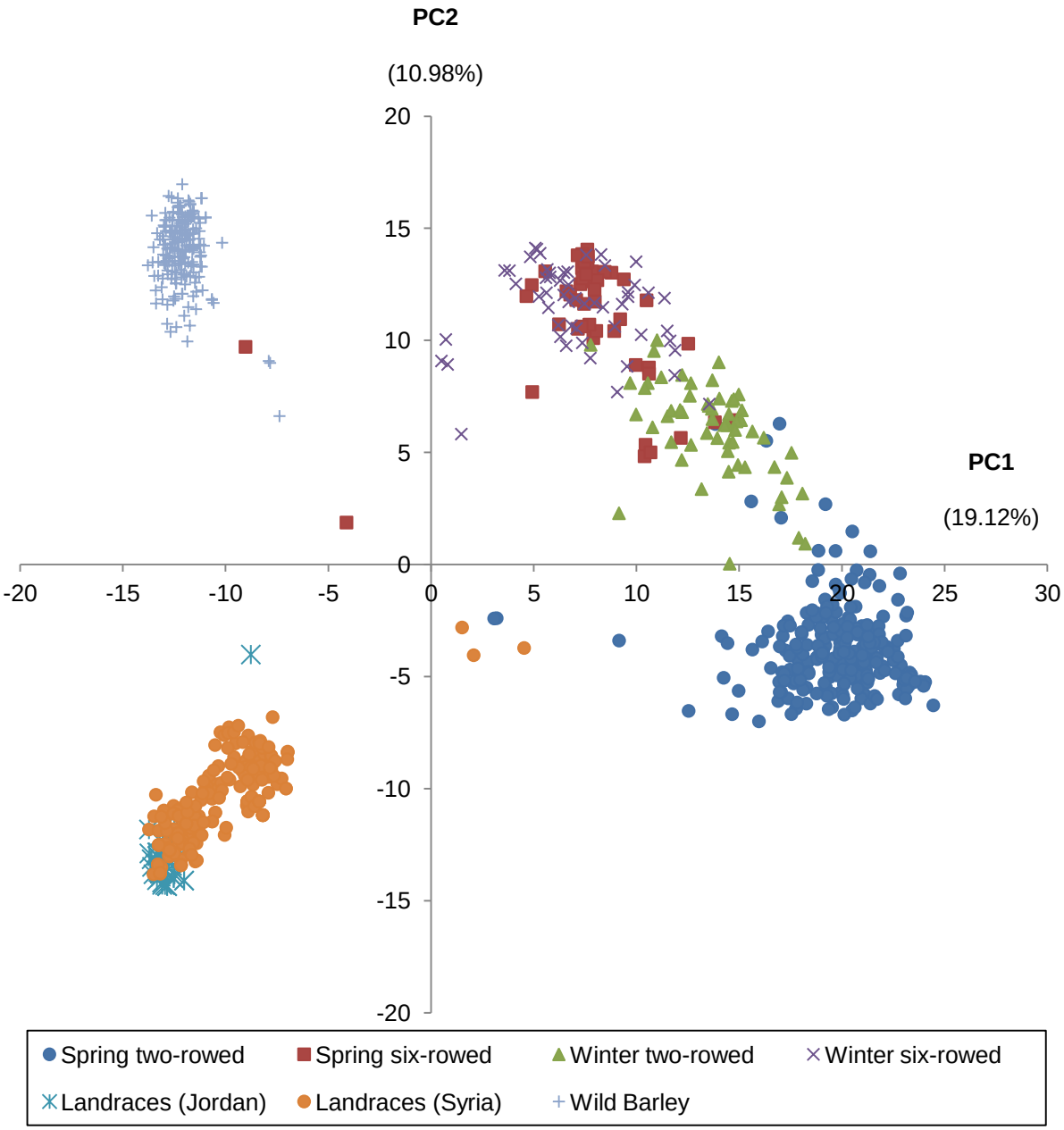



**Fig 3.1**: Population structure of different gene pools using BOPA1 SNP markers. Horizontal and vertical axis showing principal components (PC1 and PC2) explaining 19.12 and 10.98% variation. Genotypes from different gene pools are shown with different color symbols. Blue round: spring two-rowed (HVCC_S gene pool), brown squares: spring six-rowed (HVCC_S gene pool), green triangles: winter two-rowed (HVCC_W gene pool), purple cross: winter six-rowed (HVCC_W gene pool), light blue cross: Jordanian landraces (LRC gene pool), light brown circles: Syrian landraces (LRC gene pool) and light blue plus: wild barley (HSC gene pool).

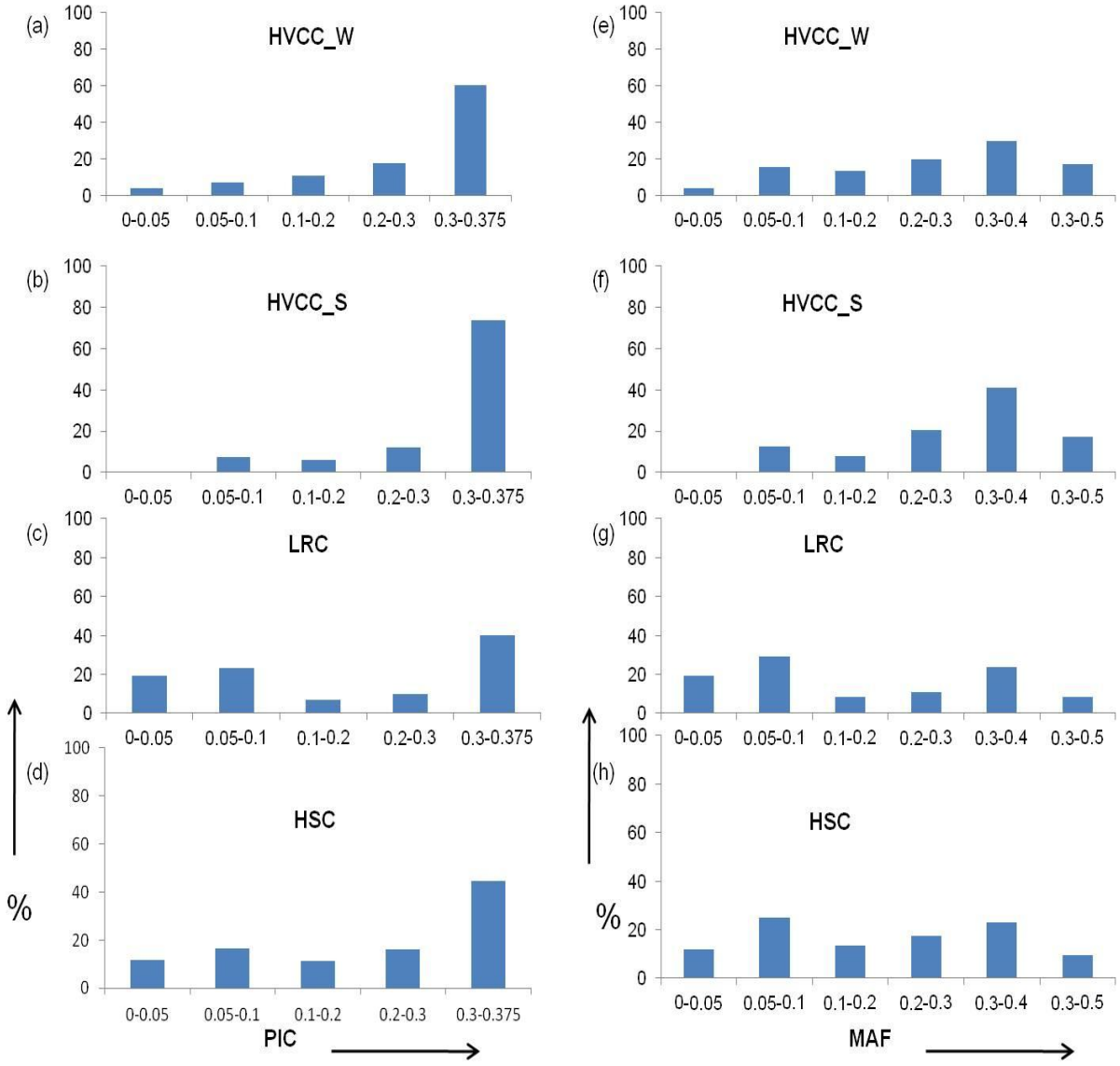



**Fig 3.2**: Summary statistics of the BOPA1 SNP markers across the gene pools. Polymorphic information content (PIC) of the winter barley (HVCC_W) **(a)**; spring barley (HVCC_S) **(b)**; landraces (LRC) **(c)**; and wild barley (HSC) **(d))** and minor allele frequencies (MAF) of winter barley (HVCC_W) **(e)**; spring barley (HVCC_S) **(f)**; landraces (LRC) **(g)**; and wild barley (HSC) **(h)** are shown in the histogram plots.

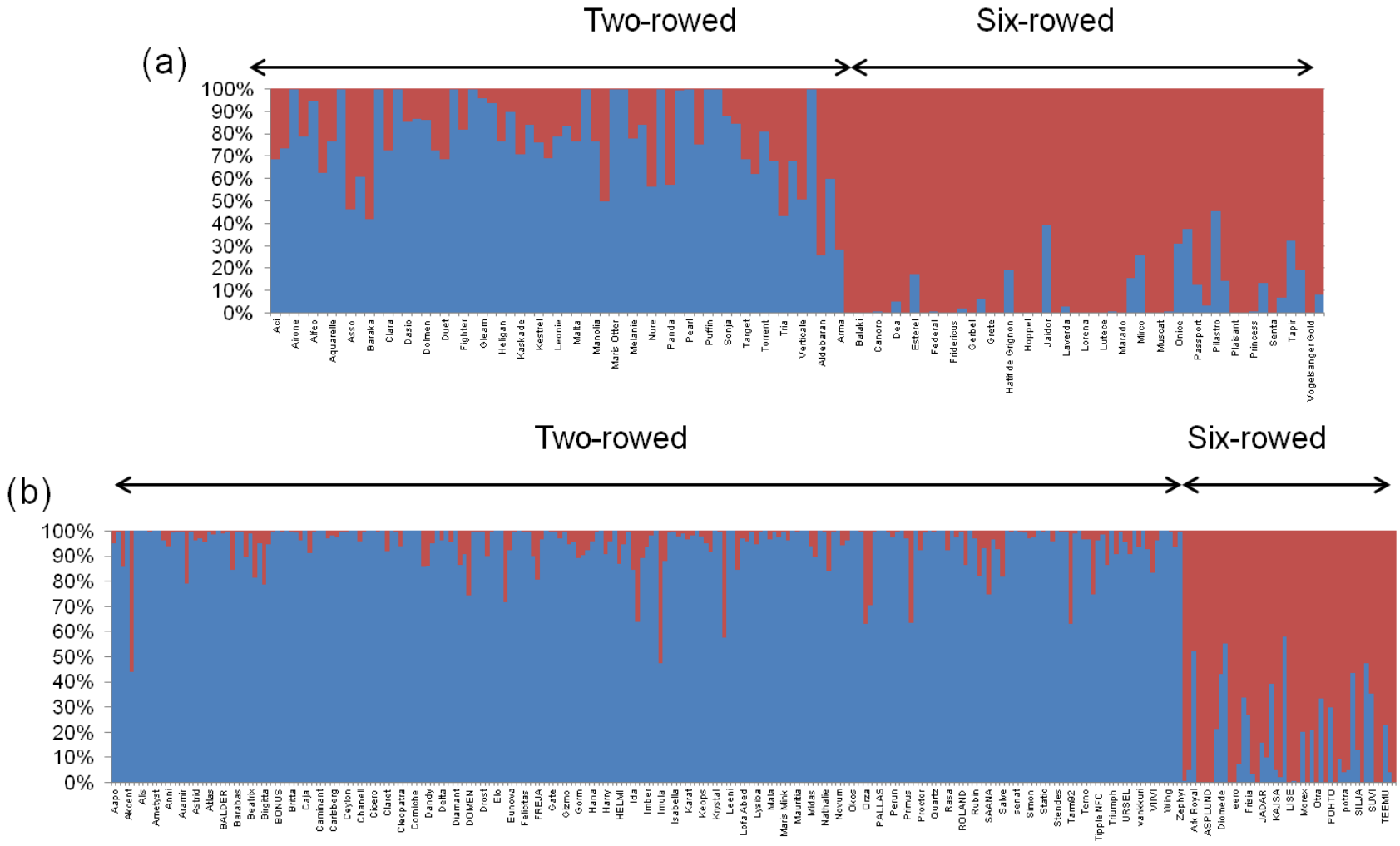



**Fig 3.3**: Population structure based on model based STRUCTURE analysis. Genotypes are represented by a vertical bar which is portioned into 2 groups (k=2; blue and brown color represent groups) and horizontal axis is the proportion of membership of each genotype to these groups. The genotypes are ordered according to the row-type (two-rowed and six-rowed type); (a) and (b) are population structure of winter (HVCC_W) and spring barley (HVCC_S) gene pools

### 3.2 Linkage disequilibrium

The extent of linkage disequilibrium (LD) in different gene pools was compared and found to be highly variable (Fig 3.4). In general, high LD regions near to the centromeres and lower LD at the telomeric end of the chromosomes were observed (Fig 3.5). Significant LD blocks were found near to the centromeric regions on chromosomes 3H and 5H in HVCC and LRC gene pools. Inter-chromosomal LD dispersal was observed across the chromosomes in cultivated spring, winter and landraces gene pools (Fig 3.5). This long range inter-chromosomal LD may be due to the complex history of the barley breeding. Inter-chromosomal LD can cause false positives in association mapping. Therefore, determining the LD pattern across the genome is very important to determine the efficiency and resolution of the association mapping population. In comparison to cultivated gene pool, the wild barley gene pool had no significant LD regions across seven chromosomes. However, few regions near to centromere on chromosome 3H, 5H and 7H in wild barley were having significant LD (Fig 3.5 c-d).

Threshold $r^2$ value for the background LD was determined for each gene pool. These values were 0.145, 0.146, 0.142 and 0.08 for HVCC_W, HVCC_S, LRC and HSC gene pools, respectively (Fig 3.4 a-d). When comparing across the gene pools, the extent of LD decay was different among the gene pools and in general LD decay was 10 cM in HVCC_S, 14 cM in HVCC_W, 6 cM in LRC and < 1 cM in HSC gene pools (Fig 3.4 a-d). The extent and the pattern of LD were variable in these gene pools. Thus, marker trait association in these gene pools is expected to be of different resolution. In general, we expect high resolution of marker-trait association in landraces and wild barley gene pools due to their low LD.

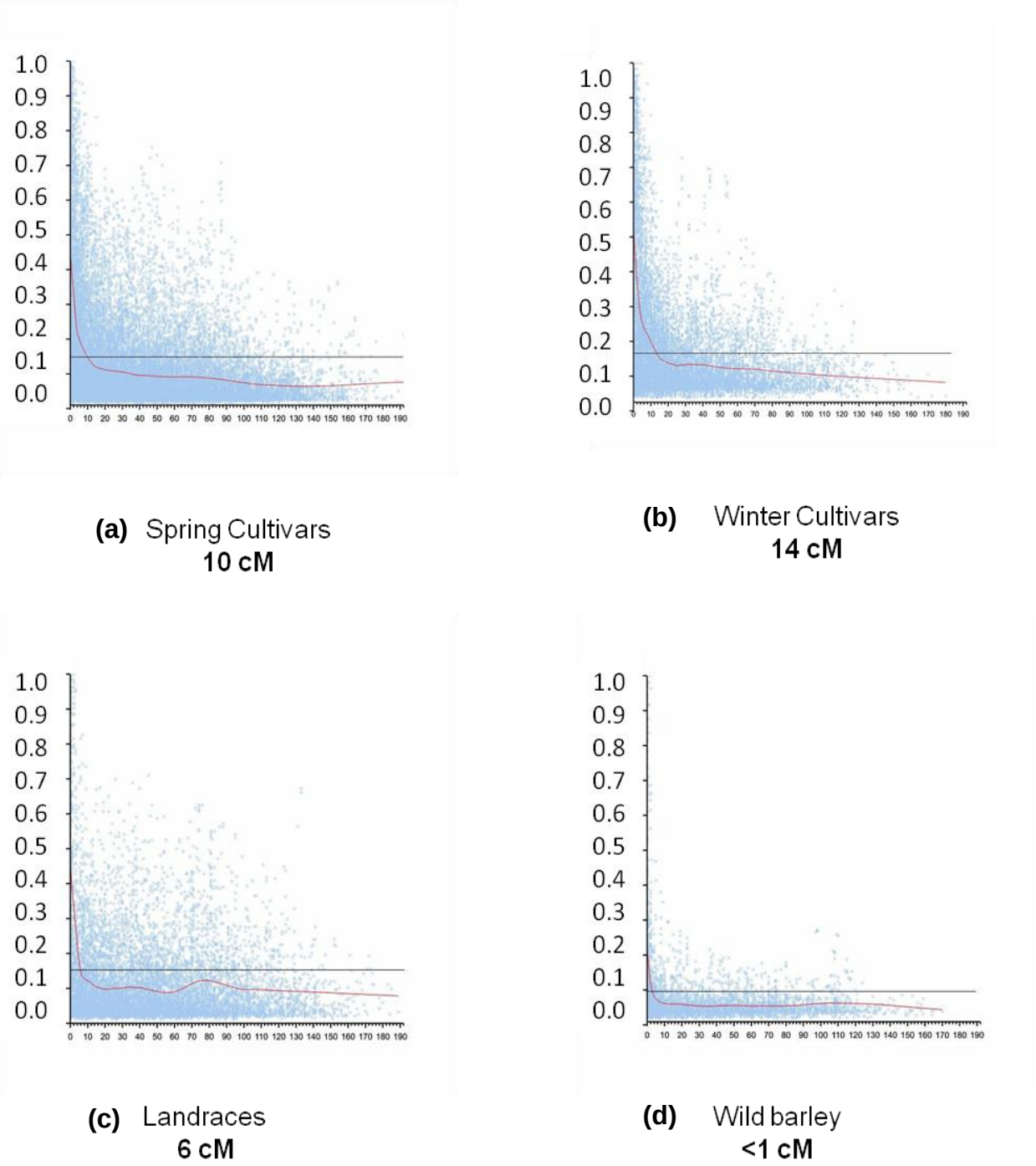



**Fig 3.4**: LD decay graph of the gene pools. Intra-chromosomal LD decay of the gene pool is shown (a-d). The horizontal axis showing genetic distances (cM) and vertical shows the $r^2$ values of the SNPs in LD. Horizontal black line and red line showing threshold and the loess curve

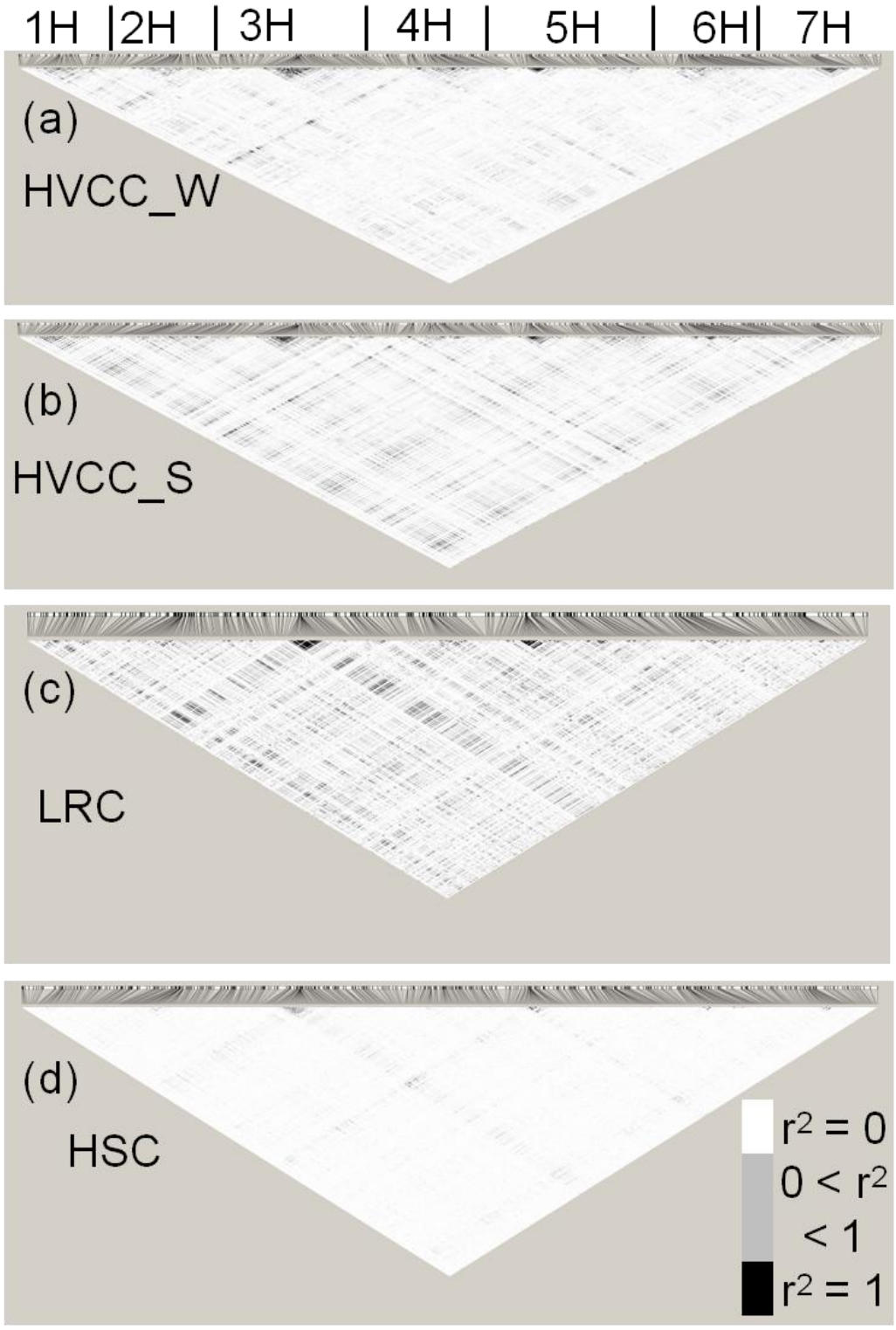



**Fig 3.5**: LD heatmap showing LD pattern across the chromosomes. High LD regions are depicted with grey and black colors across the gene pools [winter (HVCC_W) (a); spring (HVCC_S) (b) cultivars; landraces (LRC) (c) and wild barley (HSC) (d)] spanning seven chromosomes.

### 3.3 Phenotypic analysis

In order to perform GWAS, robust and reliable phenotypic data is important. Data generated in field trials was manually checked and the outliers were removed from analysis. Mean values were calculated for each phenotypic trait based on the best statistical model. Data was found to be variable across the gene pools for all traits. Minimum, maximum, standard deviation, variation, coefficient of variation and repeatability of the trait means were calculated and summarized in Table 3.1. Coefficient of variation for the thirteen traits scored for winter cultivars (HVCC_W) ranged from 1.798 to 58.54 and for twelve traits for spring cultivars (HVCC_S) gene pool ranged from 4.386 to 101.4 as shown in Table 3.1. In all cases, the repeatability of the traits was above 0.5 which reflects the robustness of phenotypic data. In general, repeatability values were high for most of the grain traits. Traits like frost tolerance and yield that are highly environmental dependent showed moderate repeatability values.

Most of the traits measured were not independent from each other and high correlations and relationships was observed between the traits (see Bi-plot analysis Sup. Figs S8, S9). As expected, powdery mildew field infection and powdery mildew detached leaf assay phenotypes clustered together in both HVCC_W and HVCC_S gene pools. Similarly, rowed type and Grains/ Ear traits were clustered together. Further, all seed traits clustered together. Interestingly ear length and heading date clustered close to each other in both HVCC_W and HVCC_S gene pools (Sup. Figs S8, S9). Seed traits displayed highly significant correlation among themselves (Table 3.2 and 3.3). Total grain yield showed significant correlations with yield component traits like grains per ear (GEar), thousand grain weight (TGW), grain length (Glength), grain width (Gwidth) and ratio of grain width by length (W_l) traits. Starch (SC) and protein content (PC) showed high coefficient of variation in LRC than in HVCC_W and HVCC_S gene pools (Table 3.1). Interestingly, heading date (Hd) was negatively correlated to GEar and positively correlated to Yld in the HVCC_S gene pool. This suggests that the late flowering cultivars have less GEar but still manage higher yield. However Yld and TGW were positively correlated also with Hd. This suggests spring barley cultivars that take more time to heading have increased yield because of increased TGW. Plant height (Ht) also showed negative correlation with the yield. This fits well with other cereals like wheat and rice, where plant architecture and yield was improved by reducing plant height. Powdery mildew under field and under controlled detached leaf assay

conditions showed positive correlation of 0.48 in HVCC_S gene pool. Interestingly Hd showed positive correlation (0.53) with the Ht in the HVCC_W gene pool (Table 3.2, 3.3).

Heritability values of 0.58 and 0.68 were observed for the powdery mildew field infection (PMfld) trait in HVCC_W and HVCC_S gene pools, respectively. Mean distribution of maximum leaf infection phenotype in powdery mildew detached leaf assay infection (PMdla) assay was highest in winter cultivars (54 %), followed by landraces (49 %) and was lowest in wild (8 %) barley gene pools (Pie Chart Fig 3.7). As expected, positive and negative controls (Ingrid BCmlo5 and Morex) of the experiments showed infection score of 0 and 3, respectively (Fig 3.6). Continuous phenotypic distribution of different infection classes (PMdlA) was observed in HVCC_S and in HVCC_Genobar collection (Fig 3.7).

**Table 3.1**: Summary statistics of traits in different gene pools. Minimum (Min.), maximum (Max.), mean, standard deviation (SD), variance (Vari.), coefficient of variation (C.V) and the heritability values of the traits are described.

| **HVCC_W** | **Min.** | **Max.** | **Mean** | **S.D.** | **Vari.** | **C.V.** | **Heritability** |
|---|---|---|---|---|---|---|---|
| Frost | 3.79 | 7.20 | 4.56 | 0.75 | 0.55 | 16.31 | 0.60 |
| Heading | 200.00 | 225.20 | 211.90 | 4.05 | 16.38 | 1.91 | 0.89 |
| Height | 55.58 | 110.10 | 84.35 | 8.23 | 67.72 | 9.76 | 0.75 |
| Spike length | 6.14 | 27.96 | 7.94 | 2.03 | 4.11 | 25.53 | 0.90 |
| Grains per Ear | 19.73 | 57.79 | 35.57 | 12.49 | 156.00 | 35.11 | 0.87 |
| Yield | 1.16 | 7.35 | 5.84 | 1.05 | 1.09 | 17.90 | 0.68 |
| Thousand Grain Weight | 40.33 | 61.67 | 49.55 | 4.89 | 23.88 | 9.86 | 0.79 |
| Grain area | 23.20 | 33.15 | 27.04 | 1.70 | 2.90 | 6.30 | 0.79 |
| Grain length | 8.26 | 12.35 | 9.48 | 0.58 | 0.34 | 6.14 | 0.81 |
| Grain Width | 3.46 | 4.30 | 3.80 | 0.18 | 0.03 | 4.72 | 0.81 |
| Grain Width/length | 0.30 | 0.48 | 0.40 | 0.03 | 0.00 | 8.02 | 0.85 |
| Protein Content | 8.85 | 11.69 | 9.77 | 0.54 | 0.30 | 5.57 | 0.56 |
| Starch Content | 49.69 | 56.86 | 52.98 | 0.95 | 907.00 | 1.80 | 0.68 |
| Powdery Mildew field | 2.50 | 75.50 | 28.19 | 15.09 | 227.60 | 53.52 | 0.59 |
| Powdery Mildew detached leaf assay | 2.50 | 75.50 | 39.81 | 23.30 | 543.00 | 58.54 | |
| **HVCC_S** | **Min.** | **Max.** | **Mean** | **S.D** | **Vari.** | **C.V** | **Heritability** |
| Heading | 64.83 | 85.71 | 77.51 | 3.40 | 11.56 | 4.39 | 0.83 |
| Height | 59.83 | 119.50 | 78.73 | 9.43 | 89.00 | 11.98 | 0.76 |
| Spike length | 3.96 | 10.37 | 7.49 | 1.15 | 1.32 | 15.34 | 0.68 |
| Grains per Ear | 17.73 | 61.69 | 28.23 | 10.95 | 120.00 | 38.80 | 0.91 |
| Yield | 2.61 | 6.62 | 5.11 | 0.68 | 0.46 | 13.22 | 0.57 |
| Grain area | 18.54 | 30.73 | 26.25 | 1.55 | 2.40 | 5.91 | 0.80 |
| Thousand Grain Weight | 33.00 | 54.72 | 45.30 | 4.08 | 16.66 | 9.01 | 0.86 |
| Grain length | 6.31 | 11.26 | 8.85 | 0.43 | 0.19 | 4.88 | 0.81 |
| Grain Width | 3.38 | 4.12 | 3.82 | 0.16 | 0.02 | 4.06 | 0.82 |
| Width/length | 0.33 | 0.58 | 0.43 | 0.03 | 0.00 | 5.91 | 0.84 |
| Protein Content | 8.27 | 11.60 | 9.48 | 0.44 | 0.19 | 4.61 | 0.65 |
| Starch Content | 50.43 | 58.96 | 54.09 | 0.79 | 0.62 | 1.45 | 0.73 |
| Powdery Mildew field | 2.50 | 75.50 | 37.57 | 22.79 | 519.30 | 60.65 | 0.66 |
| Powdery Mildew detached leaf assay | 2.50 | 75.50 | 21.73 | 22.04 | 485.60 | 101.40 | |
| **LRC** | **Min.** | **Max.** | **Mean** | **S.D** | **Vari.** | **C.V** | **Heritability** |
| Starch Content | 45.32 | 55.25 | 50.72 | 1.30 | 1.68 | 2.55 | 0.81 |
| Protein Content | 7.80 | 12.59 | 10.12 | 0.70 | 0.49 | 6.91 | 0.54 |
| Powdery Mildew detached leaf assay | 2.50 | 75.50 | 42.16 | 23.63 | 558.50 | 56.06 | |
| **HSC** | **Min.** | **Max.** | **Mean** | **S.D** | **Vari.** | **C.V** | **Heritability** |
| Powdery Mildew detached leaf assay | 2.50 | 75.50 | 9.90 | 13.67 | 186.80 | 138.10 | |
| **HVCC_Genobar** | **Min.** | **Max.** | **Mean** | **S.D** | **Vari.** | **C.V** | **Heritability** |
| Powdery Mildew detached leaf assay | 2.50 | 75.50 | 37.93 | 25.76 | 663.60 | 67.90 | |

**Table 3.2**: Correlation matrix of the traits in HVCC_W gene pool. Row type (Row), frost tolerance (Fr), heading date (Hd), height (Ht), spike length (Sl), grains per ear (GEar), thousand grain weight (TGW), grain area (Garea), grain length (Glength), grain width (Gwidth), ratio grain width/length (W_l), yield (Yld), starch content (SC), protein content (PC), powdery mildew resistance field (PMfld) and powdery mildew resistance (detached leaf assay, PMdla).

| | Row | Fr | Hd | Ht | Sl | GEar | TGW | Garea | Glength | Gwidth | W_l | Yld | SC | PC | PMfld | PMdlA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | 1.00 | | | | | | | | | | | | | | | |
| Fr | 0.01 | 1.00 | | | | | | | | | | | | | | |
| Hd | -0.31* | -0.11 | 1.00 | | | | | | | | | | | | | |
| Ht | 0.04 | -0.25* | 0.53* | 1.00 | | | | | | | | | | | | |
| Sl | -0.37* | 0.28* | 0.16 | -0.22* | 1.00 | | | | | | | | | | | |
| GEar | 0.96* | -0.08 | -0.19* | 0.15 | -0.34* | 1.00 | | | | | | | | | | |
| TGW | -0.70* | -0.13 | 0.28* | 0.15 | 0.28* | -0.67* | 1.00 | | | | | | | | | |
| Garea | -0.26* | -0.05 | 0.15 | 0.21* | 0.19* | -0.26* | 0.76* | 1.00 | | | | | | | | |
| Glength | 0.46* | 0.03 | -0.15 | 0.13 | -0.05 | 0.42* | 0.02 | 0.64* | 1.00 | | | | | | | |
| Gwidth | -0.77* | -0.16 | 0.31* | 0.15 | 0.23* | -0.73* | 0.92* | 0.61* | -0.19* | 1.00 | | | | | | |
| W_l | -0.78* | -0.09 | 0.28* | 0.00 | 0.16 | -0.74* | 0.51* | -0.11 | -0.82* | 0.71* | 1.00 | | | | | |
| Yld | 0.16 | -0.64* | 0.09 | 0.14 | -0.36* | 0.26* | 0.16 | 0.16 | 0.12 | 0.11 | -0.05 | 1.00 | | | | |
| SC | -0.62* | -0.22* | 0.22* | -0.10 | 0.15 | -0.56* | 0.18 | -0.30* | -0.72* | 0.28* | 0.68* | -0.01 | 1.00 | | | |
| PC | -0.41* | 0.36* | 0.12 | 0.11 | 0.18 | -0.48* | 0.43* | 0.31* | -0.05 | 0.42* | 0.30* | -0.50* | -0.15 | 1.00 | | |
| PMfld | 0.12 | 0.16 | -0.29* | -0.26* | -0.01 | 0.08 | -0.22* | -0.19* | -0.04 | -0.19 | -0.08 | -0.16 | -0.08 | -0.07 | 1.00 | |
| PMdlA | 0.20* | -0.03 | -0.15 | -0.07 | -0.14 | 0.20* | -0.31* | -0.32* | -0.12 | -0.24* | -0.06 | -0.07 | -0.09 | -0.10 | 0.08 | 1.00 |

*Significant at P< 0.05

**Table 3.3**: Correlation matrix of the traits in HVCC_S gene pool. Row type (Row), heading date (Hd), height (Ht), spike length (Sl), grains per ear (GEar), thousand grain weight (TGW), grain area (Garea), grain length (Glength), grain width (Gwidth), ratio grain width/length (W_l), yield (Yld), starch content (SC), protein content (PC), powdery mildew resistance field (PMfld) and powdery mildew resistance (detached leaf assay, PMdla) traits.

| | Row | Hd | Ht | Sl | TGW | GEar | Garea | Glength | Gwidth | W_l | Yld | SC | PC | PMfld | PMdlA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Row | 1.00 | | | | | | | | | | | | | | |
| Hd | -0.70* | 1.00 | | | | | | | | | | | | | |
| Ht | 0.28* | -0.22* | 1.00 | | | | | | | | | | | | |
| Sl | -0.71* | 0.62* | 0.02 | 1.00 | | | | | | | | | | | |
| TGW | -0.64* | 0.31* | -0.14 | 0.52* | 1.00 | | | | | | | | | | |
| GEar | 0.97* | -0.62* | 0.31* | -0.67* | -0.66* | 1.00 | | | | | | | | | |
| Garea | -0.31* | 0.08 | -0.09 | 0.28* | 0.83* | -0.33* | 1.00 | | | | | | | | |
| Glength | 0.13 | -0.19* | 0.06 | 0.00 | 0.43* | 0.08 | 0.83* | 1.00 | | | | | | | |
| Gwidth | -0.72* | 0.43* | -0.25* | 0.50* | 0.88* | -0.71* | 0.64* | 0.11 | 1.00 | | | | | | |
| W_l | -0.54* | 0.42* | -0.22* | 0.30* | 0.21* | -0.51* | -0.27* | -0.75* | 0.55* | 1.00 | | | | | |
| Yld | -0.66* | 0.44* | -0.40* | 0.45* | 0.58* | -0.65* | 0.39* | 0.10 | 0.59* | 0.27* | 1.00 | | | | |
| SC | -0.09 | 0.10 | -0.31* | 0.00 | -0.06 | -0.06 | -0.24* | -0.40* | 0.07 | 0.36* | 0.01 | 1.00 | | | |
| PC | 0.04 | -0.03 | 0.35* | -0.09 | -0.16 | 0.02 | -0.29* | -0.28* | -0.13 | 0.19* | -0.45* | -0.35 | 1.00 | | |
| PMfld | 0.39* | -0.27* | 0.39* | -0.31* | -0.40* | 0.40* | -0.30* | -0.09 | -0.42* | -0.21* | -0.53* | -0.06 | 0.26* | 1.00 | |
| PMdlA | 0.30* | -0.09 | 0.37* | -0.19* | -0.31* | 0.31* | -0.28* | -0.14 | -0.31* | -0.09 | -0.37 | -0.06 | 0.28* | 0.48* | 1.00 |

*Significant at P< 0.05

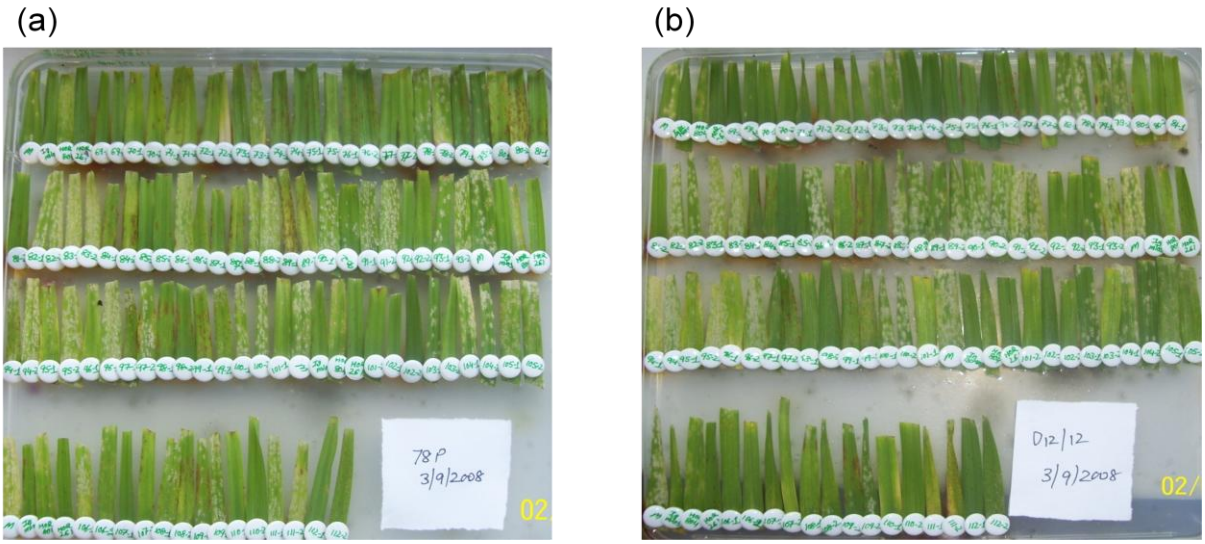



**Fig 3.6**: Visual symptoms of powdery mildew infection under detached leaf assay condition. (**a)** 78P and **(b)** D12/12 isolates infection on different genotypes of the gene pools. Different infection types are shown in the respective plates.

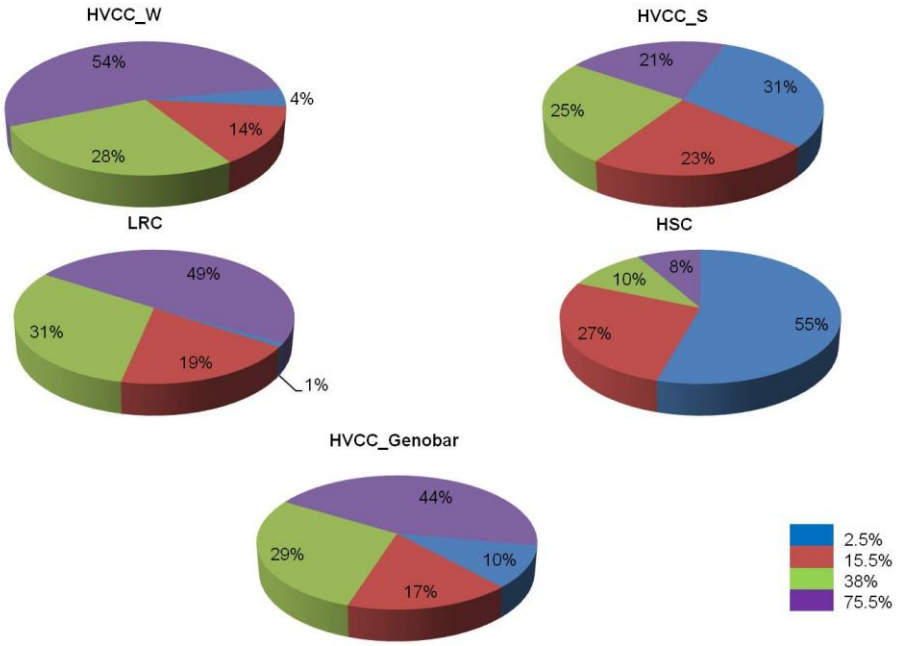



**Fig 3.7**: Phenotypic distribution of powdery mildew disease infection under detached leaf assay condition across the gene pools. The colors blue, red, green and purple (2.5, 15.5, 38 and 75.5) represents leaf infection classes. The pie charts shows percentages of genotypes in the respective classes in different gene pools; winter gene pools (HVCC_W), spring gene pool (HVCC_S), landraces gene pool (LRC), wild barley (HSC) and spring barley 224 genobar gene pool (HVCC_Genobar).

## 3.4 Genome-wide association analysis

### 3.4.1 GWA analysis of phenological traits with BOPA1 SNP markers

***Frost tolerance (Fr)***

Thirty four SNP markers were found significantly associated in GWAS with frost tolerance in winter cultivar gene pool (HVCC_W) (Sup. Table S5, Fig 3.8). SNPs with highly significant association to the trait were mapped at ~108cM on chromosome 5H. This region corresponds to the chromosomal region harbouring the *CBF* (*C-repeat binding factor*) gene family chromosomal region (Choi et al. 2002). Similarly the cernalization gene *VrnH2*, chromosomal region on chromosome 4H was also significantly associated (Yan et al. 2004). In addition significantly associated regions on chromosomes 1H long arm, 5H long arm, 6H short arm and 7H long arm were found.

***Heading date (Hd)***

GWAS for the heading date yielded twenty eight significantly associated SNPs in HVCC_W and twenty nine significantly associated SNPs in HVCC_S genepool (Sup. Table S6, Fig 3.9). No association was observed on chromosome 4H in HVCC_W gene pool and 1H and 6H in HVCC_S gene pool. Highly significant SNP markers in both gene pools ($-\log_{10} P = 6.4$ in HVCC_W and $-\log_{10} P = 5.323$ in HVCC_S) were from the 2H region near to the *eam6* locus (Comadran et al 2009). In HVCC_W, SNPs corresponding to the genetic positions of major flowering time genes were found to be significantly associated with Hd. For instance SNPs from 1H (~108cM) correspond to *HvFT3*, SNPs from 2H (~26cM) to *PpdH1*, SNPs from 2H (~52cM) to *eam6* and SNPs from 5H (~63cM) to *HvCo3* regions (Cuesta-Marcos et al. 2008; Faure et al. 2007; Griffiths et al. 2003; Turner et al. 2005). In HVCC_S, SNPs corresponding to *eam6*, *HvG1*, *HvFT1* and *HvCo1* regions were significantly associated to Hd [viz., 2H (~52cM) corresponds to *eam6*, 3H (~54cM) to *HvG1* 7H (~41.85cM) to *HvFT1* and 7H (~83cM) to *HvCo1*] (Cuesta-Marcos et al. 2008; Dunford et al. 2005; Griffiths et al. 2003; Kikuchi et al. 2009) (Sup. Table S6).

***Plant height (Ht)***

Forty nine and thirty one SNP markers were significantly associated with Ht in HVCC_W and HVCC_S gene pools, respectively (Sup. Table S7, Fig 3.10). In HVCC_W gene pool, no SNP markers were associated on chromosome 1H and highly significant SNP ($-\log_{10}$ P = 5.04) was near the *Sdw3* region on chromosome 2H (~ 63.53cM) (Gottwald et al. 2004). In HVCC_S gene pool, a highly significant SNP ($-\log_{10}$ P = 4.79) was observed on chromosome 3H (~126cM) near the *sdw1/denso* region (Jia et al. 2011).

***Spike length (Sl)***

In GWA scan, thirty five SNP markers for HVCC_W and forty one SNP markers for HVCC_S were significantly associated to spike length (Sup. Table S8, Fig 3.11). In HVCC_S gene pool, no association was observed on chromosome 6H, and the most significantly associated SNP was from 4H (51.3cM) ($-\log_{10}$ P = 5.8). However, in HVCC_W gene pool, highly significant association was observed near *eam6* on chromosome 2H (~53.53cM) (Cuesta-Marcos et al. 2008). Chromosomal regions on 2H (*Vrs1* ~ 82 cM) and 4H (*int-C* ~26.19cM) that have significance for row type were also found to be significantly associated with spike length trait (Komatsuda et al. 2007; Ramsay et al. 2011).

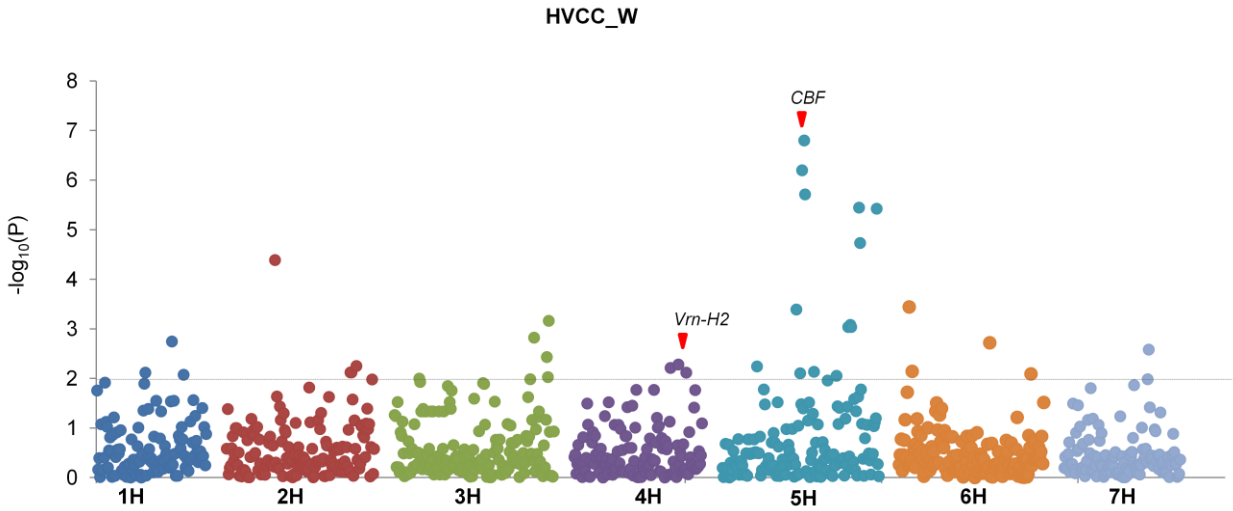



**Fig 3.8**: GWAS scan for the trait frost tolerance (Fr) in the HVCC_W gene pool. The X-axis shows seven chromosomes from short arm to long arm (left to right) and the Y-axis represents $-\log_{10}(P)$ values of the SNP markers. The horizontal grey dotted line marks the threshold level ($-\log_{10}(P) = 2$)

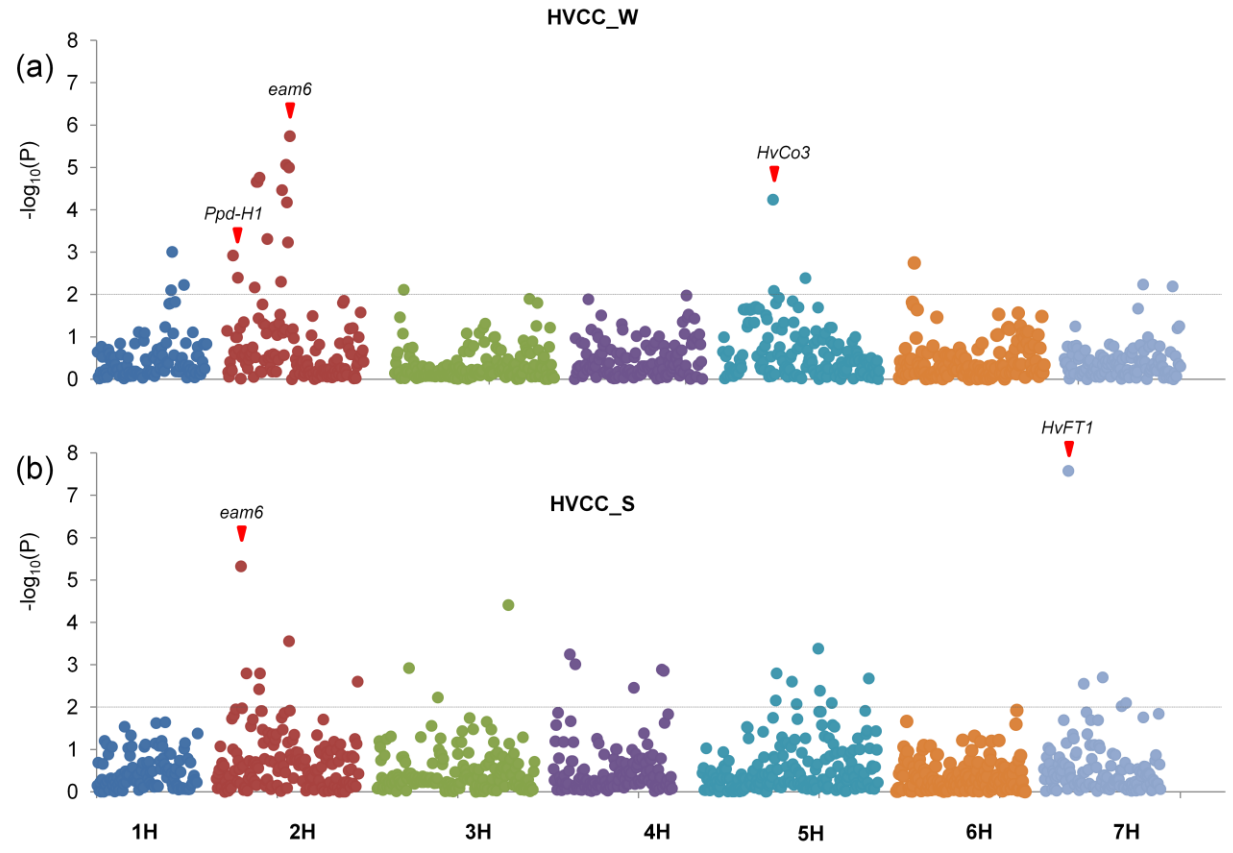



**Fig 3.9**: GWAS scan for the heading date (Hd) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

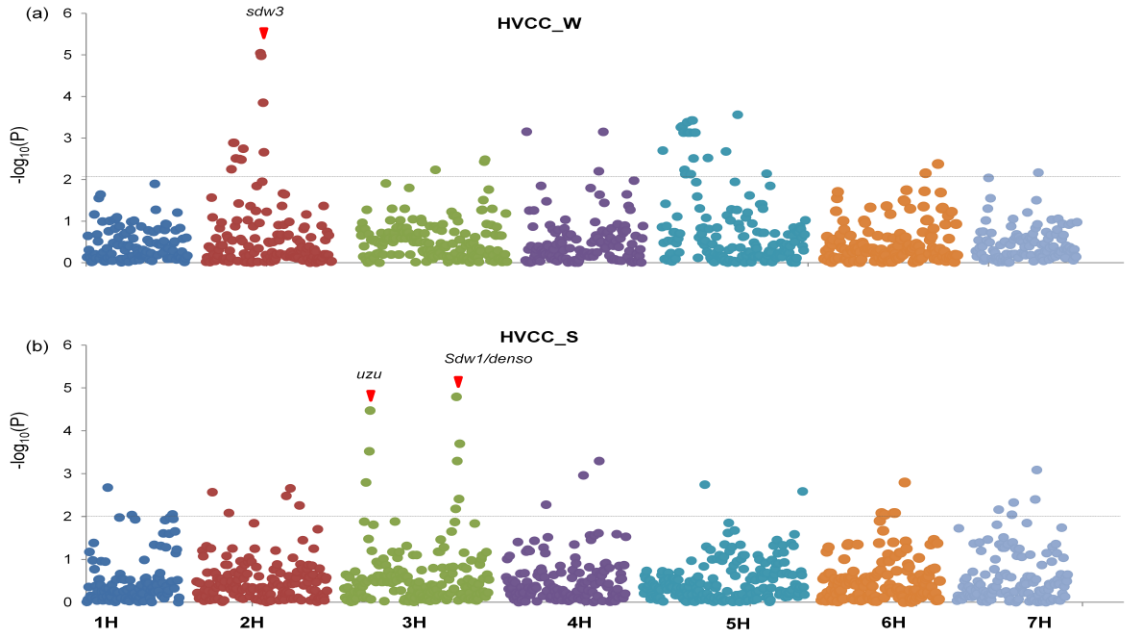



**Fig 3.10**: GWAS scan for plant height (Ht) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

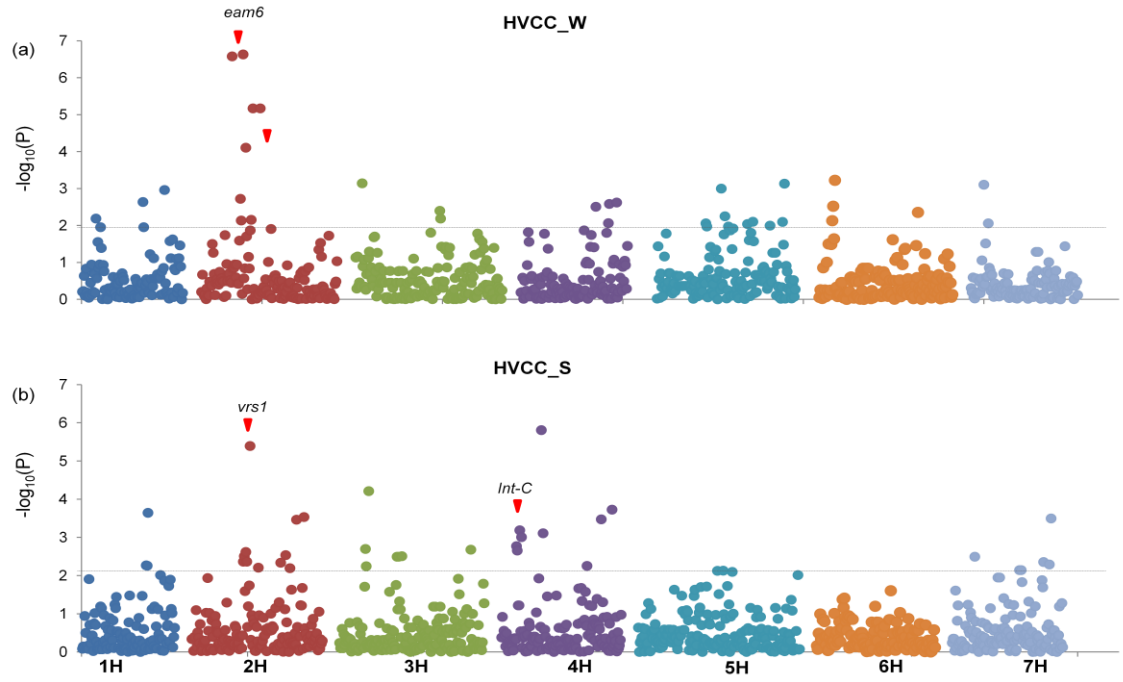



**Fig 3.11**: GWAS scan for spike length (Sl) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

### 3.4.2 GWA analysis of grain traits with BOPA1 SNP markers

***Barley spike morphology (Row)***

Two-rowed and six-rowed spike type is the major classification of the spike in addition some intermediate types are also observed in natural populations. Five independent loci controlling row-type spike phenotype in barley are known and located on chromosomes 1H (*Vrs3*), 2H (*Vrs1*), 3H (*Vrs4*), 4H (*Vrs5* also called *int-C*) and 5H (*Vrs2*). GWAS results have showed that 97 SNPs were significantly associated to row type trait in HVCC_S and 37 SNPs in HVCC_W gene pools. (Sup. Table S9, Fig 3.12). No association was observed on chromosome 5H in HVCC_W gene pool. Segregation of *Vrs3* (1H), *Vrs1* (2H) and *int-C* (4H) was observed in both spring and winter gene pools (Pourkheirandish and Komatsuda 2007). In addition *Vrs4* and *Vrs2* segregation was observed in HVCC_S gene pool. Most significant SNP markers association with significance of $-\log_{10} P = 96.86$ and 36.84 were near to *int-C* region on chromosome 4H in both gene pools. The other significant ($-\log_{10} P = 20.86$ and $-\log_{10} P = 21.20$) SNP markers regions was near to *Vrs1* chromosomal region on chromosome 2H (Komatsuda et al. 2007). Second highest significant SNP ($-\log_{10} P = 48$) in HVCC_S was from the 1H region near 96 cM. Several other significant associations were detected across the chromosomes that have significant effect on row type (Sup. Table S9, Fig 3.12).

***Grains / ear (GEar)***

GWA scan for the GEar trait detected thirty two and eighty five significantly associated SNP markers in HVCC_W and HVCC_S gene pools, respectively (Sup. Table S10, Fig 3.13). Highly significant associations in both gene pools were found at the *int-C* region on chromosome 4H (26.19cM) and *Vrs1* region on 2H (~82cM). Across the gene pools SNPs from chromosomal region on 1H at ~96cM had shown significant associations. No association was observed on chromosome 5H in HVCC_W gene pool.

***Thousand grain weight (TGW)***

Thirty one SNP markers in HVCC_W and sixty one SNP markers in HVCC_S gene pool were observed to be significant across the genome. Spike morphology controlling *Vrs1* and *int-C*

genomic regions on chromosome 2H and 4H were highly associated with TGW (Sup. Table S11, Fig 3.14). As expected chromosomal region on 1H (96.92cM) that showed significant associations with traits GEar and row type was also significant in GWAS for TGW. Moreover this region was reported to be associated for TGW in a previous study by Pasam et al. (2012).

***Grain area (Garea)***

Significant marker-trait associations (47 SNPs in HVCC_W and 51 SNPs in HVCC_S) for Garea were observed across the gene pools (Sup. Table S12, Fig 3.15). Intermedium spike morphology determining *int-C* genomic region was significantly associated with Garea in both gene pools. In HVCC_S, the *Vrs1* region and genomic regions on chromosome 1H, 4H, 5H, and 6H were highly associated with Garea in both gene pools (Sup. Table S12).

***Grain length (Glength)***

In total thirty five SNP markers in HVCC_W and fifty eight SNP markers in HVCC_S gene pools were significantly associated with Glength. Highly significant association with $-\log_{10}P$ = 3.73 in HVCC_W gene pool was from chromosome 3H at ~78cM and in HVCC_S gene pool with significance level $-\log_{10}P$ = 4.62 was from chromosome 4H (51.3cM), respectively (Sup. Table S13, Fig 3.16). Chromosome 4H genomic region ~51 cM was significantly associated to Glength in both the gene pools.

***Grain width (Gwidth)***

Across the chromosomes thirty six SNP markers in HVCC_W and forty seven SNP markers in HVCC_S were significantly associated to Gwidth trait. Row type gene *Vrs1* and *int-C* regions were significantly associated with Gwidth trait in both gene pools (Sup. Table S14, Fig 3.17). Chromosomal region ~63.53 cM on 2H is interesting for trait Gwidth as this region had significant associations in both the gene pools.

***Grain ratio width/length (W_l)***

In total twenty eight SNP markers in HVCC_W and forty two SNP markers in HVCC_S were associated at genome-wide level (Sup. Table S15, Fig 3.18). Genomic region 74.37cM on

chromosome 2H in HVCC_W gene pool was most significantly associated with W_l trait. This region was also significantly associated with GEar, Gwidth and row type traits (Sup. Table S33). No association was observed on chromosome 7H in HVCC_W gene pool. In the HVCC_S most significant association was from chromosome 4H (59.37) (Sup. Table S15, Fig 3.16). Approximately eight genomic regions were commonly significant across the gene pools (Sup. Table S15). Gwidth and Glength associations often coincide with the W_L associations in both gene pools (Sup. Table S36, S37).

***Yield (Yld)***

One of the most complex agronomic traits is grain yield. Across the chromosomes twenty five SNP markers in HVCC_W and forty nine SNP markers in HVCC_S gene pools were significantly associated. Highly significant SNP marker in HVCC_W gene pool was from chromosome 3H (169.32cM) (Sup. Table S16, Fig 3.19). This genomic region was also significantly associated with Garea and Fr traits in HVCC_W gene pool (Sup. Table S36). Although no known QTL for yield coincides at this genomic region so probably this region is new and needs further characterization in future. In HVCC_S gene pool chromosome 7H (84.92cM) was significantly associated with Yld (Sup. Table S16, Fig 3.17). Associations for agronomic traits like Ht, Hd, Garea, Glength, SC, PC were found significantly associated at this chromosomal region (Sup. Table S37). In the literature three QTL viz. *QHd.HaTr-7H*, *QHt.HaTr-7H* and *QHt.HaMo-7H* were reported for this chromosomal region (QTL workbook graingenes). Several genomic regions associated with other yield component traits (TGW, Ht, GEar, Row type, Hd) were also associated significantly with Yld (Sup. Table S36, S37).

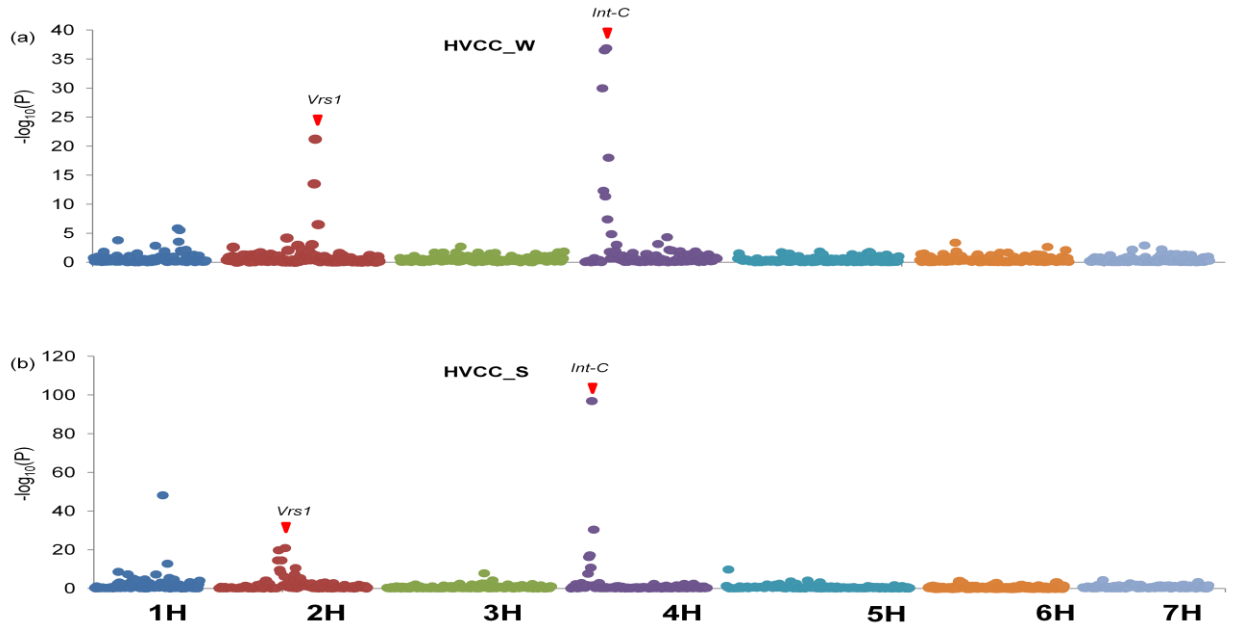



**Fig 3.12**: GWAS scan for row type in **(a)** HVCC_W and **(b)** HVCC_S gene pools

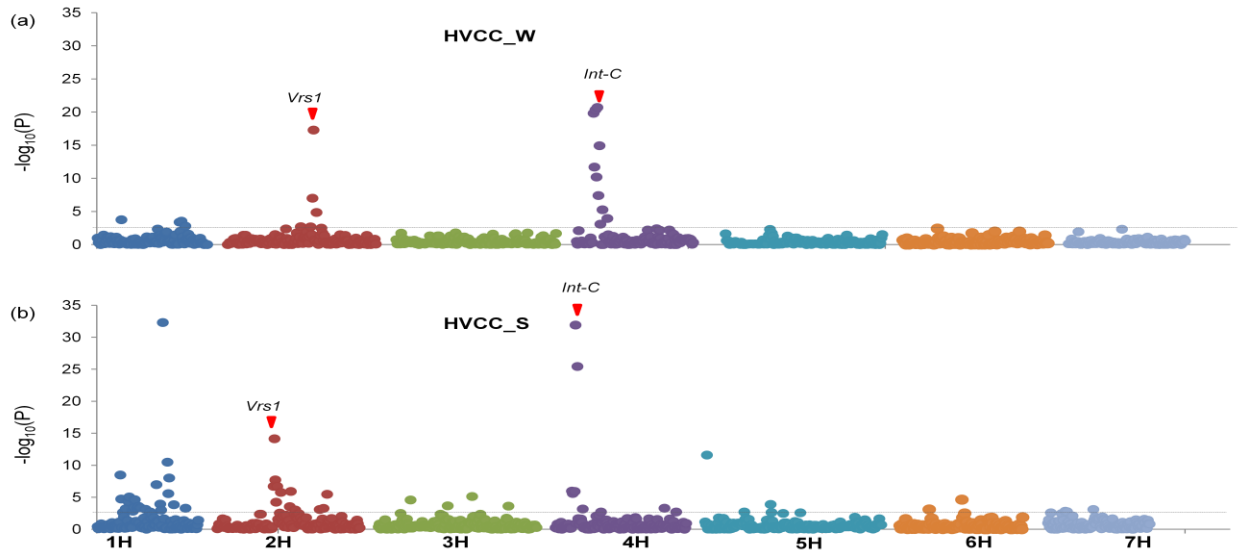



**Fig 3.13**: GWAS scan for Grains per ear (GEar) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

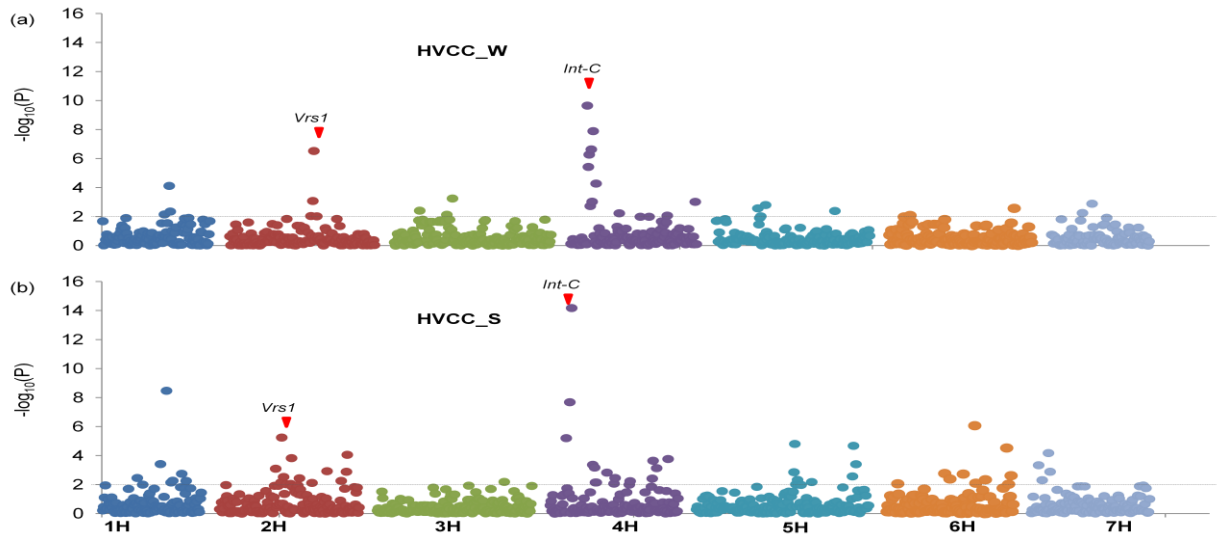



**Fig 3.14**: GWAS scan for thousand grain weight (TGW) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

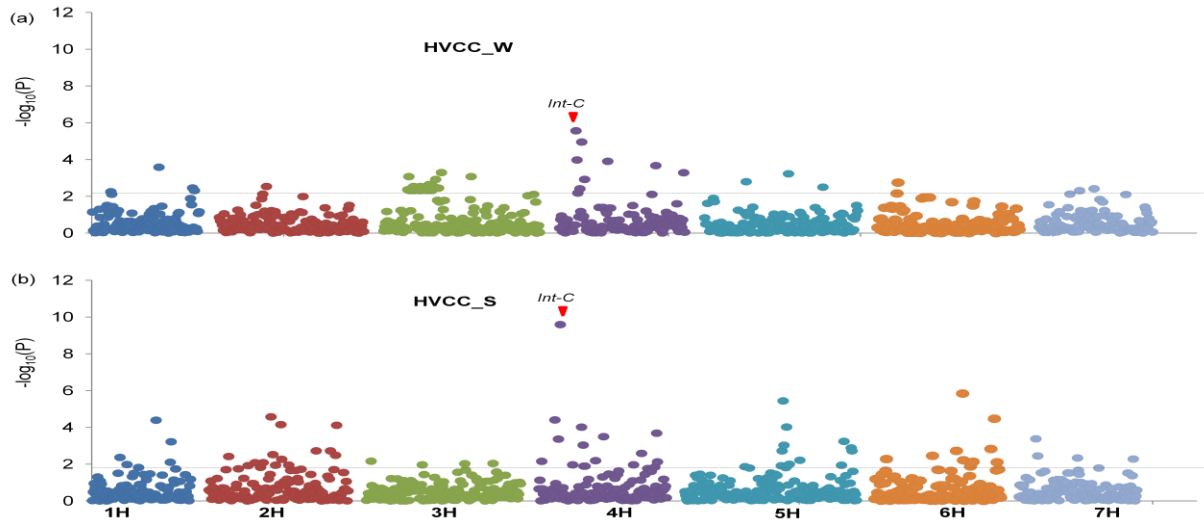



**Fig 3.15**: GWAS scan for grain area (Garea) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

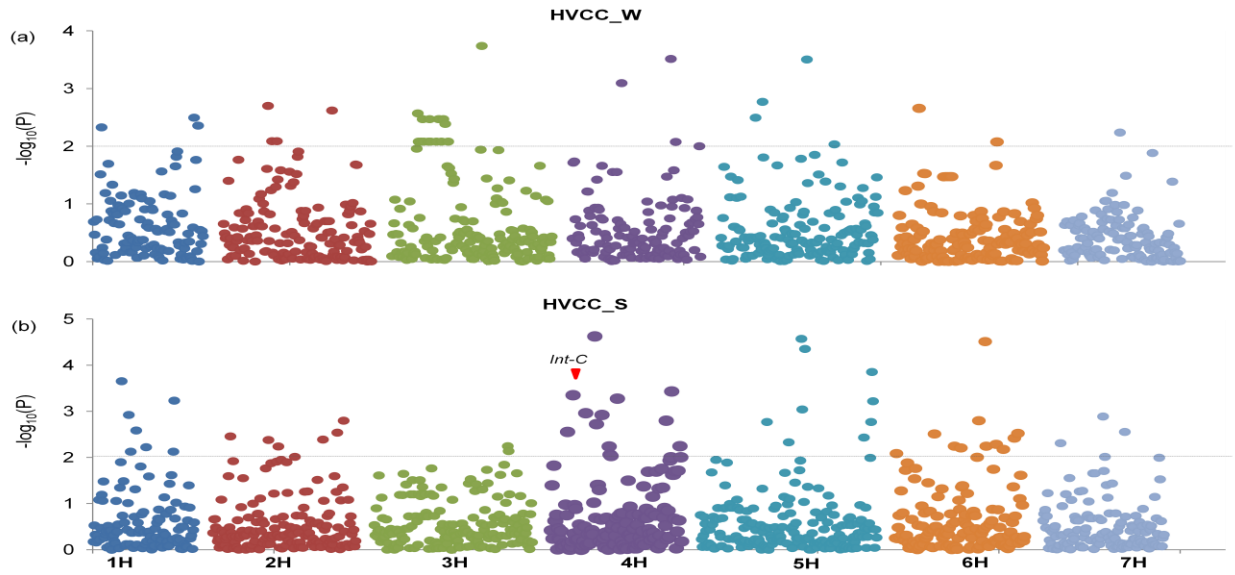



**Fig 3.16**: GWAS scan for grain length (Glength) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

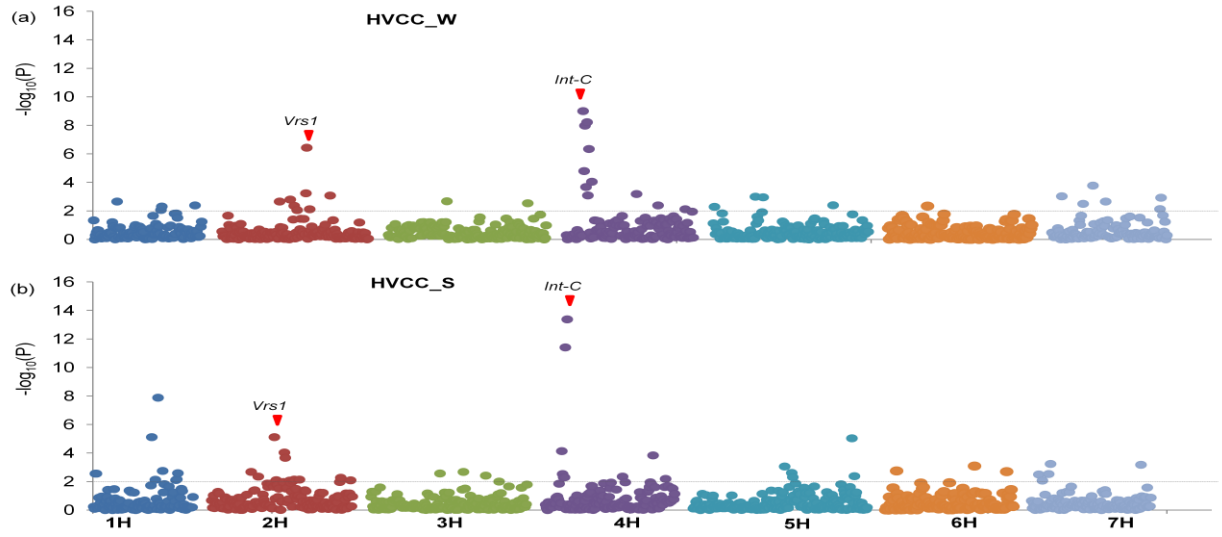



**Fig 3.17**: GWAS scan for grain width (Gwidth) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

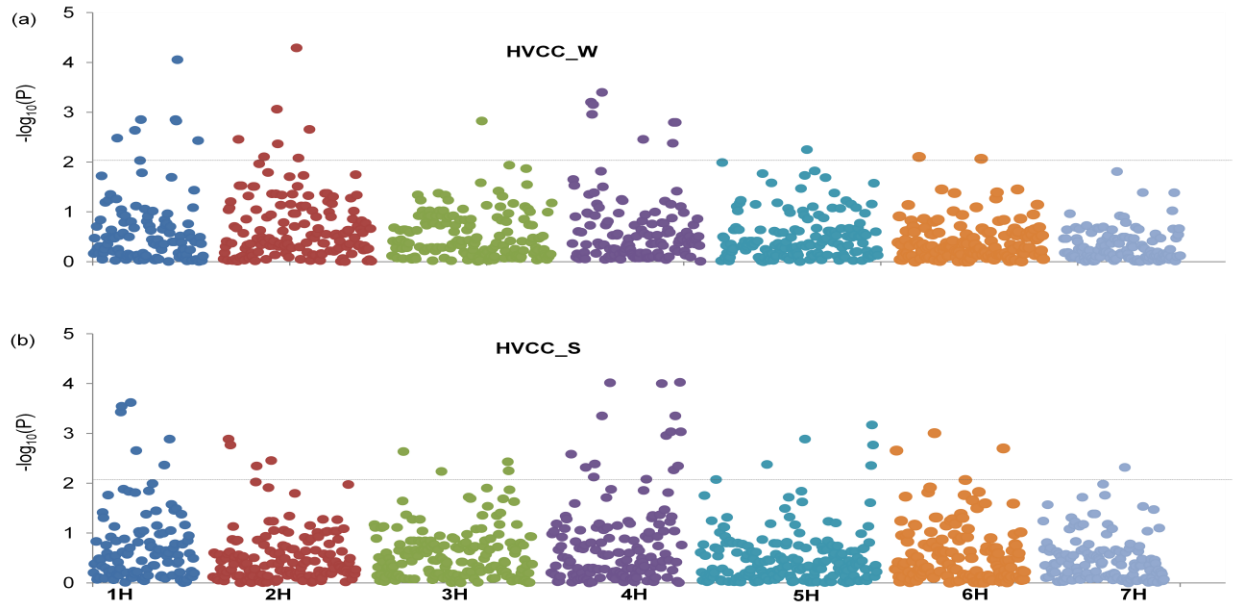



**Fig 3.18**: GWAS scan for ratio grain width/length (W_l) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

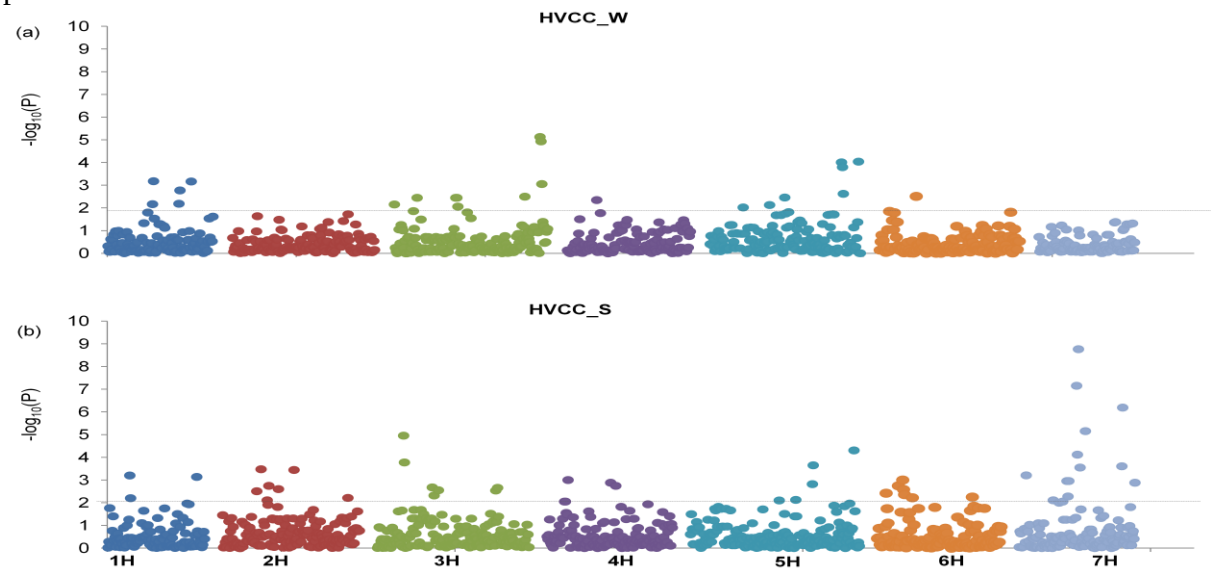



**Fig 3.19**: GWAS scan for grain yield (Yld) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

### 3.4.3 GWA analysis of grain quality traits with BOPA1 SNP markers

***Starch content (SC)***

Twelve SNP markers located on chromosome 2H, 4H, 5H and 7H in HVCC_W gene pool were significantly associated with SC. Thirty three SNPs were significantly associated in HVCC_S. Highly significant marker-trait associations were found on chromosome regions 7H (110.99cM) in HVCC_W, 7H (84.92cM) in HVCC_S and 3H (114cM) in LRC gene pool (Sup. Table S17, Fig 3.20). Only nine SNP markers were able to surpass the $-\log_{10}P = 2$ in LRC gene pool. These SNPs were located on chromosome 2H (3 SNPs), 3H (2 SNPs), 4H (1 SNP), 5H (2 SNPs) and 6H (1 SNP). Chromosome 1H and 3H did not have any association in HVCC_W gene pool. On the other hand chromosome 5H did not have any association in HVCC_S gene pool.

***Protein content (PC)***

Grain protein content is an important seed quality trait. Low PC is desirable in malting barley for beer production. In the GWA scan across the gene pools, thirty SNP markers in HVCC_W, thirty nine SNP markers in HVCC_S and thirteen SNP markers in LRC gene pools were significantly associated with PC (Sup. Table S18, Fig 3.21). These SNPs were distributed across all the chromosomes. In HVCC_W gene pool no SNP marker association was observed on chromosome 7H. As expected, genomic regions associated with PC were often significantly associated with SC (Sup. Table S36, S37).

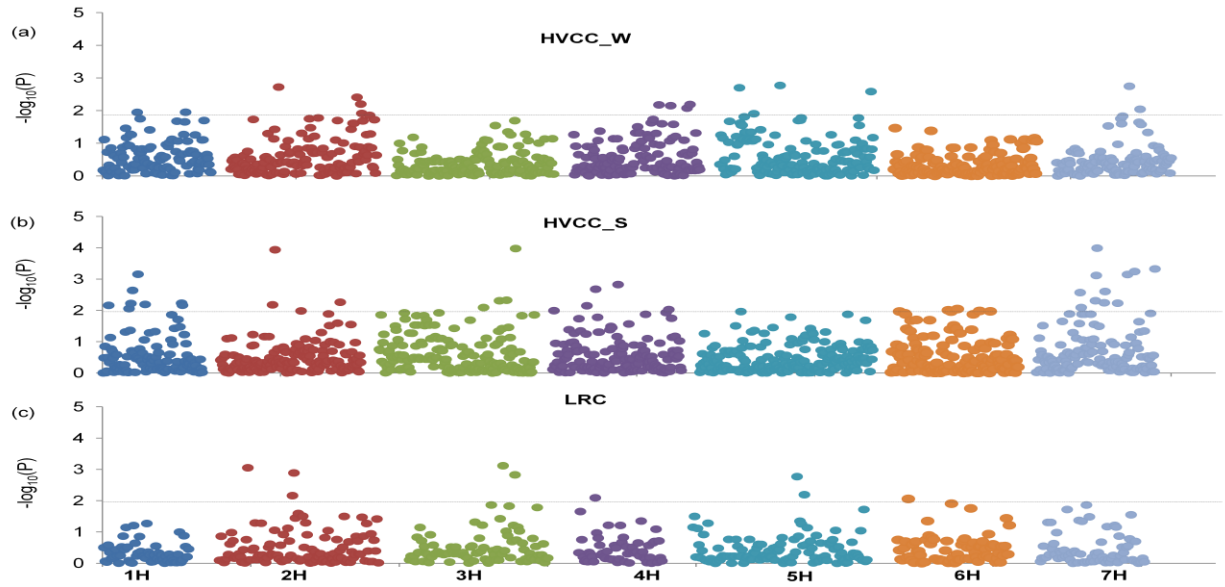



**Fig 3.20**: GWAS scan for the starch content (SC) in **(a)** winter barley (HVCC_W), **(b)** spring barley (HVCC_S) and **(c)** landraces (LRC) gene pools. The X-axis shows seven chromosomes from short arm to long arm and the Y-axis represents $-\log_{10}$ (P) values of the SNP markers. The horizontal grey dotted line shows the threshold level ($-\log_{10}$ (P) = 2)

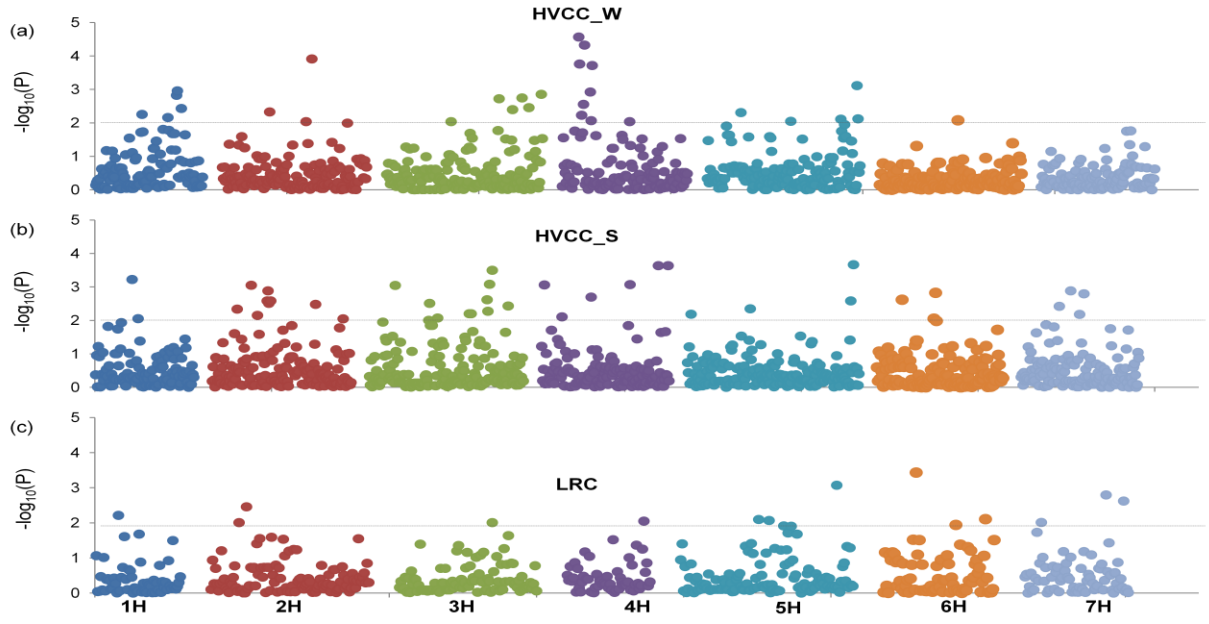



**Fig 3.21**: GWAS scan for the protein content (PC) in HVCC_W **(a)** and HVCC_S **(b)** and LRC **(c)** gene pools

### 3.4.4 GWA analysis of powdery mildew disease resistance with BOPA1 SNP markers

Powdery mildew (PM) is one of the most economically devastating diseases of barley. One of the major genes, *mlo* is known to confer non-race specific durable resistance in barley (Freialdenhoven et al. 1996). In order to characterize PM resistance independent of *mlo*, 38 HVCC_S barley genotypes and one accession BCC1411 from HVCC_Genobar were removed, as they were shown to carry *mlo* in their pedigree and the GWAS was performed excluding them (Sup. Table S40).

***Powdery mildew resistance under field condition (PMfld)***

In our GWA scan eighteen SNPs in HVCC_W and forty four SNPs in HVCC_S gene pools were significantly associated (Sup. Table S19, Fig 3.22) to PMfld. The most significant association was located on chromosome 5H (123cM) in HVCC_W and on 1H (3.75cM) in HVCC_S. Significant associations on short arm of chromosome 1H, around the resistance loci *Mla* was observed in HVCC_S gene pool. SNPs from chromosomal region near major genes *Mlhb* on 2H (26.53cM), *Mlh* on 6H (64.36cM) and *Mlf* on 7H (136.62cM) were significantly associated in HVCC_W gene pool. We also detected positive associations close to the major genes like *MlLa* on 2H (156.72cM), *Mlg* on 4H (65.05) and *Mlf* on 7H (136.62cM) (Comadran et al. 2009) (Sup. Table S19). Significant associations were observed both in winter and spring gene pools on chromosome 3H at 111.42 cM and 114 cM. Interestingly, a seedling resistance QTL *Rbgq2b* was located at this region (Aghnoum et al. 2009). Many significant marker trait associations were found across the chromosomes in our GWAS which are concurrent to positions of previously known disease resistance genes.

***Powdery mildew resistance under detached leaf assay condition (PMdla)***

Powdery mildew detached leaf assay was performed as an alternative approach to characterize the durable resistance of the *Bgh* infection. We identified seventeen SNP markers in HVCC_W, nineteen SNP markers in HVCC_S, seven SNP markers in LRC and nineteen SNP markers in HSC gene pools that were significantly associated across the genome (Sup. Table S19, Fig 3.23). In the HVCC_S gene pool and LRC gene pools, no SNP marker associations were observed on chromosome 1H and 6H. SNPs from chromosome 1H (88.23cM) in HVCC_W, 7H (77.85cM) in

HVCC_S, 2H (6.45cM) and 4HL (unmapped) in HSC gene pools were highly significant in GWAS (Sup. Table S19). Several genomic regions were significantly associated across the gene pools. Genomic region 2H (~50.49cM and ~137cM) in HVCC_S and HSC gene pools, 2H (~120cM), 6H (~65.03cM) in HVCC_W and HSC gene pools, 3H (~24cM), 7H (~74cM) in HVCC_S, LRC and HSC gene pools were found coinciding and significantly associated (Sup. Table S19). In addition, some of these significant associations were coinciding with the PMfld trait associations like SNP associations on genomic region 1H (88.23cM) was significant in the PMdla in HVCC_W and PMfld in HVCC_S gene pools (Sup. Table S19).

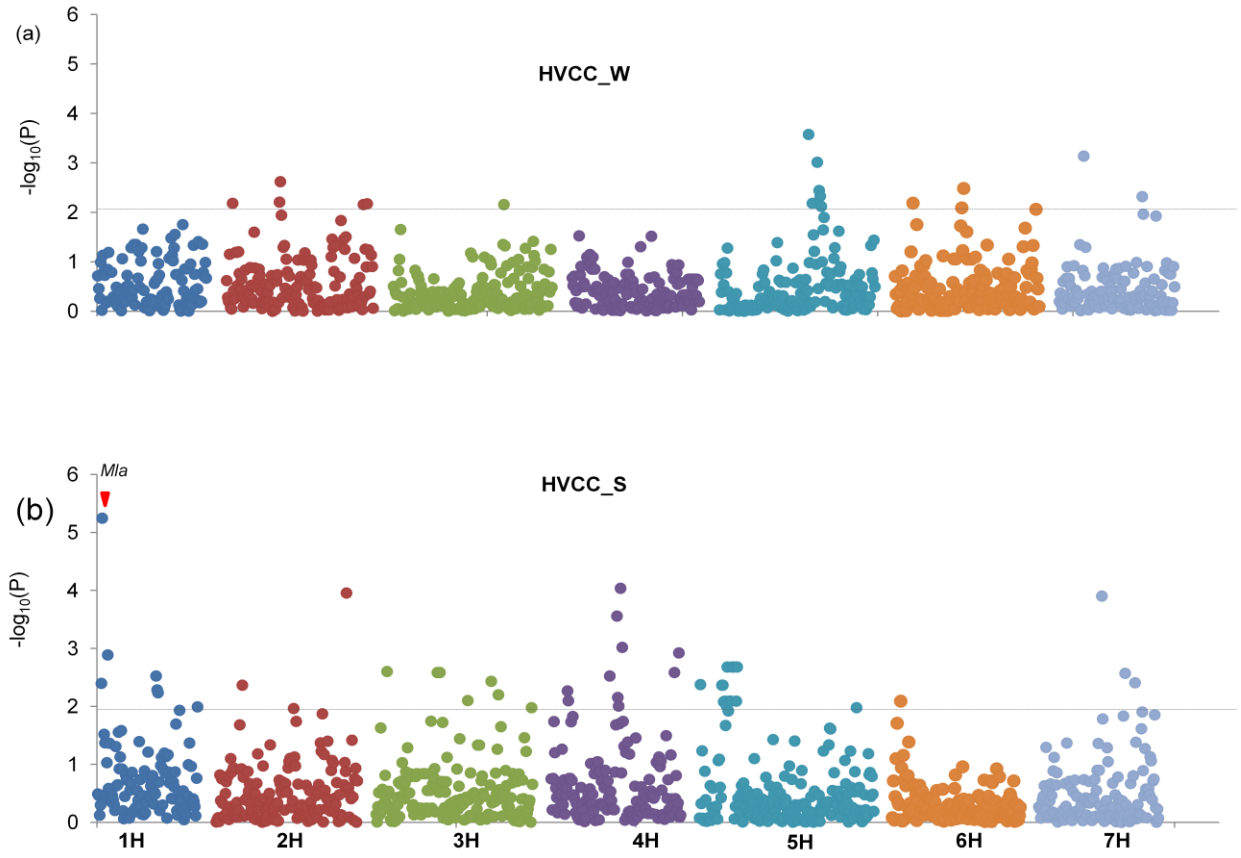



**Fig 3.22**: GWAS scan for the powdery mildew infection in field (PMfld) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

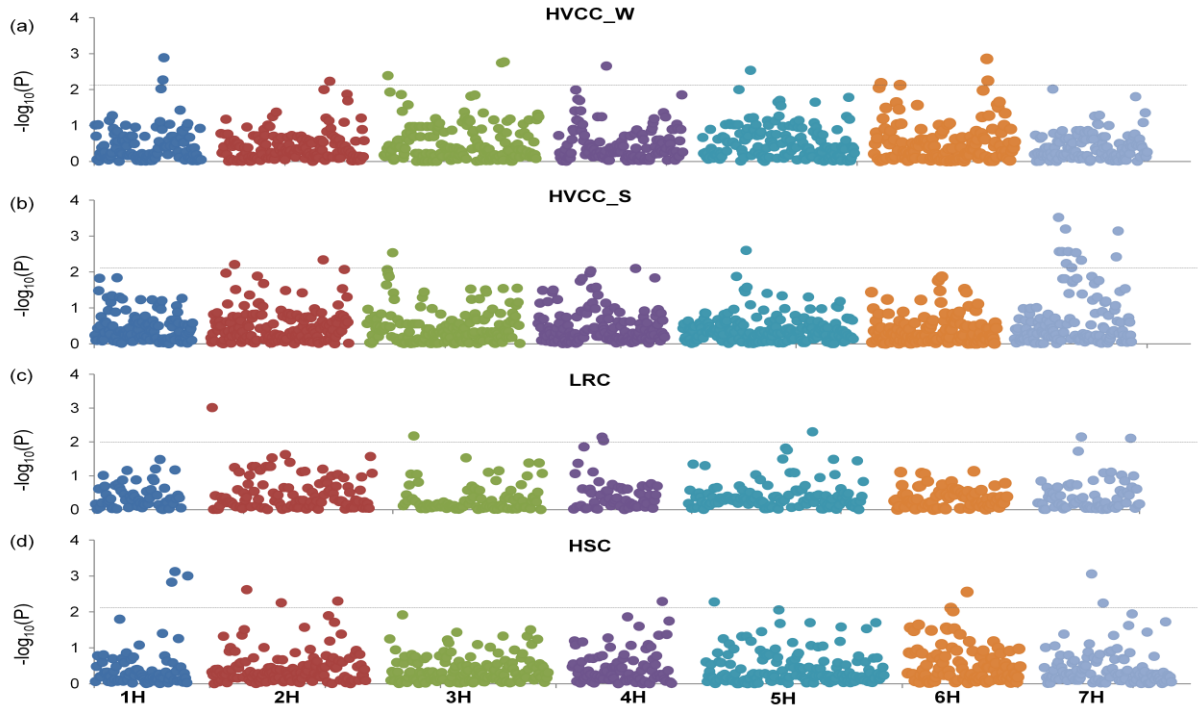



**Fig 3.23**: GWAS scan for the powdery mildew detached leaf assay (PMdla) in **(a)** winter barley HVCC_W, **(b)** spring barley HVCC_S, **(c)** landraces (LRC) and **(d)** wild barley (HSC) gene pools

### 3.4.5 Ascertainment bias of BOPA1 SNP markers

The design of the BOPA1 array has unfavorable effects on its performance. They were inefficient in landraces and wild barley . As shown in Fig 3.2, Minor Allele Frequencies (MAF) and Polymorphic Information Content (PIC) of the BOPA1 in HVCC, LRC and HSC were different across these populations. Around 19 percent of SNP markers in LRC and 12 percent in HSC have minor allele frequency (MAF) $\leq 0.05$ (Fig 3.2). The relative performance of the BOPA1 SNP markers in HVCC was considerably high. Most of the SNP markers were developed from cultivated genotypes and they performed well in cultivated (HVCC) gene pool. In comparison the BOPA1 SNPs performed badly in LRC and HSC gene pools. Relatively high MAF and PIC was observed in HSC then LRC gene pool. Nevertheless their performance in both LRC and HSC was bad in comparison to the cultivated barley populations.

This ascertainment bias of the SNP markers in LRC and HSC gene pool causes serious problems, while comparing the associations across the gene pool. In our study we removed SNP markers with ≤ 0.05 percent MAF as markers with low MAF causes spurious association. As a result many genomic regions were missed while comparing across the gene pools. In our association analysis, 984 SNP markers were used for HVCC_W, 1062 SNP markers for HVCC_S, 582 SNP markers for LRC and 813 SNP markers for HSC gene pools, respectively. Decrease in marker density becomes aggravated by the decreasing extant of LD in LRC and HSC. This difference of the marker numbers was a major obstacle while comparing the genomic regions associated with the traits across different gene pools. Recently, high resolution genetic platform (iSelect SNP chip) was developed in barley that has ~7000 SNP markers. With multifold increase in marker coverage improvement in ascertainment issues was expected. For an example, GWAS for the grain yield successfully un-raveled high number of SNP associations using iSelect chip in comparison to the BOPA1 SNP markers (Fig 3.24). In general, with high marker coverage several genetic positions were saturated with markers and when the region is significant all these markers become highly significant. Thus increasing marker coverage not only increased the significance level but several new regions were identified as an example Fig 3.24.

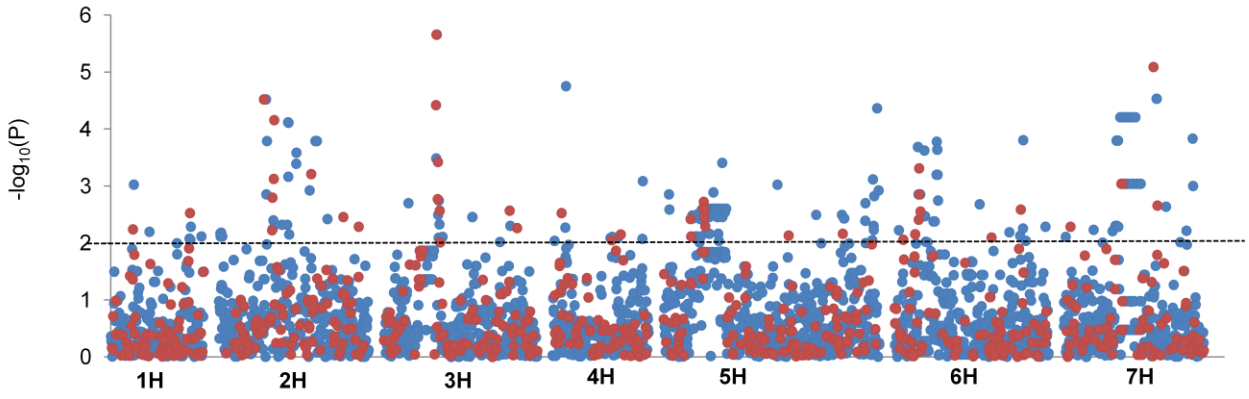



**Fig 3.24**: GWA scan for grain yield (Yld) in the spring barley cultivar panel. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes (only markers mapped in Morex X Barke RIL map are shown)

### 3.4.6 Genome-wide association studies of phenological traits with iSelect SNP markers

***Frost tolerance (Fr)***

In HVCC_W gene pool, 206 SNP markers across the genome were significantly associated with Fr (Sup. Table S20, Fig 3.25). SNPs with significantly high association to Fr were observed on chromosomal regions of 1H (~139.8cM) (mapped using LD with BOPA, Close et al. 2009, not shown in Fig 3.25) followed by 5H (99cM). SNPs corresponding to the vernalization gene regions on 4H (*VrnH2*) and on 5H long arm (*VrnH1*) were also found to be significantly associated with frost tolerance (Yan et al. 2004) (Sup. Table S20).

***Heading date (Hd)***

GWAS using iSelect SNPs for Hd revealed 112 and 282 SNPs to be significantly associated in HVCC_W and HVCC_S gene pools, respectively (Sup. Table S21, Fig 3.26). Highly significant SNP markers (Bk_12, BK_14, Bk_15 and Bk_16) were detected at 26.6cM on chromosome 2H in HVCC_W gene pool. These markers in iSelect were derived from major heading date gene *PpdH1* gene (Turner et al. 2005). In HVCC_S gene pool SNP from chromosomal region 7H was associated with high significance. This genomic region was co-localized with *HvFT*1 region (Faure et al. 2007). In both gene pools significant SNP markers associations were observed on 2H near *eam6* locus (~50cM) (Cuesta-Marcos et al. 2008).

***Plant height (Ht)***

Genome-wide scan for Ht using iSelect SNP markers yielded significant marker-trait associations across the gene pools (266 SNPs in HVCC_W and 159 SNPs in HVCC_S) (Sup. Table S22, Fig 3.27). Genomic region near *Sdw3* on chromosome 2H (57.5cM) significantly associated with Ht in HVCC_W gene pool (Gottwald et al. 2004). In the HVCC_S gene pool, highly significant marker-trait associations were observed near the *sdw1/denso* on chromosome 3H (111.1cM) (Sup. Table S22) (Jia et al. 2011). With BOPA1 SNPs also we observed that SNPs close to *Sdw3* were highly significant in HVCC_W and SNPs close to *sdw1* were highly significant in HVCC_S. Chromosome regions 2H (~39.1cM and ~70.50cM), 4H (~48.50cM and

~103.10cM), 5H (~18.0cM, ~66 cM, ~113cM, ~123cM and ~145 cM) and 7H (~88cM) were found significantly associated across the HVCC_W and HVCC_S gene pools.

***Spike length (Sl)***

Two hundred seven SNP markers in HVCC_W and one hundred thirty two SNP markers in HVCC_S were significantly associated in GWAS for Sl trait (Sup. Table S23, Fig 3.28). Highly significant association was near to *eam6* genomic region on 2H (~59cM) in HVCC_W gene pool (Cuesta-Marcos et al. 2008). Chromosomal region 7H (79.6cM) in HVCC_S gene pool was highly significant with $-\log_{10}P = 4.92$ (mapped using LD with BOPA Close et al. (2009) not shown in Fig 3.28). Several genomic regions were significantly associated across the gene pools viz. 1H (~136cM), 2H (~2cM), 3H (~114cM, ~124cM and ~142cM), 4H (~96cM and ~101cM), 5H (~84cM, ~113cM, ~145cM and ~160cM), 6H (~101cM) and 7H (~26cM) (Sup. Table S23).

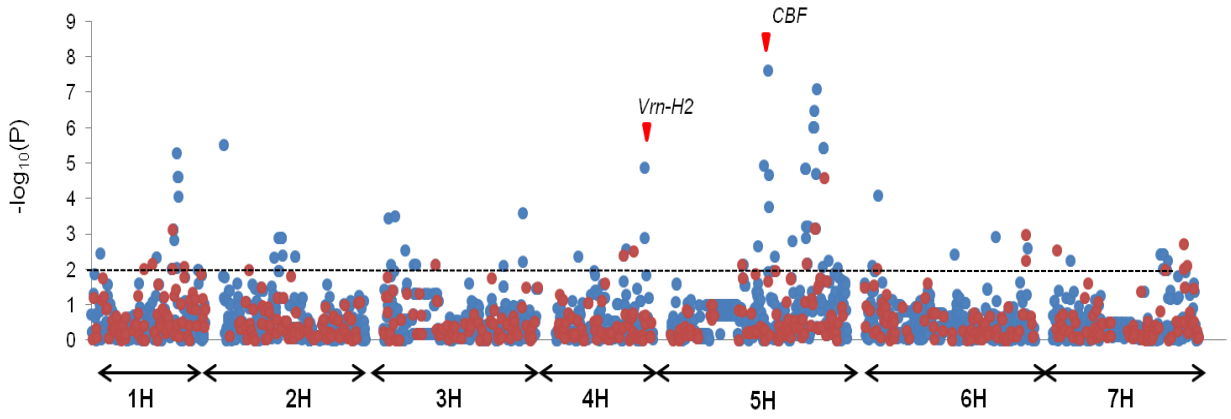



**Fig 3.25:** GWAS scan for the trait frost tolerance (Fr) in the HVCC_W gene pool. The X-axis shows seven chromosomes from short arm to long arm (left to right) and the Y-axis represents $-\log_{10}$ (P) values of the SNP markers. The horizontal dotted line marks the threshold level ($-\log_{10}$ (P) = 2). Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes (only markers mapped in Morex X Barke RIL map are shown)

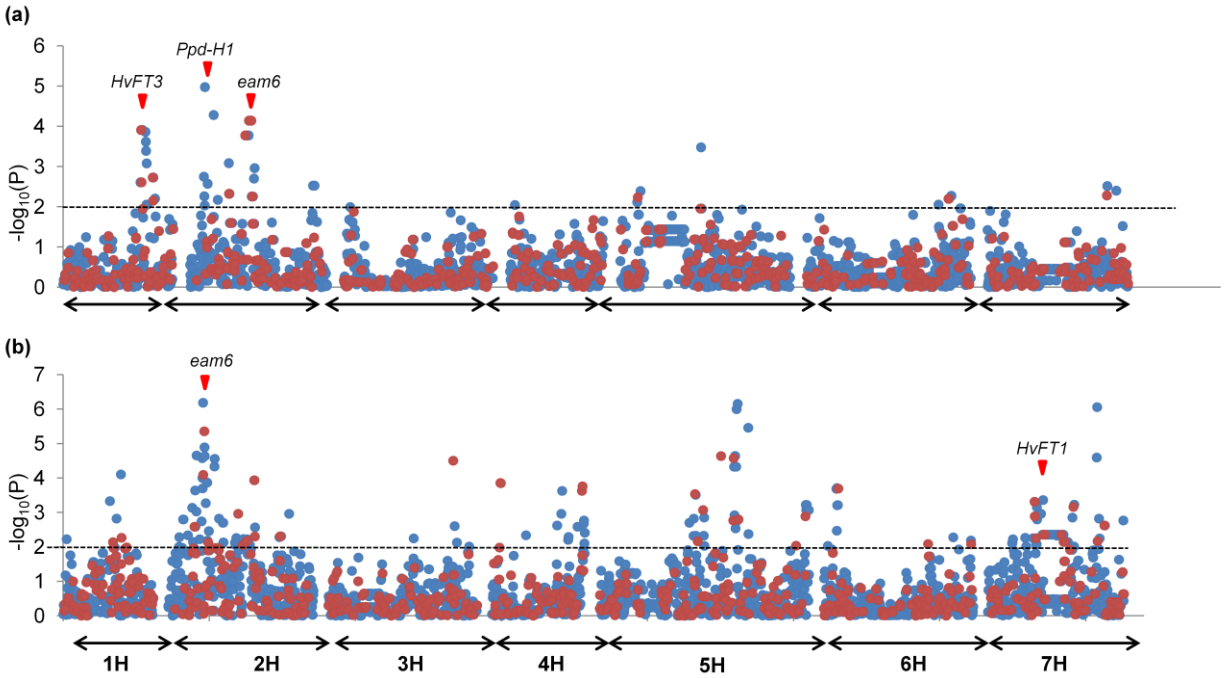



**Fig 3.26**: GWAS scan for the Heading date (Hd) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes (only markers mapped in Morex X Barke RIL map are shown)

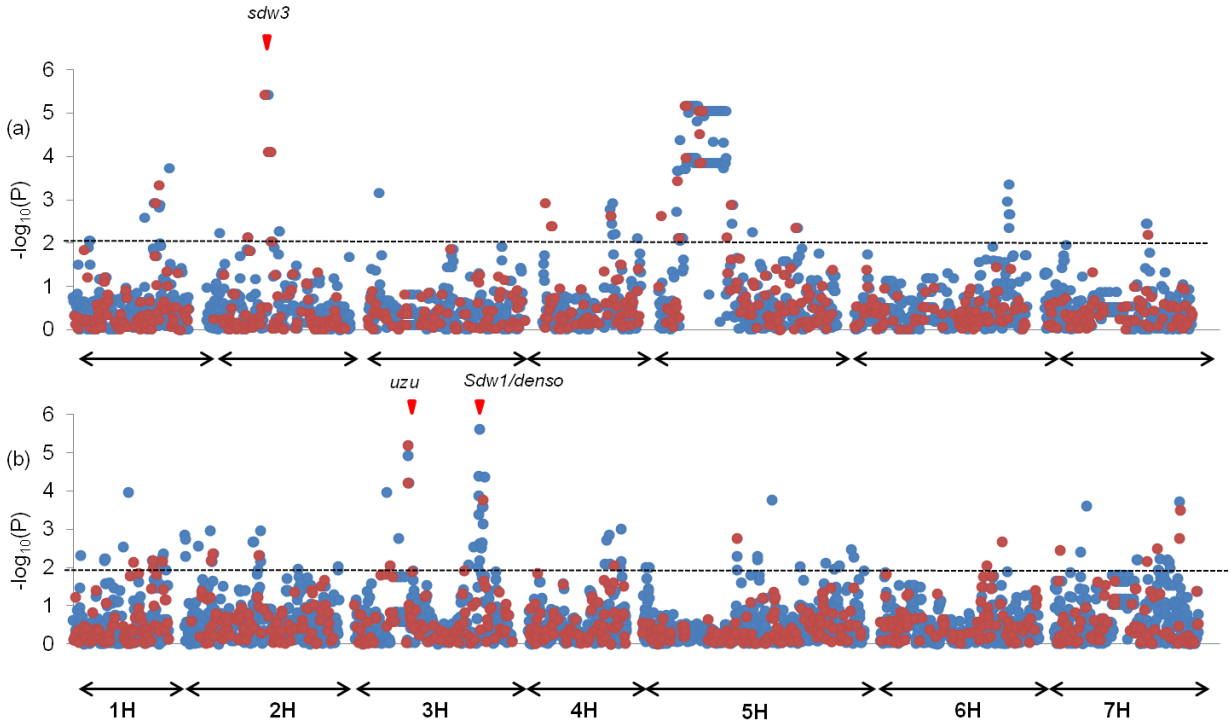



**Fig 3.27**: GWAS scan for the Plant height (Ht) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes (only markers mapped in Morex X Barke RIL map are shown)

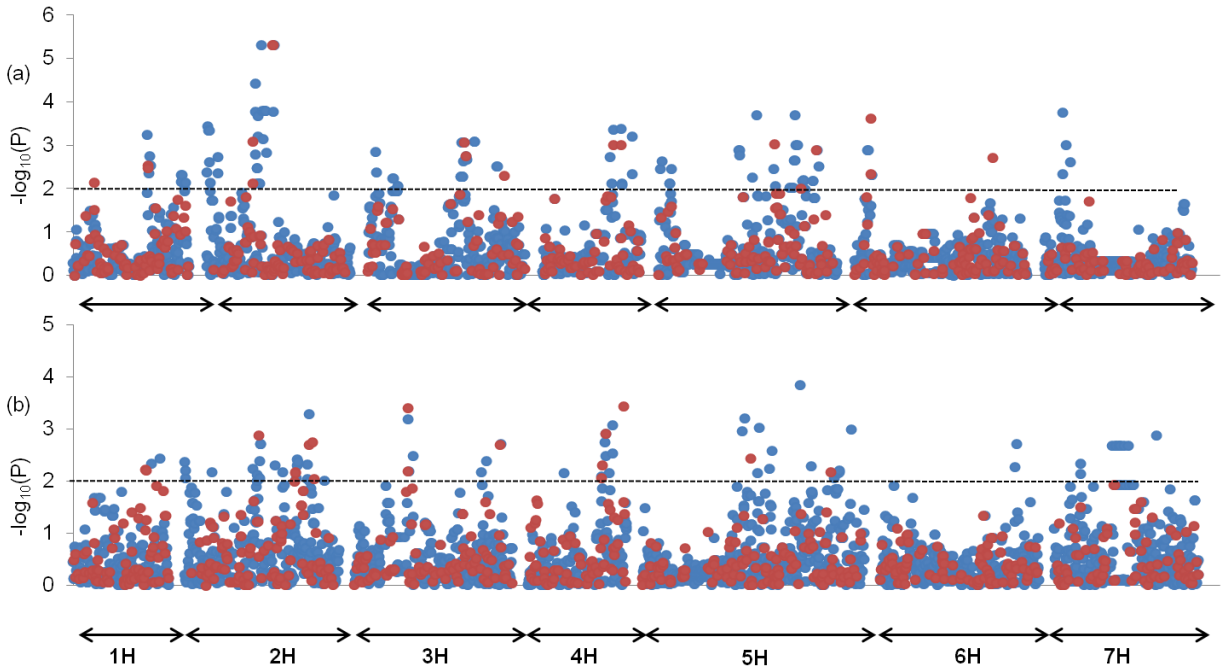



**Fig 3.28**: GWAS scan for the Spike length (Sl) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes (only markers mapped in Morex X Barke RIL map are shown)

### 3.4.7 Genome-wide association studies of grain traits with iSelect SNP markers

***Barley spike morphology (Row)***

GWA scan for row type showed 383 SNPs in HVCC_W and 349 SNPs in HVCC_S were significantly associated with the row type trait (Sup. Table S24, Fig 3.29). SNP with significantly high association to row type was on chromosome 4H (28cM) near the *int-C* region in HVCC_W gene pool. In HVCC_S gene pool, SNPs with significantly high association to row type were observed on 1H (~94cM) and on 4H (28cM) near *int-C* region. Segregation of *Vrs1* (2H), *Vrs2* (5H), *Vrs3* (1H), *Vrs4* (3H) and *int-C* (4H) genes on corresponding chromosomal regions were observed in both gene pools (Pourkheirandish and Komatsuda 2007). Highly significant associations were co-localized near *int-C* and *Vrs1* chromosomal regions. Significant association of SNP with row type on chromosome 1H (~94cM) in HVCC_S gene pool is interesting as no

gene controlling row type is characterized in this region. Several other SNPs with significant association to row type were consistently present across the gene pools (Sup. Table S24, Fig 3.29). These could be due to the genes with small effect that modify the row type phenotype.

***Grains / ear (GEar)***

In total, 223 SNP markers in HVCC_W were significantly associated to trait GEar. GWA scan in HVCC_S revealed 206 significant marker-trait associations for GEar (Sup. Table S25, Fig 3.30). Highly significant SNP marker association was observed near to *int-C* region on 4H (~28cM) in HVCC_W gene pool (Sup. Table S25). In HVCC_S gene pool, iSelect SNP located on chromosomal region 1H (94.3cM) was highly significant (Sup. Table S25). In addition, several other genomic regions associated to row type trait were also significantly associated to the trait GEar in both winter and spring gene pools (Sup. Table S25).

***Thousand grain weight (TGW)***

GWA scan with iSelect SNPs for TGW had found significant associations with 242 SNPs in HVCC_W and 216 SNPs in HVCC_S gene pools. Highly significant association in both winter and spring gene pools was close to *int-C* region on 4H (~29.7cM) (Sup. Table S26, Fig 3.31). Several regions across chromosomes were significantly associated in both the gene pools which provide an additional validation for these regions (Sup. Table S26, Fig 3.31). Chromosomal regions on 1H (~97cM, ~112cM and ~123cM), 2H (~22cM, ~26cM, ~74cM, ~80cM and ~112cM), 3H (~73cM and ~135cM), 4H (~28cM, ~51.30cM, ~76cM and ~103cM), 5H (~49.50cM), 6H (~55.0cM, ~60.20cM and ~72.50cM) and 7H (~24cM, ~74cM and ~119cM ) were significantly associated across the gene pools (Sup. Table S23).

***Grain area (Garea)***

Significant associations (245 SNPs in HVCC_W and 253 SNPs in HVCC_S) for Garea were observed across the genome in both gene pools (Sup. Table S27, Fig 3.32). Highly significant associations were observed with iSelect SNP 11_20777 (4H; ~29.7cM) in both gene pools. Several associated regions were significant across both the gene pools like 2H (~56.09cM,

~77cM and ~139cM), 3H (~56.65cM, ~63cM, ~95cM and ~155cM), 5H (~31cM, ~49cM, ~147cM and ~173cM), 6H (~56.0cM) and 7H (~76cM and ~111cM) (Sup. Table S27, Fig 3.32).

***Grain length (Glength)***

Across the genome, 114 and 307 SNPs were significantly associated to Glength trait in HVCC_W and HVCC_S gene pools, respectively. On chromosome 6H at ~97.7cM highly significant associations for Glength were observed in HVCC_W gene pool (Sup. Table S28, Fig 3.33). In HVCC_S highly significant SNP association was detected near rowed type gene *int-C* on 4H (29.7cM). Many genomic regions were significantly associated to trait Glength across the gene pools viz. 1H (~50cM), 2H (~58.90cM, and 130cM), 3H (83cM, ~95cM and ~104cM), 4H (~61cM, ~112cM), 5H (~46cM, ~129cM, ~145cM, ~176cM), 6H (~53cM, ~97cM) and 7H (~34cM and ~74cM) (Sup. Table S28).

***Grain width (Gwidth)***

Genome-wide significant associations for the trait Gwidth (329 SNPs HVCC_W and 184 SNPs in HVCC_S) were observed across the gene pools (Sup. Table S29, Fig 3.33). SNPs close to spike morphology determinant row-type gene *int-C* was most significantly associated with Gwidth trait. Significant marker-trait associations were observed across winter and spring gene pools viz. 1H (~5.09cM and ~86cM), 2H (~22.93cM, ~68cM, ~74cM, ~79cM, ~98cM and~139cM), 3H (~61cM and ~73cM), 4H (~28.66cM and ~69.50cM), 5H (~4.0cM, ~140cM and ~172cM), 6H (~54cM and ~61cM) and 7H (~64.36cM and ~144cM) , respectively.

***Grain ratio width/length (W_l)***

In total, 163 SNP markers in HVCC_W and 172 SNP markers in HVCC_S were significant in GWAS using iSelect SNPs for W_L trait (Sup. Table S30, Fig 3.34). Chromosome region 138.8cM on 2H and 88.2cM on 1H were highly significant in HVCC_W and HVCC_S gene pools. As expected, several chromosomal regions that were significant in GWAS for traits Gwidth and Glength were also significant with ratio of grain Width/length trait (Sup. Table S38, S39, Fig 3.34). Many genomic regions were significantly associated in both the gene pools like 1H (~52.61cM and ~108cM), 2H (~55cM, ~66cM, ~112cM and ~145cM), 3H (~83cM), 4H

(~28cM, ~43cM, ~51cM and ~112cM), 5H (~151cM) and 6H (~52.23cM and ~74.50cM) , respectively.

***Yield (Yld)***

Significant associations were found in HVCC_W (119 SNPs) and HVCC_S (105 SNPs) gene pools for grain yield trait across the genome (Sup. Table S31, Fig 3.35). Highly significant associations were on chromosome 5H (~99cM) in HVCC_W and on 3H (~55cM) in HVCC_S gene pools. Many significant SNP marker associated regions were coinciding in both the gene pools like 1H (~53.91cM), 2H (~2.10cM and ~100cM), 3H (~55cM, ~59cM, ~136cM), 4H (~115cM), 5H (~48.99cM, ~144.80cM and ~158cM), 6H (~81.90cM) and 7H (~0.0cM, ~46.65, ~78cM, ~115cM and ~125cM) , respectively (Sup. Table S31, Fig 3.35).

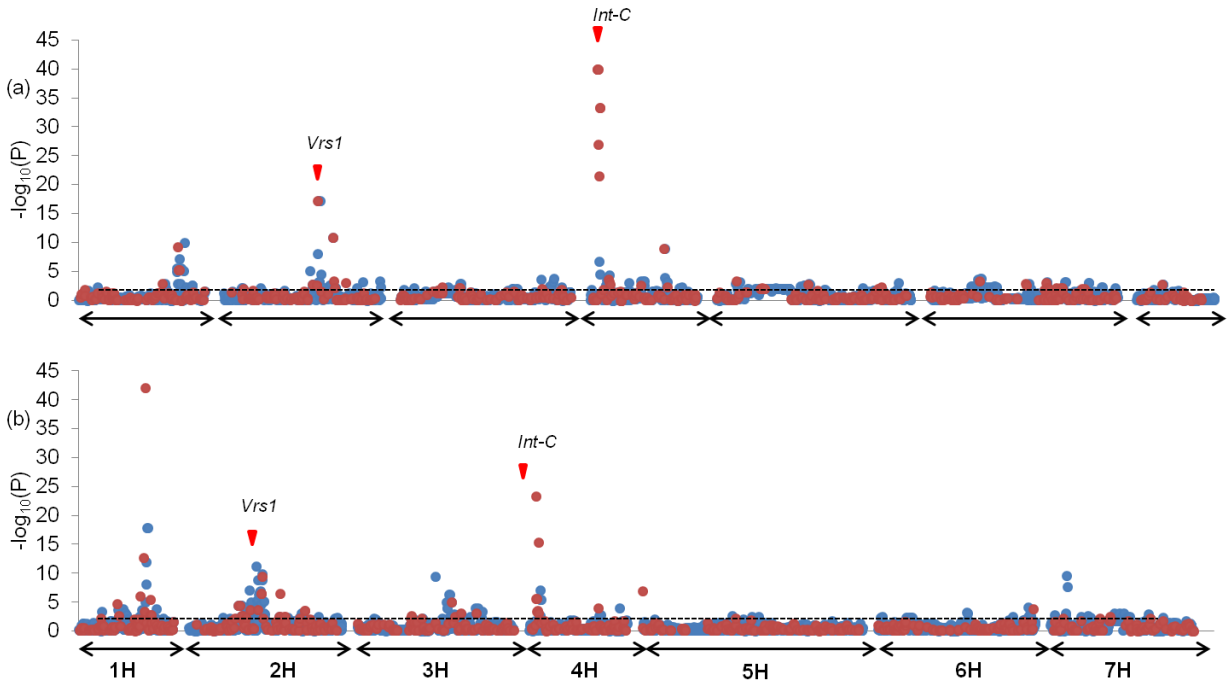



**Fig 3.29**: GWAS scan for the row-type (Row) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes (only markers mapped in Morex X Barke RIL map are shown)

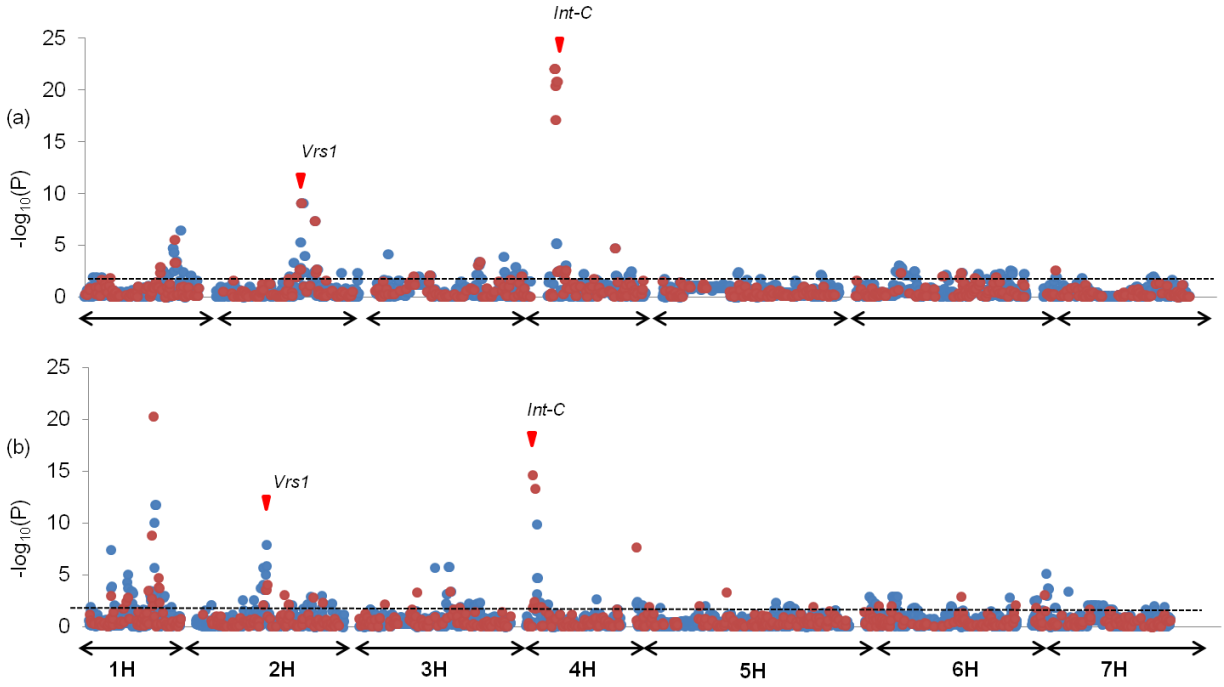



**Fig 3.30**: GWAS scan for the Grains per ear (GEar) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes

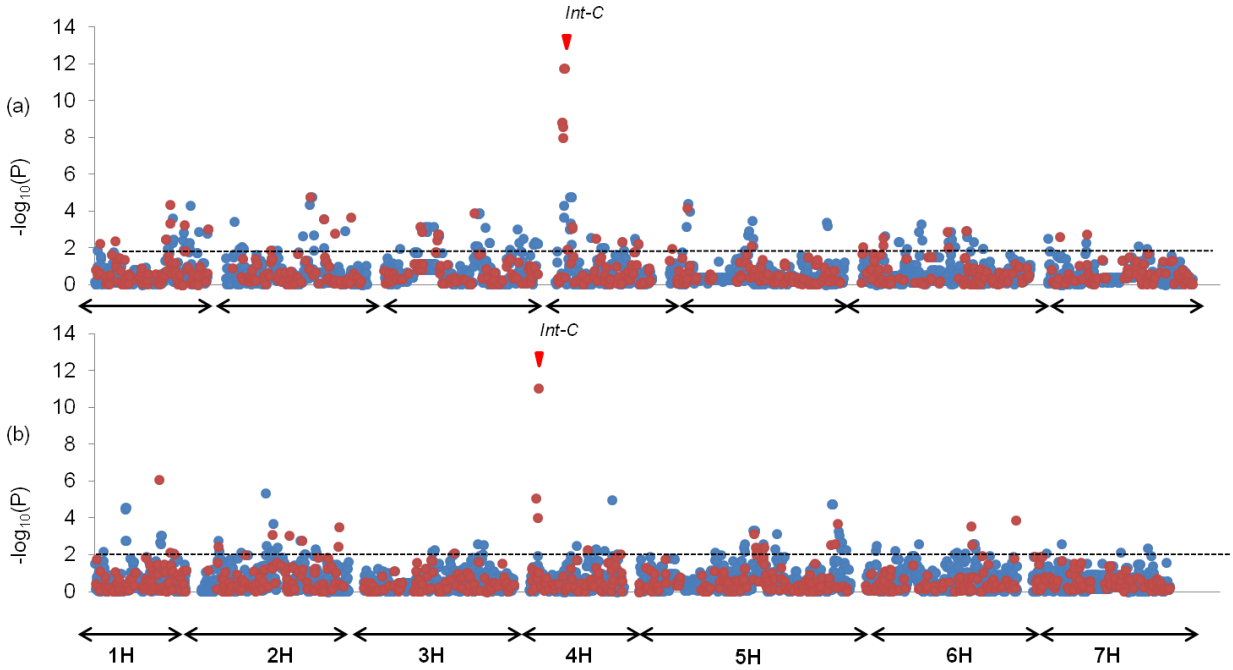



**Fig 3.31**: GWAS scan for the thousand grain weight (TGW) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes

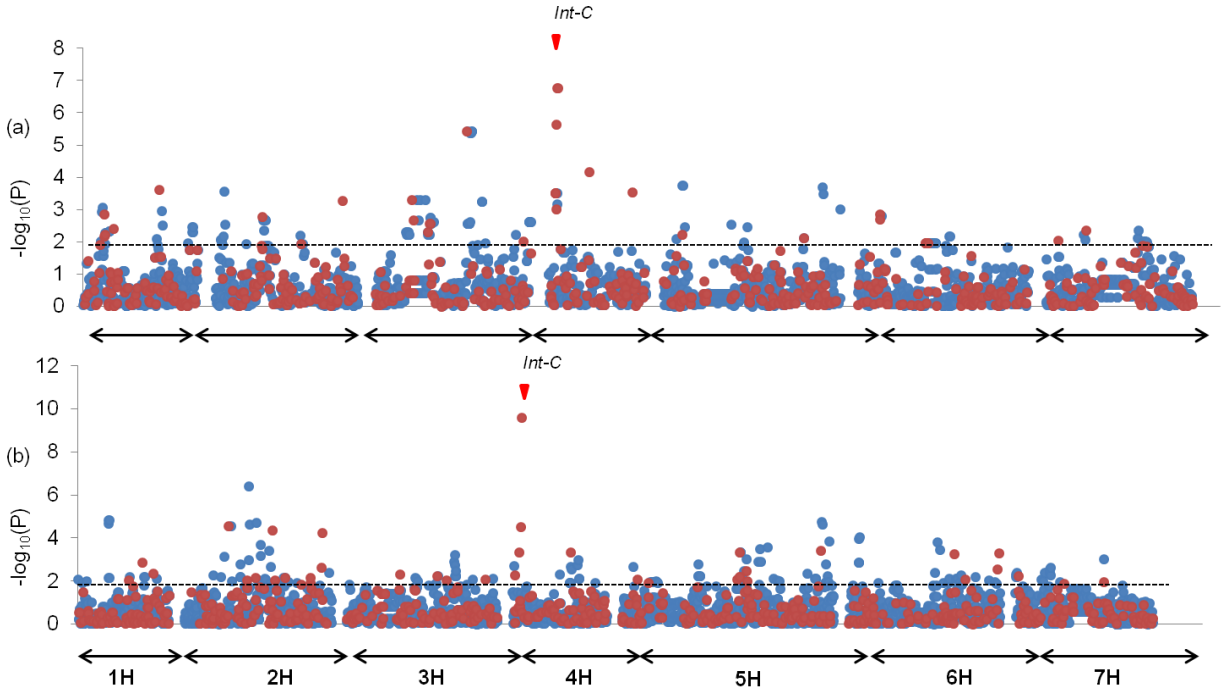



**Fig 3.32**: GWAS scan for the grain area (Garea) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes

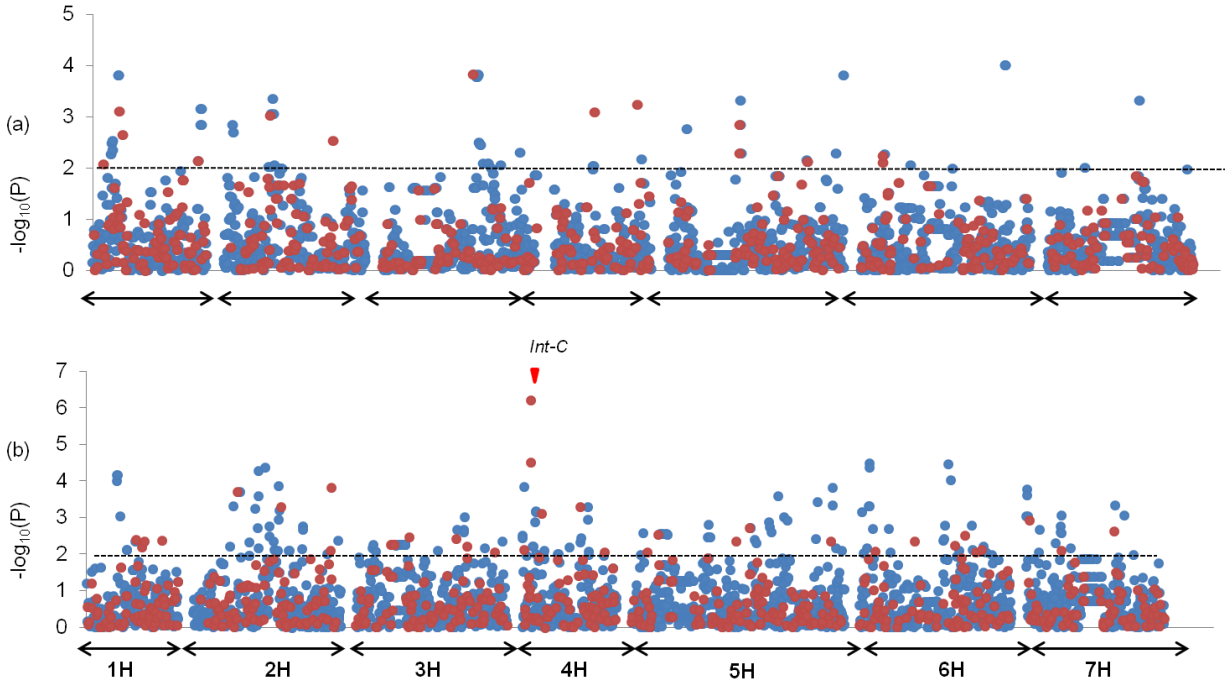



**Fig 3.33**: GWAS scan for the grain length (Glength) in **(a)** HVCC_W and **(b)** HVCC_S gene pools Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes

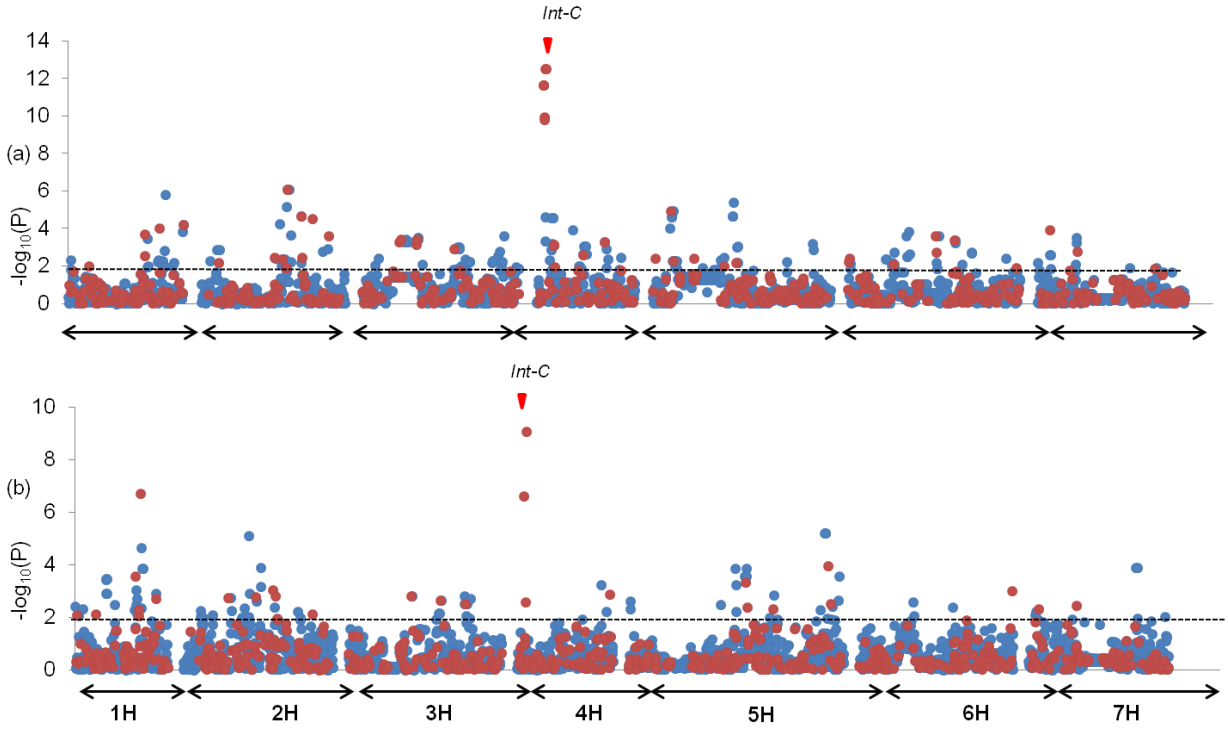



**Fig 3.34**: GWAS scan for the grain width (Gwidth) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes

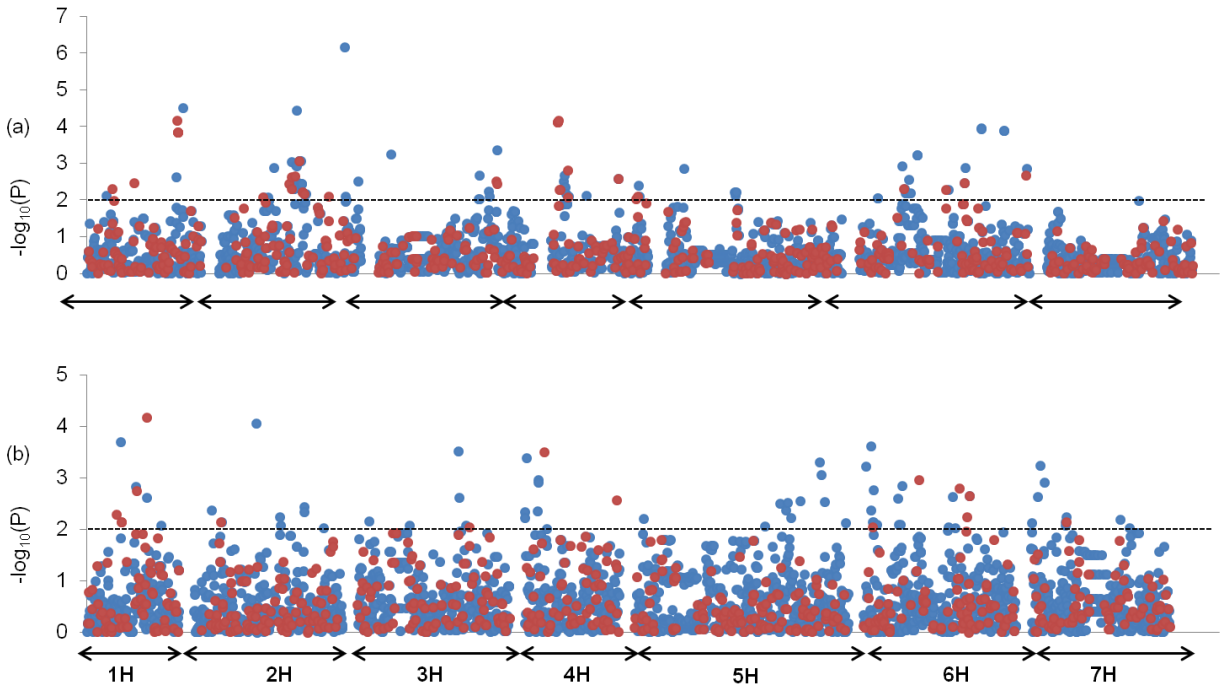



**Fig 3.35**: GWAS scan for the grain ratio width by length (W_l) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes

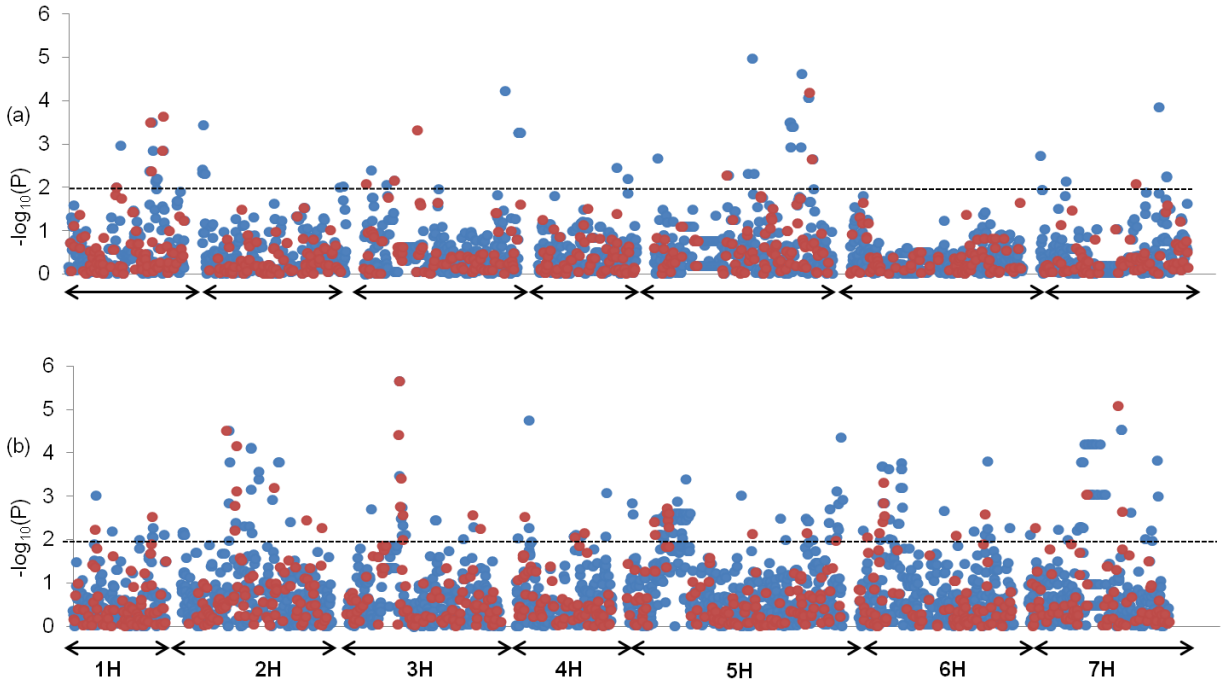



**Fig 3.36**: GWAS scan for the Grain yield (Yld) in **(a)** HVCC_W and **(b)** HVCC_S gene pools. Blue and red color represents iSelect and BOPA1 SNP markers significances across seven barley chromosomes

### 3.4.8 Genome-wide association studies of grain quality traits with iSelect SNP markers

***Starch content (SC)***

GWAS for trait SC using iSelect SNPs revealed 85 significant associations in HVCC_W and 173 significant associations in HVCC_S (Sup. Table S32, Fig 3.37). Highly significant associations were on chromosome 1H (91.6cM) in HVCC_W and on 7H (89cM) in HVCC_S gene pool. Genomic regions on 1H (~50.44cM), 2H (~58cM, ~71cM, ~83cM and ~144cM), 3H (~99cM), 4H (~51.30cM and ~81.70cM), 5H (~32cM and ~151cM) and 7H (~81.76cM) were found significantly associated in both gene pools (Sup. Table S32, Fig 3.37).

***Protein content (PC)***

In HVCC_W gene pool, 144 SNPs were significantly associated to PC and in HVCC_S 148 SNPs showed significant association to PC trait (Sup. Table S33, Fig 3.38). SNP from chromosome 2H at 87.1cM in HVCC_W and from 2H at 18.7cM in HVCC_S were highly significant (Sup. Table S33, Fig 3.38). Several genomic regions were significant across the gene

pools like like 1H (~53.91cM, ~65.90cM), 2H (~10.90cM, ~58.90cM, ~68cM, ~70.50cM and ~86.60cM), 3H (~127cM and ~141cM), 5H (~18.0cM and ~166cM), 6H (~60.20cM) and 7H (~133cM) (Sup. Table S33).

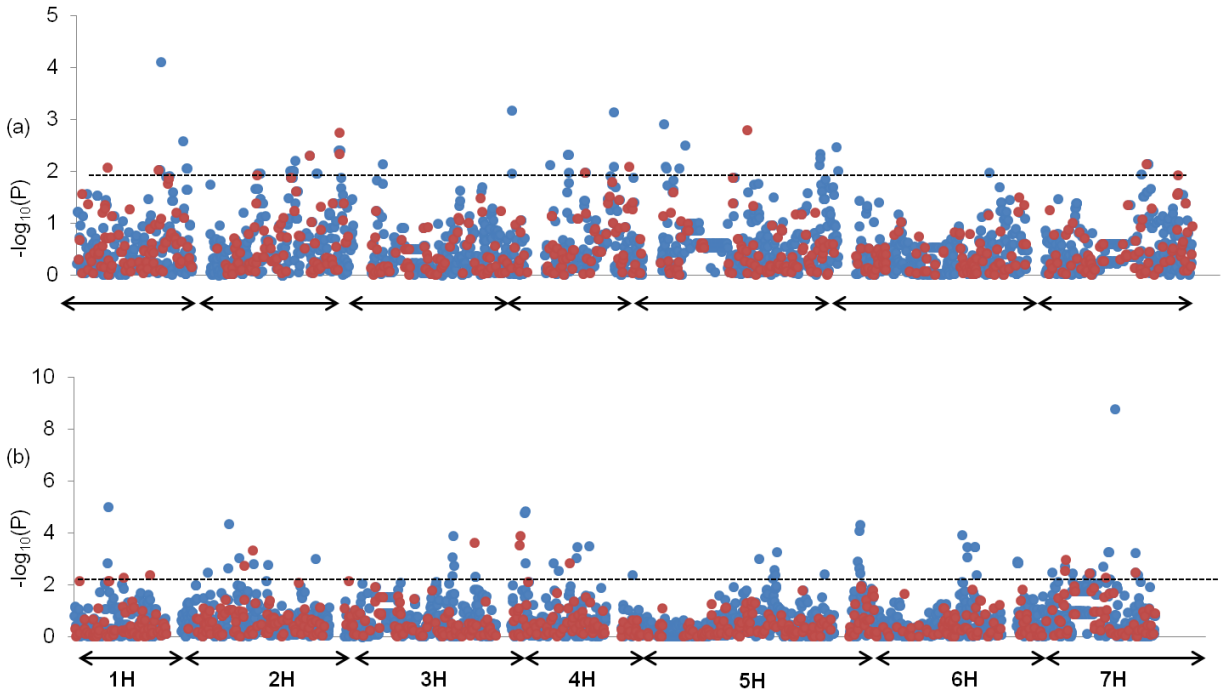



**Fig 3.37**: GWAS scan for the starch content (SC) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

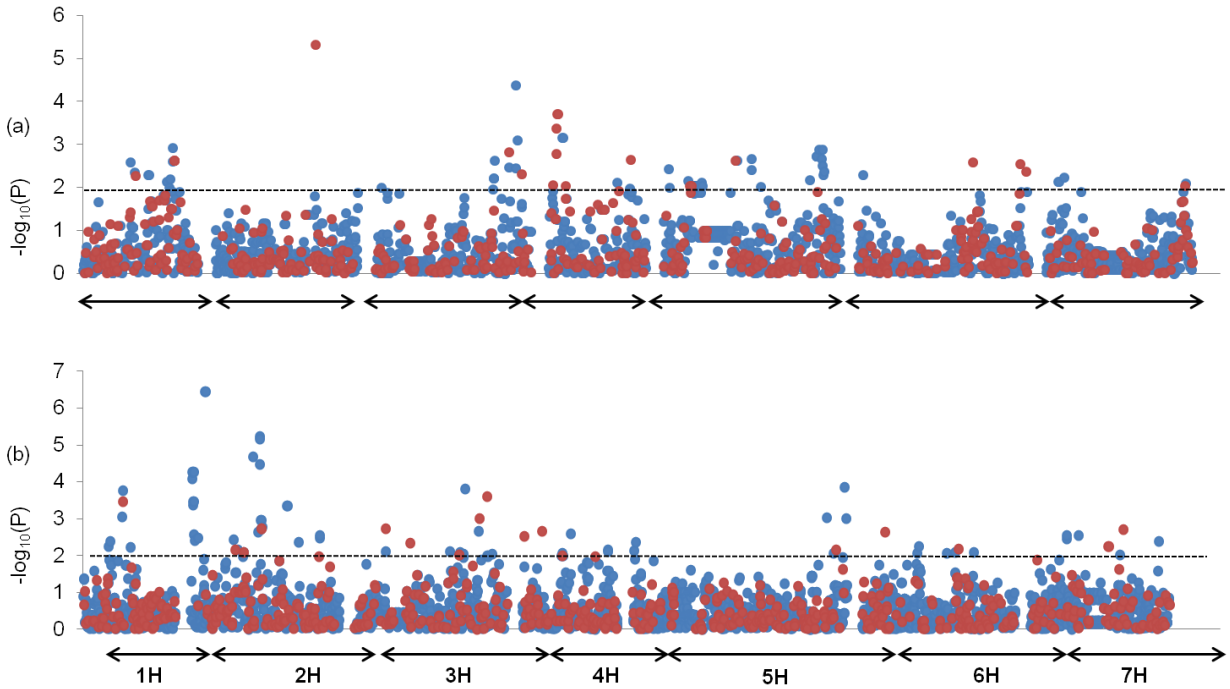



**Fig 3.38**: GWAS scan for the protein content (PC) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

### 3.4.9 Genome-wide association studies of powdery mildew disease resistance with iSelect SNP markers

***Powdery Mildew Field (PMfld)***

Genome-wide scan for powdery mildew resistance under field condition resulted in a total of 79 and 320 SNPs significantly associated in HVCC_W and HVCC_S gene pools, respectively (Sup. Table S34, Fig 3.39). Chromosomal region near the *Mla* genomic region on short arm 1H (~5cM) was significantly associated in the HVCC_S gene pool. Most significant association was on chromosome 5H (130cM) in HVCC_W and 2H (150cM) in HVCC_S. Many significant associations were commonly present in both gene pools like 1H (~55cM), 3H (~69.60cM), 4H (~96.60cM and ~103cM) and 5H (~47.02cM, ~126.50cM, ~134cM, ~155cM) and 7H (~132cM) (Sup. Table S34, Fig 3.39).

***Powdery Mildew resistace under detached leaf Assay in different gene pools (PMdla)***

In total 62, 121 and 190 SNPs in HVCC_W, HVCC_S, and HVCC_Genobar gene pools were significantly associated to the trait PMdla (Sup. Table S35, Fig 3.40). Highly significant associations were observed on chromosome 3H (123.7cM) in HVCC_W, 2H (150cM) in HVCC_S and 5H (43.8cM) in HVCC_Genobar gene pools (Sup. Table S35, Fig 3.40). Interestingly, some of the associated genomic regions had significance across three gene pools viz. 3H (~25cM, ~78cM, and ~114cM), 5H (~62.10cM, ~135cM and ~143cM) and 7H (~74.0cM) , respectively (Sup. Table S35, Fig 3.40).

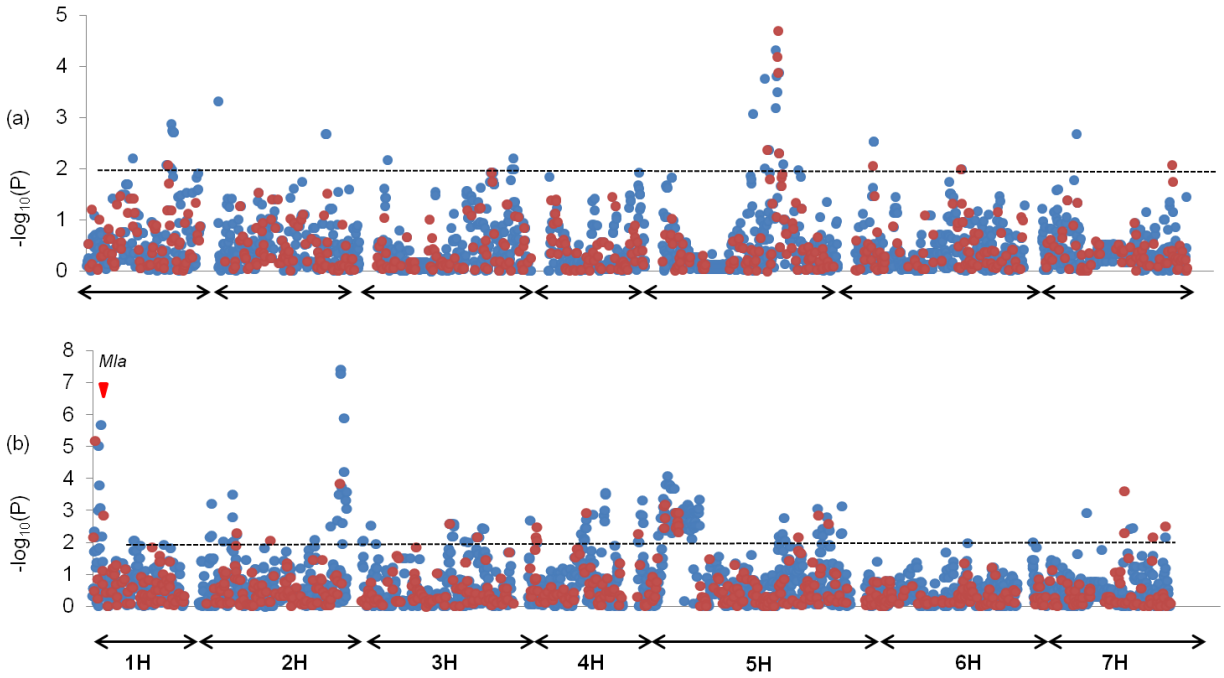



**Fig 3.39**: GWAS scan for the powdery mildew resistance under field condition (PMfld) in **(a)** HVCC_W and **(b)** HVCC_S gene pools

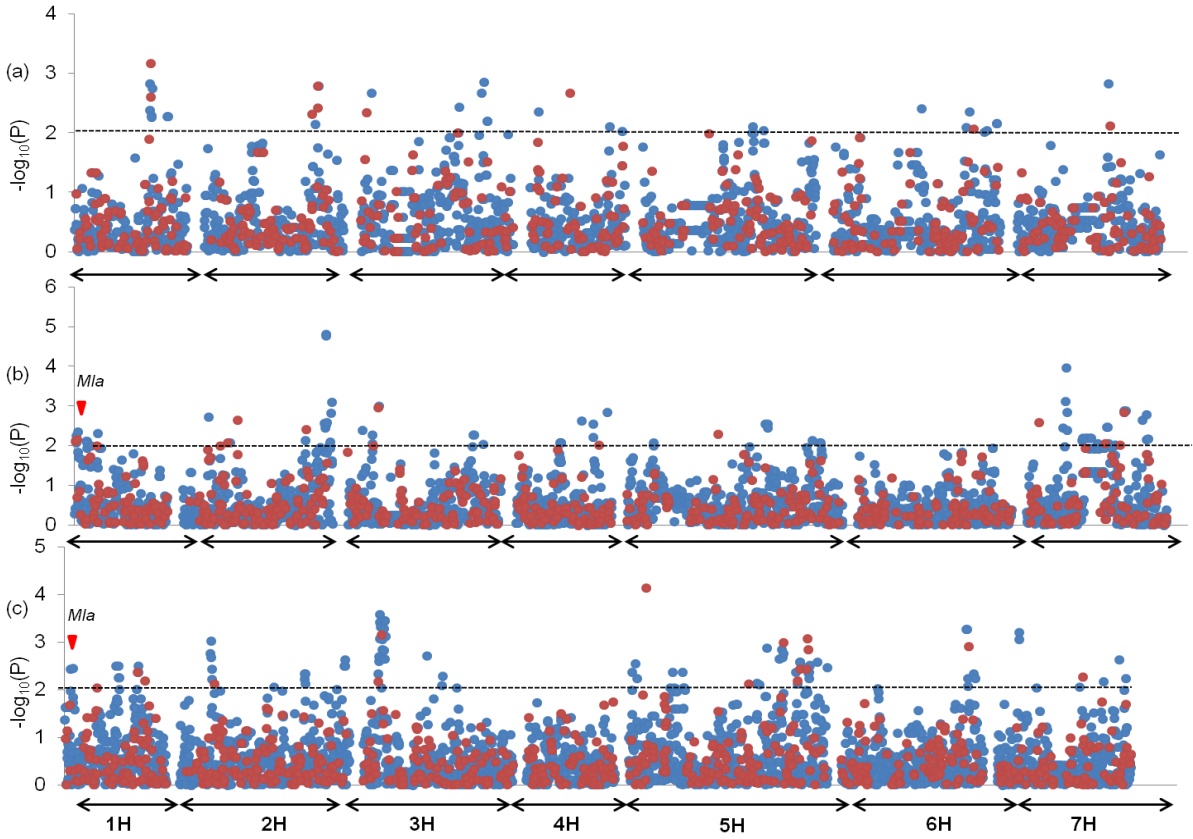



**Fig 3.40**: GWAS scan for the powdery mildew resistance under detached leaf assay (PMdla) in **(a)** HVCC_W **(b)** HVCC_S and **(c)** Genobar_224 gene pools

### 3.5 Coincidence of genome-wide associations across traits

In this study several genomic regions were found significantly associated across traits (Fig 3.41, 3.42, 3.43). Majority of the traits were found to have significant correlations at phenotypic level (Table 3.2 and Table 3.3). The positive and negative correlations observed among the traits at phenotypic level indicated that these traits are not independent. The effects of the interdependency of these traits and their overlapping pathways have resulted in multiple trait associations at some of the genomic regions and often this pleiotropic effect is reported from major genes (Sup. Table S36, S37, S38 and S39, Fig 3.42, Fig 3.43). For example *Sdw1* gene on 3H in HVCC_S gene pool was significant with yield and protein content traits (Fig 3.43). Chromosomal region 3H (~50-60cM) was significant to row type, yield, TGW, grain area, grain width, grain length, frost and Protein content traits in winter barley (Fig 3.42). Interestingly in winter barley chromosomal region on 1H at ~90cM significant associations were observed with heading date, height, yield, frost and starch content, similarly heading date and yield association were observed at 104cM (1H). SNPs from heading date *PpdH1* region on chromosome 2H (~26cM) showed strong association with grain width, TGW and ratio of grain width by length traits in winter barley gene pool (Sup. Table S36, S37, S38 and S39). In spring barley gene pool height, spike length and yield showed association at 54cM on chromosome 3H. Also *eam*6 chromosomal region showed significant association on chromosome 2H at ~ 60cM for heading, grains per ear, yield, grain length and ratio of width by length traits. In addition row-type associated regions mostly showed significant association with grains per ear, thousand grain weight, grain width and protein and sometimes with starch content in both winter and spring gene pools.

In order to see the pattern in associations of different traits, principal component analysis was performed across all the traits using associated SNPs. PCA showed that grain traits clustered together in both gene pools. In HVCC_W gene pool PC1 captured 13.33 percent of the variation of the data using BOPA1 SNP associations. On this axis grain traits like TGW, Garea, Glength and W_L were grouped differently from rest of the traits (Fig 3.41, 3.42, 3.43). Another 12.25 percent of variation was captured on PC2 and further clustering of the traits was observed. Interestingly traits like protein content, yield, powdery mildew field, frost, heading and height were clustered closer to each other. Spike morphology related trait like rowed type and grains per ear were clustered together. PCA of the SNP trait associations with iSelect SNPs did not alter the

structuring of the traits. PCA with iSelect results explained same proportion of variation, only the resolution of the separation of traits from each other changed (Fig 3.41, 3.42, 3.43).

Similarly PCA was performed with GWAS results using BOPA1 SNPs in HVCC_S gene pool. PC1 explained 15.11 percent of variation (Sup. Fig S10). Traits like rowed type, GEar, Gwidth and TGW clustered separately from the all other traits (Sup. Fig S8, S9). PC2 explaining 12.5 percent of variation, separates traits like Glength, W_L, Gwidth, TGW, GEar and row type. Traits like yield, starch content, spike length, heading, height, protein and PM field and PM detached leaf assay were clustered together (Sup. Fig S8, S9). Similar structuring of the traits was observed with PCA constructed using results of GWAS with iSelect SNPs (Sup Fig S10). Only exception observed was for the traits spike length and plant height. (Sup. Fig S8, S9).

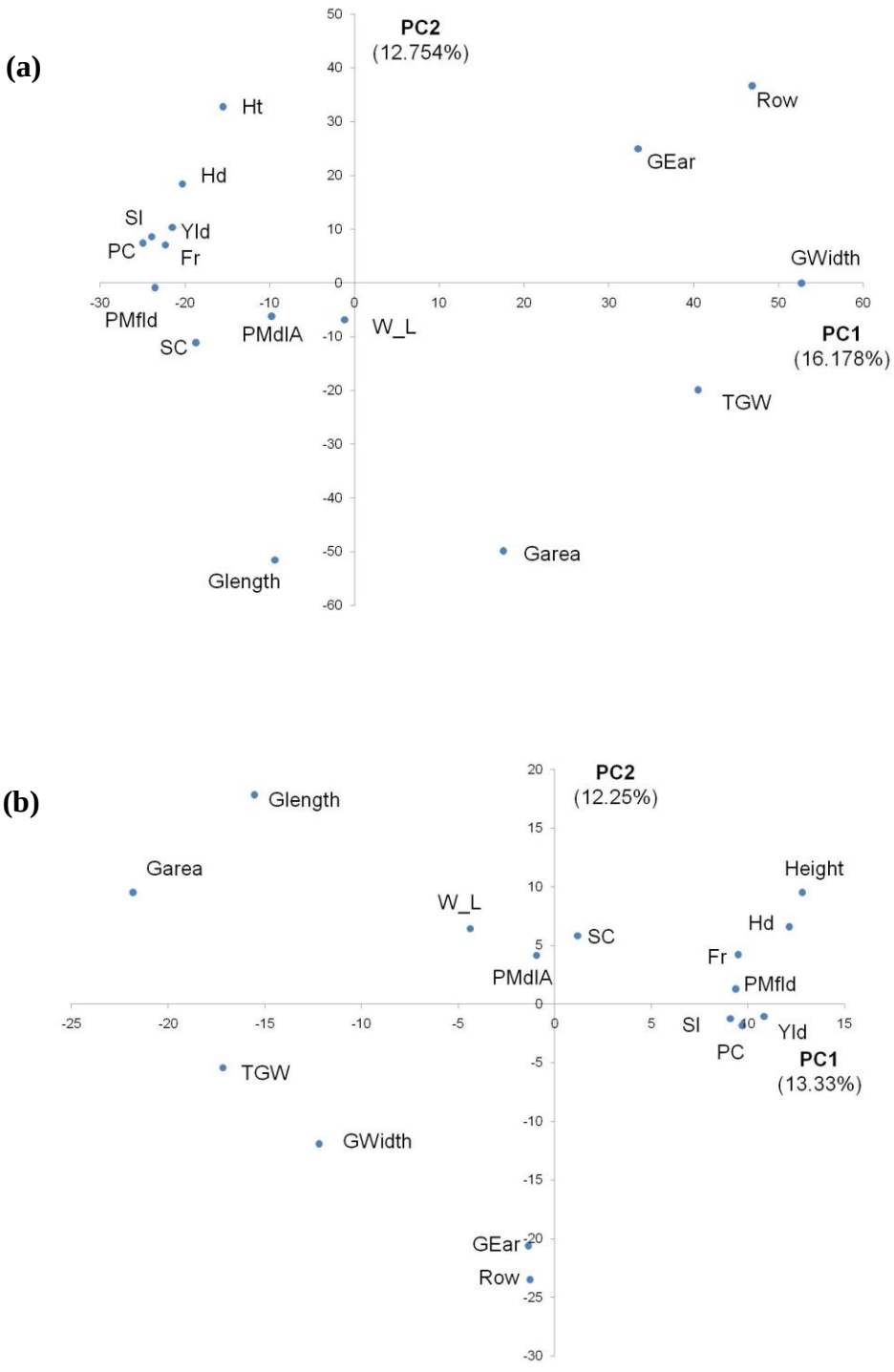
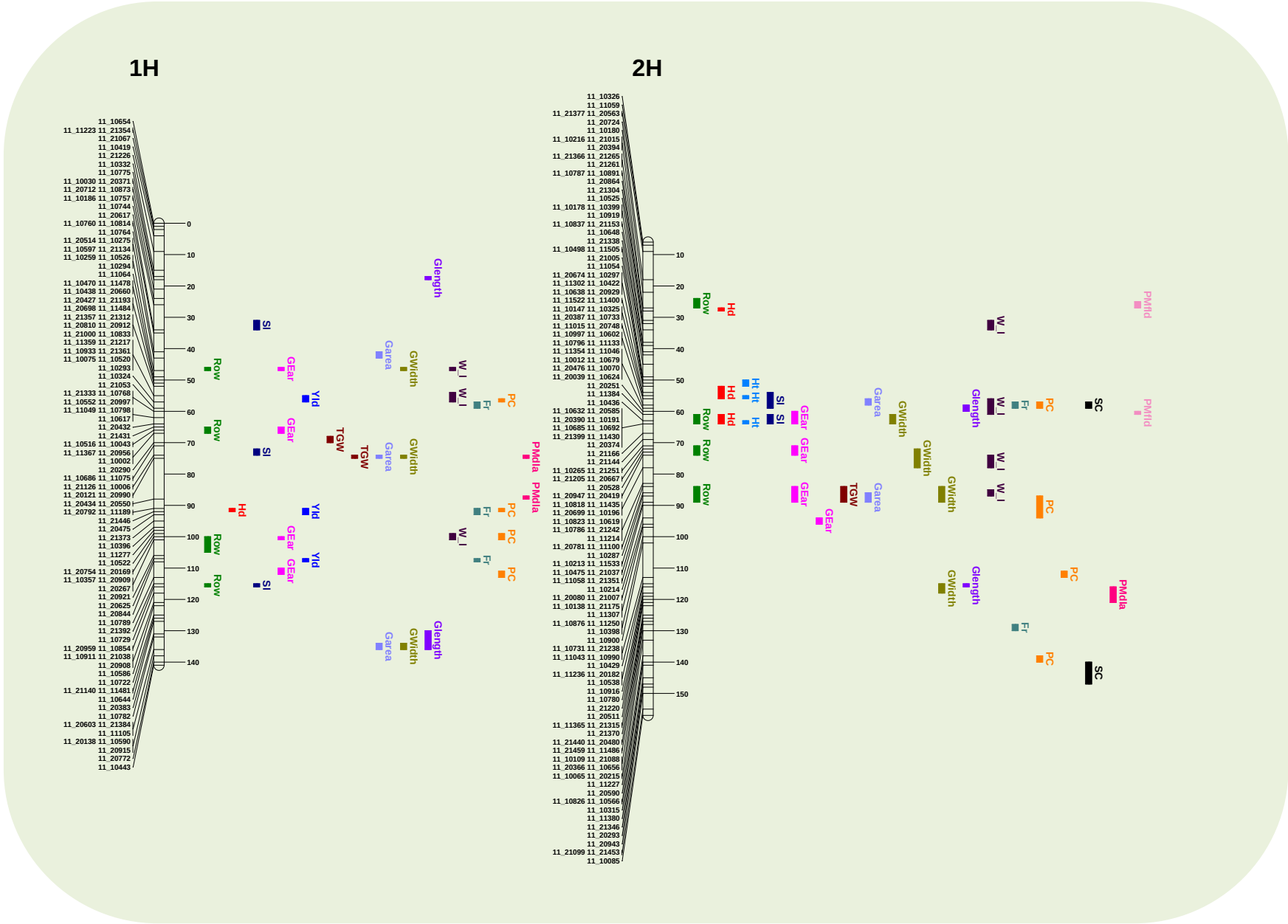
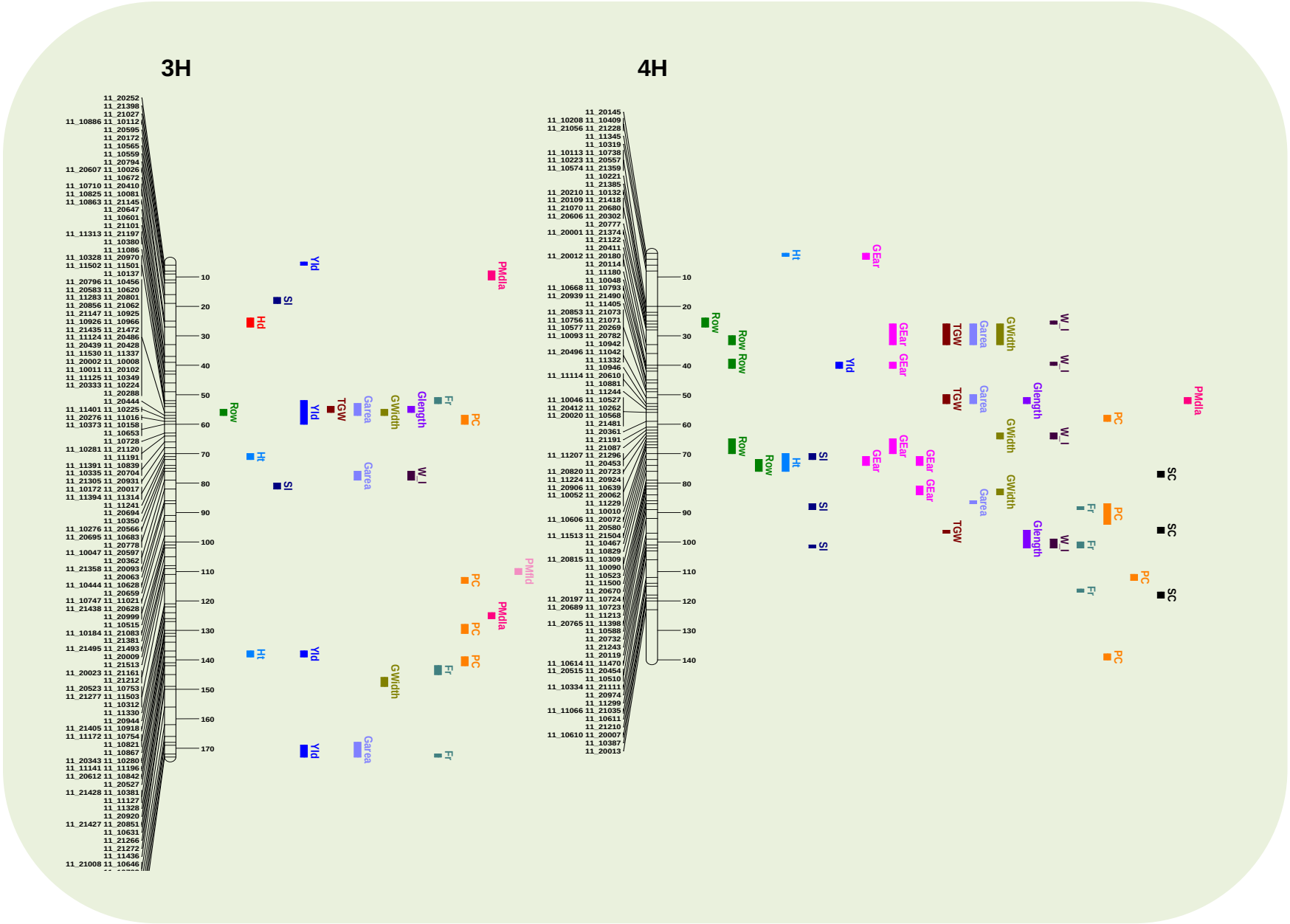
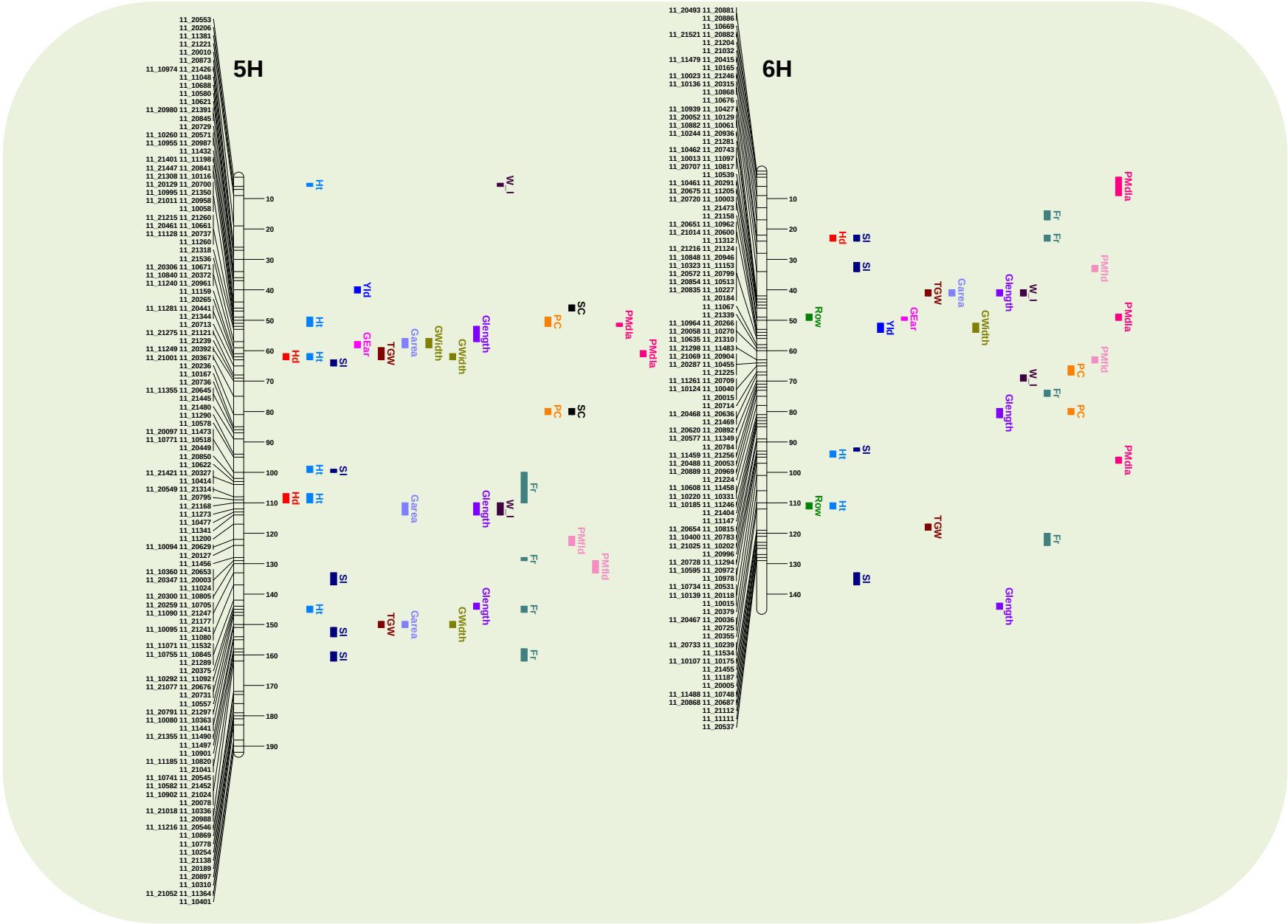



**Fig 3.41**: Principal component analysis of the associations in winter barley. Significant principal components (PC1 and PC2) are shown using **(a)** BOPA1 and **(b)** iSelect SNP markers


1H
2H


3H
4H


5H
6H

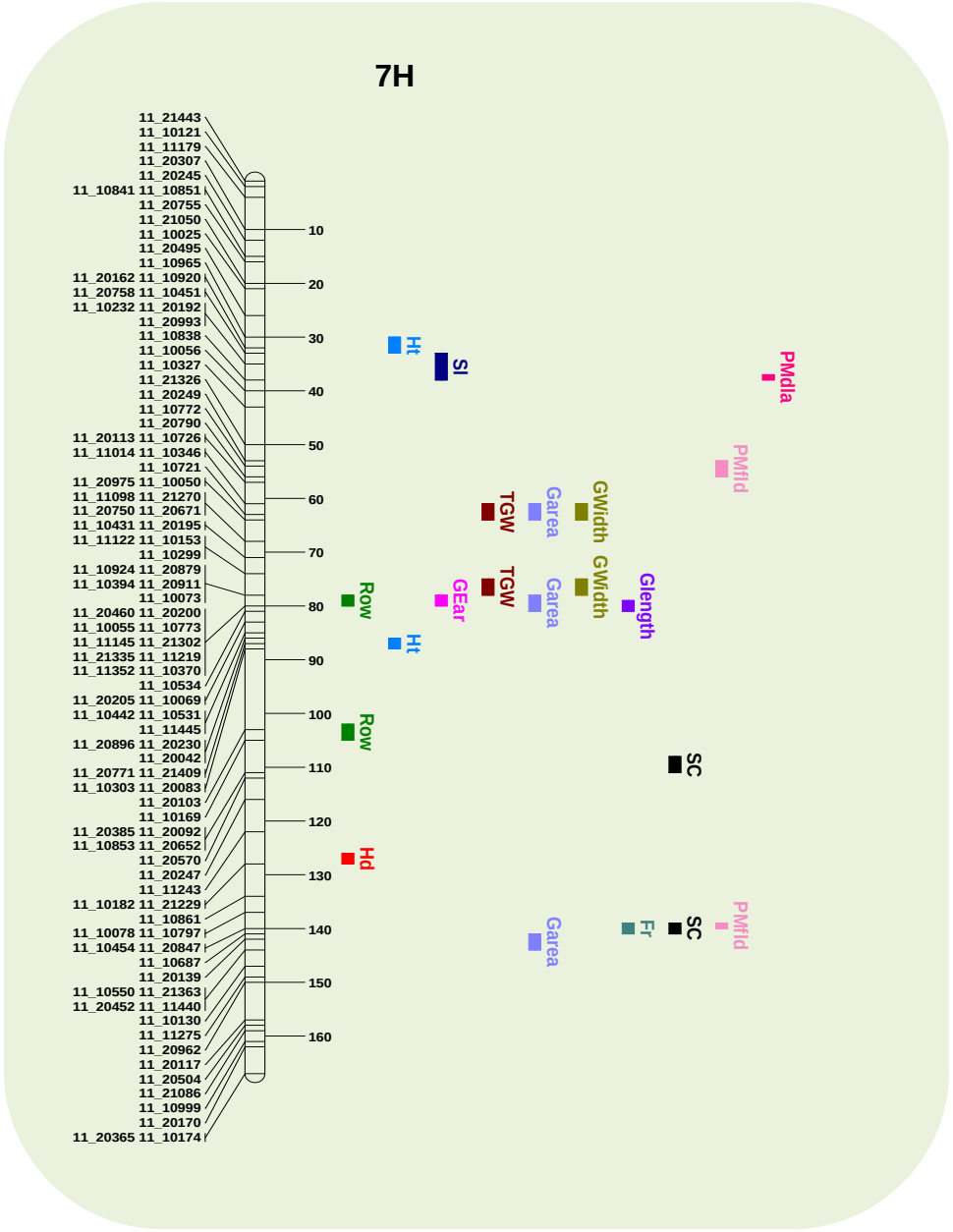
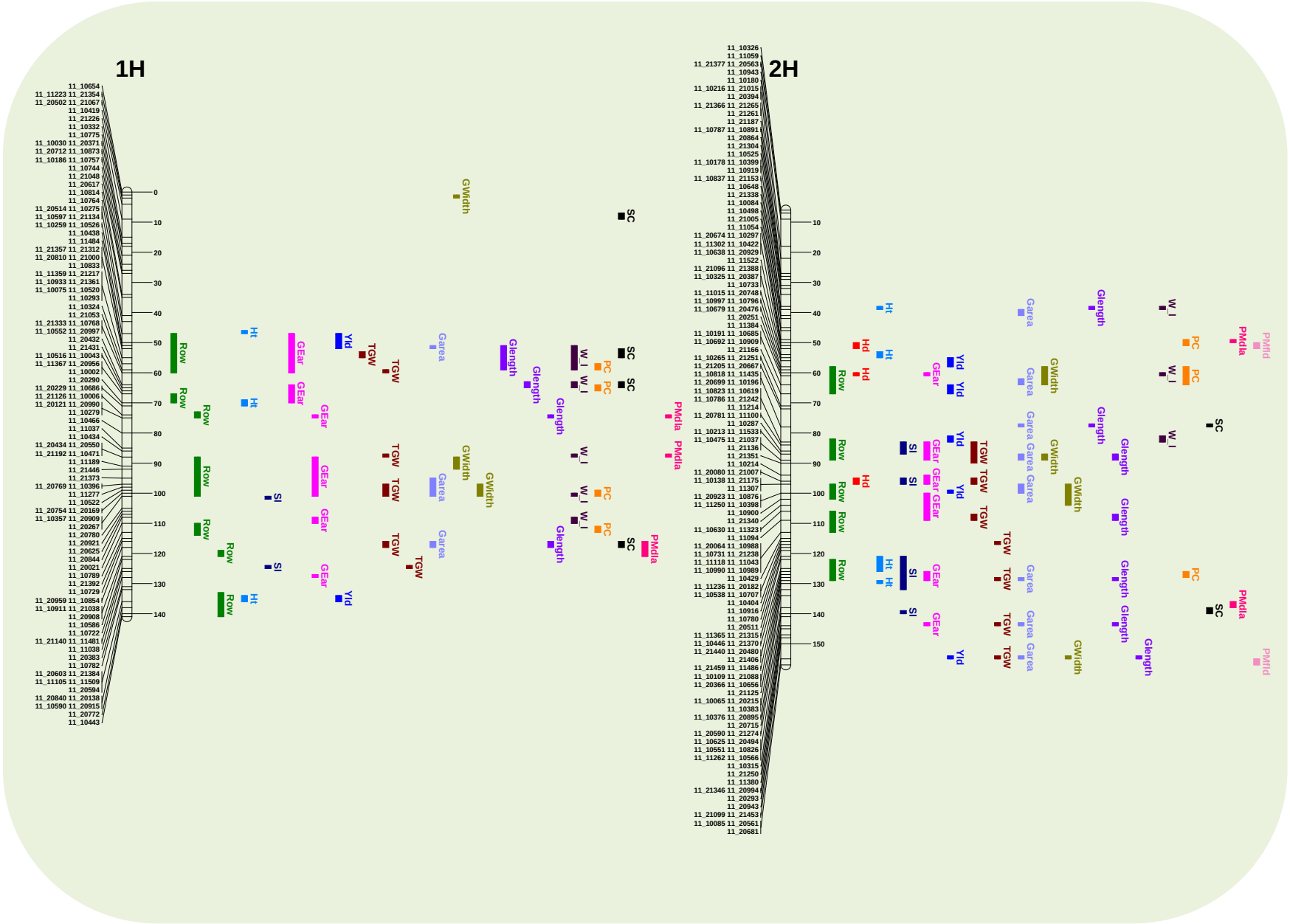
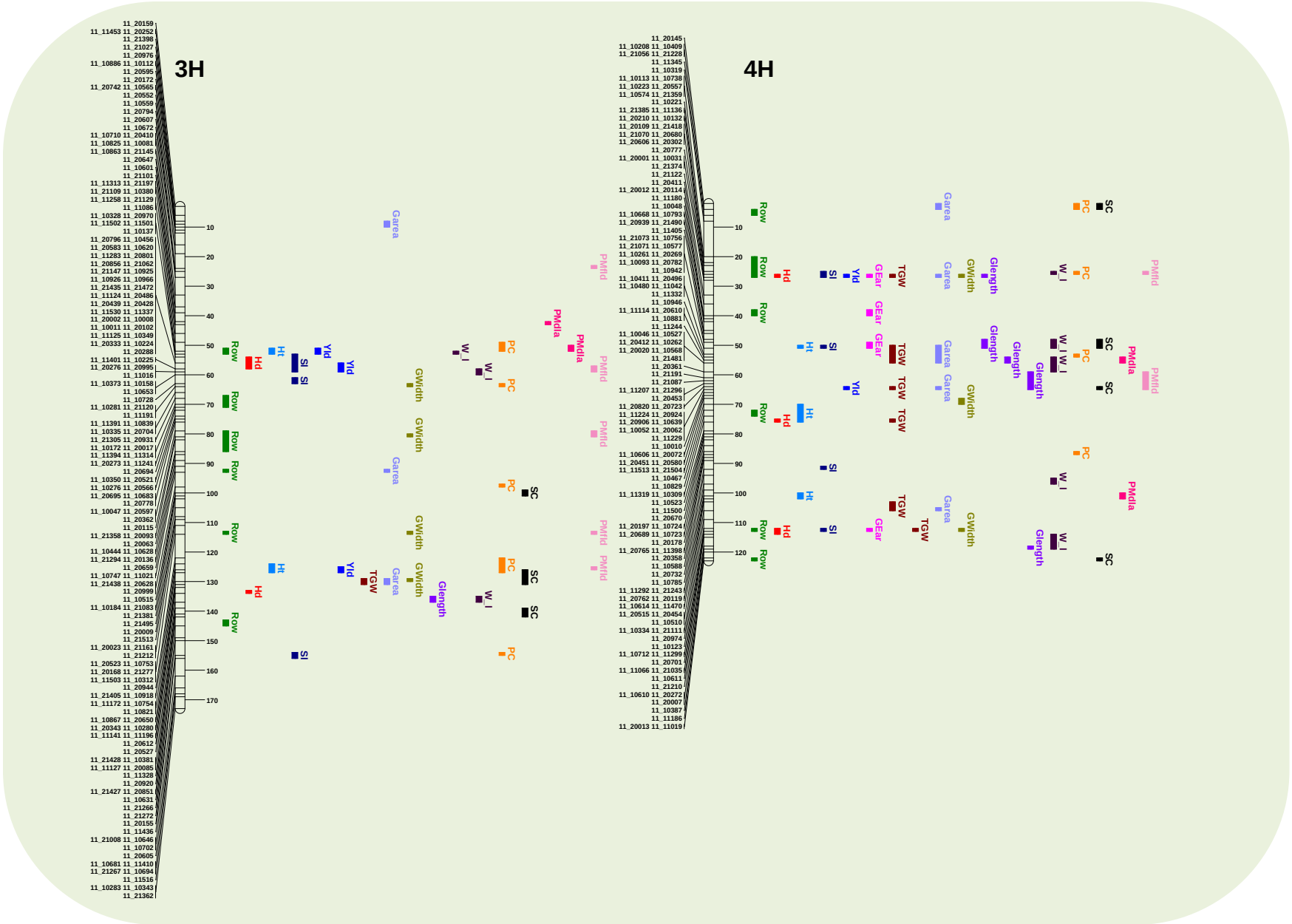
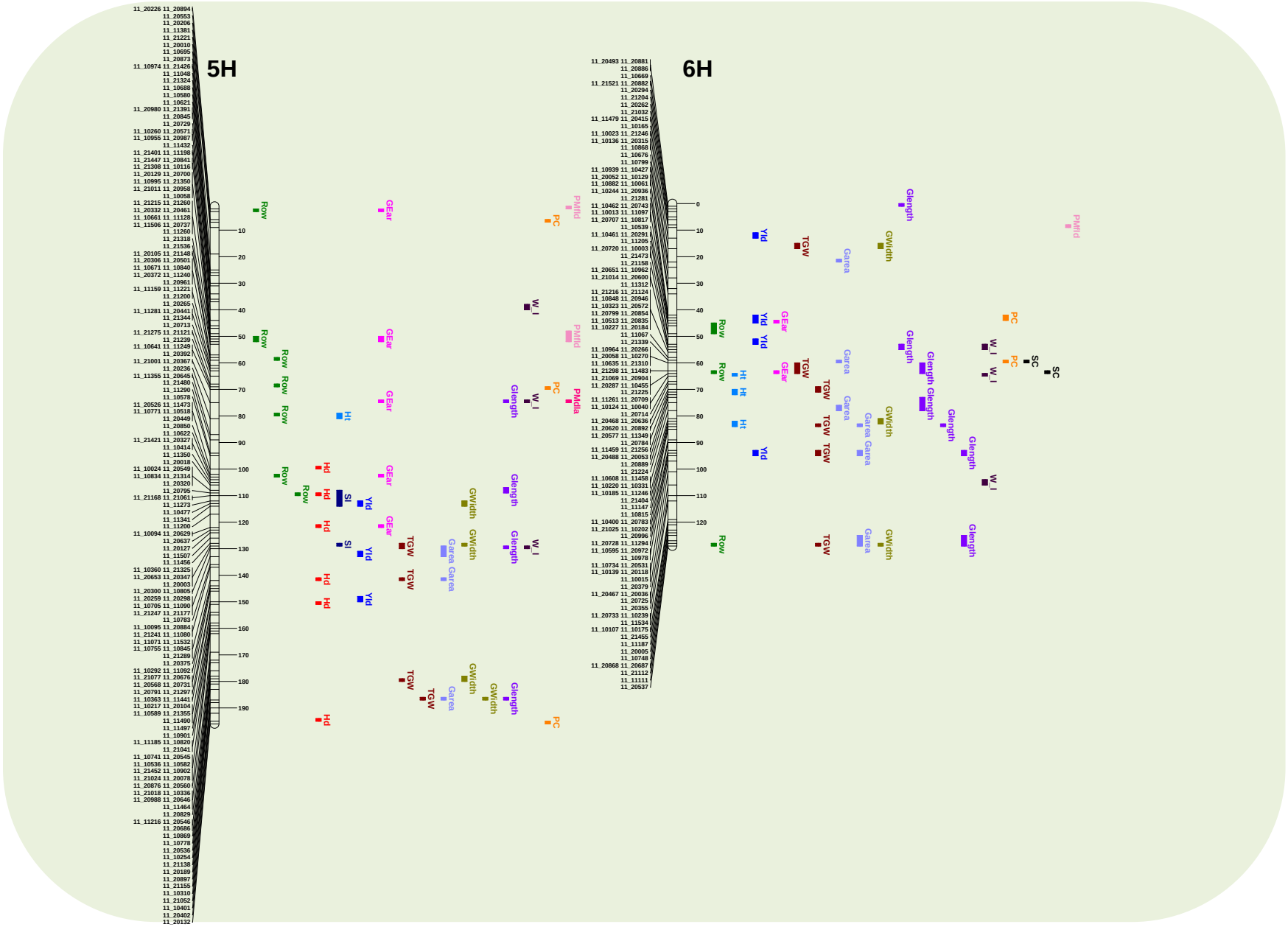



**Fig 3.42**: Genome-wide association of all traits in winter (HVCC_W) barley gene pool using BOPA1 SNP markers. Chromosome 1H-7H are shown along with the associated chromosomal regions highlighted as bars of various color according to the traits. Short abbreviations for traits are shown viz. Frost (Fr), row type (Row), height (Ht), heading (Hd), spike length (Sl), grains per ear (GEar), thousand grain-weight (TGW), grain area (Garea), grain length (Glength), grain width (Gwidth), grain width / length (W_l), starch content (SC), protein content (PC), powdery mildew resistance in field (PMfld) and powdery mildew detached leaf assay resistance (PMdla). Significant associated regions are shown ($-\log_{10} P \geq 2$) as bars of different colors


1H
2H


3H
4H


5H
6H

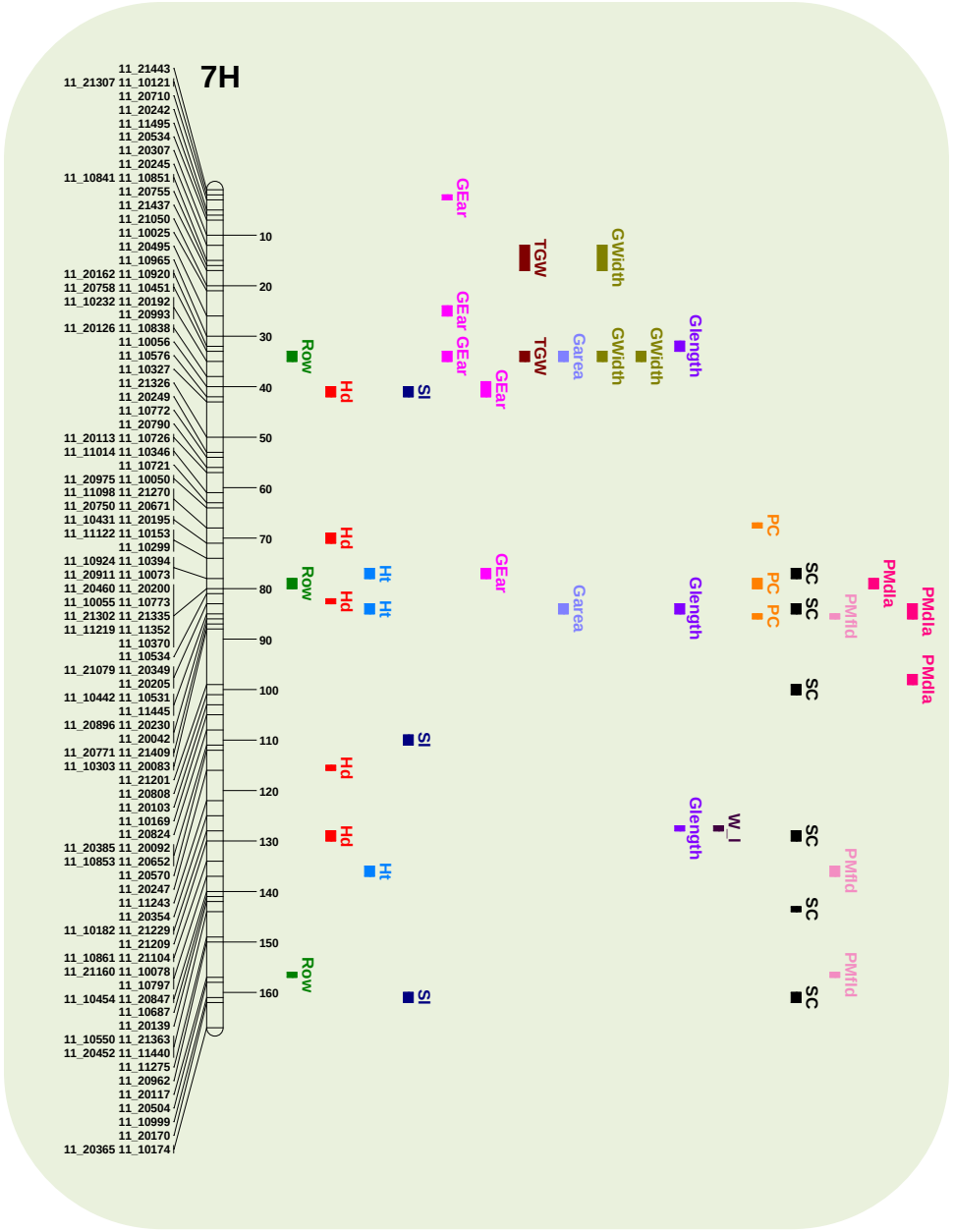



**Fig 3.43**: Genome-wide association of all traits in spring (HVCC_S) barley gene pool using BOPA1 SNP markers. Chromosome 1H-7H are shown along with the associated chromosomal regions highlighted as bars of various color according to the traits. Short abbreviations for traits are shown viz. row type (Row), height (Ht), heading (Hd), spike length (Sl), grains per ear (GEar), thousand grain-weight (TGW), grain area (Garea), grain length (Glength), grain width (Gwidth), grain width / length (W_l), starch content (SC), protein content (PC), powdery mildew resistance in field (PMfld) and powdery mildew detached leaf assay resistance (PMdla). Significant associated regions are shown ($-\log_{10} P \geq 2$) as bars

# 4 Discussion

Genome wide association (GWA) is a powerful approach to dissect and unravel the genetic architecture of complex agronomic traits. The direct utilization of available natural variation for QTL discovery in GWAS has many inherent advantages. LD mapping saves time required for establishment of mapping population which is several years in case of biparental mapping populations. However association mapping populations has to be purified by single seed descent method (SSD) and requires also time. Nevertheless association mapping has more advantage over the biparental mapping as it can be immediately adopted using pure lines. High genetic resolution can be obtained using LD mapping approaches (Rafalski 2010). It is based on many generations of recombination in the ancestral population in comparison to only few recombination events in biparental mapping. The LD decay is variable in different germplasm sets and can be exploited in discovering the QTL at minimum to high resolution (Aranzana et al. 2005; Brachi et al. 2010; Cockram et al. 2010; Morrell et al. 2005; Pasam et al. 2012; Stracke et al. 2007). GWAS in crop plants is challenging as in many cases phenotyping large collections is limiting due to the long life cycle of crop plants. Barley with large and highly repetitive genome is difficult crop to work with at genomic level. However, the recent advances in barley genomics and the establishment of vast genetic resources has opened up the feasibility of gene cloning both by employing mixed strategies map based approaches that include the use of information from sequenced model and crop genomes and GWAS (Close et al. 2009; Komatsuda et al. 2007; Mayer et al. 2011; Ramsay et al. 2011; Sato et al. 2009). The availability of a large number of SNPs from exonic regions and the recently developed parallel SNP genotyping platforms in barley has incited several GWAS both for gene cloning and QTL detection (Cockram et al. 2010; Comadran et al. 2011a; Pasam et al. 2012; Schmalenbach et al. 2011). Barley is a predominantly self pollinating diploid species and has high LD blocks and few hundred markers in LD are enough to cover all the barley genome (Brown et al. 1978; Rostoks et al. 2006). Conversely, QTL discovered in high LD germplasm have low genetic resolution as the causal gene could be far from the associated SNP marker. In recent studies, different gene pools of barley were shown to have different LD pattern and the discriminatory LD can be used to fine map QTL and locate candidate genes (Comadran et al. 2011a; Morrell et al. 2005; Russell et al. 2011).

In the present study, different gene pools were investigated at a genome wide level to discover marker associations for agronomic and biotic stress related traits. The discriminatory LD pattern of these gene pools was exploited in marker associations for the studied traits. Although several marker trait associations were detected in the gene pools, the idea of exploiting the discriminatory LD of gene pools to discover candidate gene has not worked as anticipated in the beginning of the study. Ascertainment bias of the SNP markers was the major problem while comparing associations across the gene pools. Furthermore, in some cases phenotyping of agronomic traits was not possible in all gene pools e.g. wild barley was not studied for the agronomic traits as it was very difficult and time consuming to propagate and work on it. Similarly landraces from Syria and Jordan were not adapted to the growing conditions of Europe. As a result only powdery mildew resistance under detached leaf assay was scored on wild barley and landraces under green house growing conditions. Although they were planted twice in the field but they turn out to be complete failures. During the year 2009 LRC were planted during spring but were unable to flower in two locations in Germany (IPK and Halle) (Fig 4.1a). We thought they need vernalization as they originated from Syria and Jordan where they need cold environment during their development. In order to have uniform growth and heading all LRC were planted in winter in the subsequent year 2010. It was too cold for them and most of them were unable to survive under extreme cold conditions of 2010 (Fig 4.1b).

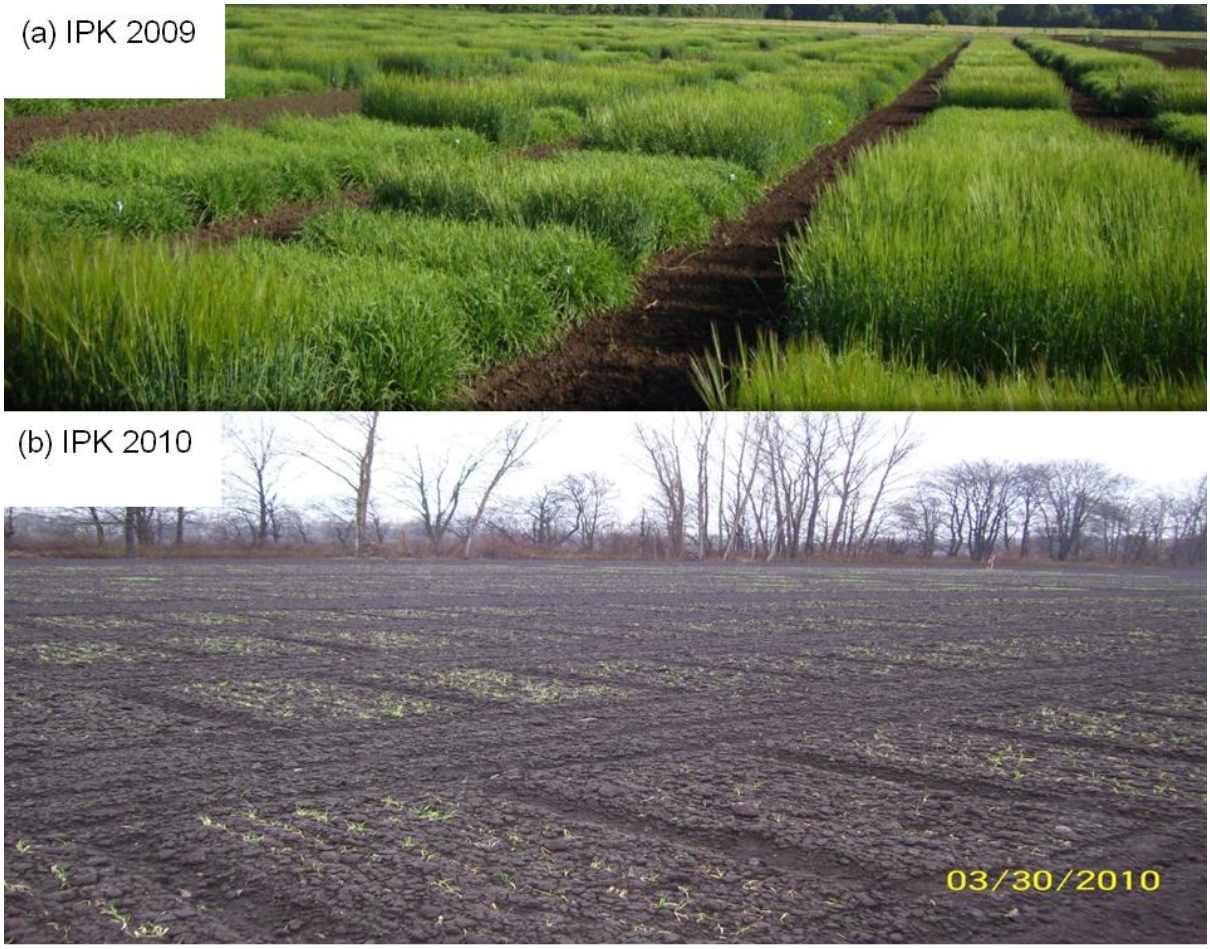



**Fig 4.1**: LRC field plots from IPK location; (a) shows heading in Jordanian genotypes and only vegetative growth in Syrian landraces from spring sowing in year 2009 IPK (b) shows complete frost damage during winter in year 2010 IPK

### 4.1 Population structure and linkage disequilibrium (LD)

A structured population is an inherent character of most of the crop plants which is the outcome of the complex history of domestication and breeding. In plants the population structure is influenced by the genetic bottlenecks, mutation, genetic drift, and migration histories. Many crops have experienced severe bottlenecks during domestication and plant breeding (Hyten et al. 2006; Kilian et al. 2006; Reif et al. 2005). Population structure is serious problem in association studies as division of the population into distinct subgroups related by kinship is the source of false positives (Rafalski 2010). If a certain trait is common in one subgroup but rare in another then any markers distributed unevenly in these groups can be positive in association test. This type of association can be located on any genomic location. In order to control false positives,

population structure has to be determined and structure should be accounted in the statistical model for association analysis.

In this study, PCA of the all available collections together has shown that the genotypes were structured according to the gene pool classification. Cultivated gene pools (HVCC_W and HVCC_S) clustered separately from landraces (LRC) and wild barley (HSC). Further, HVCC_S and HVCC_W gene pool were indicating higher degree of genetic diversity in comparison to the HSC and LRC gene pool (Fig 3.1). In addition using STRUCTURE and Neighbour Joining (NJ) tree cultivated gene pools (HVCC_W and HVCC_S) were majorly structured according to the row-type (Fig 3.3, Fig S6). As barley breeding follows separately on two-rowed and six-rowed types, the observed differences on molecular basis confirms the genetic differentiation of two and six-rowed barley types. Interestingly in PCA axis spring two vs. six rowed separation was wider in comparison to the winter two vs. six rowed barley. This probably suggests plant breeding between row-type employed more frequently in winter barley in comparison to the spring type although there is no literature to prove this point. Similarly based on the NJ cluster analysis landraces were clustered based on their geographic origin (Fig S7). The LRC gene pool was reported to be diverse and structured according to their geographical origins using SSRs and chloroplast SSR data (Russell et al. 2011; Russell et al. 2003). In current study sub-structuring of the landraces according to the geography (Syrian and Jordanian landraces) was observed but the relative differentiation was comparably small in comparison to the cultivated gene pools (Fig 3.1, S7). The wild barley gene pool (HSC) appears as small cluster in the PCA (Fig 3.1) whereas high genetic diversity and further sub-structuring of the wild barley was reported using SSR markers (Hübner et al. 2009). Also NJ cluster analyses showed that wild barley have sub-structuring according to their origin (Fig S7).

These observed differences signified that the interpretation of relative genetic diversity (based on PCA) of the gene pools is not straight forward. Ascertainment bias of the BOPA1 SNP markers is the major reason that explained the discrepancy in estimating the diversity of HSC and LRC gene pool (Fig 3.1). Although the LRC and HSC gene pools were reported to be diverse in earlier studies in the current study using BOPA1 SNP markers their diversity is underestimated (Hübner et al. 2009; Russell et al. 2003). A large number of SNP markers were not segregating in these gene pools. These SNPs were either monomorphic or with a very low MAF (see Fig

3.2). Discrete haplotype structuring (less randomly assorted haplotype) in the cultivated gene pool in comparison to the landraces and wild barley gene pool could have attributed to the observed differences in ordination plot (Russell et al. 2004; Russell et al. 2011).

Few cultivars were found in close proximity to the HSC gene pool and some to the LRC (see Fig 3.1). The HVCC_S genotype Orza clustered close to the LRC genotypes (SLB_66-71, SLB_05-076 and SLB_45-021. In the pedigree this genotype originated from the cross of cultivar Tokak and landrace 4857. The HVCC_S genotype Rondo (naked six-rowed) clustered close to the wild barley cluster. This genotype is known to have good salt tolerance and closeness of the genotype to the wild barley gene pool means probably it carries wild introgressions. The wild barley accessions B1K-22-06 and B1K-22-03 were divergent from the major cluster of HSC gene pool (see Fig 3.1). These genotypes were collected from Ramat David in the Northern part of Israel and their deviation from wild gene pool may be due to their local adaption and proximity to cultivated barley in this region. This probably means they are the admixtures of cultivated gene pool. These examples showed that except occasional admixtures, these gene pools are distinct from each other. The landraces and wild barley gene pools were sampled from geographically distinct areas so admixture of the genotypes is not naturally possible. Although introgression of these gene pools is almost impossible but admixture during propagation of the genotypes cannot be ignored. Limited degree of secondary contact was reported in landraces and wild barley using the geographically matched genotypes sampled from Jordan and Syria (Russell et al. 2011). In this study, the associations for each gene pool were separately analyzed and compared. In almost all recent association studies, kinship matrix is used in controlling population structure and was reported to perform superior compared to other models (Kang et al. 2008a; Pasam et al. 2012; Stich and Melchinger 2009; Yu et al. 2006). In general it is also obvious that genetic relatedness correction had adverse effect on the number of significant associations. In this study for almost all traits, as expected, significant marker surpassing the $-\log_{10}P = 2$ reduced dramatically using stringent mixed model analysis using kinship matrix. However further markers surpassing false discovery threshold for almost all traits were reduced drastically when stringent conservative false discovery rate (FDR) was used (Table 4.1). In few cases no SNP marker was able to surpass the threshold (Table 4.1) (Storey and Tibshirani 2003). Nevertheless, still the discovered associations above threshold $-\log_{10}P = 2$ were congruent to the published literature. Thus

although less stringent in terms of the statistics but the $-\log_{10}P = 2$ threshold in declaring significance makes sense and thus used in this study.

Population history, mutations, natural selection and chance events in small populations (genetic drift) are the forces that affect the level of LD (Stumpf and McVean 2003). Effective short distance LD is mainly impacted by recombination, on the other hand long distance LD/ inter-chromosomal LD may be due to the population structure or epistasis or trans-acting regulators (Kouyos et al. 2007; Slatkin 2008). Inter-chromosomal LD was observed in HVCC and LRC gene pools and this long range LD should be accounted in order to control false positives. The extent of LD increased in HVCC_W and HVCC_S gene pools in comparison to the LRC and HSC gene pools (Fig 3.4, 3.5). Level of LD decay in spring and winter barley gene pool in this study is comparable to other genome-wide studies but high in comparison to the world-wide spring barley collection where average LD level was reported 5-10cM (Comadran et al. 2011a; Comadran et al. 2009; Pasam et al. 2012). One of the reasons for high LD could be all these cultivars are from Europe and are diverse in terms of the phenotypic values but still closely related to each other. Morrell et al. (2005) found a low level of LD similar to the outcross species maize using re-sequencing of 18 loci in wild barley. The extent of LD decay is variable across the gene pools (see Fig 3.4). In this study extreme LD decay observed in wild barley gene pool (<1cM) and is comparable to the low LD observed in earlier studies on wild barley. This highlights the potential of wild barley associations as SNP associations could be in close proximity to the causal genes. The low LD in HSC gene pool means thousands of SNP markers are required to cover the genome for GWAS. We used the BOPA1 (1536 SNPs) and after removing the SNPs with low MAF only 813 SNPs were considered for GWAS. However, due to handling and phenotyping difficulties involved in large multi-environmental field experiments with wild barley, we did not consider wild barley for GWAS with agronomic traits. The GWAS studies in HSC were done for the trait powdery mildew detached leaf assay, which allowed us to compare the results across all the gene pools.

### 4.2 Comparing associations of BOPA1 and iSelect

GWAS in barley using 1536 BOPA1 SNPs resulted in significant associations for all the traits. However, considering the large genome size of barley (<5 Gbp) the BOPA1 marker coverage is still inadequate. The SNP coverage across the genome is not uniform with several gaps and all the regions are not uniformly covered as the BOPA1 is designed with only 1 SNP per EST (Close et al. 2009). Consequently, several genomic regions are not included into BOPA1 assay and we might have missed many genomic regions as the SNPs included in the study could be mono-morphic or with low MAF (Fig 3.2). In order to increase the marker number and achieve better coverage across the genome, cultivar collections and Genobar collection were genotyped with recently developed iSelect assay (Comadran et al. unpublished). The final successful iSelect SNPs after excluding low MAF SNPs varied among different collections (4923 SNPs in HVCC_W; 5753 SNPs in HVCC_S and 6461 SNPs in HVCC_Genobar). GWAS were performed using these iSelect SNPs for all the traits in these gene pools.

GWAS with iSelect SNPs resulted in more significant associations across the collections for all the traits in comparison to the BOPA1. The significances of the associations were also improved in almost all the traits (Fig 3.25-3.40, Sup. Table S38, S39). The redundancy of iSelect SNPs from the same genomic region has added advantage. Association of the genomic region is function of allele frequency and with BOPA1 few hundred markers (around 984 and 1026 were used for HVCC_W and HVCC_S) were successfully used for association analysis. Thus there is high chance of missing significant genomic regions when the SNP is with MAF. However seven fold increase in marker number in iSelect chip can easily overcome this problem of excluding SNPs. In the iSelect redundancy of the markers from the same regions has advantage as the probability of uniformly covering the genome is increased. For each trait using iSelect markers multifold increase in number of markers surpassing the P-value threshold (P-value $\leq$ 0.01) was observed. However, the associations observed were not exclusive as 1482 BOPA1 SNPs (from the 1536 BOPA1 SNPs) were included in iSelect design. SNP marker associations using iSelect chip were distributed across the genome (Sup. Table 36, 37, 38, 39). Across all traits significant increase in number of SNP trait associations were observed by increasing the number of SNPs from BOPA1 (~1200 SNPs) to iSelect (~7000 SNPs) (Table 4.1).

The same chromosomal regions were found significantly associated with BOPA1 and iSelect SNP markers. The best associated SNPs from BOPA1 assay remained significant across both genotyping platforms. For an instance significantly associated SNPs for row type in HVCC_W and HVCC_S gene pools were significant across the genotyping platforms. The SNP markers from BOPA1 (SNP 11_20606) at the region near to the *int-C* region on chromosome 4H, and *Vrs1* (SNP 11_10287) on chromosome 2H were still significant in GWAS with iSelect assay. Thus both marker systems are robust and reliable. Sometimes, the significance of associations changed across the platforms. For example, in the HVCC_W gene pool SNP marker 11_20606 on chromosome 4H associated with $-\log_{10}$ P of 36.85 in BOPA1 and with 39.91($-\log_{10}$ P) in iSelect platform for trait row type. Another example on 2H near *Vrs1* region, SNP marker 11_10287 significance level ($-\log_{10}$ P) in BOPA1 and iSelect platforms is 13.51 and 17.27 respectively. These differences in significance level of same SNPs across the genotyping platforms could be the result of variable numbers of genotype assayed in both the platforms (Sup. Table S1). The other reason was that some of the SNP markers were not successful across the genotyping platforms e.g. 11_10213 SNP marker was significantly associated with row-type in HVCC_W gene pool in BOPA1 platform but this SNP failed in iSelect SNP assay. In powdery mildew resistance in winter, spring and Genobar gene pool significant number of associations surpassing $-\log_{10}$ P = 2 increases e.g. four to three fold and seven to six fold in winter and spring barley gene pools for PM resistance under field and detached leaf assay (Table 4.1). Thus relative increase of significant SNP association was observed among all populations that clearly suggest that even though LD was variable in these gene pools still iSelect perform comparably better than BOPA1. Six fold increase of marker (iSelect) number has successfully captured majority of segregating QTL in this study.

**Table 4.1**: Summary of associations found using BOPA1 and iSelect SNP markers in winter and spring barley gene pools. Significant ($log_{10}$ P ≥ 2) number of SNPs are shown. In brackets number of SNPs surpassing false discovery rate α 0.05 are shown (Storey and Tibshirani 2003).

| | HVCC_W | | HVCC_S | |
|---|---|---|---|---|
| **Traits** | **BOPA1** | **iSelect** | **BOPA1** | **iSelect** |
| Row type (Row) | 37 (21) | 383 (174) | 97 (50) | 349 (153) |
| Frost (Fr) | 34 (9) | 206 (67) | | |
| Heading (Hd) | 28 (15) | 112 (45) | 29 (3) | 282 (58) |
| Spike length (Sl) | 35 (5) | 207 (6) | 41 (9) | 132 (1) |
| Height (Ht) | 49 (16) | 266 (102) | 31 (2) | 159 (25) |
| Grains per Ear (GEar) | 32 (15) | 223 (29) | 85 (68) | 206 (73) |
| Thousand grain weight (TGW) | 31 (5) | 242 (13) | 61 (25) | 216 (22) |
| Grain width (Gwidth) | 36 (18) | 329 (182) | 47 (12) | 184 (21) |
| Grain length (Glength) | 35 (-) | 114 (19) | 58 (11) | 307 (86) |
| Ratio of grain Width/length (W_l) | 78 (2) | 163 (24) | 42 (8) | 172 (-) |
| Yield (Yld) | 25 (5) | 119 (14) | 49 (20) | 305 (70) |
| Grain area (Garea) | 47 (6) | 245 (27) | 51 (19) | 253 (58) |
| Starch Content (SC) | 12 (-) | 85 (-) | 33 (3) | 163 (19) |
| Protein Content (PC) | 30 (5) | 144 (2) | 39 (-) | 148 (14) |
| PM resistance under field (PMfld) | 18 (-) | 79 (-) | 44 (5) | 320 (27) |
| PM resistance under detached leaf assay (PMdla) | 17 (-) | 62 (-) | 19 (-) | 121 (2) |

### 4.3 GWAS for agronomic and quality traits

With the development of the high-throughput SNP genotyping platforms, genome-wide association studies were conceived in barley (Close et al. 2009; Waugh et al. 2009). The success and potential of genome-wide association studies in barley germplasm collections is published recently (Cockram et al. 2010; Comadran et al. 2011b; Comadran et al. 2009; Roy et al. 2010; Wang et al. 2011; Waugh et al. 2010). In the present study GWAS in different gene pools were performed and significant associations are found across the gene pools. Some of the significantly associated SNPs are close to the known QTL and major gene regions. Multiple trait associations at many chromosomal regions have been observed in this study (see Fig 3.42, 3.43).

In present GWAS, significant marker-trait associations for agronomic and quality traits were observed across the genome in different gene pools. Several seed traits SNP associations were found located near known rowed-type genes *Vrs3* on chromosome 1H, *Vrs1* on chromosome 2H and *int-C* region on chromosome 4H (Pourkheirandish and Komatsuda 2007) (Sup. Table S36, S37, S38, S39). Even though population structure was controlled by using kinship in the association analysis, row type genes were significantly associated with several traits. Interestingly row-type associated regions showed significant association with grains per ear, thousand grain weight, grain width and sometimes with starch and protein content but not with grain yield. This means row-type can change the spike morphology hence some grain traits like grains per ear were significant but cannot influenced the grain yield. This is perhaps, because of the complex breeding history of two rowed and six rowed barley that significant associations were observed. In order to see thousand grain weight association independent from row-type, genomic regions for *Vrs* associated regions were used as cofactors in the analysis. The results were similar and as expected highly significant region on chromosome 2H and 4H *Vrs1* and *int-C* were not significant. However, the other associated regions remain significant. Thus TGW associations other then row-type regions are more promising and needs further analysis like 1H long arm.

SNP associations for trait barley spike morphology (row type) at *int*-C (4H) is much more pronounced than the associations near the major rowed type related gene *Vrs1* on 2H. In the recent past, the candidate gene of *int*-C was identified as *HvTB1* (*Teosinte Branch1*) and showed

to have effects on size and fertility of the lateral spikelets and tillering. For the European / Western barley germplasm *int-C* has more effect on quantitative variation for the grain traits in comparison to major row type *Vrs1* gene (Komatsuda et al. 2007; Ramsay et al. 2011; Waugh et al. 2010). Our HVCC collections majorly comprised of European barley and hence associations close to *int*-C were more prominent. Furthermore, the resolution of associations is also based on how the traits are classified. In the current study row type phenotype is scored only as two-rowed and six-rowed barley and used for GWAS. The *Vrs1* gene has series of alleles and *deficiens* types are controlled by a third allele but still considered as two-rowed (Saisho et al. 2009). Therefore, scoring of barley merely as two-rowed and six-rowed would dilute the exact phenotyping and hence the effects of each locus are not accurately estimated in the analysis. Exploiting the syntenic relationships between barley, rice and *Brachypodium* information we were able to identify the candidate gene for the major rowed type gene *Vrs1*. The detected candidate gene (*HvHox1*) for this association is previously cloned and characterized by Pourkheirandish et al. 2007 (Pourkheirandish and Komatsuda 2007). Significantly associated SNPs (12_30896 and 12_30897) on 2H are from the syntenic region of rice locus LOC_Os07g39320 that corresponds to the *Oshox14* candidate gene of *Vrs*. However SNPs associated for the rowed-type trait in our gene pools near to the *int-C* region were thirteen gene models away in rice from the candidate *int-C* gene on chromosome 4H (Comadran et al. 2011b). Although further refined phenotyping into two-rowed, six-rowed and *deficiens* spikes as reported from Comadran et al. (2011) improved the resolution. Similarly, several of the seed traits showed significant associations at *int-C* and *Vrs1* regions. Nevertheless, these candidate genes in these cases have been cloned recently and can only validate already cloned genes. On the other hand genome-wide association has been successful in plants but relatively few studies were fortuitous in identifying causative variant when underlying gene was not known a priori. One such example is from maize for oleic acid where 553 inbred lines and 8590 loci were used to identify fatty acid desaturase, *fad2*, candidate gene that was 2 kb from the associated genetic marker (Beló et al. 2008). In barley recently causative variant for the anthocyanin pigmentation was identified using GWA combined with comparative analysis (Cockram et al. 2010). Re-sequencing the putative anthocyanin pathway gene *HvbHLH1* identified a deletion resulting in a premature stop codon upstream of the basic helix-loop helix domain which was confirmed both in association and biparental mapping population.

Many of the significant associations for row type in HVCC_W and HVCC_S gene pools were coinciding with the known row type gene positions (Pourkheirandish and Komatsuda 2007). In addition, several significant associations were observed which did not coincide with the map positions of the previously known row type genes (Fig 3.12, 3.29, Sup. Table S9, S24, S36, S37, S38, S39). The discovered novel associations might have indirect effect on the row type morphology or they are trans-acting regulators that control the spike morphology. This also means that all these genes have escaped decades of mutagenesis efforts in barley. With genome-wide scan on natural populations novel regions can be identified that probably was not possible with mutant screening.

Interestingly, unmapped SNP (BK_01) from the gene alcohol dehydrogenase 2 (*Adh2*) had strong association with row type, heading date, grains per ear, thousand grain weight, grain width, grain length, ratio grain width/ length and starch traits in HVCC_W gene pool. Although the role of *Adh2* in aleurone layer was reported (Hanson et al. 1984), their significance on spike morphology and grain traits is not known yet. In our analysis this gene was significantly associated in the HVCC_W gene pool. This indicates that either *Adh2* gene has indirect role in grain traits or some other unknown gene placed near to *Adh2* is affecting these phenotypes.

The SNPs from 4H (~113cM) and 5H long arm (~153cM) were found to be significantly associated with frost tolerance (Sup. Table S20). These regions correspond to *VrnH2* and *VrnH1* vernalization genomic regions (Yan et al. 2004). Another interesting unmapped iSelect SNP (BK4) from the dehydrin gene *Dhn9* was found to be significantly associated with grain width/length, frost and starch content traits (Choi et al. 1999). *Dhn9* gene was previously mapped to 5H in a different biparental population (Choi et al. 1999). Another BOPA2 SNP marker 12_31050 derived from Dhn9 gene which is included into iSelect assay is mapped at 5H, 151 cM. Several other SNPs in this region were significantly associated to the trait frost tolerance (Sup. Table S5, S20). The role of dehydrin genes for cold tolerance and the drought was reported in earlier studies (Kosová et al. 2011; Van Zee et al. 1995; Zhu et al. 2000). However, significant associations with traits like starch content and the ratio W_L are probably pleiotropic effects of dehydrin genes (Ismail et al. 1999; Marzin et al. 2008). Moreover, for frost tolerance, SNP (12_30850) derived from the CBF gene on chromosome 5H was significantly associated in the HVCC_W gene pool. This chromosomal region was reported to play significant

role for frost and cold tolerance in barley (Campoli et al. 2009; Fricano et al. 2009; Stockinger et al. 2007). Number of SNPs associated with frost tolerance and their significances in GWAS improved with iSelect assay compared to BOPA1 (Sup. Table S20). Genomic regions 2H (~58cM) and 5H (~144) were previously known to affect the frost tolerance (Chen et al. 2009b). These regions were significantly associated with genome-wide association of frost tolerance in our study (Sup. Table S20).

Associations for Heading date were coinciding with the major known flowering time gene *PpdH1* (2H) and *HvFT3* (1H) in HVCC_W gene pool. However, no associations were found in HVCC_S gene pool at this region. Although functional SNPs were considered for the iSelect development, these SNPs were monomorphic for the HVCC_S gene pool. Considering *PpdH1* as the major photoperiodic gene, we assumed that this region was fixed in the European spring barleys. The mutated photoperiod insensitive allele *ppdH1* is present in most of the European spring barleys studied (Jones et al. 2008; Turner et al. 2005). Interestingly *PpdH1* was reported as strongly associated flowering time gene in the world-wide spring barley panel across four environments from Germany (Stracke et al. 2009). These differences could be due to the origin of genotypes as in 224 world-wide panel 109 genotypes are from Europe and rest are from East Asia, West Asia, North Africa and America. It suggests that majority of European spring barley have no variation at this locus and other collections like 224 world-wide collection have variation but due to the genotypes from other geographical areas. Associations with heading date were significantly improved using iSelect at the *PpdH1* region in HVCC_W gene pool. Significant associations were observed with SNPs (BK_12, BK_14, BK_15 and BK_16) derived from *PpdH1* gene on chromosome 2H in HVCC_W gene pool for heading date, grain width and grain width/length traits. *PpdH1* is a major heading date gene and has pleiotropic effect on yield and yield component traits which described why strong associations observed for heading and grain traits in this study (Li et al. 2005; Turner et al. 2005; Wang et al. 2010).

The flowering time gene *HvFT1* region on chromosome 7H showed high significance in the spring barley HVCC_S gene pool (Faure et al. 2007). However underlined SNP marker has low MAF that means few cultivars in our spring collection are segregating at this locus. Association of *eam6* genomic region on chromosome 2H in both spring and winter barley gene pool emphasized the relevance of this gene in European barley cultivars (Comadran et al. in prep.).

The chromosomal region near the *eam6* on 2H in HVCC_S gene pool was also significantly associated with many grain traits (Sup. Table 38, 39). This region was also significantly associated in the genome-wide association study of Comadran et al. (2011b). The SNPs corresponding to the genomic regions of heading date candidate genes viz. *PpdH1, HvCo1, eam6* were found significantly associated in HVCC_S gene pool (Comadran et al. 2011b; Cuesta-Marcos et al. 2008). One more strong associated chromosomal region on 5H (~138cM) in HVCC_S gene pool co-localized near the phytochrome C genomic region that is candidate of *Sgh2* rice gene homologues (Szucs et al. 2006; Szucs et al. 2007). Often significantly associated SNPs were found to have minor effects. This signifies the importance of minor effect QTL for local adaptation to different European environmental conditions in winter and spring growth habit. Flowering time is under the control of complex genetic architecture as revealed from the studies of maize (Buckler et al. 2009) and from the model plant *Arabidopsis thaliana* (Brachi et al. 2010; Ehrenreich et al. 2009; Salomé et al. 2011). In this study, in addition to the known QTL regions involved in flowering time many other QTL regions were found significant (Sup. Table S6, S21). These results emphasizes that flowering time in barley is complex and is explained by many small effect QTL.

The plant height associations on chromosome 3H are near to the candidate gene *Sdw1/denso* (Jia et al. 2009) in HVCC_S gene pool (Sup. Table S4). The plant height semi-dwarfing genes are well exploited in cereals like wheat and rice (Hedden 2003; Peng et al. 1999). On the other hand there was no report of widely exploited plant height genes in barley. Although *denso* seems to be the strongest and probably used widely in barley as it was highly significant in this study. In barley, many QTL for plant height have been described that have role in plant height differences (Barley QTL workbook in Graingene). Interestingly, in HVCC_S gene pool highly associated SNP was co-localized near candidate gene *Sdw1/denso.* The significantly associated BOPA1 SNP (11_11172) and iSelect SNP (12_31238) are 28 and 19 gene models away from *GA20* oxidase candidate gene in syntenic comparisons with rice. *GA20* oxidase candidate gene in rice was known as green revolution causing gene and also reported recently in barley affecting yield related traits (Jia et al. 2009; Jia et al. 2011). This genomic region showed significant associations with multiple traits like yield, starch content, protein content and powdery mildew scored under field conditions in this study (Sup. Table S37, S39). Multiple agronomic and

quality traits in barley were co-localized in this region and reported by other studies (Jia et al. 2011; Pasam et al. 2012).

Highly significant SNPs from the iSelect were co-localized near to *uzu* and *Sdw1* genomic regions on 3H for plant height trait (Chono et al. 2003; Jia et al. 2009) (Sup. Table S22, S39). The significance level of the association on chromosomal region 5H (~47cM) for height in the HVCC_W gene pool improved using iSelect SNP chip. Fourteen significant SNPs placed at 5H (~47cM) were significantly associated at this genomic region (Sup. Table S7, S22). This genomic region association was also reported to have significantly associated in the GWAS study of Pasam et al (2012). The significance of associations for plant height on chromosome 3H region (~111cM) has improved with iSelect SNPs in HVCC_S gene pool (Sup. Table S7, S22). SNPs from genomic region 2H (~130cM) was significantly associated for the trait height in HVCC_S gene pool. This region was reported to have associations with height in other studies (Chen et al. 2009a; Comadran et al. 2011b).

Most of the associated genomic regions for starch content were also significant in GWAS of Pasam et al 2012 like 2H (~71cM), 4H (~65cM), 5H (~47cM), 7H (~77cM) and 7H (~140cM). In addition, chromosome 2H (~52cM) and 7H (~84cM) coincides with QTL region of QStr.StMo-2H and QSch7 (Abdel-Haleem et al. 2010; Pasam et al. 2012). For protein content several associations found were coinciding with GWAS of Pasam et al 2012 viz. 1H (~47cM and ~112cM), 2H (~58cM and ~82cM), 3H (~58cM and ~127cM), 5H (~89.38cM and ~176cM), 6H (~49.4cM) and 7H (~34cM) respectively (Sup. Table S18, Fig 3.21).

Grain yield is the most important but complex trait. In this study one of the strongest associations for grain yield genomic region was on 7H (~84.92cM). This region showed significant associations to height and starch content traits in HVCC_S gene pool (Sup. Table S16, S31). The associated SNP (11_11445) at this chromosomal region showed significance level up to 6.32 ($-\log_{10}P$). Interestingly, the same SNP was also reported significant from the association study in Mediterranean barley by Comadran et al. 2011 (Comadran et al. 2011b). This region is promising and robust as many contiguous set of markers showed significant association. Another interesting region on chromosome 1H at 76.2cM had significantly associated SNP (11_20121) for the TGW, Garea, Gwidth, GEar traits in HVCC_W gene pool (Sup. Table S36, S37, S38,

S39). Genomic region on 3H at ~59cM in HVCC_W gene pool significantly associated with yield and this region co-localized with known QTLs regions of *QYld.StMo-3H.1* (Romagosa et al. 1999; Romagosa et al. 1996). On 6H significant (11_21404) association was observed at 82.56cM in HVCC_S gene pool for TGW, Garea, Glength and Gwidth traits. QTL for yield is co-localized (*QYld.BlKy-6H*) at this genomic region (Bezant et al. 1997). Another QTL on long arm near 129cM in HVCC_S barley had significance for TGW, Garea, Gwidth, Glength and row type. QTL for spike grain weight *QSgw.BlKy-6H.2* lies near to the associated region (Bezant et al. 1997).

SNPs at 1H (~100.69cM) were significantly associated with several traits in both winter and spring gene pools viz. row type, grains per ear, grain width by length and protein content in HVCC_W and row type, ear length, grains per ear and ratio of grain width by length in HVCC_S gene pools (Sup. Table S36, S37, S38, S39). Significant positive association in this region is 30 gene model away from the rice flowering time locus homologue of barley *HvFT3* that is known to affect flowering (Faure et al. 2007). Our field trials (spring barley) were conducted under long days condition in Europe and *HvFT3* affect flowering under short days condition (Kikuchi et al. 2009). Therefore we did not detect any associations in this region with the heading date phenotype in spring barley but in winter barley. As winter barley experienced short days before flowering this region has influence in winter barley. However this region was also significantly associated with several grain traits. Earlier, *HvFT3* was reported to be involved in heading of barley as an indirect flowering inducer. (Kikuchi et al. 2009). It indicates *HvFT3* possibly have major role in grain traits and heading date is indirectly affected. In addition, other possibility that a gene located close to *HvFT3* is influencing the grain traits cannot be ignored. Multiple traits were significantly associated on several genomic regions spanning chromosomes (Fig 3.42, 3.43) (Sup. Table S36, S37, S38, S39). One of these chromosomal regions on 1H (~75.45cM), in HVCC_W gene pool, coincided with thousand grain weight, grain area, grain width and powdery mildew detached leaf assay traits. In the HVCC_S gene pool, the same region coincided with rowed type, grains per ear and grain length traits. This region was also significantly associated with grains per spike trait in the recent association studies of Comadran et al (2011b) in Mediterranean barley (Comadran et al. 2011b). Significant associations for row-type, grains per ear, thousand grain weight, grain width and grain length trait were observed on chromosome 1H

(96.92cM) in HVCC_S gene pool. Co-localizing of multiple traits in some genomic regions may be because of epistatic or pleiotropic effects of major genes. Many genes in barley have been reported to have strong pleiotropic effect on many agronomic like *Gpert*, *denso* and *int-C* (Comadran et al. 2011b; Ramsay et al. 2011; Thomas et al. 1991). In the current studies apart from the above mentioned loci, several other genomic regions affecting multiple traits were detected across the genome (Fig 3.42, 3.43). Associations observed in this study have high significance in comparison to the world-wide spring barley collection (Pasam et al. 2012) (Pasam et al. unpublished). One of the reasons could be that the extensive LD and less structuring has been seen in the present study as compared to sub-structuring observed in the world-wide collection due to geography and row-type.

### 4.4 GWAS for powdery mildew resistance

In this study, powdery mildew resistance was characterized under field and under controlled conditions using the detached leaf assay in the laboratory. The powdery mildew infection was moderate to high under field conditions as reflected from the repeatability of the traits that is affected by environment (0.59 and 0.69 in HVCC_W and HVCC_S gene pools). In comparison to the cultivars few land races were resistant in the PM under detached leaf assay (Fig 3.7). The plausible reason could be that isolates (78P and D12/12) used in this study were from Europe and the Syrian and Jordanian land races probably had not experienced the hyper-virulent races of powdery mildew in their distribution range. Thus LRC were not selected in their target environment against the powdery mildew resistance. On the other hand significantly high numbers of accessions in HSC were resistant under PMdla. Wild barley is known as the rich source of resistance against powdery mildew and thus the relatively high number of resistant accessions found in this study is not astonishing (Dreiseitl et al. 2006; Jahoor and Fischbeck 1987; Kintzios and Fischbeck 1996). As it is well known from different studies, powdery mildew resistance is complex and controlled by many QTL. We found associations spanning all the chromosomes and across the genepools. Many SNPs were found to be significantly associated with PMfld in HVCC_W (79 SNPs) and HVCC_S (320 SNPs) gene pools (Sup. Table 19, 34, 35).

At some genomic regions associations coincided across the gene pools (Sup. Table 19, 34, 35). However, variable LD decay and ascertainment bias in the gene pools caused problems when comparing the associations across the gene pools (Fig 3.2, 3.4). Nevertheless, some genomic regions had significant associations across two or more gene pools: 1H (~55cM), 3H (~69.60cM), 4H (~96.60cM and ~103cM) and 5H (~47.02cM, ~126.50cM, ~134cM, ~155cM) and 7H (~132cM) for the powdery mildew field infection. Detection of the associations at same genomic regions in more than one genepool already provides validation for the detected marker trait association. SNPs from the genomic region near *Mla* cluster genes on chromosome 1H short arm were significantly associated in the HVCC_S gene pool. This signifies the importance of *Mla* alleles in European spring barley cultivars. Also recently Comadran et al. (2009) in genome-wide association study using Mediterranean barley germplasm has reported *Mla* genomic region as significantly associated. This highlights that cultivated barley germplasm at the *Mla* locus was quite often used in barley breeding. Most significant associations in the HVCC_W gene pool for the powdery mildew field trait were at the long arm of chromosome 5H (123cM). This region corresponds to powdery mildew disease resistance QTL *Rbgq3* (Aghnoum et al. 2009; Schweizer and Stein 2011) (Sup. Table S19, Sup. Table S34). The same region was found significantly associated to powdery mildew detached leaf assay resistance in the LRC gene pool. Incidentally, another major locus *Ml*(*TR*) is also co-localized in the same region on 5H as reported by Maroof et al. (1994). Thus this region is important and needs further evaluation in order to discover the candidate gene at this associated region.

Another interesting and highly significant association in HVCC_S gene pool was at ~150cM on chromosome 2H. This region had significant associations for both powdery mildew under field as well as under detached leaf assay conditions. Moreover, this region was significantly associated in the HVCC_Genobar gene pool. This region corresponds to the resistance *Rbgq9* QTL which was reported from different populations and was detected to confer resistance both at seedling and adult plant stage (Aghnoum et al. 2009). Major gene *MlLa* also reported from the same chromosomal region. However it was highly susceptible with the virulent isolates used in this study (see Table 2). Thus the observed resistance plausibly originated from the QTL and most probably *Rbgq9* QTL. Nevertheless still the possibility of new *MlLa* allele more effective against the virulent isolates cannot be excluded.

Associations found for powdery mildew detached leaf assay (PMdla) screening condition did not coincide to any known major powdery mildew gene. This agrees well with the design of the powdery mildew detached leaf assay experiment as this assay was solely designed to allow screening of non-race specific basal resistance. However slightly significant associations observed at the *Mla* genomic regions in HVCC_S gene pool. Even though virulent isolates used in this study showed susceptibility with *Mla* containing differentials (see Table 2). Thus the significance showed at this region of chromosome 1H could be due to some un-described QTL or *Mla* loci from this region.

LD declines within 1cM interval in the HSC gene pool (Fig 3.4) and thus relatively more markers are required to cover all the regions and detect significant associations. Despite the fact that wild barley gene pool had significantly reduced LD we still were able to detect significant associations for trait PMdla in wild barley gene pools using BOPA1 SNPs. These associations with BOPA1 SNPs could be the result of the strong LD between the causal or functional SNPs present near to the associated chromosomal regions. The observed significant associations in wild barley could be very near to the candidate genes or else the associated genomic region is itself a candidate gene as visually no significant LD was observed in the wild barley populations. In the current GWAS for PMlda with BOPA SNPs in wild barley collection (HSC) revealed 19 significant marker trait associations. Recently Roy et al (2010) also reported high number of QTL (13) for spot blotch resistance using 318 wild barley using DArT and SNP markers. Interestingly, some associated SNPs in wild barley (HSC) for the PMdla were significant across the gene pools. For instance, SNPs at genomic regions 2H (~54cM), 2H (~139cM), 6H (~65.03cM), 7H (~79cM) were significant in both HSC and HVCC_S; SNPs from genomic regions 2H (129cM) are commonly significant in HSC and HVCC_W; and SNPs from genomic regions 3H (28cM), 7H (74cM) are significant across HSC and LRC gene pools (see Sup. Table S19). These associated regions can be used to identify the candidate genes. Furthermore, increasing SNP markers or re-sequencing of the genes near to the physical interval could be investigated to get closer to the candidate genes.

Powdery mildew resistance under field and detached leaf assays yielded several significant associations across the gene pools using iSelect chip. On chromosome 1H at the telomeric end (~2-10cM) significant associations were observed for both traits (PMfld and PMdla) in the

HVCC_S gene pool. The same region was found significantly associated with powdery mildew detached leaf assay condition in HVCC_Genobar gene pool. *Mla* gene cluster is located at this hot spot and is known to confer approximately 30 distinct resistance specificities against barley *Bgh* interactions (Jørgensen 1994; Schulze-Lefert and Vogel 2000; Wei et al. 1999). This region was known to confer complete resistance to intermediate resistance (such as *Mla7*, *Mla10*, and *Mla12)* (Boyd et al. 1995; Caldo et al. 2004). As discussed earlier virulence of the isolates to the *Mla* containing differentials means possibly not known *Mla* gene but some QTL at this locus could impart resistance in PMdla (See Table 2). Still significant association observed in this study could be due to the qualitative and quantitative resistant locus *RMo1* that confers multiple pathogen resistance as reported recently (Inukai et al. 2006). However this chromosomal region was significantly associated with field powdery mildew infection. This means *Mla* alleles are effective to unknown isolates of *Bgh* which are probably widespread under field conditions at IPK. Effectiveness of *Mla* under field condition but not in detached leaf assay confirms that race specific and non-race specific condition resistance are under control of different genetic mechanisms. Significant marker trait associations for the powdery mildew resistance under field conditions were observed on chromosome region 1H (~50cM) in HVCC_W, HVCC_S gene pools. This region was also significant in the HVCC_S and the HVCC_Genobar gene pools with the detached leaf screening (PMdla). *Ror1* gene and *Rbgq6* basal resistance QTL are located nearby at this genomic region on 1H (Aghnoum et al. 2009; Freialdenhoven et al. 1996).

In the HVCC_W gene pool, significant marker trait associations were observed under both PMfld and PMdla on 1H (~95cM). This region was also significantly associated in HVCC_Genbar gene pool in PMdla condition. The *Mlga* resistance gene coinciding at this genomic region and could be plausible candidate for this association (Jørgensen 1994). Significant marker trait associations were found under the powdery mildew field trait screening at the telomeric region of 2H in the HVCC_W and the HVCC_S gene pools. This region corresponds to the meta-QTL for PM disease resistance located on the telomeric end of the 2H short arm (Schweizer and Stein 2011). One of the most interesting associations in HVCC_S gene pool was on chromosome 2H (~150cM) under PMfld and PMdla conditions (Sup. Table S19, S34, S35). The same region was significantly associated in HVCC_Genobar gene pool. This region co-localized near to the QTL *Rbgq9* that was found as a QTL controlling basal resistance

in seedling and adult plant stages (Aghnoum et al. 2009). This genomic region also corresponds near to the meta-QTL position on 2H long arm (Schweizer and Stein 2011).

On chromosome 3H at ~25cM, significant SNP association was observed in the HVCC_W gene pool for both PMdla and PMfld (Sup. Table S19, S34, S35). The region remained significant for powdery mildew screening under detached leaf assay in HVCC_S and the HVCC_Genobar gene pools. This region co-localized to the QTL *Rbgq10* that was known to confer basal seedling resistance (Aghnoum et al. 2009). Significant marker trait associations on the same chromosome 3H at ~74cM were observed for both powdery mildew traits in the HVCC_S gene pool. In addition, this genomic region was significantly associated in the HVCC_Genobar under powdery mildew detached leaf assay. The QTL *Rbgq2b is* co-localized at this region and confers basal resistance (Aghnoum et al. 2009). On the long arm of 3H at ~140cM significant SNP marker association for PMdla was observed in the HVCC_S and the HVCC_Genobar gene pools. QTL for powdery mildew resistance under seedling and adult plant stage was reported form this region (Backes et al. 2003).

Significant association was detected on chromosome 4H (~54cM) for PMdla in HVCC_S and HVCC_Genobar gene pools. Powdery mildew resistance QTL *Rbgq12* and *Mlg* were known from this region of genome (Aghnoum et al. 2009). However in our study *Mlg* resistance can be excluded as PM isolates was known to break this resistance (Table 2), so most probably the other QTL conferred the resistance. On the long arm on chromosome 4H (~108cM), significant marker trait association in HVCC_S gene pool for both powdery mildew traits was observed. Also for PMdla in HVCC_W gene pool this genomic region was found significantly associated. This region co-localized within the interval of QTL reported from von Korff et al. (2005) study using wild barley introgression. On chromosome 5H at position ~46cM, significant associations were found for both traits in HVCC_S gene pool. This region was significantly associated for PMdla trait in HVCC_Genobar gene pool. This chromosomal region co-localized with the QTL *Rbgq14* that caused adult plant resistance (Aghnoum et al. 2009). Additionally in many QTL studies this region is reported to co-segragate and is interesting hot-spot for powdery mildew resistance (Schweizer and Stein 2011). On the long arm of chromosome 5H (~165cM) in HVCC_S gene pool significant SNP associations were observed for both powdery mildew traits. Additionally in HVCC_Genobar gene pool this region was significantly associated for PMdla. The basal

resistance QTL *Rbgq16* co-localized at near to this chromosomal region on 5H (Aghnoum et al. 2009). The *Ror2* loci that mediated *mlo* basal resistance was also reported from this region of genome (Collins et al. 2003; Freialdenhoven et al. 1996). However in the study of Aghnoum et al. (2009) *Ror2* was considered as candidate for the basal resistance but the resistance found in their study did not co-segregate with *Ror2* candidate gene. Nevertheless, in our study associated genomic regions can extend to the larger cM distances depending upon the LD decay so we cannot exclude the *Ror2* gene as candidate for this region. It is not still clear, if the associations observed at this location are due to *Ror2* or some un-known gene in near vicinity on the chromosome 5H at this region. Chromosomal region on 6H (~123cM) in the HVCC_W gene pool was significantly associated under field and powdery mildew detached leaf assay condition. Also significant association was observed at this genomic region in HVCC_Genobar gene pool under powdery mildew detached leaf assay condition. This region is co-localized near to the basal resistance QTL *Rbgq19* that conferred resistance at seedling stage (Aghnoum et al. 2009).

Interestingly, some associated SNPs in wild barley (HSC) for the PM detached leaf assay were significant across the gene pools. For instance, SNPs at genomic regions 2H (~54cM), 2H (~139cM), 6H (~65.03cM), 7H (~79cM) were significant in both HSC and HVCC_S; SNPs from genomic regions 2H (129cM) are commonly significant in HSC and HVCC_W; and SNPs from genomic regions 3H (28cM), 7H (74cM) are significant across HSC and LRC gene pools(see Sup. Table S19). Recently Roy et al (2010) also reported thirteen QTL for spot blotch resistance using 318 wild barley using DArT and SNP markers. The observed significant associations in wild barley could be very near to the candidate genes as visually no significant LD was observed in the wild barley populations (Fig 3.4).

## 4.5 Conclusions and Outlook

The discovery of Quantitative Trait Loci (QTL) has been revolutionized using Genome-Wide Association Scans (GWAS). The multifold advantage of association mapping over traditional bi-parental mapping, like high resolution and use of natural populations, has added great advantage over traditional bi-parental mapping. However, the power to capture the common variant of the phenotype at moderate to high resolution depends upon the extent of linkage disequilibrium (LD) in the population. In the present study, diverse barley panels were analyzed. The extent of LD found in these panels was variable. LD ranged between <1cM in wild barley to 14cM in the winter barley panel. In total, 16 traits were scored in winter barleys and 15 traits in spring barleys and significant marker-trait associations were identified by GWAS. Co-localization of multiple marker-trait associations at major genes genomic regions indicated pleiotropic effects of these associations. Associations of several agronomic traits at the major row-type loci, *Vrs1* and *int-C* suggested that breeding efforts were mostly confined to row-type and growth habit. Statistical inclusion of kinship estimates should correct for population structure related associations at these loci. However significant marker-trait associations at these loci indicate the use of different sources of row-type in the breeding germplasm. The discovery of segregating significant associations, specific to different gene pools, reflected the allelic diversity at these genomic regions in the barley gene pools. In most cases, several significant marker-trait associations were observed over all seven barley chromosomes. This result suggested that yield and yield related traits are under the control of several QTL as expected for quantitative traits.

Another interesting finding was the low LD in wild barley and landraces. This observation has great potential and can be exploited in candidate gene discovery. High marker coverage is required in wild barley and landraces considering such a low LD pattern. However, only a set of 1536 SNP markers were used in this study. During the analysis, significant numbers of SNPs were excluded from the analysis because of their low minor allelic frequencies (MAF). Therefore, only an inadequate amount of SNP markers comprising 582 SNPs in LRC and 813 SNPs in HSC were considered for genome-wide association analysis. This led to ascertainment bias of the BOPA1 markers in determining the associations at the genome-wide level. Based on such biased marker coverage, the intended comparison of associated regions across the gene pools was not feasible any more. Ascertainment bias evidently affected the results as

significantly associated SNP markers were eliminated from the analysis due to their low MAF. Several association studies have emphasized that the favorable alleles are mostly rare and regular SNP genotyping methods are prone to miss these allele effects as rare alleles tends to be with low MAF ($\leq 0.05$). In order to remove such biasness, a relatively high number of SNP markers is needed to cover the genome in order to identify candidate genes in populations with extremely low LD like wild barley. Considering the total genetic distance of the seven chromosomes of 1085.77cM and taking into account the LD decay of < 1cM in wild barley - approximately 2170 equally spaced polymorphic SNP markers would be required in order to identify all significant marker-trait associations in wild barley.

Using iSelect SNP markers (in total 7842 SNPs on the array) in winter and spring barleys significantly increased the number of associations and the significance level of marker-trait associations. Thus high marker coverage using iSelect was advantageous in identifying marker-trait associations at genome-wide level. The marker-trait associations observed in this study showed a high level of significance compared to the findings in the world-wide spring barley collection. Reasons could be I) more extensive LD, and II) less population structure in the present study - as compared to extensive sub-structuring due to geographic origin and row-type observed in the world-wide spring collection.

With the availability of next generation sequencing platforms and high throughput multiplexing of genotypes, the cost of sequencing reduced several fold in comparison to Sanger sequencing (Metzker 2010). In addition, improvement in enrichment techniques and massively parallel sequencing approaches will make genotyping by sequencing an cost effective and powerful approach (Mamanova et al. 2010). Such technologies definitely increase the power of detection in low LD gene pools. Few successful examples from the sequenced *Arabidopsis* and rice genome elucidated recently that genotyping by sequencing has the power to unravel complex quantitative traits also in cereals (Atwell et al. 2010; Cao et al. 2011; Huang et al. 2010b; Huang et al. 2011).

In the future, associations discovered in this study could be validated through carefully generated bi-parental mapping populations. Then the underlined candidate gene(s) can be characterized using mutants and transgenic approaches. In several instances, candidate genes can be discovered

using syntenic and colinearity information from sequenced grass genomes like rice, *Brachypodium* and *Arabidopsis*.

As shown in this study, the extent of LD differs significantly in the gene pools. Such information can be used to zoom in at the candidate genes. For example, low LD in wild barley gene pools could be exploited if the phenotype is also segregating in wild barley. Genotyping and detailed phenotype data generated for cultivars in this study can be used for genomic selection of related germplasm in plant breeding. Moreover, from the breeding perspective it is important to identify the genomic regions that significantly contribute to the phenotype of interest. Studies at the genome-wide level uncovered several gene pool specific associations that provide the gateway in order to the efficient use of genetic resources for breeding that were not tapped yet. Diversity in the studied gene pools can be used for allele mining strategies as shown by (Bhullar et al. 2009). In addition, re-sequencing of wild barley and cultivated populations and comparing their sequences can discover candidate genes that are diverse in wild but are fixed in cultivated germplasm due to genetic bottlenecks (Tenaillon et al. 2004; Wright et al. 2005; Xu et al. 2012; Yamasaki et al. 2005). Sequence-informed conservation and utilization approaches of Genbank accessions can be applied to identify novel alleles for sustainable crop improvement (Kilian and Graner 2012). The future of the crop improvement relies on the re-sequencing and comparing different related plant genomes and developing new genetic mapping strategies and evolutionary analysis to discover the genetic variation (Morrell et al. 2012). As the barley genome remains un-sequenced, high resolution genetic platforms like Illumina 9K iSelect array are one of the best alternatives that can be used for GWAS. However, custom made SNP markers from wild barley populations could be much more effective in the future - as they may not have low MAF problems. In addition, the physical map of barley is envisioned (International Barley Sequencing Consortium) and will be available soon. This will definitely improve the candidate gene discovery and marker development in barley.

## 5 Summary

Genome-wide association scan (GWAS) is a powerful new tool for investigating the genetic architecture of complex traits in crops. The basic strategy of GWAS includes extensive phenotyping, genotyping (genome-wide) and statistical analyses - to relate genotype and phenotype. Despite this promising and straightforward approach, one of the major concerns in association mapping is population structure that is inherent in almost all natural populations and the extent of LD. Population structure, if not addressed properly, causes false positives. The extent of LD estimates the resolution of associations and is mostly population dependent.

The present study was conducted at the genome-wide level to identify marker-trait associations of agronomic traits with a special focus on powdery mildew disease resistance in barley. In this study, diverse barley gene pools were selected comprising of cultivars from Europe [winters (112 genotypes) and springs (282 genotypes)], landraces from Syria and Jordan (238) and wild barleys (216) from Israel. These genotypes were genotyped using a set of 1536 biallelic SNP markers (BOPA1) spanning all seven chromosomes. To study population structure, the principal component analysis (PCA) was used (among others) for all genotypes of all populations. The major clustering was based on the population type (viz. cultivars, landraces and wild barley). Furthermore, landraces and wild barley were well separated from cultivars. Within the cultivars, further clustering was observed based on growth habit and spike morphology (row type). In addition, within landraces (comprising only two-row type) sub-structuring was evident due to the origin (viz. Syria and Jordan). Wild barleys were completely distant from all other panels. Only low diversity within the wild barley panel was detected using the BOPA1 array. This observation is in contrast to the level of diversity found in this panel using microsatellite markers. Ascertainment bias of SNP markers is expected to be the main reason.

The resolution of association mapping is based on the extent of LD within the panel. Population structure, selection, drifts, or physical linkages are the major forces that determine the nature and extent of LD. In the cultivated gene pool, high LD was observed in comparison to landraces and wild barleys. Interestingly, trends of high LD near centromeres and low LD at the telomeric regions of the chromosomes were observed. In comparison to the cultivars (winter and spring), lower LD was observed in landraces, reaching values below <1cM in wild barley. Significant inter-chromosomal LD was observed in cultivated and landraces that could affect the association

results and can cause false positives if not accounted for. Thus, comparatively robust and reliable mixed model analysis using kinship estimates from markers was used for association analysis.

Phenotypic data was generated in multi-environmental field trials that were conducted at four locations (CRA – Fiorenzuola d'Arda, Italy; JHI – Dundee, UK foremerly SCRI; IPK – Gatersleben, Germany; UniHalle – Halle, Germany) with two replications - in two consecutive years (2008-09 and 2009-10). In total, 16 traits in winter and 15 traits in spring barley gene pools were scored. High variation and moderate to high heritability was observed for the measured traits that showed the robustness of the phenotype. Initially, GWAS were performed using 1536 SNP markers (BOPA1) that yielded significant marker-trait associations. In a second step the iSelect array was applied for association scans (in total 7842 SNPs on the chip). Significant associations were observed for row-type at the *Vrs3*, *Vrs1* and *int-C* genomic regions on chromosome 1H, 2H and 4H. In addition to the previously known *Vrs* loci, highly significant associations were observed on chromosome 3H (long arm), 1H (long arm) and 7H (long arm). Associations at *Vrs* loci specifically at *Vrs1* and *int-C* were also highly significant for the traits grains per ear, thousand grain weight, grain area, grain width, protein content and starch content - but not for total yield. Chromosomal regions near *CBF* genes (5H) showed significant associations for frost tolerance in winter barleys. Highly significant marker-trait associations for heading date were observed at the *Ppd-H1* locus in winter barley. This region also showed significant associations with grain related traits like grain width and ratio of grain width and length traits. Another interesting region near *HvFT3* on 1H was significantly associated with heading date and several grain traits in winter barley. Interestingly, several SNPs at *PpdH1* were monomorphic in spring barley and thus *Ppd-H1* was not showing any significance here. This indicates that the mutated (insensitive) *ppd-H1* allele is fixed in the European spring barley gene pool studied here. Interestingly, the *HvFT1* chromosomal region on 7H was highly significant although the associated marker showed low allele frequency. This indicates that few cultivars in our panel were segregating at this locus. Another highly significant association in spring barley was found at ~ 108cM on 7H using iSelect SNP markers. This region was not significant using BOPA1 SNP markers. The number of significantly associated SNPs increased using iSelect markers and in most cases also the level of significance increased. Thus highly dense marker coverage (using iSelect) was quite successful in identifying significant associations at several chromosomal regions that we missed during genome-wide scan using 1536 BOPA1 markers.

The significance level near centromere on chromosome 2H for heading date increased at the genomic region of *eam6* in spring barley. One highly significant associated region for height was located near the *Sdw3* region on 2H (~ 63.53cM) in winter barley and 3H (~126cM) near to the *sdw1/denso* genomic regions. One highly significant association for grain yield, height and spike length was observed on chromosome 3H at ~54cM in spring barley.

Powdery mildew is a major barley disease that causes severe yield losses worldwide. Breeders often deployed race specific resistance genes in breeding but these are not durable. However, the less studied race non specific basal resistance is durable, has quantitative inheritance and cannot be overcome easily by pathogens. In order to characterize the powdery mildew basal resistance at a genome-wide level, we screened 282 spring and 112 winter barley cultivars, 317 landraces, 216 wild barleys and a world-wide collection of 224 spring barleys. In order to detect basal resistance in a robust manner, a detached leaf assay was performed using two poly-virulent races having overcome 44 major resistance genes. More resistant genotypes were observed in spring (54%) and wild (82%) barley gene pools than in the winters (18%), landraces (20%) and the 224 world-wide springs (27%). In addition, powdery mildew resistance was monitored under field conditions in two years at IPK. A number of significant associations (P-value $\leq 0.01$) was found on all seven chromosomes across the gene pools using BOPA1 SNPs – thus suggesting a complex genetic set-up of broad-spectrum basal resistance. In some cases, associated chromosomal regions mapped in proximity to known QTL for disease resistance - underlining the accuracy of the association mapping approach for basal resistance. Interestingly, wild barley yielded significant associations even though LD was extremely low but the associated regions were with low significance level. In some regions, the denser SNP coverage led to increased significances of the associations. On chromosome 2H (~149cM) one significant association in spring barley under field and detached leaf assay conditions was found. This region also showed marker-trait significance in the 224 world-wide spring barley panel. Finally, genome-wide associations in diverse gene pools of barley were successfully performed and population specific significant associations were observed. In several cases, candidate genes underlying the association can be identified using comparative syntenic information from sequenced grass genomes like rice and *Brachypodium*. In future, several of the identified regions can be dissected in detail using e.g. re-sequencing or mutant screening approaches.

## 6 Zusammenfassung

Genomweite Assoziationsstudien (GWAS) stellen einen neuen, leistungsstarken Ansatz dar, die genetische Architektur komplexer Merkmale in Kulturpflanzen zu untersuchen. Die Basis der GWAS besteht aus der extensiven Phänotypisierung und Genotypisierung (genomweit) sowie der statistischen Analyse, um Genotyp und Phänotyp miteinander zu korrelieren. Die vielversprechenden GWAS beinhalten aber auch zwei bedeutende und wichtige Herausforderungen – die Populationsstruktur sowie das Ausmaß des *Linkage Disequilibriums* (LD). Wird die Populationsstruktur nicht adäquat einbezogen, kann sie bei GWAS Falsch-Positive verursachen. Das Ausmaß des LD wiederum bedingt die Auflösung der Assoziationen und ist größtenteils populationsabhängig.

Die vorliegende Studie enthält die Identifizierung von Marker-Merkmals-Assoziationen für agronomische Merkmale mit speziellem Focus auf die Mehltauresistenz in Gerste. Hierfür wurden verschiedene Genpools ausgewählt, die europäische Sorten (112 Wintergenotypen), Sommergersten (282), Landrassen aus Syrien und Jordanien (238) sowie Wildgersten aus Israel (216) umfassen. Diese Genotypen wurden mit 1536 biallelischen SNP-Markern (BOPA1) genotypisiert. Die Hauptkomponentenanalyse (PCA) wurde angewandt, um die Populationsstruktur innerhalb der Gesamtheit an Genotypen und in den verschiedenen Genpools zu untersuchen. Die Trennung basiert demnach hauptsächlich auf dem Populationstyp (Sorten, Landrassen und Wildgerste). Innerhalb der Sorten wurde eine weitere Differenzierung beruhend auf dem Wuchstyp und der Ährenmorphologie (Zeiligkeit) beobachtet. Die Wildgersten waren von allen anderen Populationen am weitesten entfernt. Bei ihnen wurde nur eine geringe genetische Diversität mit dem BOPA1-Array detektiert. Diese Beobachtung steht im Gegensatz zum Ausmaß der Diversität, die in demselben Panel mit Mikrosatellitenmarkern gefunden wurde. Der Hauptgrund für diese geringe Diversität liegt im sog. ‚*ascertainment bias*' der SNP-Marker in den Wildgersten.

Die Auflösung bei GWAS hängt ab vom Ausmaß des LD im ausgewählten Panel. Populationsstruktur, Selektion, Drift oder echter physischer Kopplung sind die Hauptkräfte, die die Natur und das Ausmaß des LD bestimmen. Im Genpool der Sorten war ein hohes LD im Vergleich zu Landrassen und Wildgerste vorhanden. Dabei wurde ein Trend zu hohem LD in den Centromer- und in den Telomerregionen der Chromosomen festgestellt. Im Vergleich zu Kultursorten wurde in den Landrassen ein geringeres LD beobachtet, in Wildgerste erreichte es

Werte <1cM. In Landrassen und Sorten wurde signifikantes inter-chromosomales LD detektiert, welches die Ergebnisse der Assoziation beeinflussen und ebenfalls zu Falsch-Positiven führen könnte. Deshalb wurde die robuste Methode der gemischten Modelle mit Einbezug der Verwandtschaftsbeziehungen (‚*kinship*'), basierend auf den Markerinformationen, zur Analyse der Assoziationen genutzt.

Die phänotypischen Daten wurden in Feldversuchen an vier Standorten (Fiorenzuola d'Arda, Italien; Dundee, UK; Gatersleben, Deutschland; Halle, Deutschland) in zwei aufeinanderfolgenden Jahren (2008-09 und 2009-10) erhoben. Insgesamt wurden 16 Merkmale im Winter- und 15 im Sommergenpool untersucht. Es war eine hohe Variation und eine moderate bis hohe Heritabilität für die untersuchten Merkmale zu beobachten. Als Erstes wurden die GWAS mit 1536 SNP Markern (BOPA1) berechnet, was signifikante Marker-Merkmals-Assoziationen ergab. In einem zweiten Schritt wurden die SNPs des iSelect-Arrays für die GWAS genutzt (insgesamt 7842 SNPs). Signifikante Assoziationen für Zeiligkeit wurden für die *Vrs3*, *Vrs1* und *int-C* Genomregionen gefunden. Zusätzlich zu den schon bekannten *Vrs*-Loci, wurden hochsignifikante Assoziationen auf 3HL, 1HL und 7HL gefunden. Assoziationen an den *Vrs* Loci und speziell für *Vrs1* und *int-C* waren ebenfalls hoch signifikant für das Korngewicht der Ähre, Tausendkorngewicht, Kornfläche und –breite, Protein- und Stärkegehalt, nicht jedoch mit dem Ertrag an sich. Regionen nahe den *CBF*-Genen (5H) zeigten signifikante Assoziationen für Frosttoleranz in Wintergerste. Hoch signifikante Assoziationen für den Zeitpunkt des Schossens wurden für den *Ppd-H1*-Locus in Wintergerste gefunden. In dieser Region gab es weitere signifikante Assoziationen zu Kornmerkmalen wie Kornbreite und das Verhältnis Breite zu Länge. Eine weitere interessante Region nahe *HvFT3* auf 1H war signifikant mit dem Zeitpunkt des Schossens und mehreren Kornmerkmalen in Wintergerste assoziiert. Interessanterweise waren verschiedene SNPs von *PpdH1* in Sommergerste monomorph, weshalb *Ppd-H1* in dieser Population nicht signifikant assoziiert war. Das weist darauf hin, dass das insensitive *ppd-H1*-Allel in den hier untersuchten europäischen Sommergersten fixiert ist. Weiterhin war auch die Region um *HvFT1* auf 7H hochsignifikant assoziiert, obwohl der Marker nur eine geringe Allelfrequenz aufwies. Daraus lässt sich schlussfolgern, dass nur wenige Sorten in diesem Material für diesen Locus segregierten. In Sommergerste wurde mit den iSelect SNPs eine weitere hochsignifikante Assoziation bei ~ 108cM auf 7H detektiert. Mit den BOPA1 SNPs konnten dort keine signifikanten Assoziationen gefunden wurden. Die Zahl signifikant

assoziierter SNP-Marker erhöhte sich durch die Nutzung des iSelect-Arrays und in den meisten Fällen auch das Signifikanzniveau. Die hohe Markerabdeckung mit iSelect erwies sich somit als sehr erfolgreich für die Identifikation signifikanter Assoziationen. Für den Zeitpunkt des Schossens erhöhte sich z.B. das Signifikanzniveau in der Sommergerstenkollektion nahe des *eam6* Locus‘s. Eine mit der Pflanzenhöhe hochsignifikant assoziierte Region wurde in der Nähe der *Sdw3*-Region auf 2H (~ 63.53cM) in Wintergerste lokalisiert, eine weitere auf 3H (~126cM) nahe der *sdw1/denso*-Genomregion. Für die Merkmale Korngewicht, Pflanzenhöhe sowie Ährenlänge wurde eine hochsignifikante Assoziation auf 3H bei ~54cM gefunden.

Mehltau ist in Gerste eine bedeutende Krankheit, die weltweit zu hohen Ernteverlusten führt. Züchter führten oft rassespezifische Resistenzgene in den Zuchtprozess ein, die jedoch nicht dauerhaft sind. Die weniger untersuchte nicht-rassespezifische Basalresistenz ist dagegen dauerhaft, hat eine quantitative Vererbung und kann von den Pathogenen nur schwer überwunden werden. Für die genomweite Charakterisierung der Basalresistenz wurden 282 Sommer- und 112 Wintergerstensorten, 317 Landrassen, 246 Wildgersten sowie eine weltweite Kollektion von 224 Sommergersten untersucht. Um die Basalresistenz verlässlich zu untersuchen, wurde ein sog. ‚*detached leaf assay*’ mit zwei poly-virulenten Rassen genutzt, welche bereits 44 Hauptresistenzgene überwunden haben. Dabei erwiesen sich die Sommertypen der Kultivare (54%) und die Wildgersten (82%) am resistentesten. Weiterhin wurde die Mehltauresistenz unter Feldbedingungen in zwei Jahren am Standort IPK untersucht. Signifikante Assoziationen (P-Wert≤ 0,01) wurden auf allen sieben Chromosomen innerhalb der Genpools mit den BOPA1 Markern detektiert, was die komplexe genetische Struktur der breiten Basalresistenz zeigt. Auch wenn das LD in Wildgerste nur sehr gering war, konnten auch hier signifikante Assoziationen detektiert werden, diese hatten allerdings nur ein geringes Signifikanzniveau.

Abschließend lässt sich sagen, dass genomweite Assoziationen in diversen Genpools erfolgreich detektiert wurden und oft populationsspezifische Assoziationen erhalten wurden. In vielen Fällen können die dahinterliegenden Kandidatengene durch vergleichende Syntänie-Informationen sequenzierter Grasgenome wie Reis und *Brachypodium* identifiziert werden. Zukünftig können viele der identifizierten Regionen im Detail genau untersucht werden, z.B. durch Re-sequenzierung oder dem Screening von Mutanten.

# References


Abdel-Haleem H, Bowman J, Giroux M, Kanazin V, Talbert H, Surber L, Blake T (2010) Quantitative trait loci of acid detergent fiber and grain chemical composition in hulled × hull-less barley population. Euphytica 172:405-418

Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279-292

Aghnoum R, Marcel TC, Johrde A, Pecchioni N, Schweizer P, Niks RE (2009) Basal host resistance of barley to powdery mildew: connecting quantitative trait loci and candidate genes. Molecular Plant-Microbe Interactions 23:91-102

Andrew C (2008) The Gladiator Diet, How to eat, exercise, and die a violent death Archaeology Magazine: 61

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:e60

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627-631

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

Backes G, Madsen LH, Jaiser H, Stougaard J, Herz M, Mohler V, Jahoor A (2003) Localisation of genes for resistance against *Blumeria graminis* f.sp. *hordei* and *Puccinia graminis* in a cross between a barley cultivar and a wild barley (Hordeum vulgare ssp. spontaneum) line. Theor Appl Genet 106:353-362

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

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Beló A, Zheng P, Luck S, Shen B, Meyer D, Li B, Tingey S, Rafalski A (2008) Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. Molecular Genetics and Genomics 279:1-10

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Jornal of the Royal Statistical Society B 57:289-300

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc Lond B Biol Sci 274:227-274

Bezant J, Laurie D, Pratchett N, Chojecki J, Kearsey M (1997) Mapping QTL controlling yield and yield components in a spring barley (*Hordeum vulgare* L.) cross using marker regression. Molecular Breeding 3:29-38

Bhullar NK, Street K, Mackay M, Yahiaoui N, Keller B (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus. Proc Natl Acad Sci USA 106:9519-9524

Bindschedler LV, Burgis TA, Mills DJ, Ho JT, Cramer R, Spanu PD (2009) In planta proteomics and proteogenomics of the biotrophic barley fungal pathogen *Blumeria graminis* f. sp. *hordei*. Mol Cell Proteomics 8:2368-2381

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

Blattner FR (2004) Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. Molecular Phylogenetics and Evolution 33:289-299

Bockelman HE, Valkoun J (2010) Barley Germplasm Conservation and Resources. Barley. Wiley-Blackwell, pp 144-159

Boyd LA, Smith PH, Foster EM, Brown JKM (1995) The effects of allelic variation at the *Mla* resistance locus in barley on the early development of *Erysiphe graminis* f.sp. *hordei* and host responses. The Plant Journal 7:959-968

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet 6

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

Brown AHD, Zohary D, Nevo E (1978) Outcrossing rates and heterozygosity in natural populations of *Hordeum spontaneum* Koch in Israel. Heredity 41:49-62

Brown JK (1994) Chance and selection in the evolution of barley mildew. Trends Microbiol 2:470-475

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Sofia da Silva H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714-718

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695-705

Caldo RA, Nettleton D, Wise RP (2004) Interaction-dependent gene expression in *Mla*-specified response to barley powdery mildew. The Plant Cell Online 16:2514-2528

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.). The Plant Journal 40:143-150

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, Matus-Cádiz MA, Pozniak CJ, Cattivelli L, Fowler DB (2009) Comparative expression of Cbf genes in the Triticeae under different acclimation induction temperatures. Mol Genet Genomics 282:141-152

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43:956-963

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Current Opinion in Plant Biology 11:215-221

Chandler PM, Marion-Poll A, Ellis M, Gubler F (2002) Mutants at the *Slender1* locus of barley cv Himalaya. molecular and physiological characterization. Plant Physiology 129:181-190

Chen A, Baumann U, Fincher G, Collins N (2009a) *Flt-2L* , a locus in barley controlling flowering time, spike density, and plant height Functional & Integrative Genomics 9:243-254

Chen A, Gusta L, Brûlé-Babel A, Leach R, Baumann U, Fincher G, Collins N (2009b) Varietal and chromosome 2H locus-specific frost tolerance in reproductive tissues of barley (Hordeum vulgare L.) detected using a frost simulation chamber. Theor Appl Genet119:685-694

Ching A, Caldwell K, Jung M, Dolan M, Smith O, Tingey S, Morgante M, Rafalski A (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:19

Choi DW, Zhu B, Close TJ (1999) The barley (*Hordeum vulgare* L.) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 *Dhn* genes of cv Dicktoo. Theor Appl Genet 98:1234-1247

Chono M, Honda I, Zeniya H, Yoneyama K, Saisho D, Takeda K, Takatsuto S, Hoshino T, Watanabe Y (2003) A semidwarf phenotype of barley *uzu* results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. Plant Physiology 133:1209-1219

Christiansen MJ, Andersen SB, Ortiz R (2002) Diversity changes in an intensively bred wheat germplasm during the 20(th) century. Molecular Breeding 9:1-11

Clarke B, Liang R, Morell MK, Bird AR, Jenkins CLD, Li Z (2008) Gene expression in a starch synthase IIa mutant of barley: changes in the level of gene transcription and grain composition. Functional & Integrative Genomics 8:211-221

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, 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 10:582

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. Journal of Experimental Botany 58:1231-1244

Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W, O'Sullivan DM (2008) Association mapping of partitioning loci in barley. BMC Genet 9:16

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

Collins NC, Thordal-Christensen H, Lipka V, Bau S, Kombrink E, Qiu JL, Huckelhoven R, Stein M, Freialdenhoven A, Somerville SC, Schulze-Lefert P (2003) SNARE-protein-mediated disease resistance at the plant cell wall. Nature 425:973-977

Comadran J, Thomas WT, 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

Comadran J, Ramsay L, MacKenzie K, Hayes P, Close TJ, Muehlbauer G, Stein N, Waugh R (2011a) Patterns of polymorphism and linkage disequilibrium in cultivated barley. Theor Appl Genet 122:523-531

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 WT, Romagosa I (2011b) Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in Hordeum vulgare in Mediterranean environments. Theor Appl Genet 122:1363-1373

Cox DJ, Frey KJ (1984) Combining Ability and the Selection of Parents for Interspecific Oat Matings. Crop Sci 24:963-967

Cuesta-Marcos A, Igartua E, Ciudad F, Codesal P, Russell J, Molina-Cano J, Moralejo M, Szűcs P, Gracia M, Lasa J, Casas A (2008) Heading date QTL in a spring × winter barley cross evaluated in Mediterranean environments. Molecular Breeding 21:455-471

deVicente MC, Tanksley SD (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585-596

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

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155-162

Devos KM (2010) Grass genome organization and evolution. Curr Opin Plant Biol 13:139-145

Diamond J (1997) Guns, Germs and Steel. Random House, London

Doležel J, Greilhuber J, Lucretti S, Meister A, Lysák MA, Nardi L, Obermayer R (1998) Plant genome size estimation by flow cytometry: Inter-laboratory Comparison. Annals of Botany 82:17-26

Donini P, Law JR, Koebner RMD, Reeves JC, Cooke RJ (2000) Temporal trends in the diversity of UK wheat. Theor Appl Genet 100:912-917

Dreiseitl A, Dinoor A (2004) Phenotypic diversity of barley powdery mildew resistance sources. Genet Resour Crop Ev:251-258

Dreiseitl A, Dinoor A, Kosman E (2006) Virulence and diversity of *Blumeria graminis* f. sp. *hordei* in Israel and in the Czech Republic. Plant Disease 90:1031-1038

Druka A, Franckowiak J, Lundqvist U, Bonar N, Alexander J, Houston k, Radovic S, Shahinnia F, Vendramin V, Morgante M, Stein N, Waugh R (2011) Genetic dissection of barley morphology and development. Plant Physiol. 155: 617-627

Dunford RP, Griffiths S, Christodoulou V, Laurie DA (2005) Characterisation of a barley (*Hordeum vulgare* L.) homologue of the *Arabidopsis* flowering time regulator GIGANTEA. Theor Appl Genet 110:925-931

Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, Purugganan MD (2009) Candidate gene association mapping of *Arabidopsis* flowering time. Genetics 183:325-335

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-1587

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

Felker FC, Peterson DM, Nelson OE (1983) Growth characteristics, grain filling, and assimilate transport in a shrunken endosperm mutant of barley. Plant Physiology 72:679 -684

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

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. Science 333:86-89

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85-88

Freialdenhoven A, Peterhansel C, Kurth J, Kreuzaler F, Schulze-Lefert P (1996) Identification of genes required for the function of non-race-specific *mlo* resistance to powdery mildew in barley. Plant Cell 8:5-14

Fricano A, Rizza F, Faccioli P, Pagani D, Pavan P, Stella A, Rossini L, Piffanelli P, Cattivelli L (2009) Genetic variants of *HvCbf14* are statistically associated with frost tolerance in a European germplasm collection of Hordeum vulgare. Theor Appl Genet 119:1335-1348

Friedt W, Horsley RD, Harvey BL, Poulsen DME, Lance RCM, Ceccarelli S, Grando S, Capettini F (2010) Barley Breeding History, Progress, Objectives, and Technology. Barley. Wiley-Blackwell, pp 160-220

Godfrey D, Bohlenius H, Pedersen C, Zhang Z, Emmersen J, Thordal-Christensen H (2010) Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. BMC Genomics 11:317

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 2:258

Gottwald S, Stein N, Borner A, Sasaki T, Graner A (2004) The gibberellic-acid insensitive dwarfing gene sdw3 of barley is located on chromosome 2HS in a region that shows high colinearity with rice chromosome 7L. Molecular Genetics and Genomics 271:426-436

Grando S, Macpherson H (eds) (2005) Food Barley: Importance, Uses and Local Knowledge. ICARDA, Aleppo, Syria

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

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

Hackett CA, Ellis RP, Forster BP, McNicol JW, Macaulay M (1992) Statistical analysis of a linkage experiment in barley involving quantitative trait loci for height and ear-emergence time and two genetic markers on chromosome 4. Theor Appl Genet 85:120-126

Halterman D, Zhou F, Wei F, Wise RP, Schulze-Lefert P (2001) The MLA6 coiled-coil, NBS-LRR protein confers AvrMla6-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J 25:335-348

Hammer K (1984) Das Domestikationssyndrom. Kulturpflanze 32:11-34

Hanson AD, Jacobsen JV, Zwar JA (1984) Regulated expression of three alcohol dehydrogenase genes in barley aleurone layers 1. Plant Physiol 75:573-581

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

Harlan JR (1971) Agricultural Origins - Centers and Noncenters. Science 174:468

Harlan JR (1995) Agricultural origins and crop domestication in the Mediterranean region. Diversity 11: 14-16.

Harlan JR and Zohary D (1966) Distribution of wild wheats and barley. Science 153:1074–1080.

Haseneyer G, Ravel C, Dardevet M, Balfourier F, Sourdille P, Charmet G, Brunel D, Sauer S, Geiger HH, Graner A, Stracke S (2008) High level of conservation between genes coding for the

GAMYB transcription factor in barley (*Hordeum vulgare* L.) and bread wheat (*Triticum aestivum* L.) collections. Theor Appl Genet 117:321-331

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

Haseneyer G, Stracke S, Piepho HP, Sauer S, Geiger HH, Graner A (2010b) DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. BMC Plant Biology 10:5

Hedden P (2003) The genes of the Green Revolution. Trends in Genetics 19:5-9

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

Hillman GC, Madeyska E and Hather J (1989) Wild plant foods and diet at Late Paleolithic Wadi Kubbaniya: the evidence from charred remains. In F. Wendorf, R. Schild and A. E. Close (ed.) The prehistory of Wadi Kubbaniya. Volume 2. Stratigraphy, paleoeconomy, and environment, 162-242. Dallas, TX, Southern Methodist University Press.

Houston K, Druka A, Bonar N, Macaulay M, Lundqvist U, Franckowiak J, Morgante M, Stein N, Waugh R (2012) Analysis of the barley bract suppression gene *Trd1*. Theor Appl Genet 125:33-45

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F, Li J, Han B (2010a) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961-967

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010b) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961-967

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet advance online publication

Hübner S, Höffken M, Oren E, Haseneyer G, Stein N, Graner A, Schmid K, E F (2009) Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. Molecular Ecology 18:1523-1536

Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103:16666-16671

International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851-861

Inukai T, Vales MI, Hori K, Sato K, Hayes PM (2006) RMo1 Confers Blast Resistance in Barley and Is Located within the Complex of Resistance Genes Containing *Mla*, a Powdery Mildew Resistance Gene. Molecular Plant-Microbe Interactions 19:1034-1041

Ismail AM, Hall AE, Close TJ (1999) Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proc Natl Acad Sci USA 96:13566-13570

Jahoor A and Fischbeck G (1987) Localization of resistance genes against powdery mildew in Hordeum spontaneum C. Koch.Barley Genetics 1V, Proc. 4th Int. Barley Genetics Symp., Okayama

Jakob SS, Blattner FR (2006) A Chloroplast Genealogy of Hordeum (Poaceae): Long-Term Persisting Haplotypes, Incomplete Lineage Sorting, Regional Extinction, and the Consequences for Phylogenetic Inference. Mol Biol Evol 23:1602-1612

Jansen RC, Jannink J-L, Beavis WD (2003) Mapping Quantitative Trait Loci In Plant Breeding Populations. Crop Sci 43:829-834

Jia Q, Zhang J, Westcott S, Zhang XQ, Bellgard M, Lance R, Li C (2009) GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. Funct Integr Genomics 9:255-262

Jia Q, Zhang X-Q, Westcott S, Broughton S, Cakir M, Yang J, Lance R, Li C (2011) Expression level of a gibberellin 20-oxidase gene is associated with multiple agronomic and quality traits in barley. Theor Appl Genet 122:1451-1460

Jones H, Leigh FJ, Mackay I, Bower MA, Smith LM, Charles MP, Jones G, Jones MK, Brown TA, Powell W (2008) Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. Mol Biol Evol 25:2211-2219

Jones JD (2001) Putting knowledge of plant disease resistance genes to work. Curr Opin Plant Biol 4:281-287

Jones JD, Dangl JL (2006) The plant immune system. Nature 444:323-329

Jørgensen IH (1992) Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. Euphytica 63:141-152

Jørgensen JH (1994) Genetics of powdery mildew resistance in barley. Crit Rev Plant Sc:97-119

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

Kasha KJ, Kao KN (1970) High frequency haploid production in barley (Hordeum vulgare L.). Nature 225:874-876

Khlestkina EK, Huang XQ, Quenum FJB, Chebotar S, Roder MS, Borner A (2004) Genetic diversity in cultivated plants - loss or stability? Theor Appl Genet 108:1466-1472

Kiesling R (1985) The disease of barley. In 'Barley' (Ed DC Rasmusson) American Society of Agronomy: Madison, WI: 269–312.

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 Physiology 149:1341-1353

Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in genebanks. Briefings in Functional Genomics 11:38-50

Kilian B, Ozkan H, Kohl J, Haseler AV, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F (2006) Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. Mol Genet Genom 276:230-241.

Kilian B, Özkan H, Pozzi C, Salamini F (2009) Domestication of the Triticeae in the Fertile Crescent in the Fertile Crescent Genetics and Genomics of the Triticeae. In: Muehlbauer GJ, Feuillet C (eds). Springer New York, pp 81-119

Kintzios S, Fischbeck G (1996) Genetic studies on the powdery mildew resistance of winter barley lines derived from *Hordeum spontaneum* accessions collected in Israel in: Genetic resources and crop evolution , ISSN 1573-5109, Vol. 43, p. 471-479

Kislev ME, Nadel D, Carmi I (1992) Epipalaeolithic (19,000 BP) cereal and fruit diet at Ohalo II, Sea of Galilee. Isr Rev Palaeobot Palynol 73:161–166

Kleinhofs A, Kilian A, Maroof MAS, Biyashev RM, Hayes P, Chen FQ, Lapitan N, Fenwick A, Blake TK, Kanazin V, Ananiev E, Dahleen L, Kudrna D, Bollinger J, Knapp SJ, Liu B, Sorrells M, Heun M, Franckowiak JD, Hoffman D, Skadsen R, Steffenson BJ (1993) A Molecular, Isozyme and Morphological Map of the Barley (Hordeum-Vulgare) Genome. Theor Appl Genet 86:705-712

Knupffer H, Terentyeva I, .Hammer K, O. K, Sato K (2003) Ecogeographical diversity—a Vavilovian approach In: R. von Bothmer, Th. van Hintum, H. Knüpffer and K. Sato (eds), Diversity in Barley(Hordeum vulgare). Elsevier B.V., pp 53-76

Koebner RMD, Donini P, Reeves JC, Cooke RJ, Law JR (2003) Temporal flux in the morphological and molecular diversity of UK barley. Theor Appl Genet 106:550-558

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

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

Kosová K, Vítámvás P, Prášil IT (2011) Expression of dehydrins in wheat and barley under different temperatures. Plant Science 180:46-52

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). Molecular Genetics and Genomics 270:24-33

Kou Y, Wang S (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. Current Opinion in Plant Biology 13:181-185

Kouyos RD, Silander OK, Bonhoeffer S (2007) Epistasis between deleterious mutations and the evolution of recombination. TRENDS in Ecology and Evolution 22:308-315

Kraakman AT, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446

Lancashire PD, Bleiholder H, Boom TVD, Langeluddecke P, Stauss R, Weber E, Witzenberger A (1991) A uniform decimal code for growth stages of crops and weeds. Annals of Applied Biology 119:561-601

Li JZ, Huang XQ, Heinrichs F, Ganal MW, Röder MS (2005) Analysis of QTLs for yield, yield components, and malting quality in a BC3-DH population of spring barley. Theor Appl Genet 110:356-363

Linde-Laursen I, Heslop-Harrison JS, Shepherd KW, Taketa S (1997) The barley Genome and its Relationship with the Wheat Genomes. A Survey with an Internationally Agreed Recommendation for Barley Chromosome Nomenclature. Hereditas 126:1-16

Lundqvist U, Franckowiak J, Konishi T (1996) New and revised descriptions of barley genes. Barley Genet Newsl 26:22–43

Lundqvist U, Franckowiak JD, Konishi T (1997) New and revised descriptions of barley genes. Barley Genet Newsl 26:22–516

Lundqvist U, Lundqvist A (1987) An intermedium gene present in a commercial 6-row variety of barley. Hereditas 107:131-135

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends in Plant Science 12:57-63

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 7:6

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. Nat Meth 7:111-118

Marcel TC, Varshney RK, Barbieri M, Jafary H, de Kock MJ, Graner A, Niks RE (2007) A high-density consensus map of barley to compare the distribution of QTLs for partial resistance to Puccinia hordei and of defence gene homologues. Theor Appl Genet 114:487-500

Maroof MAS, Biyashev RM, Zhang Q (1994) Molecular marker analyses of powdery mildew resistance in barley. Theor Appl Genet 88:733-740

Marzin S, Mihaly R, Pauk J, Schweizer P (2008) A transient assay system for the assessment of cell-autonomous gene function in dehydration-stressed barley. Journal of Experimental Botany 59:3359-3369

Mathre DE (1997) Compendium of barley diseases, second edition. Compendium of barley diseases, second edition

Mayer KF, Martis M, Hedley PE, Simkova H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubalakova M, Suchankova P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249-1263

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. Nat Biotech 18:455-457

Metzker ML (2010) Sequencing technologies-the next generation. Nat Rev Genet 11:31-46

Molina-Cano JL, Fra-Mon P, Salcedo G, Aragoncillo C, Togores FR, García-Olmedo F (1987) Morocco as a possible domestication center for barley: biochemical and agromorphological evidence. TAG Theor Appl Genet 73:531-536

Molina-Cano JL, Russell JR, Moralejo MA, Escacena JL, Arias G, Powell W (2005) Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. Theor Appl Genet110:613-619

Moragues M, Comadran J, Waugh R, Flavell A, Russell J (2010) Effects of ascertainment bias and marker number on estimations of barley diversity from high throughput SNP genotype data. Theoretical and Applied Genetics 120: 1525-1534

Morell MK, Kosar-Hashemi B, Cmiel M, Samuel MS, Chandler P, Rahman S, Buleon A, Batey IL, Li Z (2003) Barley *sex6* mutants lack starch synthase IIa activity and contain a starch with novel properties. Plant J 34:173-185

Morrell PL and Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) East of the Fertile Crescent. Proc Natl Acad Sci USA 104:3289-3294

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

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

Muller KJ, Romano N, Gerstner O, Garcia-Maroto F, Pozzi C, Salamini F, Rohde W (1995) The barley Hooded mutation caused by a duplication in a homeobox gene intron. Nature 374:727-730

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. Plant Cell 21:2194-2202

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J (2005) The Pattern of Polymorphism in Arabidopsis thaliana. PLoS Biol 3:e196

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

Ordon F, Ahlemeyer J, Werner K, Kohler W, Friedt W (2005) Molecular assessment of genetic diversity in winter barley and its use in breeding. Euphytica 146:21-28

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 12:16

Payne RW (2009) GenStat. Wiley Interdisciplinary Reviews: Computational Statistics 1:255-258

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256-261

Pickering R, Johnston PA (2005) Recent progress in barley improvement using wild species of Hordeum. Cytogenet Genome Res 109:344-349

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

Pourkheirandish M, Komatsuda T (2007) The importance of barley genetics and domestication in a global perspective. Ann Bot 100:999-1008

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 (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

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

Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WT, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N, Hayes PM, Lundqvist U, Franckowiak JD, Close TJ, Muehlbauer GJ, Waugh R (2011) *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED* 1. Nat Genet 43:169-172

Reif JC, Hamrit S, Heckenberger M, Schipprack W, Maurer HP, Bohn M, Melchinger AE (2005) Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. Theor Appl Genet 111:838-845

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ESt (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479-11484

Riehl S, Benz M, Conard N, Darabi H, Deckers K, Nashli H, Zeidi-Kulehparcheh M (2011) Plant use in three Pre-Pottery Neolithic sites of the northern and eastern Fertile Crescent: a preliminary report. Vegetation History and Archaeobotany:1-12

Roder MS, Kaiser C, Weschke W (2006) Molecular mapping of the shrunken endosperm genes *seg8* and *sex1* in barley (Hordeum vulgare L.). Genome 49:1209-1214

Rodriguez M, Rau D, O'Sullivan D, Brown AH, Papa R, Attene G (2012) Genetic structure and linkage disequilibrium in landrace populations of barley in Sardinia. Theor Appl Genet 125:171-84

Rollefson G (1985) The 1983 Season at the Early Neolithic Site of 'Ain Ghazal. National Geographic Research 1:44-62

Romagosa I, Han F, Ullrich S, Hayes P, Wesenberg D (1999) Verification of yield QTL through realized molecular marker- assisted selection responses in a barley cross. Molecular Breeding 5:143-152

Romagosa I, Ullrich S, Han F, Hayes P (1996) Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. Theor Appl Genet 93:30-37

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Genet Genomics 274:515-527

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

Rostoks N, Schmierer D, Mudie S, Drader T, Brueggeman R, Caldwell DG, Waugh R, Kleinhofs A (2006b) Barley necrotic locus nec1 encodes the cyclic nucleotide-gated ion channel 4 homologous to the Arabidopsis HLM1. Mol Genet Genomics 275:159-168

Roussel V, Koenig J, Beckert M, Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. Theor Appl Genet108:920-930

Roy J, Smith K, Muehlbauer G, Chao S, Close T, Steffenson B (2010) Association mapping of spot blotch resistance in wild barley. Molecular Breeding 26:243-256

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 Phytologist 191:564-578

Russell JR, Booth A, Fuller JD, Baum M, Ceccarelli S, Grando S, Powell W (2003) Patterns of polymorphism detected in the chloroplast and nuclear genomes of barley landraces sampled from Syria and Jordan. Theor Appl Genet 107:413-421

Russell JR, Roger P. Ellis, Thomas WTB, Waugh R, Allan JP, Booth A, Fuller J, Lawrence P, George Y, Wayne P (2000) A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Molecular Breeding 6:553–568

Saisho D, Pourkheirandish M, Kanamori H, Matsumoto T, Komatsuda T (2009) Allelic variation of row type gene *Vrs1* in barley and implication of the functional divergence. Breeding Science 59:621-628

Saisho D, Purugganan MD (2007) Molecular Phylogeography of Domesticated Barley Traces Expansion of Agriculture in the Old World. Genetics 177:1765-1776

Saisho D, Takeda K (2011) Barley: Emergence as a New Research Material of Crop Science. Plant and Cell Physiology 52:724-727

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 3:429-441

Salomé PA, Bomblies K, Laitinen RAE, Yant L, Mott R, Weigel D (2011) Genetic Architecture of Flowering-Time Variation in Arabidopsis thaliana. Genetics 188:421-433

Sanei M, Pickering R, Kumke K, Nasuda S, Houben A (2011) Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proc Natl Acad Sci USA 108:E498-505

Sato K, Nankaku N, Takeda K (2009) A high-density transcript linkage map of barley derived from a single population. Heredity 103:110-117

Schmalenbach I, March TJ, Bringezu T, Waugh R, Pillen K (2011) High-Resolution Genotyping of Wild Barley Introgression Lines and Fine-Mapping of the Threshability Locus thresh-1 Using the Illumina GoldenGate Assay. G3 (Bethesda) 1:187-196

Schulze-Lefert P, Panstruga R (2003) Establishment of biotrophy by parasitic fungi and reprogramming of host cells for disease resistance. Annu Rev Phytopathol 41:641-667

Schulze-Lefert P, Vogel J (2000) Closing the ranks to attack by powdery mildew. Trends Plant Sci 5:343-348

Schweizer P, Stein N (2011) Large-Scale Data Integration Reveals Colocalization of Gene Functional Groups with Meta-QTL for Multiple Disease Resistance in Barley. Molecular Plant-Microbe Interactions 24:1492-1501

Slatkin M (2008) Linkage disequilibrium-understanding the evolutionary past and mapping the medical future Nat Rev Genet 9:477-485

Somers DA, Kuo T-M, Kleinhofs A, Warner RL (1983) Nitrate reductase-deficient mutants in Barley : Immunoelectrophoretic Characterization. Plant Physiology 71:145-149

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stuber K, Ver Loren van Themaat E, Brown JK, Butcher SA, Gurr SJ, Lebrun MH, Ridout CJ, Schulze-Lefert P, Talbot NJ, Ahmadinejad N, Ametz C, Barton GR, Benjdia M, Bidzinski P, Bindschedler LV, Both M, Brewer MT, Cadle-Davidson L, Cadle-Davidson MM, Collemare J, Cramer R, Frenkel O, Godfrey D, Harriman J, Hoede C, King BC, Klages S, Kleemann J, Knoll D, Koti PS, Kreplak J, Lopez-Ruiz FJ, Lu X, Maekawa T, Mahanil S, Micali C, Milgroom MG, Montana G, Noir S, O'Connell RJ, Oberhaensli S, Parlange F, Pedersen C, Quesneville H, Reinhardt R, Rott M, Sacristan S, Schmidt SM, Schon M, Skamnioti P, Sommer H, Stephens A, Takahara H, Thordal-Christensen H, Vigouroux M, Wessling R, Wicker T, Panstruga R (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330:1543-1546

Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney RK, Perovic D, Grosse I, Graner A (2007) A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet 114:823-839

Stich B, Melchinger A, Piepho H-P, Heckenberger M, Maurer H, Reif J (2006) A new test for family-based association mapping with inbred lines from plant breeding programs. Theor Appl Genet113:1121-1130

Stich B, Melchinger AE (2009) Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics 10: 94

Stich B, Möhring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of Mixed-Model Approaches for Association Mapping. Genetics 178:1745-1754

Stockinger EJ, Skinner JS, Gardner KG, Francia E, Pecchioni N (2007) Expression levels of barley Cbf genes at the Frost resistance-H2 locus are dependent upon alleles at Fr-H1 and Fr-H2. Plant J 51:308-321

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

Stracke S, Haseneyer G, Veyrieras J-B, Geiger H, Sauer S, Graner A, Piepho H-P (2009) Association mapping reveals gene action and interactions in the determination of flowering time in barley. TAG Theor Appl Genet118:259-273

Stracke S, Presterl T, Stein N, Perovic D, Ordon F, Graner A (2007) Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding Bymovirus resistance in barley. Genetics 175:805-817

Stumpf MP, McVean GA (2003) Estimating recombination rates from population-genetic data. Nat Rev Genet 4:959-968

Syvanen AC (2005) Toward genome-wide SNP genotyping. Nat Genet 37 Suppl:S5-10

Szucs P, Karsai I, von Zitzewitz J, Meszaros K, Cooper LL, Gu YQ, Chen TH, Hayes PM, Skinner JS (2006) Positional relationships between photoperiod response QTL and photoreceptor and vernalization genes in barley. Theor Appl Genet 112:1277-1285

Szucs P, Skinner JS, Karsai I, Cuesta-Marcos A, Haggard KG, Corey AE, Chen TH, Hayes PM (2007) Validation of the VRN-H2/VRN-H1 epistatic model in barley reveals that intron length variation in VRN-H1 may account for a continuum of vernalization sensitivity. Mol Genet Genomics 277:249-261

Takahashi R (1955) The origin and evolution of cultivated barley. In: Demerec M (ed) Advances in Genetics, vol 7. Academic, New York,:227-266

Takahashi R, Hayashi J, Yasuda S, Hiura U (1963) Characteristics of the wild and cultivated barleys from Afghanistan and its neighboring regions. Ber Ohara Inst Landw Biol, Okayama Univ 12:1-23

Takahashi R, Yasuda S (1971) Genetics of earliness and growth habit in barley. Proceedings of the 2nd Int Barley Genet Symp, Washington State University:388-408

Takeda K, Frey KJ (1987) Improving grain yield in backcross populations from Avena sativa × A. sterilis matings by using independent culling for harvest index and vegetative growth index or unit straw weight Theor Appl Genet 74:659-665

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. Proc Natl Acad Sci USA 105:4062-4067

Talamè V, Bovina R, Sanguineti MC, Tuberosa R, Lundqvist U, Salvi S (2008) TILLMore, a resource for the discovery of chemically induced mutants in barley. Plant Biotechnology Journal 6:477-485

Tanksley SD, McCouch SR (1997) Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. Science 277:1063-1066

Tanno K-i, Willcox G (2011) Distinguishing wild and domestic wheat and barley spikelets from early Holocene sites in the Near East. Vegetation History and Archaeobotany:1-9

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.). Proc Natl Acad Sci USA 98:9161-9166

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection Versus Demography: A Multilocus Investigation of the Domestication Process in Maize. Molecular Biology and Evolution 21:1214-1225

The International Hapmap Consortium (2005) A haplotype map of the human genome. Nature 437:1299-1320

Thomas WTB, Powell W, Swanston JS (1991) The effects of major genes on quantitatively varying characters in barley. 4. The GPert and denso loci and quality characters. Heredity 66:381-389

Thomas WTB, Powell W, Waugh R, Chalmers KJ, Barua UM, Jack P, Lea V, Forster BP, Swanston JS, Ellis RP, Hanson PR, Lance RCM (1995) Detection of quantitative trait loci for agronomic, yield, grain and disease characters in spring barley (*Hordeum vulgare* L.). Theor Appl Genet 91:1037-1047

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

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

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

Van Zee K, Chen FQ, Hayes PM, Close TJ, Chen T (1995) Cold-Specific Induction of a Dehydrin Gene Family Member in Barley. Plant Physiol 108:1233-1239

van Zeist W (1970) The Oriental Institute Excavations at Mureybit, Syria: Preliminary Report on the 1965 Campaign. Part III: The Paleobotany. Journal of Near Eastern Studies 29:167-176.

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

Vavilov NI (1926) Studies on the origin of cultivated platns. Bull Appl Bot Gen Plant Breed USSR 16:1–248

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

von Bothmer R, Komatsuda T, Yasuda S, Fischbeck G (eds) (2003a) The domestication of cultivated barley. Elsevier Science B.V., Amsterdam, The Netherlands

von Bothmer R, Sato K, Komatsuda T, Yasuda S, G F (2003b) The domestication of cultivated barley. In: R. von Bothmer, Th. van Hintum, H. Knüpffer and K. Sato (eds), Diversity in Barley (Hordeum vulgare). In: Roland von Bothmer TvH, Helmut Knüpffer and Kazuhiro Sato (ed) Diversity in Barley (Hordeum vulgare). Elsevier B.V., pp 9-27

von Korff M, Wang H, Leon J, Pillen K (2004) Development of candidate introgression lines using an exotic barley accession ( Hordeum vulgare ssp. spontaneum) as donor. Theor Appl Genet 109:1736-1745

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

von Wettstein D (2007) From analysis of mutants to genetic engineering. Annu Rev Plant Biol 58:1-19

Wang G, Schmalenbach I, von Korff M, Leon J, Kilian B, Rode J, Pillen K (2010) Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines. Theor Appl Genet 120:1559-1574

Wang M, Jiang N, Jia T, Leach L, Cockram J, Waugh R, Ramsay L, Thomas B, Luo Z (2011) Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. Theor Appl Genet 124:233-46

Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. Curr Opin Plant Biol 12:218-222

Waugh R, Marshall D, Thomas B, Comadran J, Russell J, Close T, Stein N, Hayes P, Muehlbauer G, Cockram J, O'Sullivan D, Mackay I, Flavell A, Ramsay L (2010) Whole-genome association mapping in elite inbred crop varieties. Genome 53:967-972

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) The *Mla* (Powdery Mildew) Resistance Cluster Is Associated With Three NBS-LRR Gene Families and Suppressed Recombination Within a 240-kb DNA Interval on Chromosome 5S (1HS) of Barley. Genetics 153:1929-1948

Weir BS (1979) Inferences about linkage disequilibrium. Biometrics 35:235-254

Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661-678

Weltzien E (1988) Evaluation of Barley (Hordeum vulgare L.) Landrace Populations Originating from Different Growing Regions in the Near East. Plant Breeding 101:95-106

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesna J, Cakir M, Poulsen D, Wang J, 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 7:206

Wicker T, Narechania A, Sabot F, Stein J, Vu G, 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 9:518

Willcox G (1997) Archeobotanical evidence for the beginnings of agriculture in Southwest Asia: The origins of agriculture and crop domestication: Proc. Harlan Symp. Aleppo, Syria, 10–14 May, 1997. P. 25–38.

Wolfe MS, McDermott JM (1994) Population genetics of plant pathogen interactions: The example of the Erysiphe graminis-Hordeum vulgare pathosystem. Annual Review of Phytopathology 32:89-112

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The Effects of Artificial Selection on the Maize Genome. Science 308:1310-1314

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L (1996) Genes from wild rice improve yield. Nature 384:223-224

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotech 30:105-111

Yamasaki M, Tenaillon MI, Vroh Bi I, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. The Plant Cell Online 17:2859-2872

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of FT. Proceedings of the National Academy of Sciences 103:19581-19586

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

Yu J, Pressoir G, Briggs WH, Vroh Bi I, 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

Zeller FJ (1998) Nutzung des genetischen Potentials der Hordeum-Wildarten zur Verbesserung der Kulturgerste (Hordeum vulgare L.). Angewandte Botanik 72:162-167

Zhang L, Lavery L, Gill U, Gill K, Steffenson B, Yan GP, Chen XM, Kleinhofs A (2009a) A cation/proton-exchanging protein is a candidate for the barley *NecS1* gene controlling necrosis and enhanced defense response to stem rust. Theor Appl Genet 118:385-397

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

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355-360

Zhou F, Kurth J, Wei F, Elliott C, Valè G, Yahiaoui N, Keller B, Somerville S, Wise R, Schulze-Lefert P (2001) Cell-Autonomous Expression of Barley Mla1 Confers Race-Specific Resistance to the Powdery Mildew Fungus via a Rar1-Independent Signaling Pathway. The Plant Cell 13:337-350

Zhu B, Choi DW, Fenton R, Close TJ (2000) Expression of the barley dehydrin multigene family and the development of freezing tolerance. Molecular and General Genetics 264:145-153

Zhu C, Yu J (2009) Nonmetric Multidimensional Scaling Corrects for Population Structure in Association Mapping With Different Sample Types. Genetics 182:875-888

Zohary D (1999) Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. Genet Resour Crop Evol 46:133-142

Zohary D, Hopf M (2000) Domestication of Plants in the Old World. Oxford Univ. Press, Oxford

# Supplementary Information

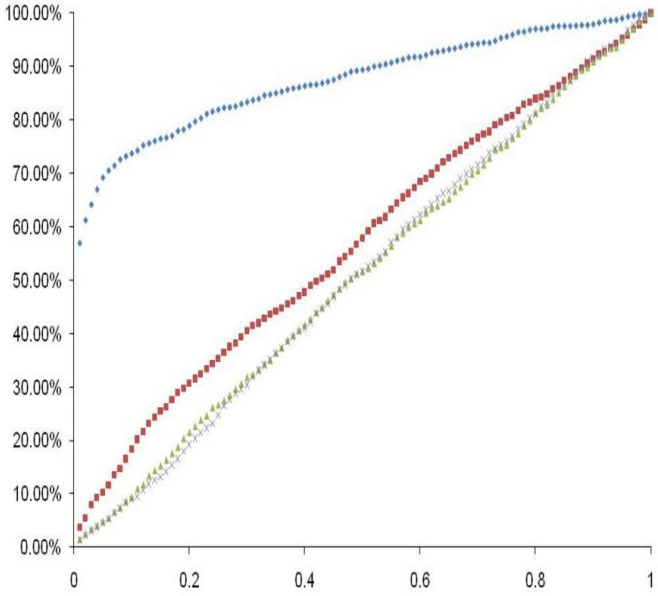



**Fig S1**: Model comparison with powdery mildew detched leaf assay.Cumulative observed p-values in percentages for naïve model in blue color (without correction for population), general linear model (GLM) using Q matrix in red color, mixed model (MLM) in green color (using principal components for population structure control), purple color MLM using kinship matrix to control population structure are shown. Model that control false positive should approximate to a uniform distribution of the p-value

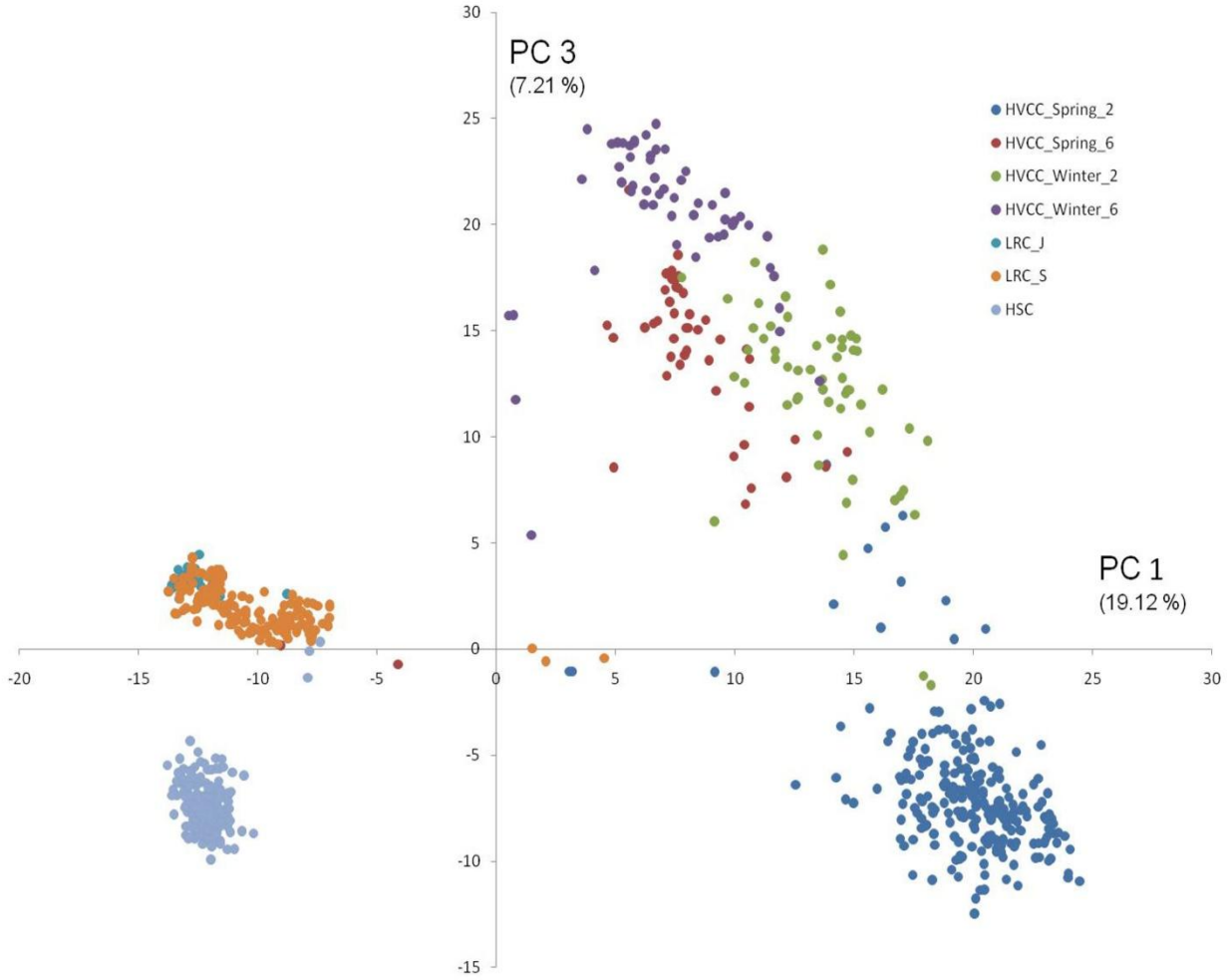



**Fig S2**: Population structure of barley gene pools with BOPA1 SNP markers. PC1 and PC3 show the distribution of the genotypes. Different colors are assigned to groups of genotypes. HVCC_Spring_2 (spring two rowed varieties); HVCC_Spring_6 (spring six rowed varieties); HVCC_Winter_2 (winter two rowed varieties); HVCC_Winter_6 (winter six rowed varieties); LRC_J (landraces from Jordan); LRC_S (landraces from Syria) and HSC (Wild barley) gene pool are shown

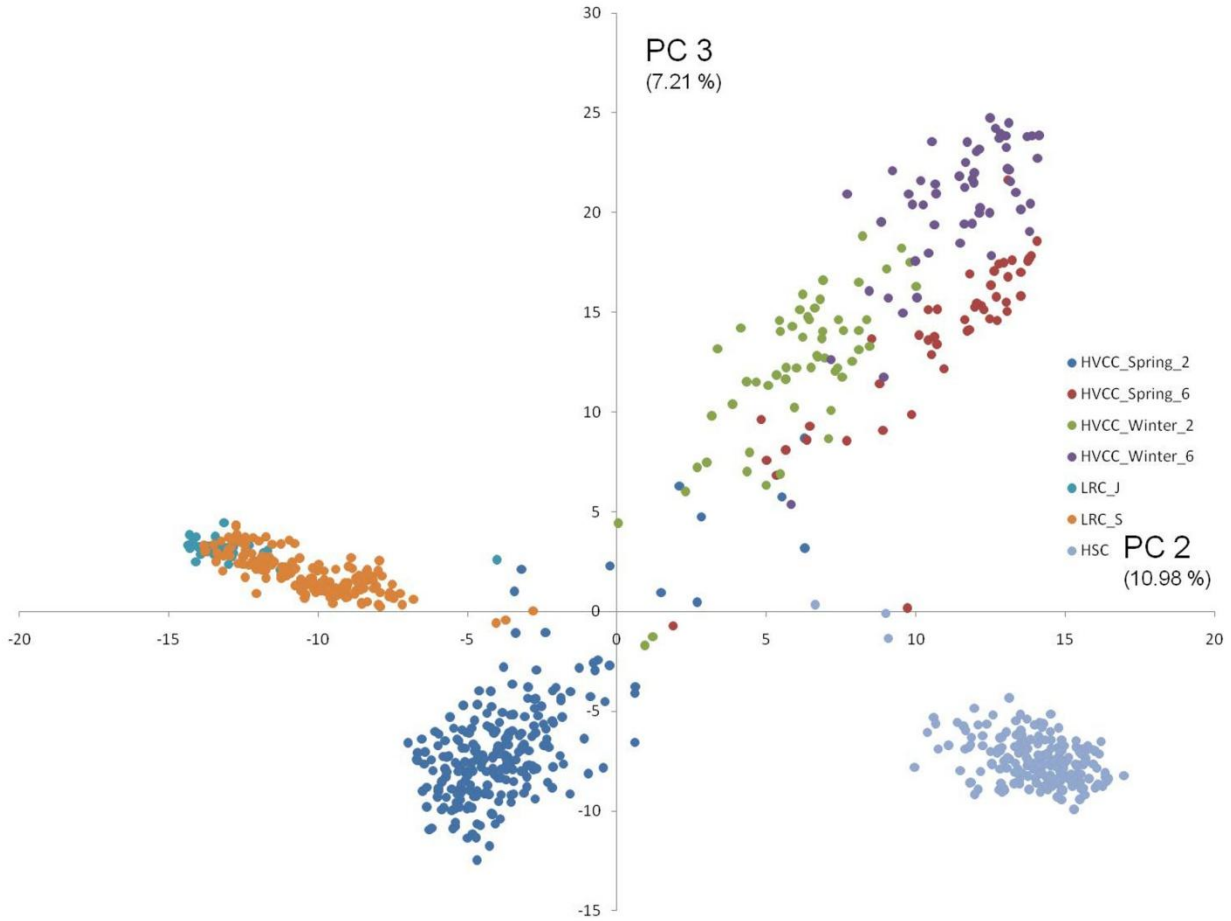



**Fig S3**: Population structure of barley gene pools using BOPA1 SNP markers. PC2 and PC3 are shown. Different colors are assigned to groups of genotypes. HVCC_Spring_2 (spring two rowed varieties); HVCC_Spring_6 (spring six rowed varieties); HVCC_Winter_2 (winter two rowed varieties); HVCC_Winter_6 (winter six rowed varieties); LRC_J (landraces from Jordan); LRC_S (landraces from Syria) and HSC (Wild barley) gene pool are shown

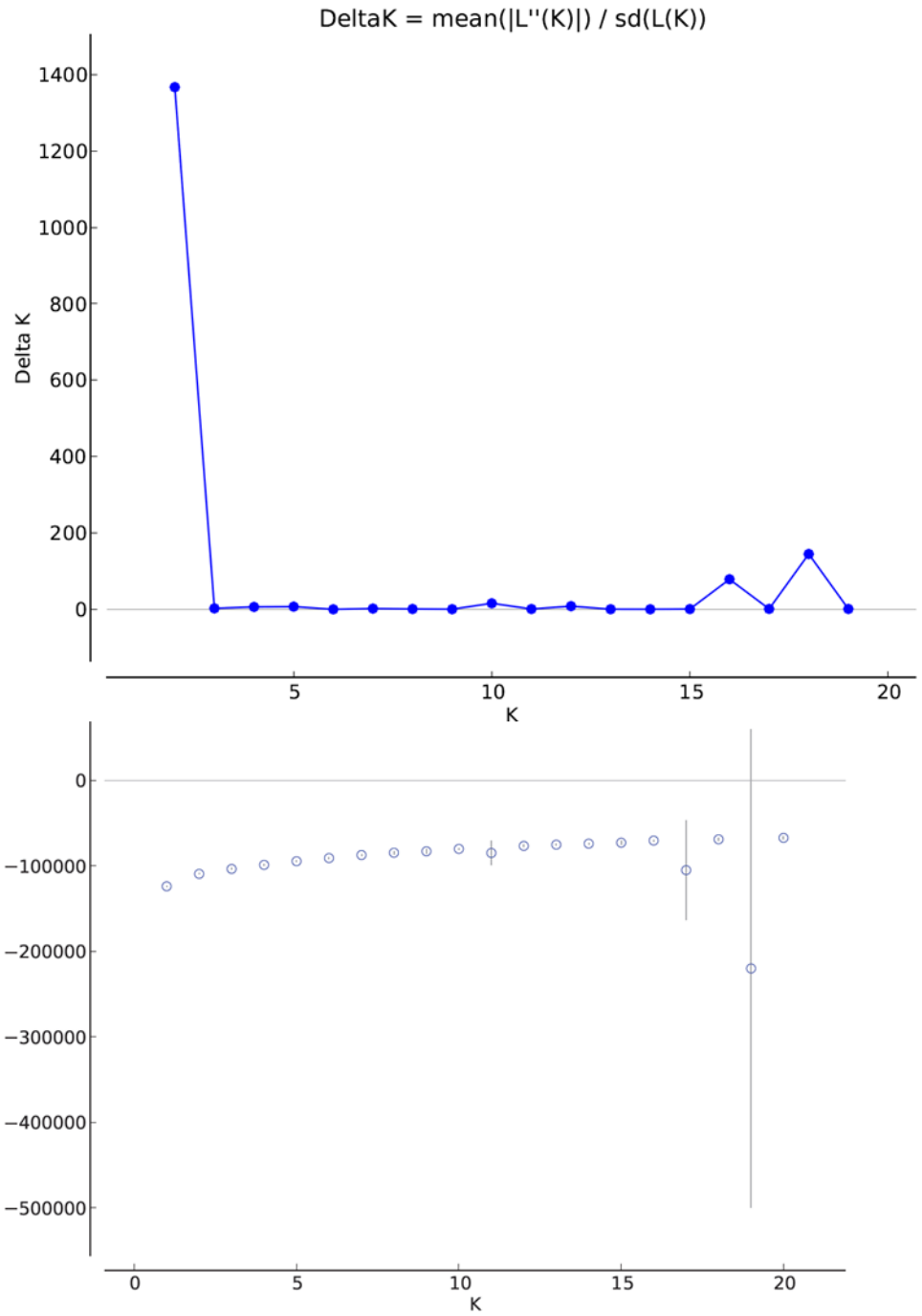



**Fig S4**: STRUCTURE results in winter barley. On top Delta k vs. k (number of assumed populations) plotted as proposed by Evanno et al. (2005) and graph showing maximum change of delta k at k=2; as an alternative approach on down log probability data (LnP(D) mean ± SD) as function of k (number of clusters) from the STRUCTURE run plotted for each structure run and the graph plateau at k 2 (proposed by Pritchard et al. 2000)

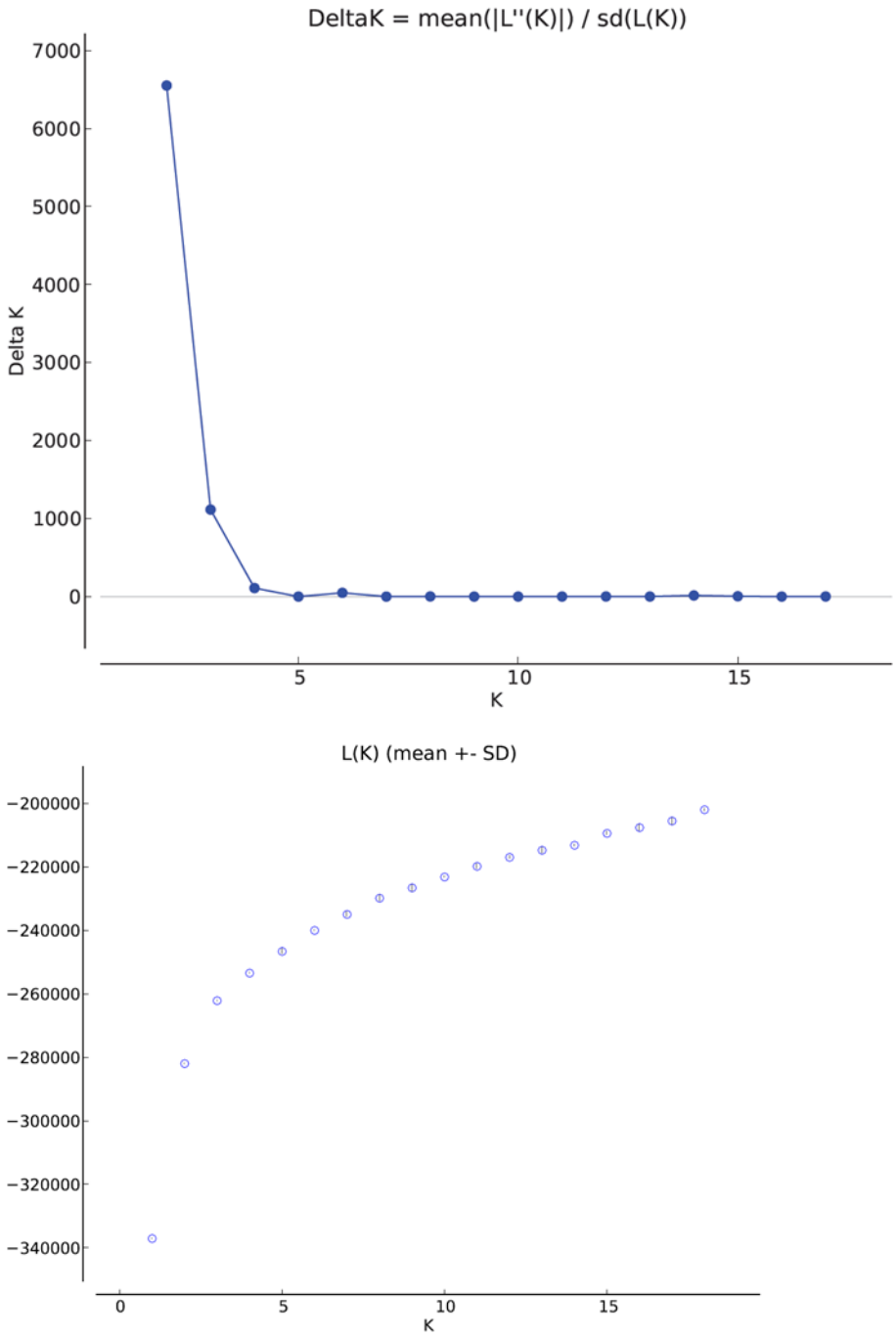
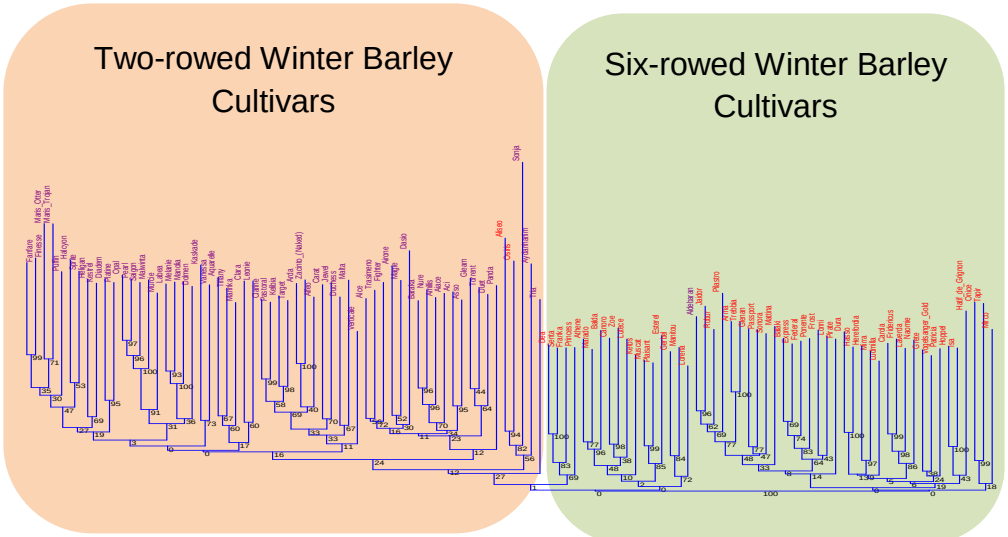



**Fig S5**: STRUCTURE results in spring barley. On top Delta k vs. k (number of assumed populations) plotted as proposed by Evanno et al. (2005) and graph showing maximum change of delta k at k=2; as an alternative approach on down log probability data (LnP(D)) as function of k (number of clusters) from the STRUCTURE run plotted for each structure run and the graph plateau at k 2 (proposed by Pritchard et al. 2000)

**(a)**




**(b)**

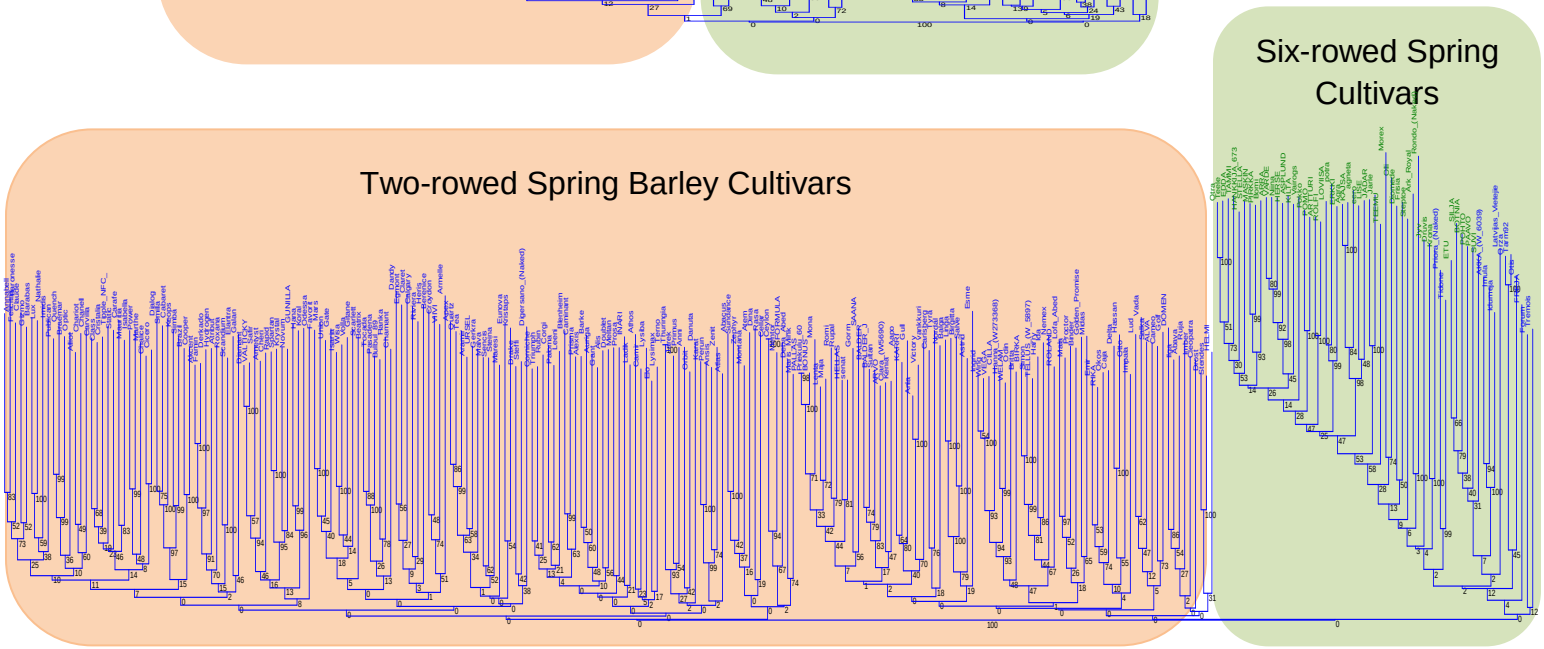



**Fig S6**: Neighbour Joining (NJ) clustering of the cultivars (HVCC) **(a)** Two-rowed and six-rowed winter barley (HVCC_W) are shown **(b)** Two-rowed and six-rowed spring barley (HVCC_S) are shown

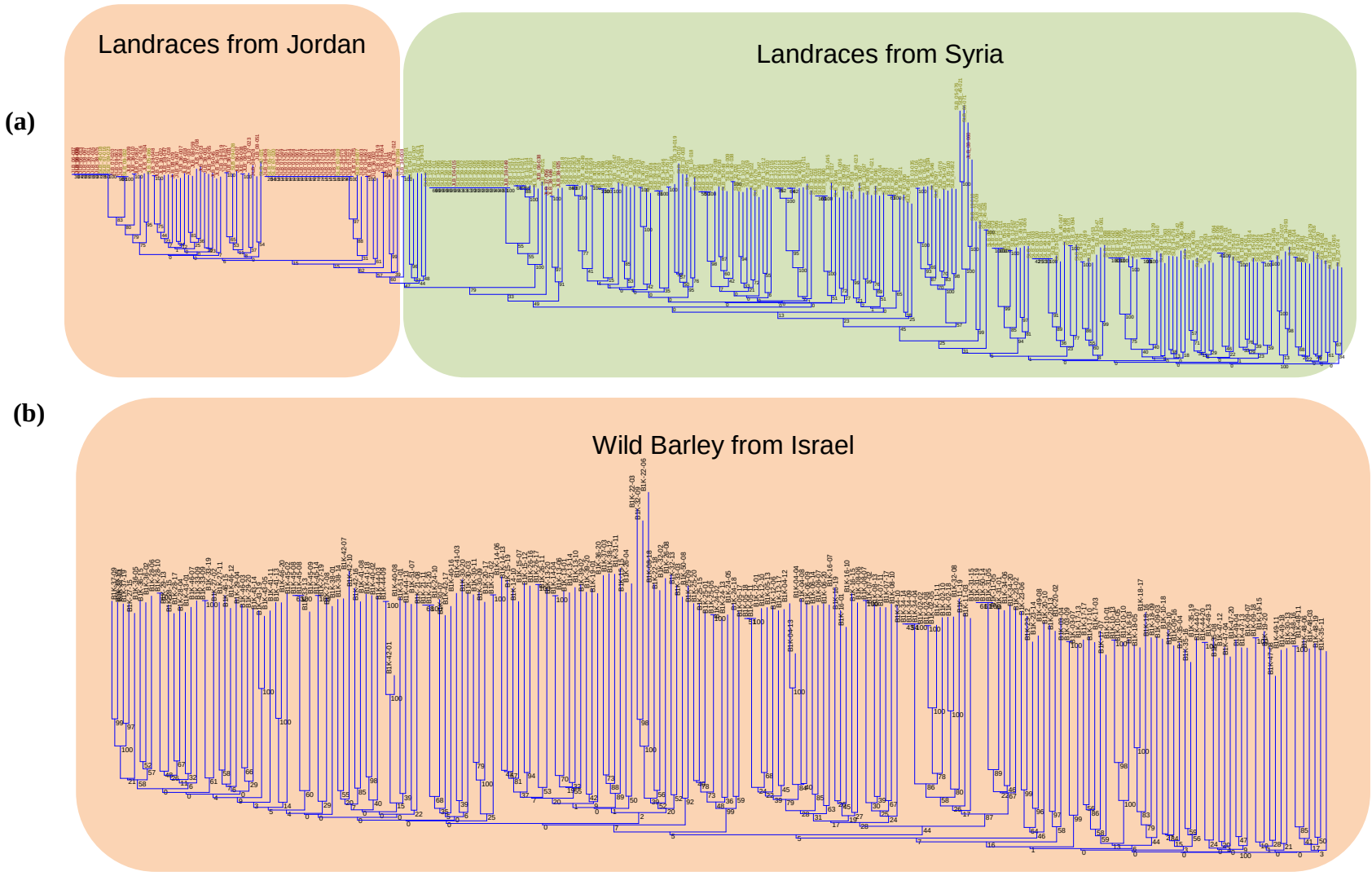



**Fig S7**: Neighbour Joining (NJ) clustering of the landraces and wild barley **(a)** Jordanian (LRC_J) and Syrian (LRC_S) landraces are shown **(b)** Wild barley (HSC) from Israel are shown

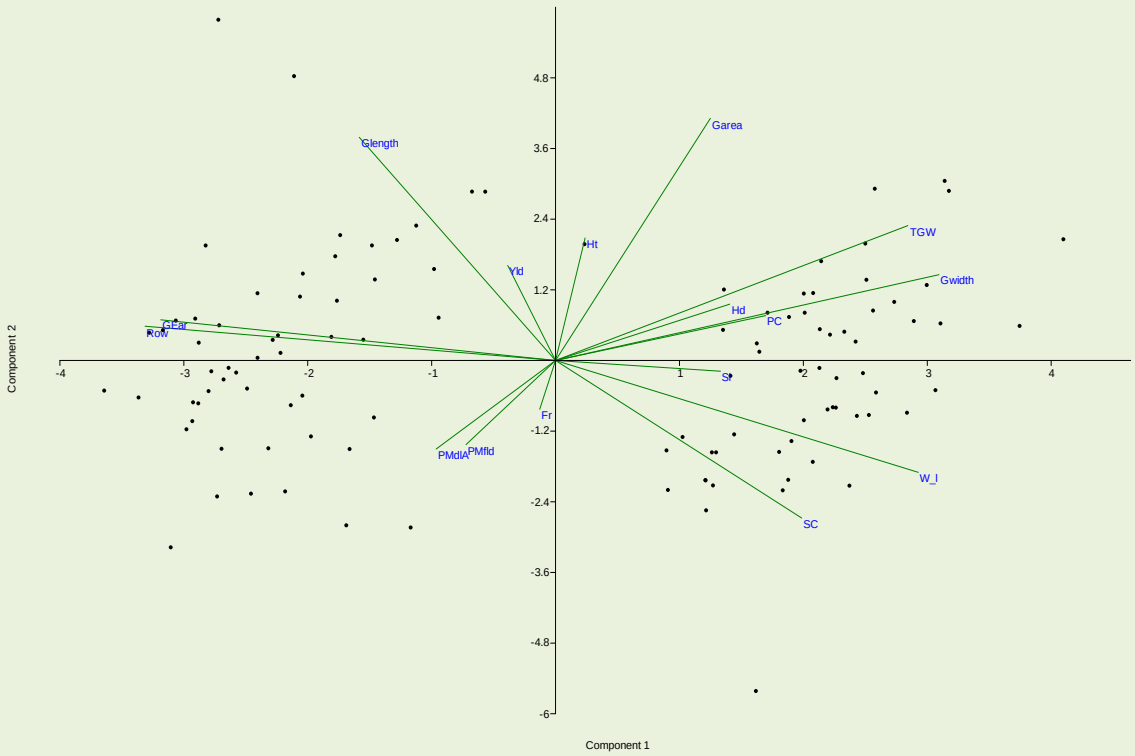



**Fig S8**: Bi-plot analysis of the phenotypic traits in HVCC_W gene pool. Component 1 explained 32.63 percent and component 2, 17.74 percent of the total variation

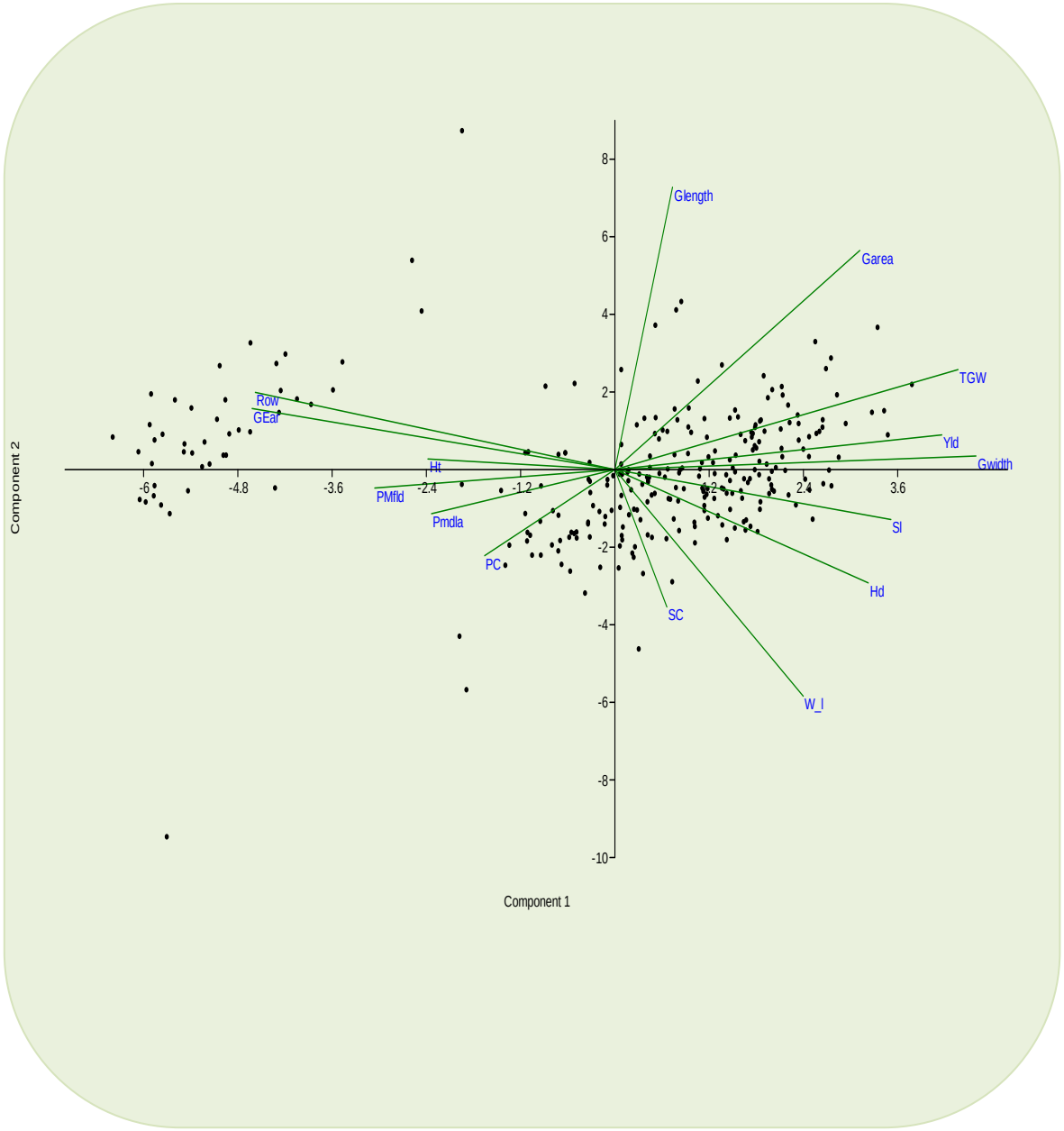
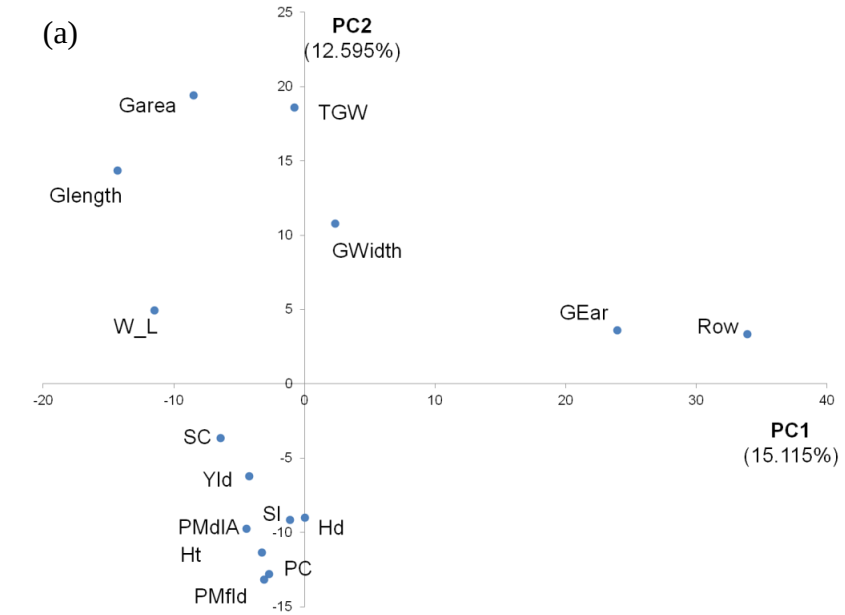



**Fig S9**: Bi-plot analysis of the phenotypic traits in the HVCC_S gene pool. Component 1 explained 38.97 percent and component 2 explained 18.94 percent of the total variation



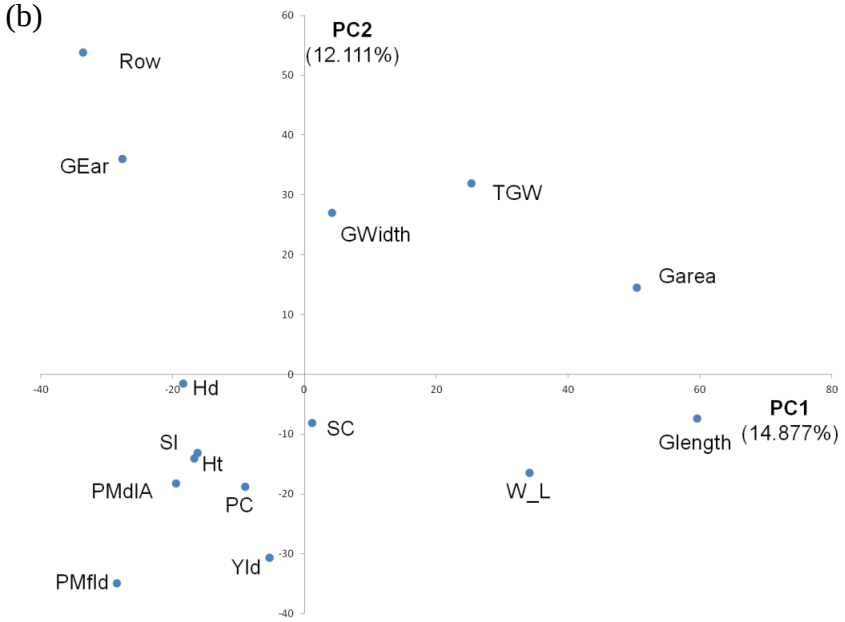




**Fig S10:** Principal component analysis using associations values from all traits in spring barley. Significant principal components (PC1 and PC2) are shown using **(a)** BOPA1 and **(b)** iSelect SNP markers

# ACKNOWLEDGEMENTS

I am grateful to my major supervisor Prof. Andreas Graner, for providing me this interesting PhD research project. His encouragement, guidance and discussions helped me during the period of investigation and preparation of the thesis.

I would like to extend my gratitude to DFG (Deutsche Forschungsgemeinschaft) and IPK for financially supporting me.

I express my sincere thanks to all EXBARDIV members of the project, who were providing data that was used in this study. Special thanks to Prof. Andrew Flavell (University of Dundee) coordinator of the project to supervise the project. Also thanks are due to Dr. William Thomas (James Hutton Institute) who was instrumental in coordinating and helping in analyzing the data.

My cordial thanks are due to Dr. Benjamin Kilian, for valuable comments and supervising the project progress during the whole period of this study.

Also I would like to acknowledge the guidance of Prof. Klaus Pillen (Martin-Luther-Universtität Halle-Wittenberg) who gave valuable suggestions that helps in improving and compiling my research work.

I would like to thanks Prof. Søren K. Rasmussen (University of Copenhagen) who gave critical valuable comments during the course of the project.

I am very thankful to Dr. Patrick Schweizer, for providing necessary facilities and encouragement during the part of study. His always positive approach and dedication towards science was inspiring. It was great experience for me.

I acknowledge Peter Schreiber and his staff for extending great help in successfully conducting the field experiments without their untidy support huge field experiments would not be possible.

I would like to acknowledge Dr. Kerstin Neumann, Dr. Ruonan Zhou, Dr. Ruvini Ariyadasa, Dr. Stephan Weise for their feedbacks, inputs and support.

I extend my sincere thanks to Dr. Nils Stein and all the Genome Diversity group members that have been supportive throughout my research work. I would like to thank Manuela Knauft for her support during powdery mildew resistance screening. Also I would like to thank Karin Lipfert for printing and editing support.

I would like to thank my late grandparents for their revered love and blessings that they bestowed on me. I owe deep gratitude to my loving and supporting parents, and for my sister "Poonam", her husband Tarkeshwar Amol, my Bhanja Tejas and their family who always shared the burden of my despair. I feel short of words to thank them for their ungrudging help, unaparallel affection and blessings.

I would like to acknowledge Raj Kishore Pasam my colleague for his friendship and ever willing help. He was always there when I need him and together we enjoyed great moments of life.

Special thanks are due to my dear friends Harshavardhan, Geetha, Shailendra, Shiveta, Rajesh, Pooja, Sidram and Naser for their memorable company, ever willing help and cooperation during the trailing period of my life.

I would like to thank Dr. Britt Leps who is always supportive and guiding me in solving all official affairs.

Last but not least I would like to thank my wife Nisha for her patience and loving support that motivates me during my work and her family for morale support. All may not be mentioned, but none is forgotten.

Rajiv Sharma

Gatersleben

## CURRICULUM VITAE

Rajiv Sharma
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Genebank/ Genome Diversity
Correnstrasse 3 06466 Gatersleben, Germany
Email: sharmar@ipk-gatersleben.de

## ACADEMIC QUALIFICATION

**PhD student (June 2008 to present):** Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Group Genome Diversity, Gatersleben, GERMANY.

Title: "*Genome-wide association studies in diverse gene pools of barley*"

**Research Fellow (April 2007-June2008):** School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, INDIA.

Title: "*Water use efficiency in rice through molecular marker interventions*"

**Master of Sciences in Plant Breeding and Genetics (2004-2006):** Punjab Agricultural University, Ludhiana, INDIA.

Title: "*Inheritance and molecular mapping of bacterial blight resistance gene(s) in Oryza rufipogon*"

**Bachelor of Sciences Agriculture (Honours) in the discipline of Crop Science (2000-2004):** Punjab Agricultural University, Ludhiana, INDIA.

## PUBLICATIONS

8. Raj K Pasam, **Rajiv Sharma**, Marcos Malosetti, Benjamin Kilian, Andreas Graner (2012) Genome-wide association studies for agronomical traits in a world-wide spring barley collection. **BMC Plant Biology** 12: 16

7. Kilian B, Özkan H, Shaaf S, Hübner S, Pasam RK, **Sharma R**, Neumann K, Weißgerber W, Konovalov FA, Keilwagen J, Friedel S, Knüpffer H, von Korff M, Coupland G, Graner, A (2012) Comparing genetic diversity within a crop and its wild progenitor: a case study for barley. In: Maxted N, Dulloo ME, Ford-Lloyd BV, Frese L, Iriondo JM, Pinheiro de Carvalho MAA (editors). *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces*. CABI pp.186-192

6. Dharminder Bhatia, **Rajiv Sharma**, Yogesh Vikal, G.S. Mangat, Ritu Mahajan, Neerja Sharma, Jagjeet Singh Lore, Naveen Singh, Tajinder S. Bharaj and Kuldeep Singh (2011)

Marker-assisted development of bacterial blight resistant, dwarf, and high yielding versions of two traditional basmati rice. **Crop Science** 51:759-770

5. Benjamin Kilian, K Mammen, Eiten Millet, **Rajiv Sharma**, Andreas Graner, Francisco Salamini, Karl Hammer, Hakan Özkan (2011) Aegilops L. KOLE, C : **Wealth of wild crop relatives: Genetic, genomic and breeding resources**. Springer Publishers, New York pp. 1-76

4. Graner A, Haseneyer G, Kilian B, Pasam RK, **Sharma R** & Stein N (2011) Nutzung genetischer Vielfalt: Herausforderung und Perspektiven. Agrar Spectrum 44: 45-61

3. Shailendra Sharma and **Rajiv Sharma** (2010) Biotechnology and sustainable agriculture marker assisted selection chapter 18 in Stable Food Production and Sustainable Agriculture : A Challenge Ahead in 21st Century by R S Sengar and A K Sharma, Studium press 2010, xx, 570 p, ISBN : 938001214

2. Kuljit K Cheema, N Grewal, Y Vikal, **Rajiv Sharma**, J S Lore, A Das, D Bhatia, R Mahajan, V Gupta, T S Bharaj and Kuldeep Singh (2008) A novel bacterial blight resistance gene from Oryza nivara mapped to 38 kb region on chromosome 4L and transferred to Oryza sativa L. **Genetics Research (Cambridge University Press)** 90: 397-407

1. Dharminder Bhatia, **Rajiv Sharma**, Vikas Gupta, Yogesh Vikal and Kuldeep Singh (2007) Rapid method for extraction of ready to use fine quality DNA **Indian J. Crop Science** 2 (2): 407-409

## TALKS AND POSTERS

A Tondelli, X Xu, F Schnaithmann, **R Sharma** et al.: Structural and Temporal Variation in the Genetic Diversity of a European Collection of Barley Cultivars and Utility for Association Mapping of Quantitative Traits. **Plant and Animal Genome Conference** 2012 XX P0346

Raj K Pasam, **Rajiv Sharma**, Benjamin Kilian, Andreas Graner: Genome wide association scans to identify the candidate agronomic loci for crop improvement in spring barley. 11th **GABI-Status seminar** in Potsdam, March 2011

Ritu Mahajan, Dharminder Bhatia, **Rajiv Sharma**, Kuljeet Cheema, Vikas Gupta, Yogesh Vikal and kuldeep Singh: Inheritance and molecular mapping of a bacterial blight resistance gene from *oryza glaberrima*. **International Conference on preparing agricultural for climate change** February, 6-8 2011, at PAU, Ludhiana, India

Raj K Pasam, **Rajiv Sharma**, Grit Haseneyer, Benjamin Kilian, Andreas Graner: QTL identification for grain quality traits in barley using LD mapping in natural populations. **10th Gatersleben Research Conference**, Gatersleben, 23-24.11.2010

Raj K Pasam, **Rajiv Sharma**, Grit Haseneyer, Benjamin Kilian, Andreas Graner: QTL identification for grain quality traits in barley using LD mapping in natural populations. **Genomics Based Breeding Conference**, Giessen, 26-28.10.2010

**Rajiv Sharma**, Raj K Pasam, Benjamin Kilian, Andreas Graner: Genome-wide association studies for plant architecture traits in different barley gene pools. **Genomics Based Breeding Conference**, Giessen, 26-28.10.2010

**Rajiv Sharma**, Raj K Pasam, Benjamin Kilian, Andreas Graner: Genome-wide association studies for plant architecture traits in barley. **6th Plant Science Student Conference**, Gatersleben, 15-16.06.2010

Flavell AJ, X Xu , J Comadran , W Thomas, D Marshall, J Russell , R. Waugh, A Tondelli , L Cattivelli, F Schnaithmann , K Pillen , A Schulman , C Ingvardsen , S Rasmussen , **R Sharma**, B Kilian: An incremental approach for association mapping of useful traits in barley - The EXBARDIV Project. **2nd Intl Symposium on Genomics of Plant Resources** 24-27.04-2010 Bologna, Italy

Raj K Pasam, **Rajiv Sharma**, Benjamin Kilian, Andreas Graner: Whole Genome Association Mapping in Spring Barley Cultivars. 10. **GABI-Status seminar** in Potsdam, 9.-11. March 2010

Raj K Pasam, **Rajiv Sharma**, Grit Haseneyer, B Kilian, Andreas Graner: Whole Genome Association Mapping in Spring Barley. **GPZ-Tagung** in Freising, 15. - 17. March 2010

Harleen Chawla, Yogesh Vikal, **Rajiv Sharma**, J S Lore , C Veera Cruz and Kuldeep Singh: Molecular Mapping of Bacterial Blight Resistance Gene *xa8* in rice**. International Colloquium on Biotechnology**, DAV College, Jalandhar, Punjab, 27-28November, 2009

Yogesh Vikal, **Rajiv Sharma**, Harleen Chawla, J S Lore, C Veera Cruz and Kuldeep Singh: Towards fine mapping of bacterial blight resistance gene *xa8* in rice. **The International Rice Genetics Symposium and 7th International Symposium on Rice Functional Genomics**, Manila Hotel, Philippines, 16-19th November, 2009

Kuldeep Singh, Yogesh Vikal, Kuljit K Cheema, **Rajiv Sharma**, Jagjeet S Lore, Aparna Das, Dharminder Bhatia, and Vikas Gupta: A Novel Bacterial Blight Resistance Gene From Oryza nivara Mapped To 38 Kbp Region On Chromosome 4l And Transferred To *O. sativa* L. **Plant and Animal Genome Conference** 2009 XVII (P242)

Kuldeep Singh, Yogesh Vikal, Ritu Mahajan, Kuljit K Cheema, Dharminder Bhatia, **Rajiv Sharma**, Jagjeet S Lore and TS Bharaj: Three novel bacterial blight resistance genes identified, mapped and transfer to cultivated rice *O. sativa* L. abstract presented in **The 2nd International Conference on Bacterial Blight of Rice** Nanjing, China Oct 1-3, 2007 page76-77

## PERSONAL DETAILS

Nationality: Indian
Gender: Male
Date of Birth: April 14, 1982
Marital Status: Married

# ERKLÄRUNG

Herr Rajiv Sharma

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

**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 studies in diverse gene pools of barley*" 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 …….

Rajiv Sharma