Aus dem Institut für Agrar- und Ernährungswissenschaften
der Naturwissenschaftlichen Fakultät III
der
Martin-Luther-Universität Halle-Wittenberg

DEVELOPMENT OF STRATIFIED BARLEY POPULATIONS FOR ASSOCIATION MAPPING STUDIES

Dissertation

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

vorgelegt von
Raj Kishore Pasam M.Sc.
geb. am 15. August 1982 in India

Gutachter:

Prof. Dr. Andreas Graner (Gatersleben)
Prof. Dr. Klaus Pillen (Halle)
Prof. Dr. Frank Ordon (Quedlinburg)

Verteidigung am: 06th August, 2012

Halle/Saale 2012

Table of Contents

List of Abbreviations

%	percent
Mha	Million hectares
cM	centimorgan
cm	centimeter
Gb	Giga base pairs
Kb	Kilo base pairs
A_e	Allelic richness
AB-QTL	Advanced Backcross QTL
AFLP	Amplified Fragment Length Polymorphism
AM	Association mapping
AMOVA	Analysis of molecular variance
AMT	Annual mean temperature
AN	Average allele number
APT	Annual precipitation
BCC	Barley core collection
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear unbiased Predictor
BOPA	Barley Oligonucleotide Pool Assay
CPC	Crude protein content
CV	Coefficient of variation
DArT	Diversity Array Technology
EA	East Asia
EST	Expressed Sequence Tag
EU	European Union
F_{st}	Fixation index
FDR	False Discovery Rate
GD	Average gene diversity
GLM	General Linear Model
GWAS	Genome-Wide Association Studies
GxE	Genotype by environment
H_e	Heterozygosity
HD	Heading date
IBSC	International Barley Sequencing Consortium
LD	Linkage Disequilibrium
LE	Linkage Equilibrium
MAF	Minor allelic Frequency
MAS	Marker Assisted Selection

MDR	Mean diurnal temperature range
MLM	Mixed Linear Model
MTA	Marker Trait Association
MTW	Mean temperature of warmest Quarter
MxB	Morex x Barke population
NIRS	Near Infrared Reflectance Spectrometer
NJ	Neighbor-joining
OPA	Oligonucleotide Pool Assay
PCR	Polymerase chain reaction
PCA	Principal Component analysis
PCoA	Principal Co-ordinate Analysis
PHT	Plant height
PIC	Polymorphic Information Content
POPA	Preliminary Oligonucleotide Pool Assay
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
REML	Residual Maximum Likelihood
RFLP	Restriction Fragment Length Polymorphism
RT	Row type
SC	Starch content
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeats or Microsatellites
TGW	Thousand grain weight
WANA	West Asia and North Africa

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Supplementary material

Chapter 2

Chapter 3

Chapter 4

CHAPTER ONE: General Introduction

1.1 Barley history and importance

The genus *Hordeum* belongs to the *Triticeae* tribe comprising of 32 species and 45 taxa including diploid, polyploid, annual and perennial types and shows a wide geographical distribution throughout the world (Bothmer et al., 2003). Cultivated barley *Hordeum vulgare* L., the domesticated form of *Hordeum spontaneum* C. Koch, is one of the oldest known domesticated cereal crops. The immediate ancestor of cultivated barley described as *Hordeum spontaneum* was discovered by German botanist Carl Koch in Turkey (Bothmer et al., 1995). The domestication of crops is marked as epochal in the evolution of human civilizations. Barley is considered one of the founder crops of the Old World agriculture which played a major role in the development of agrarian civilizations (Diamond, 2002). Wild barley seeds have been found in many pre-agricultural sites, supporting the hypothesis that wild barley seeds have been collected from nature long before domestication (Fuller, 2007; Kilian et al., 2009). Archeobotanical evidences indicate the presence of early domesticated barley in many civilizations throughout the Middle East, Mediterranean, North Africa, East Asia. Subsequently domesticated barley spread to Europe and the Americas (Clark, 1967; Newman and Newman, 2006; Smith, 1927). The origin of domesticated barley, whether monophyletic or polyphyletic, is still a subject of constant debate (Kilian et al., 2006; Saisho and Purugganan, 2007). The discovery of wild barley in regions other than the Fertile Crescent such as Morocco, Libya, Egypt, Crete, Tibet and the vast genetic diversity of Ethiopian barley support the theory of multicentric origin of barley (Molina-Cano et al., 2005). The proposed centers of origin of barley are within Fertile Crescent region (Badr et al., 2000; Kilian et al., 2006), 1500-3000 km east of the Fertile Crescent (Morrell and Clegg, 2007), Ethiopia (Orabi et al., 2007), and Tibet (Brücher and Åberg, 1950). The importance of barley in the old world is evident from the history and also from several studies in the past 150 years that investigated barley domestication and migration patterns of various agrarian civilizations (Bothmer et al., 2003).

Barley is a diploid (2n=14) and predominantly self-pollinated crop. Barley withstands warm, dry, marginal soil environments and to some extent salinity and a broad range of soil pH conditions. Because of these features barley was grown as principal grain crop in many areas and was an important constituent of the human diet in the past (Zohary and Hopf, 2000). Cultivated and wild barley are adapted to a wide spectrum of ecological environments ranging from arctic to desert climate and can be grown in different habitats (Nevo et al., 1992). Today, barley is grown from 70° N in Norway to 46° S in Chile.

Barley consists of different morphological and adaptational forms encompassing two-rowed, six-rowed, naked, hulled, hooded, spring and winter types. Based on its end use it can be classified as feed, malting and food barley. The morphological, physiological and functional variation in barley is a reflection of the underlying large genetic diversity which eases the environmental adaptation of barley (Graner et al., 2003). Consequently, the primary genepool of barley comprises hundreds of modern cultivars and thousands of varieties and landraces.

1.2 Economic importance of barley

Barley was initially used as food grain in various forms, but later on for feed, malting and brewing purposes. Barley was an energy food and a preferred diet for building strength in ancient times. Such was the significance of barley that ancient Roman gladiators were popularly known as '*hordearii'* meaning barley men (Grando and Gómez Macpherson, 2005). However, today barley is primarily used for feed (55%-60%), secondly for malting (30%-40%) and in some areas for human consumption (2%-3%) and 5% for seed purposes (Baik and Ullrich, 2008). From the barley usage statistics, it is evident that barley is of vital importance to animal feed and for malting and brewing industries. However, recently again there is an improved interest in barley for human consumption as functional foods (Newman and Newman, 2006). Any food in its natural or processed form that in addition to the nutrients also provides substances that improve human health is considered as functional food. It has been demonstrated that barley has hypocholesterolemic effects and lowers blood sugar levels. Barley grain is a good source of both β-glucan which helps in lowering cholesterol levels and blood glucose levels; and tocols which also lower the total cholesterol levels

(Baik and Ullrich, 2008; Pins and Kaur, 2006; Wang et al., 1993). Considering the broad adaptability and health benefits of barley, developing of highly nutritive food barley to cope with changing climatic conditions can help in providing food security to humankind in future.

Table 1.1 Global barley production and cultivated area over the last 50 years. Area is area in million hectares and Prod is production in million tonnes

		Wheat	Maize	Rice	Barley	Sorghum	Millet	Oats	Rye	Triticale
1961	Area	204.2	105.6	115.4	54.5	46.0	43.4	38.3	30.3	0.0
	Prod	222.4	205.0	215.6	72.4	40.9	25.7	49.6	35.1	0.0
1971	Area	213.9	118.2	134.5	67.7	50.1	43.5	29.3	20.0	0.0
	Prod	347.5	313.6	317.7	131.2	61.9	29.7	54.5	31.7	0.0
1981	Area	239.2	127.9	145.0	81.5	45.9	37.4	26.3	15.1	0.1
	Prod	449.6	446.8	410.1	149.6	73.3	27.0	40.3	24.9	0.1
1991	Area	223.3	134.0	146.7	76.3	42.8	36.7	20.1	14.3	1.3
	Prod	546.9	494.4	518.7	169.8	55.7	24.9	33.5	29.0	4.7
2001	Area	214.6	137.5	151.9	56.2	43.4	35.0	13.1	9.9	2.9
	Prod	589.8	615.5	598.3	144.0	59.7	29.0	27.3	23.3	10.8
2009	Area	225.6	158.6	158.3	54.1	40.0	33.7	10.2	6.6	4.3
	Prod	685.6	818.8	685.2	152.1	56.1	26.7	23.3	18.2	15.7

Source: FAO (2009)

Today, barley is the fourth major cereal crop of the world after wheat, rice and maize. Barley is cultivated over 54 million hectares with an estimated yield of 152 million tons (**Table 1.1**) (2009, FAO). The distribution of barley cultivation and the global production estimates overview is provided in **Fig 1.2.1a** and **Fig 1.2.2b**. Europe is leading in barley cultivation with 51% (27 Mha) of total barley cultivated area, followed by Asia with 21%, Africa with 10%, Oceania with 7.7%, North America with 7.7% and South America with 1.9% area. Europe with 63% (95 Mt) of the total barley production is leading the world in barley production followed by Asia (14.04%), North America (9.51%), Africa (6.08%), Oceania (5.61%) and South America (1.58%) (2009, FAO). This data demonstrates the wide distribution and adaptation of barley and exemplifies the future scope of increasing barley production by extending into new areas and by increasing the overall productivity. Barley breeding for improved high yielding cultivars with environmental adaptability is one of the major approaches for increasing barley productivity and production. In a broader perspective, barley breeding for improved cultivars implies assembling of various alleles of the genes that interact among and produce optimum combinations of desired quantitative and qualitative traits of agronomic and economical importance.

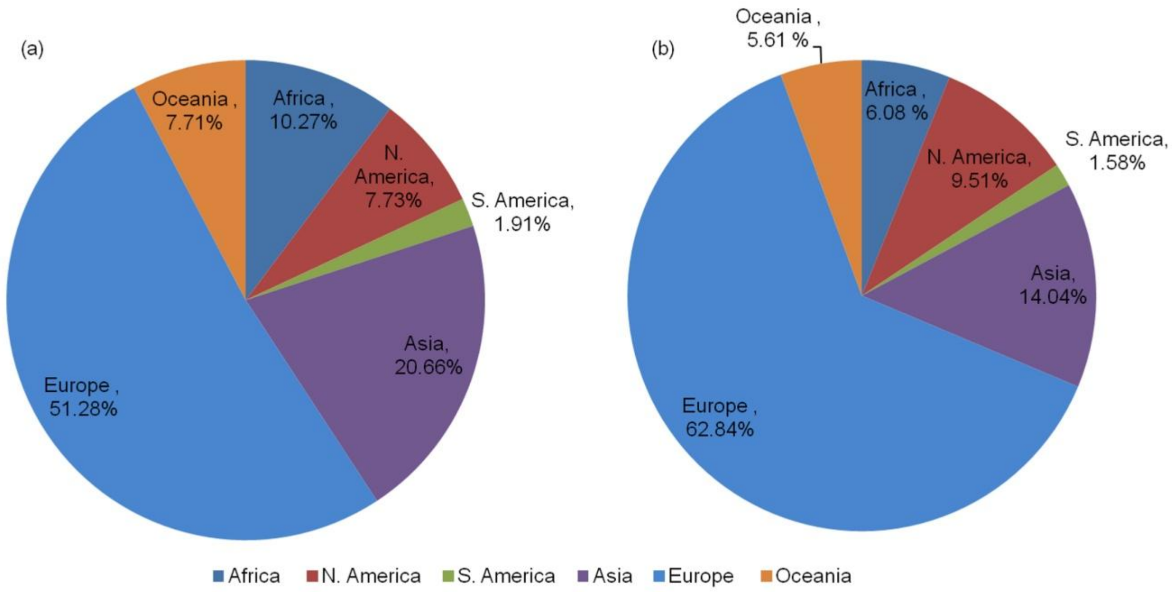


Fig 1.2.1 Barley worldwide cultivated area and production. **(a)** Worldwide distribution of barley cultivated area in percentage across the continents in 2009, and **(b)** Worldwide distribution of barley production in percentage across the continents in 2009

1.3 Barley genomic resources

In addition to its agricultural importance, the barley genome is considered as a model for other crop species of the Triticeae tribe including wheat and rye (Hayes and Szucs, 2006; Schulte et al., 2009). Barley is one of the premiere choices in plant research, but especially has been a favorite in genetic experiments. The prominence of barley in genetics is attributed to its diploid nature, low chromosome number, large chromosomes, self fertility, high degree of natural and easily inducible variation, easy hybridization, wide adaptability and relatively less space requirements (Qi et al., 1996). Its only drawback is the relative large size of the genome exceeding 5 Gbp (Bennett and Smith, 1976). Nevertheless, multiple studies of trait mapping have been published for barley using genetic maps constructed by conventional approaches to the latest molecular and physical mapping approaches and are reviewed elsewhere (Graner et al., 2010; Ullrich, 2010). The molecular era in barley emanated almost two decades ago with the publishing of first comprehensive molecular maps in barley using RFLP markers (Graner et al., 1991; Heun et al., 1991; Kleinhofs et al., 1993). Subsequently, AFLP markers were used for developing several genetic maps (Powell et al., 1997).

The advent of a second generation of molecular markers, especially the most favored simple sequence repeats (SSRs) has advanced the map building in plants. SSR markers are abundant in the genome, codominant in nature, provide high information content, have potential for automation, easy to use and readily transferable among diverse crosses (Gupta and Varshney, 2000). In barley, SSR markers have been extensively used for genetic diversity studies (Malysheva-Otto et al., 2006), for developing linkage maps (Ramsay et al., 2000) and for quantitative trait loci (QTL) studies (Li et al., 2006). New SSRs and Single Nucleotide Polymorphism (SNP) marker resources were developed from EST databases and used in barley genetic studies (Close et al., 2009; Pillen et al., 2000; Thiel et al., 2003). High throughput genotyping platforms like DArT array (Wenzl et al., 2004) and Illumina GoldenGate SNP assay (Close et al., 2009) that can simultaneously screen thousands of markers were developed and used extensively for whole genome screening purposes. Furthermore, integrated high density consensus maps were developed using multiple mapping populations and multiple marker types (Sato et al., 2009; Stein et al., 2007; Varshney et al., 2007a; Wenzl et al., 2006). Despite the large size of the barley genome, consistent efforts to the establishment of a whole genome physical map and complete genome sequence of barley were initiated by the International Barley Sequencing consortium (IBSC; http://barleygenome.org/) (Schulte et al., 2009). The whole genome sequence information for barley is in progress but still not publically available. Nevertheless, the syntenic relationships of barley with other grass genomes can be exploited by comparative genomic approaches, along with the use of available extensive genetic resources for efficient ways of gene identification and their uses in further plant research and breeding (Feuillet et al., 2008; Mayer et al., 2009; Mayer et al., 2011). The genome zipper based linear gene order model provides ample scope for tracing the genes of importance in barley and exploring the polymorphism and diversity of majority of the barley genes (Mayer et al., 2011). These new resources will accelerate identification of genes underlying the traits of interest. Use of molecular markers, genetic maps and localized quantitative trait loci (QTL) information in barley breeding can help in obtaining the desired genotypes faster and with more precision.

1.4 Barley genepools and diversity

1.4.1 Barley genepools

The concept of genepools was introduced into crop diversity studies by Harlan and de Wet (1971). The genepool concept has been used to describe the available genetic diversity within a genus based on their reproductive crossability. Three genepool models were described for barley: i) The primary genepool consists of cultivated, landraces, and includes weedy and wild forms of the crop among which there are no sterility barriers and no hindrances for gene transfer. ii) The secondary genepool consists of all taxa that can be crossed with the crop but fertile hybrids emerge only in rare cases. iii) The tertiary genepool consists of taxa from which gene transfer by pollination does not occur due to strong sterility barriers (Harlan and de Wet, 1971; Maxted et al., 2006). In barley, the outlines of the distinct genepools are presented in **Fig. 1.4.1** (Brown, 1992). Elite cultivars, varieties, landraces and the barley progenitor *Hordeum spontaneum* belong to the primary genepool of barley. The wild progenitor of barley is included in primary genepool as no crossing barriers were observed between the wild and crop forms. The secondary genepool consists only one species *Hordeum bulbosum* L, which crosses to barley with some difficulties (Pickering et al., 1994). All other *Hordeum* species are grouped under the tertiary genepool (Bothmer et al., 2003). In general, the primary genepool is given high importance in plant breeding due to the high cross ability among the taxa in the genepool.

Both early domestication and later crop improvement induced several genetic bottlenecks that resulted in reduced levels of genetic diversity in modern cultivars (Caldwell et al., 2006; Kilian, 2006). Unlike modern barley cultivars, landraces and particularly wild barley reveal ample genetic variability as they were subjected to lower extent of selection pressure. Constraints imposed by the lack of a diverse genetic base in breeding materials can be overcome by increasing the use of wild ancestors, wild relatives and landrace collections in plant breeding using appropriate strategies (Tanksley and McCouch, 1997).

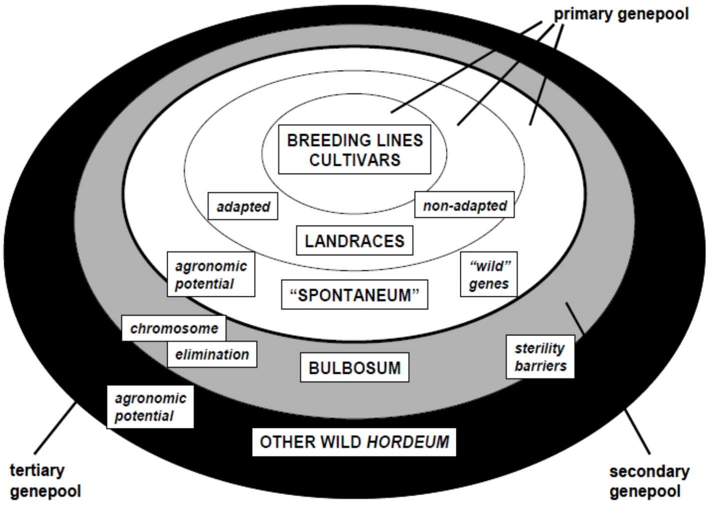


Fig 1.4.1 Schematic diagram of primary, secondary and tertiary genepools in barley (adopted from Brown 1992).

1.4.2 Wild barley

Several evidences indicate that *Hordeum spontaneum* is the progenitor of cultivated barley (Kilian et al., 2009; Nevo, 2006; Zohary and Hopf, 2000). Primitive landraces resemble very closely to wild barley and are difficult to distinguish except few special characteristics of wild barley. Characters like two-rowed spike, brittle rachis, rough awn, small kernels and seed dormancy are typical identifiers for wild barley. However, crossing between wild barley and landraces is not uncommon in regions where they are growing together. Hence the wild progenitor is also included into the primary genepool of barley (Bothmer et al., 2003; Salamini, 2002). The extent of outcrossing was found to be relatively high and variable among different wild barley populations (Abdel-Ghani et al., 2004; Brown et al., 1978). The high level of genetic diversity and low levels of linkage disequilibrium (LD) in wild barley offers a rich and largely untapped source of unique alleles for crop improvement (Caldwell et al., 2006; Morrell et al., 2005).

1.4.3 Barley landraces

Landraces are early domesticates of crops improved by local farmers over generations mainly by mass selection techniques. Early in the 20th century landraces were increasingly replaced by cultivars that were developed by cross breeding. Nevertheless, cultivation of barley landraces persisted in some regions in Europe, Asia and North Africa where harsh growing conditions prevail and where no systematic breeding activities had been established (Fischbeck, 2003; Jones et al., 2011). Early barley cultivars were still derived from direct selections among landraces or descended from genetic recombination of their parents of different landrace origin. Since then, barley breeding is mainly revolving around the use of accessions from elite genepools. Consequently, the basis for genetic diversity in present barley breeding materials has rather declined and is limited (Fischbeck, 2003).

Most of the existing vast diversity in locally adapted barley landraces and exotic germplasm is either abandoned or stacked in the genebank vaults. Landraces represent the largest part of barley germplasm conserved in genebanks worldwide. Among the total known type of barley germplasm stored in genebanks, 1,28,870 accessions (44%) represent landraces (Annonymus, 2008). Landraces are unexplored repositories of allelic diversity and contain useful alleles for crop improvement under both biotic (Silvar et al., 2011) and abiotic stress environments. Studies in the past showed that landraces performed better than cultivars under stress environments; while modern genotypes were better under stress free environments (Ceccarelli and Grando, 1996; Pswarayi et al., 2008). Knowledge of genetic diversity in landraces will help in better understanding of the genetic basis of the environmental adaptation and for efficient exploitation of underlying natural variation. This deeper understanding serves as a prerequisite for effective utilization of landraces in future breeding programs to achieve long term gains in agriculture.

1.4.4 Barley diversity

Different molecular genetics studies in barley have been reported using different markers like AFLPs (Badr et al., 2000; Varshney et al., 2007c), RFLPs (Graner et al., 1994; Graner et al., 1990), RAPDs (Russell et al., 1997), SSRs (Malysheva-Otto et al., 2006; Matus and Hayes, 2002; Pillen et al., 2000), DArTs (Zhang et al., 2009) and SNPs (Russell et al., 2011; Varshney et al., 2007c). An ever increasing reserve of these markers can be efficiently utilized for barley genetic and diversity studies (Close et al., 2009; Varshney et al., 2007a; Wenzl et al., 2006). Several genetic diversity studies were performed in barley using different germplasm collections. Malysheva-Otto et al. (2006) surveyed the genetic variation in a collection of 953 barleys using 48 SSRs. Hamblin et al. (Hamblin et al., 2010) studied the population structure and diversity in 1816 barley lines from the United States breeding programs using 1536 SNP markers. Parzies et al. (Parzies et al., 2000) evaluated Syrian landraces stored for various periods in genebanks and compared them with recently sampled Syrian landraces using morphological and isozyme markers. Pandey et al. (Pandey et al., 2006) studied 107 landraces collected from Himalayan ranges of Nepal for population structure using 44 SSRs. Yahiaoui et al. (Yahiaoui et al., 2008) evaluated the genetic diversity of 159 Spanish landraces and 66 European cultivars using 64 SSRs and investigated the association of population structure with geographic and climatic factors. Gong et al. (Gong et al., 2009) used 52 SSRs and assessed the genetic diversity among 33 wild barley accessions from Qinghai-Tibet region and 56 landraces from China. Hübner et al. (Hubner et al., 2009) studied the genetic diversity of 1010 wild barley accessions from Israel using 42 SSR markers and described the pivotal role of temperature and precipitation in shaping the current population structure of wild barley.

1.5 Barley breeding and genetic mapping of traits

1.5.1 Barley breeding

Plant breeding can benefit from the developments in genomics through i) genetic characterization of available germplasm resources, ii) tagging, cloning, and introgressing genes and or Quantitative trait loci (QTL) useful for enhancing the target trait, and iii) manipulating genetic variation in breeding populations (Xu and Crouch, 2008). The genetic variation of a quantitative trait is assumed to be controlled by collective effects (additive, dominance) of quantitative trait loci, epistatic effects between the QTL, environment and interaction between the QTL and environment. Genetic mapping of a complex quantitative traits provides knowledge about their inheritance and genetic architecture and besides, identifies markers that can be used as selection tools in plant breeding (Bernardo, 2008). DNA markers tightly linked to the gene/QTL can be used as molecular tools for marker-assisted selection (MAS). One of the best examples for the use of MAS in practical barley breeding is resistance pyramiding against the barley yellow mosaic virus complex using markers closely linked to *rym*-5 and *rym*-4 loci (Ordon et al., 2003). Currently, more molecular markers are being used to track the loci for traits like stress tolerance, yield and quality parameters in practical barley breeding programs (Collard and Mackill, 2008; Rae et al., 2007; Schmierer et al., 2005; Varshney et al., 2007b).

1.5.2 Value of wild and landraces for crop improvement

In crop species, genetic bottlenecks occurring during the transition from wild to domesticated germplasm, and from early domesticated to modern cultivars has resulted in loss of diversity and left behind potentially useful alleles (Tanksley and McCouch, 1997). The understanding of the dynamics of genetic variation in cultivated crops helps in germplasm conservation, germplasm enhancement and efficient resource utilization (Hamblin et al., 2011). This understanding in general has initiated the programs for germplasm collection and conservation for food security and agriculture in the start of last century (Vavilov, 1940) resulting in establishment of genebanks and germplasm collections. However, till now the use of wild and landrace genepools for crop improvement and modern breeding programs is still unfledged. Assessment of the genetic variation and genetic relationships present among accessions are important considerations for plant breeding and can aid in maintaining biodiversity in breeding materials.

The shifting paradigm in plant breeding research in recent years is undoubtedly benefiting from the population genetics framework imputed with linkage mapping, association mapping and comparative genomics approaches. The detection of QTL for economic traits and introgression QTL alleles using both elite and exotic materials was proposed to be a potential approach (Collard and Mackill, 2008; Prada, 2009). Up to now there are some success stories of fine mapping, isolating, cloning and characterizing new genes/QTL and are discussed elsewhere (Salvi and Tuberosa, 2005). Most of these studies demonstrated the importance of wild and exotic germplasm in contributing useful alleles towards improvement of cultivated genepools (Hoisington et al., 1999), which endures the hope to discover novel alleles by allele mining approaches. Examples include: *fw2.2* in tomato (Frary et al., 2000); seven resistant alleles of the powdery mildew resistance gene *pm3* in a wheat landrace collection (Bhullar et al., 2010); the successful transfer of powdery mildew resistance gene from *H. bulbosum* to barley (Pickering et al., 1995); and more than 30 disease resistance genes from wild introgressions are used in wheat breeding today (Hoisington et al., 1999).

In barley, several studies have been reported where wild and landrace materials have been used to introgress useful alleles into the elite germplasm (Feuillet et al., 2008). AB-QTL methods to discover and mobilize useful alleles from wild into cultivated were successfully implemented (Pillen et al., 2003; Pillen et al., 2004; von Korff et al., 2010). Superior alleles for disease resistance against powdery mildew, leaf rust and scald were introgressed from wild in to cultivated barley (von Korff et al., 2005). Identification of favorable agronomic QTL and alleles useful for improvement of malting quality from wild were reported in AB-QTL studies (von Korff et al., 2006; von Korff et al., 2008). In cultivated barley powdery mildew resistance is provided by alleles from the cloned *Mlo* gene. The naturally occurring allele *mlo-11* is the major *mlo* resistance allele in barley and is retrieved from Ethiopian barley landraces (Piffanelli et al., 2004). Boron tolerance gene (*bot1*) identified as boron-toxicity tolerance gene in barley was isolated by map based cloning approach. The favorable tolerance alleles for *bot1* are derived from Algerian landrace Sahara (Sutton et al., 2007). These examples demonstrate that wild barley and landraces can be employed to enrich the diversity in the cultivated elite germplasm. The successive articulation of the evolving genomic and genetic techniques will step-up the chances for better utilization of genetic variation stagnating in genebank shelves.

1.6 QTL mapping

The concept of detecting QTL started in the early decades of 20th century (Sax, 1923). However, the advent of the marker technologies and availability of powerful biometric methods in later decades has enabled the generation of linkage maps in many crops and consequently numerous QTL studies were reported (Asíns, 2002). QTL mapping is a key tool for assessing the genetic architecture of the underlying complex traits and facilitating estimation of number of genomic regions affecting the trait (James B, 2007). The detection of genes or QTL is mainly possible due to genetic linkage analysis which is based on recombination during meiosis (Tanksley, 1993). Both linkage mapping and linkage disequilibrium mapping strategies exploit the fact that recombination breaks up the genome into small fragments that can be correlated to the phenotype (Myles et al., 2009).

1.6.1 Linkage mapping or bi-parental QTL mapping

Most of the agronomically important traits are quantitative, resulting in difficulty for discerning genetic differences underlying the phenotype of interest. Currently, linkage mapping (analysis) is the most common approach in plants to detect quantitative trait loci (QTL) corresponding to complex traits. In linkage mapping, segregating populations are established by crossing two parental lines. The co-segregation of alleles of mapped marker and phenotypic traits allows the identification of markers linked to the trait. Due to the restricted number of meiotic events that are captured in a biparental mapping population, the genetic resolution of QTL maps often remains confined to a range of 10-30 cM (Flint-Garcia et al., 2003; Zhu et al., 2008). Moreover, linkage analysis can only sample a small fraction of all possible alleles in a population from which the parents originated. Several QTL studies for agronomic, biotic resistance, abiotic tolerance and quality traits using bi-parental approach have been reported in barley and are reviewed elsewhere (Hayes et al., 2003).

1.6.2 Association mapping (AM) or Linkage Disequilibrium (LD) mapping

An alternative approach, association mapping (AM) known as LD mapping relies on existing natural populations or designed populations of crop plant species to overcome the constraints inherent to linkage mapping. Two terms used in population genetics to describe linkage relationships are linkage equilibrium (LE) and linkage disequilibrium (LD). LE is random association of alleles at different loci. LD is the non-random association of alleles at separate loci or can also be referred as the historically reduced level of the recombination of specific alleles at different loci (Flint-Garcia et al., 2003; Hill and Robertson, 1968). Association mapping is a population based method used to identify marker trait associations (MTA) based on LD (Mackay and Powell, 2007). LD mapping exploits all ancestral recombination events that occurred in the population and takes into account all major alleles present in the population to identify significant marker-phenotype associations. LD mapping was first introduced in genetic mapping studies in humans (Hastbacka et al., 1992; Lander and Schork, 1994) and has been recently considered for plant research (Flint-Garcia et al., 2003). By exploiting non-random associations of alleles at nearby loci (LD), it is possible to scoop out significantly associated genomic regions with a set of mapped markers. Success of mapping depends on the quality of phenotypic data, population size and the degree of LD present in a population (Flint-Garcia et al., 2005; Mackay and Powell, 2007). In general, the power of association studies depends on the degree of LD between genotyped markers and the functional polymorphisms. The decay of LD varies greatly i) between species (Gupta et al., 2005), ii) among different populations within one species (Caldwell et al., 2006; Tenaillon et al., 2001), and iii) also among different loci within a given genome.

LD mapping is based on two strategies: i) re-sequencing of selected candidate genes and ii) genome-wide association which exploits marker polymorphisms across all chromosomes (Hirschhorn and Daly, 2005). Genome-wide association studies (GWAS) have become increasingly popular and powerful over the last few years in human and animal genetics. The emergence of more cost-effective, high-throughput genotyping platforms have rendered AM an attractive approach for QTL mapping in plants (Atwell et al., 2010). In the last few years, an increasing number of association studies based on the analysis of candidate genes have been published (reviewed in Gupta et al. 2005). These include e.g. the *Dwarf8* (Thornsberry et al., 2001) and the *phytoene synthase* locus in maize (Palaisa et al., 2003), flowering time genes in barley (Stracke et al., 2009), the *PsyI-AI* locus in wheat (Singh et al., 2009), the *rhg-1* gene in soybean (Li et al., 2009); and a series of candidate genes in *Arabidopsis* (Ehrenreich et al., 2009; Zhao et al., 2007). Over the last few years, candidate genes based AM studies were reported for barley (Caldwell et al., 2006; Haseneyer et al., 2010a; Stracke et al., 2007). GWAS with dense marker coverage are not yet conducted routinely for barley, albeit the potential of this approach has been demonstrated in some studies (Cockram et al., 2010; Ramsay et al., 2011; Rode et al., 2011; Waugh et al., 2009).

1.6.3 Statistical methods for LD mapping

In association mapping, the complex genetic relatedness among individuals and the population structure affects the mapping of the phenotype as the allele frequencies are highly biased between sub populations and are correlated to the phenotype variation between the populations. As a result of this genotype-phenotype covariance, spurious associations between markers and phenotype are observed (Flint-Garcia et al., 2005; Myles et al., 2009). Inbreeding crops such as barley are characterized by a high level of population structure caused by the impact of non random mating and subsequent selection. This is exemplified by two-rowed and six-rowed barley cultivars which form distinct subpopulations, because the corresponding breeding programs rely on different progenitors. The same applies to the subpopulations of spring and winter barley (Thiel et al., 2003). Occurrence of type I and type II errors is higher in AM than in biparental QTL analysis due to the confounding effect of population structure in the panel (Breseghello and Sorrells, 2006; Myles et al., 2009; Zhu et al., 2008). Specific statistical approaches have been proposed to account for population structure in AM (Price et al., 2006; Pritchard et al., 2000b). Yu et al. (Yu et al., 2006) described a mixed-linear model (MLM) approach which performs better than previous models (Stich and Melchinger, 2009). Still these models have their individual shortcomings and care needs to be taken in controlling for population structure and balancing the rate of false positives and false negatives in the analysis.

1.6.4 Prospects of LD mapping in plant crops

Potential advantages of LD mapping or GWAS are: i) increased mapping resolution, ii) breeding lines can be directly used for mapping studies, iii) diverse and relevant plant materials are phenotyped and genotyped, and iv) even genes with a small to modest effect can be detected (Myles et al 2009; Zhu et al 2008). There are also few potential drawbacks for GWAS approach. In general, GWAS requires a large number of markers depending on the genome size and the expected LD decay in the population. If the LD decays within 5kb across the genome then the optimum SNP requirement to cover the whole genome is predicted to be as high as 93,200 SNPs for rice, 147,000 for sorghum, 480,000 for maize, 1.1 million for barley and 3.2 million for wheat. Even if the LD decay is assumed to extend to 100 kb, the optimum SNP requirement will still be 4.660 for rice, 7,350 for sorghum, 24,000 for maize, 57,000 for barley and 160,000 for wheat (Semagn et al., 2010). Such an exorbitant density of markers is possible by genotyping by sequencing platforms which are only used in few crops till now (Huang et al., 2010; Lai et al., 2010). Nevertheless, most of the GWAS reported in barley till now used the available SNP marker resources which have yielded good results (Waugh et al. 2009; Ramsay et al. 2011). However further research is needed to determine the optimum marker density and population size for reliable GWAS in barley. In this regard, an ever increasing repertoire of marker and sequence resources has been developed for barley which can be efficiently utilized (Close et al., 2009; Graner et al., 2010; Rostoks et al., 2006; Wenzl et al., 2004).

Genetic diversity, relatedness within the population, population stratification, genome-wide LD extent, sample size, allelic penetrance, and allele frequency distribution determine the credibility, resolution and power of LD mapping (Flint-Garcia et al. 2003; Mackay and Powell 2007; Zhu et al. 2008). Selection of germplasm is a critical consideration for success of association mapping studies. As a consequence of genetic bottlenecks in the course of domestication and consequent selection, the allele frequencies are altered resulting in increased LD and reduced genetic variation. The extent of LD decreases gradually from modern cultivars to landraces to wild genepools and inverse trend is observed in case of allelic diversity. The price of higher LD is low resolution in GWAS studies (Hamblin et al., 2011). To fine map selected QTLs, staggered patterns of LD decay observed for different genepools of barley (cultivars, landraces, wild barley) may be exploited (Waugh et al. 2009; Caldwell et al. 2006). Several association mapping panels are available for GWAS in barley, however most of them are either cultivar collections or landraces from specific regions (Comadran et al., 2011; Massman et al., 2011; Wang et al., 2012). Up to now extended genepools of barley were neither characterized for their diversity nor explored for GWAS.

1.7 Research objectives and outline of the thesis

1.7.1 Objectives

The present thesis is aimed at three broad goals: i) to investigate different association mapping methods for understanding the genetic complexity underlying agronomic traits in spring barley. Phenotypic data from multi-environment locations were analyzed to identify marker trait associations for the traits of interest, ii) to investigate the effects of marker density on QTL detection using LD mapping approaches, and iii) to establish a spring barley landrace panel for association mapping and to characterize the genetic diversity and population structure in spring barley landrace collection. The detailed objectives of each goal are provided below.

Chapter 2:

To study the suitability of worldwide spring barley collection for GWAS, and to evaluate different GWAS methods using 918 SNP markers is described in chapter 2.

1. One of our main objectives was to map genetic polymorphisms underlying complex agronomic traits such as heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and crude protein content (CPC) in spring barley using GWAS.
2. To investigate the diverse spring barley collection comprising 224 accessions from 52 countries for phenotypic and genotypic variation.
3. To provide a comprehensive overview on population structure and genetic diversity as well as their effects on GWAS.
4. To study the dynamics of LD decay across the seven barley chromosomes.
5. To investigate different statistical approaches for GWAS.
6. To evaluate the suitability of the population for GWAS studies
7. To identify and locate QTL for the traits investigated and confirm from the previously known QTL positions.

Chapter 3:

The impact of increased marker number on QTL detection in worldwide spring barley collection by using GWAS approach is described in chapter 3.

1. GWAS of agronomic traits using the same panel of cultivars as in chapter 1 but applying 7000 SNP markers.
2. To investigate the influence of different kinship matrices based on different SNP marker sets on GWAS results.
3. To investigate the effect of marker density on the QTL discovery.

Chapter 4:

Chapter 4 describes the establishment and SSR fingerprinting of a barley landrace collection with the following objectives:

1. To study the genetic diversity of landraces originating from various geographical and climatic regions.
2. To provide insight into the population structure and subgroups of the collection.
3. To investigate the eco-geographical distribution and diversity of these landraces.
4. To study the suitability of the collection for GWAS as whole population or sub-sampled small populations.
5. To construct small core groups based on the genetic diversity and to compare the diversity of these core groups to the whole collection.

1.7.2 Thesis outline

This thesis is divided into five major chapters. In addition to the Introduction (chapter 1) and the Discussion (chapter 5) the Results presented in chapters 2, 3 and 4 are written as research articles. Therefore each of these chapters follows the scheme of a scientific paper, i.e. is subdivided into Introduction, Materials and Methods, Results, and Discussion. As these chapters are treated as independent research articles, when gathered into a single thesis there is bound to be some repetition which is always associated with the general focus of the thesis.

CHAPTER TWO: Genome-Wide Association Studies for Agronomical Traits in a World Wide Spring Barley Collection

2.1. Introduction

Genome-wide association studies (GWAS) based on linkage disequilibrium (LD) provide a promising tool for the detection and fine mapping of quantitative trait loci (QTL) underlying complex agronomic traits. In this study the genetic basis of variation for the traits heading date, plant height, thousand grain weight, starch content and crude protein content was investigated in a diverse collection of 224 spring barleys of worldwide origin. The whole panel was genotyped with an oligonucleotide pool assay containing 1536 SNPs using Illumina's GoldenGate technology (Close et al., 2009) and later with an Illumina iSelect assay containing 7864 SNPs (Comadran et al., unpublished). The morphological trait "row type" (two-rowed spike vs. six-rowed spike) was used to confirm the high level of selectivity and sensitivity of the approach. This study describes the detection of QTL for the above mentioned agronomic traits by GWAS. Different statistical models were tested to control spurious LD caused by population structure and to calculate the *P*-value of marker-trait associations. The results demonstrate that the described diverse barley panel can be efficiently used for GWAS of various quantitative traits, provided that population structure is appropriately taken into account. The observed significant marker trait associations provide a refined insight into the genetic architecture of important agronomic traits in barley. However, individual QTL account only for a small portion of phenotypic variation, which may be due to insufficient marker coverage and/or the elimination of rare alleles prior to analysis. The fact that combined SNP effects fall short of explaining the complete phenotypic variance may support the hypothesis that the expression of a quantitative trait is caused by a large number of very small effects that escape detection.

2.2 Materials and Methods

2.2.1 Association mapping panel

The association mapping panel consists of 224 spring barley accessions selected from the Barley Core Collection (BCC) (Knüpffer and van Hintum, 2003) and the barley collection maintained at the IPK Genebank Gatersleben, Germany. The panel comprises 128 two-rowed and 96 six-rowed genotypes, and among them 109 accessions originates from Europe (EU), 45 from West Asia and North Africa (WANA), 40 from East Asia (EA) and 30 from the Americas (AM). Most of the accessions are improved cultivars (149), some accessions are landraces (57) or breeder's lines (18). Further information on the germplasm can be obtained from the European Barley Database (EBDB, http://barley.ipk-gatersleben.de/ebdb.php3). This panel has been considered and described in detail by Haseneyer et al. (Haseneyer et al., 2010b). Each accession has been single-seed descended, selfed for two generations under greenhouse conditions and subsequently propagated in the field.

2.2.2 Phenotypic evaluation

The accessions were planted in a 25 x 15 lattice design with three replications in the years 2004 and 2005 at the following locations: Stuttgart (Southwest Germany), Irlbach (Southeast Germany) and Wohlde (Northern Germany). Heading date (HD) and plant height (PHT) were scored in field plots. Thousand grain weight (TGW) was estimated from sampled grains per plot. Starch content (SC) and crude protein content (CPC) were estimated using a near infrared reflectance spectrometer (NIRS) from ground seed samples from all environments. In order to convert the nitrogen content to crude protein values, a factor of 6.25 was considered. The methods described in Naumann and Bassler (Naumann and Bassler, 2004) were fallowed to estimate the starch content and nitrogen content. Phenotypic data were analyzed using REML (Residual Maximum Likelihood) implemented in GenStat 9 software (Payne, 2006). Variance components were calculated by fitting a mixed linear model (MLM) to multi-environment data. Heritabilities were estimated for all traits considering the percentages of genotypic variance, over the total phenotypic variance including genotype (G) by environment (E) variance and error variance components. Phenotypic mean

BLUEs (Best Linear Unbiased Estimates) were estimated taking into account the GxE variance and were used for association studies. Further information on phenotypic data can be obtained from Haseneyer et al. (Haseneyer et al., 2010b).

2.2.3 Genome-wide marker profiling

Illumina GoldenGate assay (1536 SNPs)

DNA for SNP genotyping was extracted for each accession from bulked leaf samples of eight 2-weeks old seedlings. A customized oligonucleotide pool assay (IPK-OPA, unpubl) containing 1536 allele specific oligos was used to genotype the panel by Illumina's GoldenGate technology (Illumina, San Diego, CA). The IPK-OPA has been mainly built on a selection of markers from two pilot assays (pOPA1, pOPA2) that are polymorphic between the two barley cultivars 'Barke' and 'Morex'. More than 95% of the 1536 SNP markers of the IPK-OPA have been included in a barley consensus map (**Supp Table 2.1**; Close et al. 2009). The SNP genotyping was performed at University of California (Southern California Genotyping Consortium, UCLA) following the protocol of Fan et al. (Fan et al., 2006; Fan et al., 2003). More details about the successful SNP markers considered for GWAS are available as supplemental information (**Supp Table 2.1**).

Scoring SNP data was done using the Illumina Beadstudio software package (Genotyping module 3.2.32; Genome viewer 3.2.9; Illumina, San Diego, CA) that can process the raw hybridization intensity data and thereby cluster the data. The normalization procedure implemented in the Beadstudio genotyping module includes outlier removal, background correction and scaling. The algorithm included uses a Bayesian model to assign normalized intensity values to one of the three possible homozygous and heterozygous genotype clusters. Stringent threshold scores (Call Rate > 0.9 and GenTrain Score > 0.7) were used to identify ambiguous results. SNPs that failed to show two-group clustering were strictly excluded from the analysis. From a total of 1536 SNP markers, 985 markers yielded good quality genotypic calls. Among the 985 successful SNP markers only 957 markers are genetically mapped and these 957 markers were used for analysis (**Supp Table 2.1**). Among the 224 accessions in the panel of genotypes, 12 genotypes performed badly in the assay (**Supp Table 2.2**). For these 12 genotypes more than 90% of the SNP markers data is missing, hence were excluded from subsequent analysis.

2.2.4 Genotypic data analysis and population structure

Polymorphic information content (PIC) values were calculated for each SNP using Powermarker 3.25. (Liu and Muse, 2005). Major allele frequency, minor allele frequency (MAF), gene diversity and Nei's genetic distance (*d*) (Nei, 1972) were calculated and a NJ (Neighbor-Joining) dendrogram (data not shown) based on *d* was computed. From the 957 SNPs, a final set comprising 918 SNPs with MAF larger than 0.05 was used for analysis of population structure, LD and marker trait associations. Polymorphism Information Content (PIC) values are determined according to Botstein et al. (Botstein et al., 1980) using the formula:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{k} p_{i^2} - \mathit{\Sigma}_{i=1}^{k-1} \mathit{\Sigma}_{j=i+1}^{k} 2p_i^2 p_j^2$$

Where p_i and p_j are the frequencies of alleles *i* and *j* respectively.

To estimate the number of subgroups in the panel, different methodologies and different software packages were employed and compared in order to determine the appropriate population structure in the collection. For the quantitative assessment of the number of groups in the panel, a Bayesian clustering analysis was performed using a model based approach implemented in the software package STRUCTUREv 2.2 (Falush et al., 2003; Pritchard et al., 2000a). This approach uses multi-locus genotypic data to assign individuals to clusters or groups (*k*) without prior knowledge of their population affinities and assumes loci in Hardy-Weinberg equilibrium. The program was run with 918 SNP markers for *k*-values 1 to 15 (hypothetical number of subgroups), with 10.0000 burnin iterations followed by 50.000 MCMC (Markov Chain Monte Carlo) iterations for accurate parameter estimates. To verify the consistency of the results five independent runs were performed for each *k*. An admixture model with correlated allele frequencies was used. The most probable number of groups was determined by plotting the estimated likelihood values [LnP(D)] obtained from STRUCTURE runs against *k*. LnP(D) is the log likelihood of the observed genotype distribution in *k* clusters and is an output by STRUCTURE simulation. The *k* value best describes the population structure based on the criteria of maximizing the log probability of data or in other words the value at which LnP(D) reaches a plateau (Pritchard et al., 2000a). STRUCTURE results with the SNP marker dataset were confirmed with the results from STRUCTURE runs using a set of Diversity Array Technology (DArT) markers (Pasam et al. unpubl, **Supp Fig 2.1**). In a second approach principal coordinate analysis (PCoA) based on the dissimilarity matrix was performed using DARwin (Diversity Analysis and Representation for windows) (Perrier and Jacquemound-Collet, 2006). In a third approach a NJ dendrogram based on Nei’s genetic distance matrix was constructed. The substructure in the collection using different methodologies was compared and the final *k* value using STRUCTURE was ascertained. For this *k* value, the Q-matrix (population membership estimates) was extracted from STRUCTURE runs. This matrix provides the estimated membership coefficients for each accession in each of the subgroups.

2.2.5 Linkage disequilibrium analysis

The extent of LD affects both the number of markers required for GWAS and the resolution of mapping the trait. LD is in many cases influenced by population structure resulting from the demographic and breeding history of the accessions. Genome-wide LD analysis was performed among the panel and subgroups by pair wise comparisons among the SNP markers using HAPLOVIEW (Barrett et al., 2005). LD was estimated by using squared allele frequency correlations (r^2) between the pairs of loci (Weir, 1996). The loci were considered to be in significant LD when $P < 0.001$, the remaining r^2 values were not considered as informative. The pattern and distribution of intra-chromosomal LD was visualized and studied from LD plots generated for each chromosome by HAPLOVIEW. To investigate the average LD decay in the whole genome among the panel, significant intra-chromosomal r^2 values were plotted against the genetic distance (cM) between markers. The smothering second degree LOESS curve was fitted using GENSTAT (Payne, 2006). A critical value for r^2 was estimated by square root transforming of unlinked r^2 values to obtain a normally distributed random variable, and the parametric 95th percentile of that distribution was taken as a critical r^2 value (Breseghello and Sorrells, 2006).

Unlinked r^2 refers to marker loci with a map distance greater than 50 cM or on independent linkage groups.

2.2.6 Association analysis

Different statistical models were used to calculate P-values for associating each marker with the trait of interest, along with accounting for population structure to avoid spurious associations by TASSEL v.2.1 (www.maizegenetics.net). We followed the formula $y = X\beta + M\alpha + Zu + e$, where y is a response vector for phenotypic values, β is a vector of fixed effects regarding population structure, α is the vector of fixed effect for marker effects, u is the vector of random effects for co-ancestry and e is the vector of residuals. X can be either the Q-matrix or the PCs from Principal Component Analysis (PCA), M denotes the genotypes at the marker and Z is an identity matrix. Six models comprising both general linear models (GLM) and mixed linear models (MLM) were selected to test the marker-trait-associations (MTA). Results were compared to determine the best model for our analysis. PCA was conducted with TASSEL. The first ten significant PCs explained 43% of the cumulative variance of all markers. A kinship matrix (K-matrix), the pair-wise relationship matrix which is further used for population correction in the association models was calculated with 918 SNP markers using TASSEL (Bradbury et al., 2007). The following models were tested: i) Naive model: GLM without any correction for population structure; ii) Q-model: GLM with Q-matrix as correction for population structure; iii) P-model: GLM with PCs as correction for population structure; iv) QK-model: MLM with Q-matrix and K-matrix as correction for population structure; v) PK-model: MLM with PCs and K-matrix as correction for population structure and vi) K-model: MLM with K-matrix as correction for population structure (Kang et al., 2008; Pritchard et al., 2000b; Stich et al., 2008; Yu et al., 2006). All SNP markers were re-mapped by association mapping to determine the mapping resolution of the panel as suggested by Rostoks et al. (Rostoks et al., 2006). The critical P-values for assessing the significance of MTAs were calculated based on a false discovery rate (FDR) separately for each trait (Benjamini and Hochberg, 1995), which was found to be highly stringent. Considering the stringency of the model used for accounting for population structure, most of the false positives were inherently controlled.

Thus, a more liberal approach as proposed by Chan et al. (Chan et al., 2010) was considered for determining the threshold level for significant MTAs. It was suggested that the bottom 0.1 percentile distribution of the *P*-values can be considered as significant, which in our analysis resulted in threshold levels of 0.05 to 0.09 for individual traits. Alternatively, as a compromise between the two approaches an arbitrary threshold P-value of 0.03 was used for all traits and all models. This rather rough estimate was obtained by arranging -$\log_{10}$ P-values in a descending order, and the value at which the curve starts to flatten is determined as the threshold value. All association models with all traits were re-analyzed using GENSTAT (Payne, 2006) to check for any discrepancy.

2.3 Results

2.3.1 Phenotypic data

Large phenotypic variation was observed for all traits. Outliers in the data were identified based on the residuals derived from the data of all environments and were removed from further analysis. For the trait heading date, data from the year 2004 were excluded from the analysis due to differences in scoring this trait between the individual locations. Variance components were calculated by REML. The results confirmed that the genotypic variance was significant for all traits ($P < 0.001$). GxE interactions were also significant ($P < 0.001$) but represented only a small fraction of the total variance. Heritabilities ranged between 0.90-0.95 indicating the robustness of the data and the low error rate. Year-wise means, ranges and heritabilities over all environments for the traits HD, PHT, TGW, SC and CPC are presented in **Table 2.3.1** and their frequency distributions are illustrated in **Supp Fig 2.2**. The correlation exhibited by the agronomic traits between each other is outlined in **Table 2.3.2.** The traits SC and CPC are highly correlated (-0.7) and other traits showed moderate to weak correlation among each other. PHT was shown to be weakly correlated with HD and also with SC and CPC. TGW is found to be positively correlated with SC and negatively correlated with CPC. Substantial phenotypic differences were reported between two-rowed and six-rowed genotypes. The means for all traits were significantly different between the two groups (**Supp Table 2.3**).

Table 2.3.1 Estimation of mean, minimum (Min), maximum (Max) and heritabilities (h^2) of traits. Heritabilities were calculated on entry mean basis.

	2004			2005			h^2(%)
Trait	Min	Max	Mean	Min	Max	Mean	GxE
Plant height	20	120	77.04	30	120	73.69	92.82
Heading date	*	*	*	56	81	68.26	92.5
Thousand grain weight	17.77	67.23	44.92	20.1	62.6	42.43	92.9
Starch content	40.8	64.58	56.85	44.01	65.31	56.91	96.3
Protein content	9.74	25.74	14.88	10.35	25.18	14.90	92.1

Table 2.3.2 Correlation coefficients among different traits

	HD	PHT	CPC	SC
HD				
PHT	0.29**			
CPC	-0.43**	-0.25**		
SC	0.43**	0.17*	-0.76**	
TGW	0.04	-0.09	-0.30**	0.33**

**highly significant at *P* < 0.001, * significant at *P* < 0.01, rest are not significant

The variation observed was larger for all traits in six-rowed barleys than in two-rowed barleys. The greatest influence of spike morphology (two-rowed vs. six-rowed) on phenotypic variation was seen for TGW, whereas the greatest influence of population structure was observed for PHT (**Supp Table 2.4**).Best Linear Unbiased Estimates (BLUEs) of genotypic means were calculated from the fixed genotypic effects to avail unbiased mean estimates. Using Best Linear Unbiased Predictors

(BLUPs) is less suitable as it would cause double shrinking (Smith et al., 2001). Henceforth BLUEs were used for further analysis. However, comparison of both BLUPs and BLUEs revealed very high concordance between both estimates, which is a direct consequence of the high heritabilities (**Supp Fig 2.3**).

2.3.2 Population structure and genetic diversity

From the high quality 985 SNPs, 957 markers had been genetically mapped and therefore were considered for this study. Of these, 39 SNPs (4%) were excluded because of a MAF below 0.05. Majority of the remaining SNPs showed MAF from 0.1 to 0.5 (**Fig. 2.3.1**). These SNP markers were distributed over all seven chromosomes with an average spacing of 1.18 cM. The distribution of SNP markers is not exactly uniform and varies within and among chromosomes with a minimum of 105 markers on chromosome 4H and a maximum of 164 markers on 5H (**Table 2.3.3**). PIC values for SNPs ranged from 0.09 to 0.5 with an average of 0.30. Most of the markers (726) displayed PIC values exceeding 0.25, demonstrating the informativeness of these markers in current panel. The average PIC values of the markers on each chromosome ranged between 0.29 (5H) to 0.33 (6H).

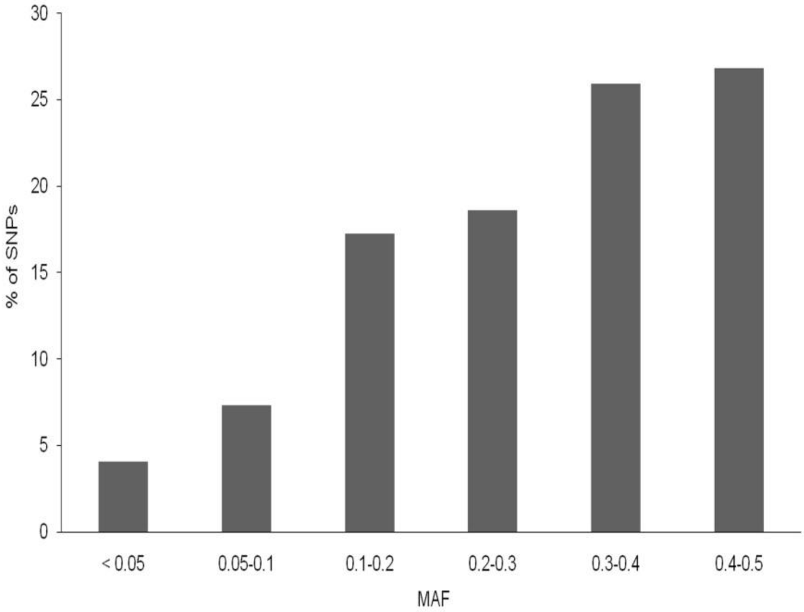


Fig. 2.3.1 SNP marker efficiency in the panel. Distribution of SNPs in the panel according to the minor allele frequency (MAF). SNPs with MAF < 0.05 were excluded from the analysis

Table 2.3.3 SNP coverage and distribution across all chromosomes. Average PIC values for all SNPs on each chromosome are represented

Chromosome	cM	Markers	Marker coverage	PIC
1H	139.79	117	1.19	0.31
2H	156.72	146	1.07	0.29
3H	173.17	151	1.15	0.32
4H	123.29	105	1.17	0.32
5H	195.42	164	1.19	0.29
6H	129.38	119	1.09	0.33
7H	166.56	116	1.44	0.30
Total	1084.33	918	1.18	0.31

The mean gene diversity value for the whole panel was 0.39 and spread within a range of 0.09 to 0.5. It was reported in several studies that the stratification of barley cultivars is concordant with spike morphology, mainly as a result of breeding history (Malysheva-Otto et al., 2006; Zhang et al., 2009). Therefore, similar molecular diversity statistics were generated separately for two-rowed and six-rowed barley groups within th panel and for the six subgroups. Observed mean PIC values are higher for the six-rowed group (0.31) than for two-rowed barleys (0.27). Similarly, average gene diversity estimated was higher in six-rowed (0.38) than in two-rowed accessions (0.33).

The population structure in the panel of 212 barley accessions was analyzed using 918 SNP markers and a model based approach in STRUCTURE. The LnP(D) appeared to be an increasing function of *k* for all the values observed. But the most significant increase of LnP(D) was observed when *k* was increased from 1 to 2 (**Fig. 2.3.2**). At *k* = 2 the panel is clearly categorized into two-rowed and six-rowed barleys with few exceptions. The two main groups were further divided yielding six subgroups in total as LnP(D) values nearly reached a plateau at *k* = 6. Hence, we chose a value of *k* = 6 for our analysis as minimum number of groups present in the panel. Different values of *k* are still possible but will not qualitatively affect the results. An accession was assigned to a subgroup if at least 50% of the genome information was estimated to belong to this group.

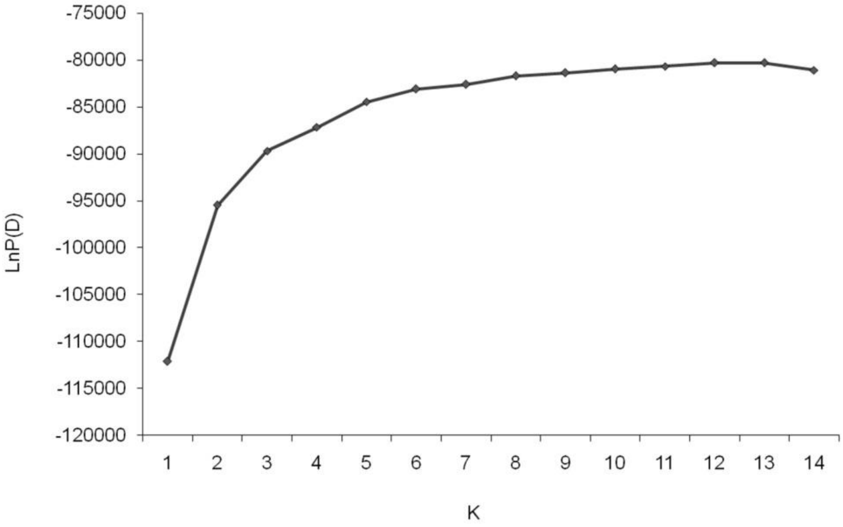
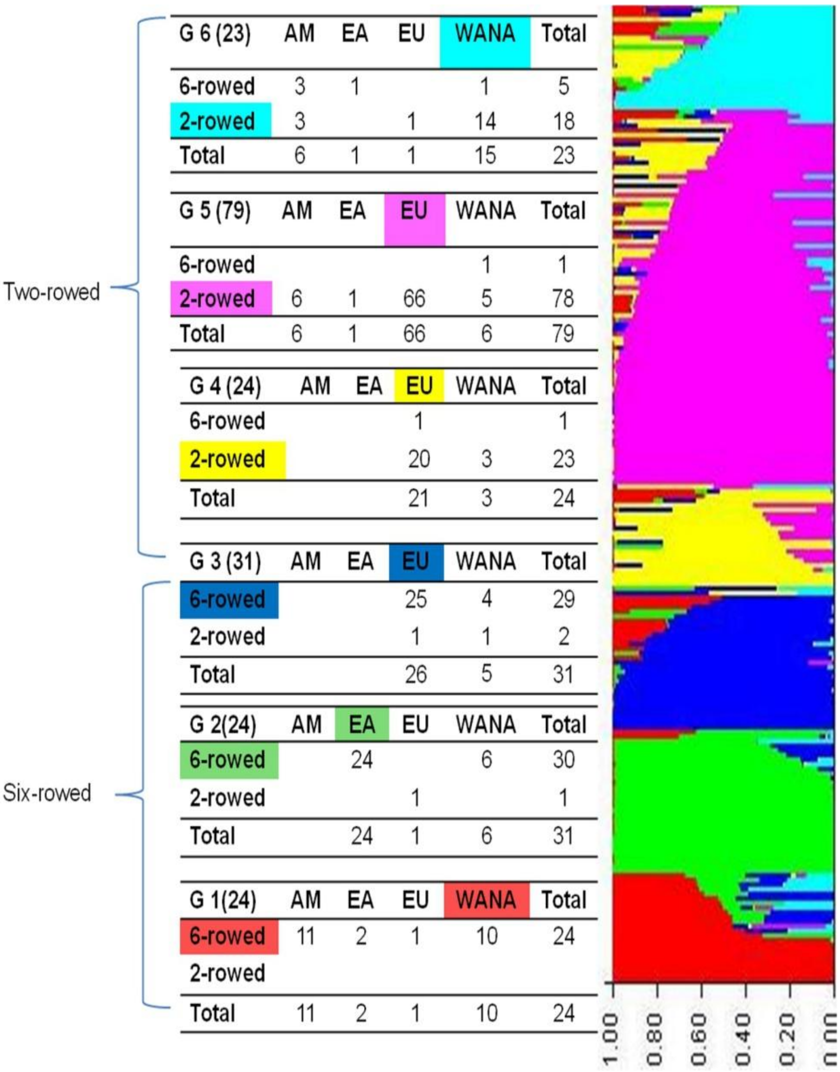


Fig. 2.3.2 STRUCTURE results using 918 SNPs. Log probability data (LnP(D)) as function of *k* (number of clusters) from the STRUCTURE run. The plateau of the graph at *K*=6 indicates the minimum number of subgroups possible in the panel



G 6 (23)	AM	EA	EU	WANA	Total
6-rowed	3	1		1	5
2-rowed	3		1	14	18
Total	6	1	1	15	23

G 5 (79)	AM	EA	EU	WANA	Total
6-rowed				1	1
2-rowed	6	1	66	5	78
Total	6	1	66	6	79

G 4 (24)	AM	EA	EU	WANA	Total
6-rowed			1		1
2-rowed			20	3	23
Total			21	3	24

G 3 (31)	AM	EA	EU	WANA	Total
6-rowed			25	4	29
2-rowed			1	1	2
Total			26	5	31

G 2(24)	AM	EA	EU	WANA	Total
6-rowed		24		6	30
2-rowed			1		1
Total		24	1	6	31

G 1(24)	AM	EA	EU	WANA	Total
6-rowed	11	2	1	10	24
2-rowed					
Total	11	2	1	10	24

Fig. 2.3.3: Population sub-structuring in the panel. Bayesian clustering of the 212 barley accessions into six defined groups (G1, G2, G3, G4, G5, G6) based on 918 SNP markers. The number of accessions per group and their respective geographical origin and row type is presented

Table 2.3.4 Summary of molecular diversity and polymorphism information for the whole panel. PIC values are given as the mean values of the corresponding marker partitions

Group	Average major allele frequency	No. genotypes	Gene diversity	PIC
Whole panel	0.6978	212	0.3904	0.3079
2-rowed group	0.7551	122	0.3325	0.2714
6-rowed group	0.7064	90	0.3852	0.3108
G1	0.7359	24	0.3551	0.2903
G2	0.7933	31	0.2844	0.2338
G3	0.7773	31	0.3060	0.2497
G4	0.7746	24	0.3106	0.2536
G5	0.7976	79	0.2791	0.2297
G6	0.7547	23	0.3296	0.2681

Table 2.3.5 Estimation of average genetic distance between different groups

Group	G 1	G 2	G 3	G 4	G 5
G 2	0.24				
G 3	0.24	0.30			
G 4	0.27	0.30	0.29		
G 5	0.31	0.36	0.35	0.17	
G 6	0.21	0.24	0.27	0.21	0.26

The accessions clustered into groups mostly according to their spike morphology and their geographical origin. The six groups are defined as: Group 1 (G1): 24 six-rowed barleys mostly from AM and WANA; G2: 31 accessions mostly six-rowed barley from EA; G3: 31 accessions mostly six-rowed barleys from EU; G4: 24 accessions mostly two-rowed from EU; G5: 79 accessions mostly two-rowed barleys from EU; G6: 23 accessions mostly two-rowed from WANA and AM (**Fig. 2.3.3**). The dominant stratification of the population according to spike morphology is confirmed by PCoA (**Supp Fig. 2.4**) and NJ dendrogram (not shown). In the PCoA, it is obvious that the primary axis separates the accessions based on row type and further grouping is related to the region of origin. Overall, the clustering of accessions was consistent among various methods and the genetic diversity within these groups was further explored.

The summary statistics for each group with 918 SNP markers is reported in **Table 2.3.4**. Observed gene diversity values ranged from 0.27 in G5 to 0.35 in G1; PIC values ranged from 0.22 in G5 to 0.29 in G1. Pairwise genetic distances ranged from 0.006 to 0.628, with an overall mean of 0.39. The average overall genetic distance between groups has been calculated, and the largest genetic distance of 0.36 was observed between the groups G2 (six-rowed, EA) and G5 (two-rowed, EU). Similarly G4 (six-rowed, EU) and G5 (six-rowed, EU) are found to be closely related groups with an average genetic distance of 0.17 (**Table 2.3.5**).

2.3.3 Linkage disequilibrium

LD analysis was performed using 918 SNPs for i) entire panel, ii) separately for two-rowed and six-rowed barleys, and iii) each of the six subgroups. Pairwise LD was estimated using the squared-allele frequency correlations (r^2) and was found to decay rapidly with the genetic distance. Different aspects of LD was studied in current panel and observed that LD varies along the chromosomes with regions of high LD interspersed with regions of low LD (**Supp Fig. 2.5**). A critical value of r^2, or basal LD, was calculated from LD analysis of unlinked pairs of loci and is estimated to be 0.2 beyond which LD is assumed to be caused by genetic linkage. The point at which the LOESS curve intercepts the critical r^2 is determined as the average LD decay of the population. Based on these criteria the intra-chromosomal LD decayed between 5-10 cM for individual chromosomes and average LD decay of the whole genome was observed to be at 7 cM (**Fig. 2.3.4**). Extensive variability in the magnitude of r^2 at a given genetic distance was detected reflecting the wide local variation in the extent of LD across the chromosomes. The correlation between r^2 and marker distance was found to be significantly negative (r = -0.40) for markers below a distance of 10 cM, whereas marker pairs with larger distance showed no significant correlation with r^2.

Significant intra-chromosomal r^2 values ($P < 0.001$) ranged from 0.02 to 1 with an average of 0.12 for the whole panel. Among all significant loci in LD, 13.7% of the loci are above the critical r^2 value of 0.2 in the whole panel. Pairs of loci are classified into 4 groups based on the inter-marker genetic distance: 0-10 cM (tightly linked markers), 11-20 cM (moderately linked markers), 21-50 (loosely linked markers) and >50 (independent markers) (Maccaferri et al., 2005).

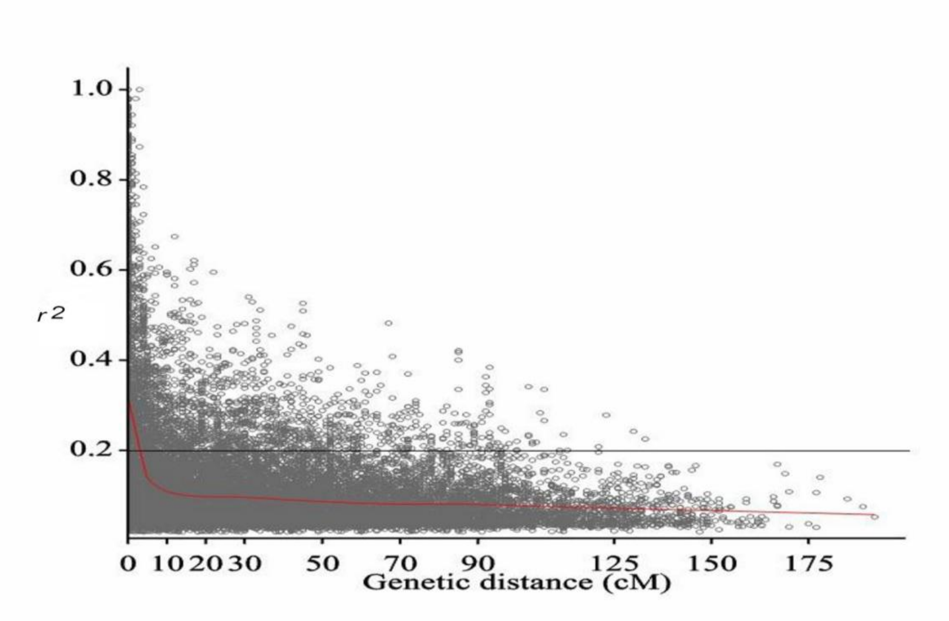


Fig. 2.3.4 Intra-chromosomal LD (r^2) decay of marker pairs over all chromosomes as a function of genetic distance (cM). The horizontal line indicates the 95th percentile distribution of unlinked r^2. The LOESS fitting curve (red line) illustrates the LD decay

Table 2.3.6 LD overview for the whole panel and the subgroups of two-rowed and six-rowed barley. LD statistics are given for the total number of locus pairs and for different marker linkage classes (for details see text).

		Total pairs	% signifi-cant	Significant pairs	Mean r^2	Pairs in complete LD	% of pairs in LD > 0.2	Mean r^2> 0.2
Whole panel	total	62222	39.4	24567	0.12	59	13.72	0.36
	0-10 cM	10602	62.2	6554	0.20	59	33.70	0.42
	11-20cM	8028	45.1	3626	0.10	0	10.21	0.29
	21-50 cM	19066	38.3	7310	0.09	0	8.40	0.27
	>50 cM	24526	28.5	7077	0.08	0	4.00	0.25
2-rowed	total	48803	21.6	10544	0.18	94	23.74	0.43
	0-10 cM	8183	50.0	4098	0.29	94	48.00	0.48
	11-20cM	6244	29.9	1869	0.13	0	13.31	0.30
	21-50 cM	15066	17.2	2601	0.12	0	9.59	0.28
	>50 cM	19310	10.2	1976	0.10	0	3.81	0.26
6-rowed	total	58356	20.2	11801	0.17	95	22.40	0.36
	0-10 cM	9947	36.8	3661	0.24	95	40.37	0.43
	11-20cM	7439	21.1	1569	0.14	0	18.26	0.27
	21-50 cM	17768	16.9	3016	0.14	0	15.04	0.27
	>50 cM	23202	15.3	3555	0.13	0	12.14	0.25

The percentages of significant loci pairs and mean r^2 values for all classes of markers in the whole panel and different subgroups are presented in **Table 2.3.6**. Among all loci pairs, only 39.4% were in significant LD in the whole panel. The percentage of significant loci pairs decreased with the distance between loci; 62.2% of the tightly linked markers showed significant r^2. Similarly 45.1%, of the moderately linked markers 38.3% of the loosely linked markers and 28.5% of independent markers were in significant LD. The portion of r^2 values exceeding the basal LD level of 0.2 decreased from 33.7% in the group of tightly linked markers to 10% for moderately linked markers to less than 4% for independent markers. Mean r^2 values decreased from 0.2 for closely linked marker loci to 0.08 for unlinked marker pairs. All loci pairs being in complete LD are spaced by map distances < 5 cM.

2.3.3.1 Patterns of linkage disequilibrium within subgroups

At the intra-chromosomal level, mean r^2 values for two-rowed and six-rowed barley groups ranged between 0.18 and 0.17, which is slightly more than the mean r^2 of the whole panel. The percentages of significant r^2 values were higher in the two-rowed than in the six-rowed sub-group for all classes of marker pairs except for the independent markers. This pattern is also similar to LD values above the basal level of 0.2, and a slightly slower LD decay was observed for two-rowed barley compared to the group of six-rowed types and to the whole panel. Similarly, the mean r^2 values were estimated for individual subgroups where they ranged from 0.3 (G5) to 0.49 (G4). LD decay in individual subgroups was much slower than in the whole panel. In **Fig. 2.3.5**, binned r^2 values are mapped against the map distance (cM) across the genome. In the whole panel the average LD decays below a basal level (0.2) within 5 cM, while in the two-rowed and six-rowed groups the basal level is reached between 10-15 cM and with LD in six-rowed barley decaying faster than in two-rowed barley. Within G5 LD decays to the basal level within 20-25 cM, while it does not reach the basal level in the remaining subgroups (G1, 2, 3, 4, 6). Average LD decay graphs for each group showed different patterns. Specifically, in the subgroups G4 and G5 at distances 45 and 74 cM larger LD peaks were observed. Scrutinizing these peaks revealed that high LD in these regions was caused by markers with low allele frequencies.

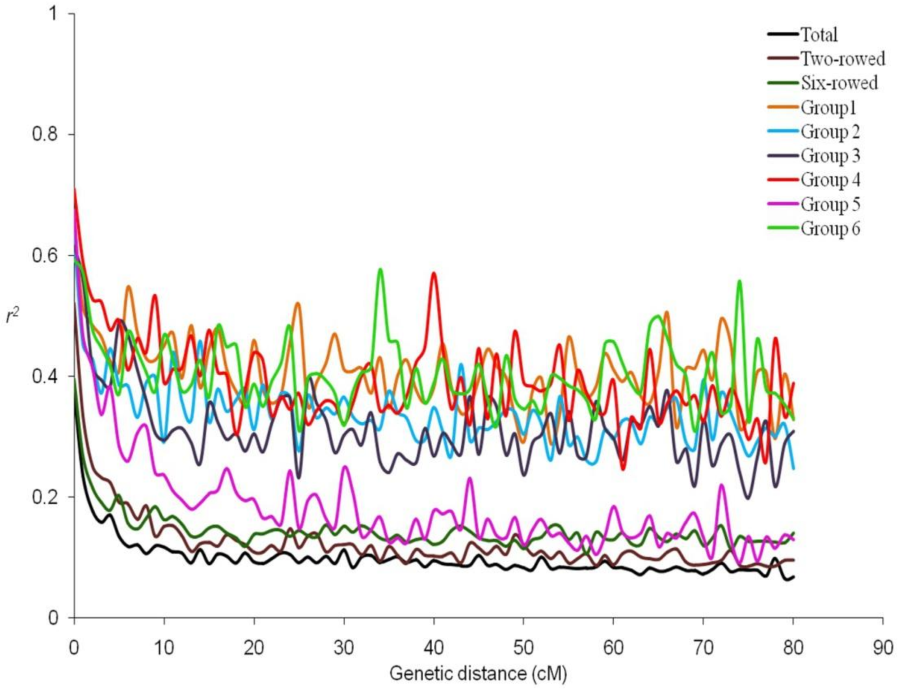


Fig. 2.3.5 Comparison of LD patterns and LD decay in the whole panel and subgroups. Mean r^2 values are plotted against the genetic distance for different groups

The consequence of the reduced population size of the individual subgroups is that the presence of allele in four accessions already might show a MAF above the critical threshold. Varying patterns of LD decay in different sub-populations are likely reflecting their breeding histories (Flint-Garcia et al., 2003) and may impinge on the QTL mapping resolution of the panel. However, there is a chance that smaller group size can sometimes result in overestimation of the LD.

2.3.4 Evaluation of the association panel

All 918 SNPs were re-mapped using an LD approach. A model with kinship accounting for population structure was used for estimating the genetic map position of the markers. Each marker was used as an individual trait and the analysis was run with the remaining SNPs to find the most significantly associated markers. The map distance between the target marker in question and the most highly associated marker was used to evaluate the resolution of the panel. More than 85% of the SNP markers had their genetic map position within 0-10 cM distance of their original map position and the majority of them re-mapped at the same position (**Fig. 2.3.6**). The original map positions used here were the consensus map positions obtained by using three mapping populations (Close et al. 2009). This re-mapping of markers shows that the resolution of QTL captured by AM approach in current panel will be within a range of 5-10 cM.

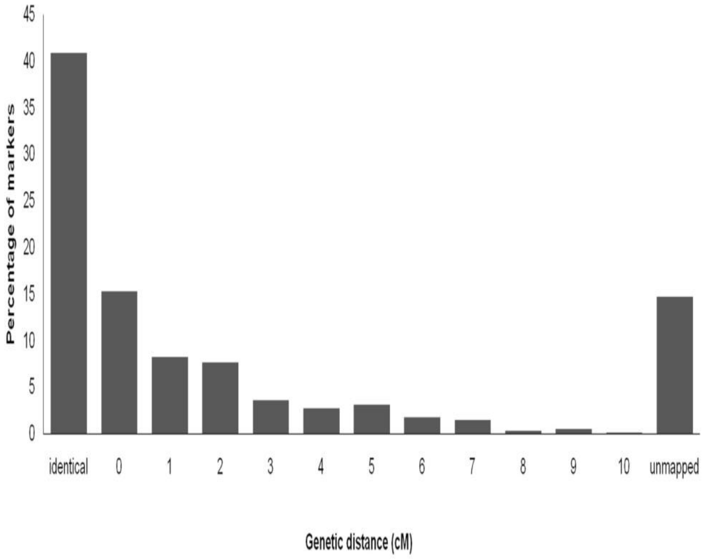


Fig. 2.3.6 Evaluating the mapping resolution of the panel. Distribution of SNPs according to their re-mapped distances using the K-model of genome-wide association approach. The group 'identical' refers to the SNPs that mapped at exactly the same position and the group '0' refers to the SNPs that mapped within a distance of 0.01 to 0.99 cM. The group unmapped refers to the SNPs that mapped beyond 10 cM of their original map distance

2.3.5 Association analysis

2.3.5.1 Comparison of models

Several models were tested to detect associations between SNP markers and agronomic traits. Owing to the complexity and the considerable amount of population structure present in our panel, Numerous spurious associations were observed when using the naive (simple) model for AM. Hence, the usefulness of various linear models to account for population structure was assessed by comparing their ability to reduce the inflation of false positive associations. To this end, ranked P-values from GWAS were plotted in a cumulative way for each model by using spike morphology as phenotypic trait (**Fig. 2.3.7**). As demonstrated by Kang et al. (Kang et al., 2008) the distribution of P-values ideally should follow a uniform distribution with less deviation from the expected P-values. The models QK, PK and K showed a good fit for P-values, while the other models were characterized by the excess of small P-values which is tantamount to an abundance of spurious associations. This is particularly obvious in the case of the “naive" model, where nearly half of the P-values are smaller than 0.01. On the other hand, the K-model performed similar to the PK and QK model in displaying a highly uniform distribution of P-values and at the same time requiring less computational time. Irrespective of the model, major marker trait associations were constantly detected. However, the more stringent the model was the less spurious background associations were detected. All models considered for GWAS are presented for the trait spike morphology (**Supp Fig. 2.6**). For all other traits only results from the K-model will be presented and discussed.

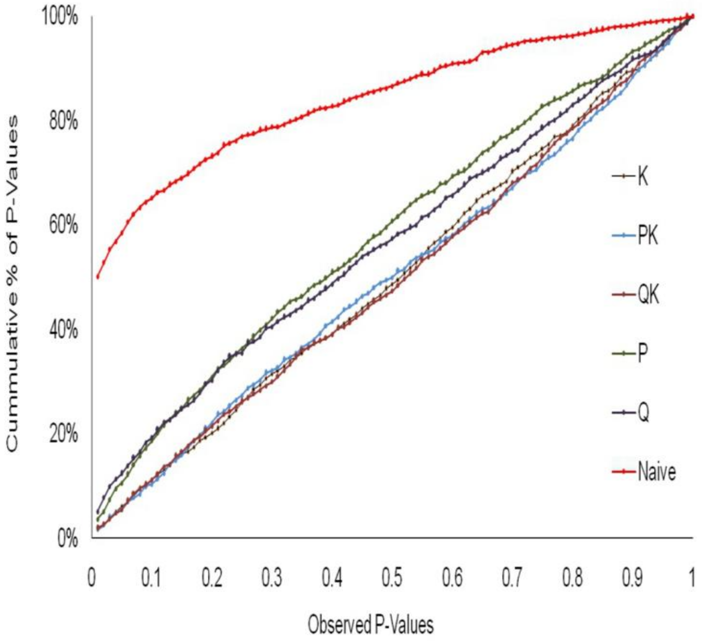


Fig. 2.3.7 Comparison of different GWA models. Cumulative distributions of P-values computed by GWAS approach for trait row-type using 918 SNPs with different association models are presented. The more uniform the distribution of P-values, the better is the model.

2.3.5.2 GWAS results

Barley spike morphology (row type)

Apart from comparing different AM models, we aimed to examine spike morphology as a proof of concept for GWAS and to evaluate the resolution of the association panel. According to its spike morphology barley is classified as two-rowed and six-rowed types and the genes for this trait have been well documented with some of them already cloned (Pourkheirandish and Komatsuda, 2007; Ramsay et al., 2011; Waugh et al., 2009). The row type character was scored in the panel and considered 918 markers for AM using all models. A marker trait association was considered when the marker main effect was significant at 0.03 [$-\log_{10}(0.03) = 1.5$].

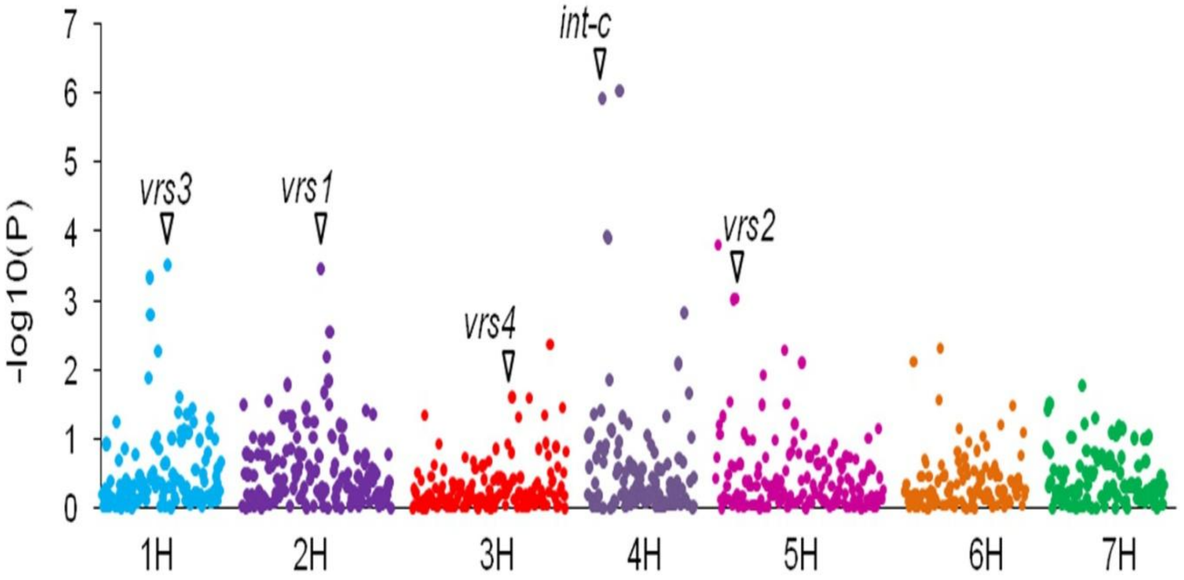


Fig. 2.3.8 GWA scan for the trait row type using 918 SNPs with K-model for statistical correction of population structure. Vertical axis represents –log10(P) values of the *P*-value of the marker trait association. SNPs in the vicinity of the genes *vrs1*. *vrs2*. *vrs3*, *vrs4* and *int-c* are marked with arrows

This resulted in a total of 34 markers that are significantly associated with the trait row type by using the K-model. Some of the results are congruent with previous row type studies (see **Fig. 2.3.8**).

Heading date

Thirty-four markers were found to be significantly associated with heading date (HD). These were grouped into 19 QTL located on all chromosomes. Significant marker trait associations within a genetic distance of 5-10 cM are delineated into a single QTL. Chromosome 2H harbors the maximum number of markers associated with the trait (**Fig. 2.3.9a**). Some of these association results with the SNP markers effectively correspond to genomic regions of previously mapped flowering time QTL. These include genomic regions of various prominent flowering pathway genes like *Ppd-H1*, *HvFT1*, *HvCO1* and *HvCO3* (see **Table 2.3.7**).

Plant height

Thirty-two markers displayed significant associations with plant height (PHT). These markers detected 19 QTL (**Table 2.3.8**). Except for chromosome 1H, significantly associated markers were found on all chromosomes with the majority located on 2H and 3H (**Fig. 2.3.9b**).

Thousand grain weight

Thirty-six markers yielding 21 QTL were significantly associated with Thousand Grain Weight (TGW, **Fig. 2.3.9c**). Markers significantly associated with the trait were present on all chromosomes. As expected some of the TGW related QTL overlapped with QTL for spike morphology. The markers SNP56, SNP215, SNP385 and SNP458 are co-localized to the same region as *Vrs3*, *Vrs1*, *Vrs4* and *Int-c* genomic regions (**Table 2.3.9**).

Starch content

Thirty-five markers were found to be significantly associated with the trait Starch Content (SC). These markers formed a total of 25 QTL (**Fig. 2.3.9d**). Significantly associated markers for starch content were present on all chromosomes. Similar to TGW markers corresponding to the *Vrs3* region (SNP56 & SNP66) are significantly associated with starch content. Several significant markers, co-localized with previously mapped genes and QTL for SC (**Table 2.3.10**).

2.3.5.7. Protein content

Thirty-four markers were found to be significantly associated with crude protein content (CPC). These markers detected a total of 23 QTL (**Fig. 2.3.9e**) and were distributed over all chromosomes. Some of the QTL for protein content overlapped with the QTL regions identified for CPC (**Table 2.3.11**).

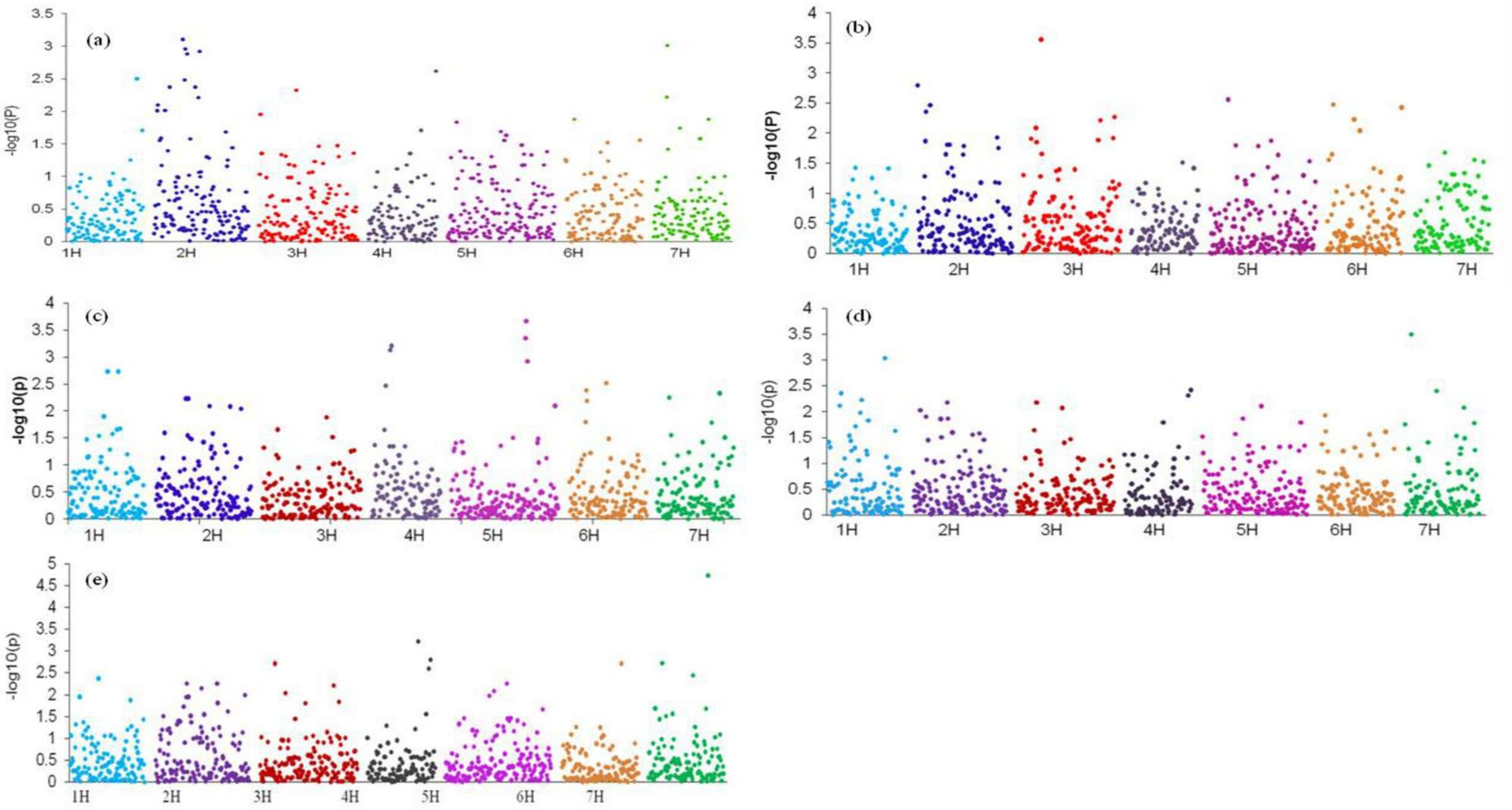


Fig. 2.3.9 GWA scans for traits HD (9a), PHT (9b), TGW (9c), SC (9d) and CPC (9e) using 918 SNPs and the K-model. Vertical axis represents –log10(P) values of the *P*-value of the marker trait association. The peaks above minimum threshold of 1.5 (*P*-value = 0.03) can be considered as significantly associated

Table 2.3.7 GWAS results for trait heading date. Significant markers associated for heading date with K-model, corresponding MAF, *P*-value of association, variance explained by marker (R^2), effect of the most significant marker within the QTL interval, name of the QTL, and the reference QTL or gene from literature

SNP	Chr	Position	MAF	P-value	-log10(P)	R^2 (%)	Effect	QTL	Reference QTL	Literature
SNP111	1H	128.14	0.19	0.0032	2.49	0. 63	-2.51	QTL1_HD	*HvFT3*	Wang et al. 2010
SNP119	1H	138.92	0.23	0.0198	1.70	0. 33				
SNP129	2H	27.29	0.37	0.0099	2.00	0. 38	-1.29	QTL2_HD	*PpdH1*	Laurie et al. 1995; Wang et al. 2010
SNP130	2H	28.44	0.36	0.0080	2.10	0. 39				
SNP133	2H	31.02	0.38	0.0280	1.55	0. 39				
SNP135	2H	33.73	0.10	0.0262	1.58	0. 51				
SNP142	2H	41.66	0.27	0.0096	2.02	0. 40	-1.37	QTL3_HD		
SNP148	2H	53.53	0.34	0.0043	2.37	0. 47				

SNP170	2H	63.53	0.32	0.0007	3.10	0. 88	-2.32	QTL4_HD	*HvFT4/ eam6*	Faure et al. 2007;
										Wang et al. 2010;
										Comadran et al. 2011

SNP174	2H	63.53	0.41	0.0011	2.96	0. 68				
SNP177	2H	63.53	0.32	0.0013	2.89	0. 55				
SNP173	2H	63.53	0.33	0.0033	2.48	0.53				
SNP183	2H	66.83	0.44	0.0265	1.58	0. 39				
SNP191	2H	71.12	0.44	0.0043	2.37	0. 4	2.53	QTL5_HD	*eps2*	Laurie et al. 1995
SNP196	2H	73.04	0.10	0.0061	2.21	0. 65				
SNP199	2H	73.75	0.14	0.0012	2.92	0. 49				
SNP242	2H	115.78	0.39	0.0207	1.68	0. 54	-1.98	QTL6_HD		
SNP284	3H	8.23	0.24	0.0111	1.95	0. 42	-1.45	QTL7_HD		
SNP340	3H	59.89	0.35	0.0047	2.33	0. 38	1.80	QTL8_HD	*HvGI*	Wang et al. 2010
SNP520	4H	82.42	0.32	0.0198	1.70	0. 47	-1.15	QTL9_HD		
SNP543	4H	123.29	0.26	0.0024	2.62	0. 67	-1.59	QTL10_HD		
SNP559	5H	39.97	0.46	0.0146	1.84	0. 32	-1.20	QTL11_HD	*HvCO3*	Griffiths et al. 2003
SNP630	5H	100.28	0.20	0.0203	1.69	0. 58	-2.02	QTL12_HD		
SNP635	5H	102.06	0.13	0.0278	1.56	0. 34				
SNP636	5H	103.92	0.28	0.0278	1.56	0. 35				
SNP639	5H	108.18	0.38	0.0236	1.63	0. 53				
SNP728	6H	28.39	0.43	0.0132	1.88	0. 28	-1.43	QTL13_HD		
SNP778	6H	60.23	0.49	0.0306	1.51	0. 37	-2.15	QTL14_HD		
SNP829	6H	124.85	0.33	0.0281	1.55	0. 37	1.52	QTL15_HD		
SNP854	7H	37.55	0.36	0.0060	2.22	0. 57	2.50	QTL16_HD	*HvFT1*	Faure et al. 2007; Wang et al. 2010
SNP855	7H	38.32	0.35	0.0009	3.01	0. 54				
SNP875	7H	68.46	0.06	0.0180	1.74	0. 48		QTL17_HD		
SNP908	7H	84.92	0.35	0.0266	1.58	0. 37	1.66	QTL18_HD	*HvCO1*	Griffiths et al. 2003;
SNP921	7H	104.78	0.35	0.0131	1.88	0. 59		QTL19_HD		

Table 2.3.8 GWAS results for trait plant height. Significant markers associated for trait plant height with K-model

Marker	Chr	Position	MAF	P-value	-log10(P)	R^2 (%)	Effect	QTL	Reference QTL	Literature
SNP122	2H	8.57	0.13	0.0016	2.80	0.94	-5.64	QTL1_PHT		
SNP136	2H	33.74	0.35	0.0138	1.86	0.54	-5.23	QTL2_PHT	*Ph2*	Qi et al. 1998
SNP137	2H	38.03	0.35	0.0044	2.36	0.84				
SNP168	2H	59.21	0.16	0.0229	1.64	0.46	6.73	QTL3_PHT		
SNP171	2H	63.53	0.10	0.0155	1.81	0.51				
SNP175	2H	63.53	0.10	0.0155	1.81	0.51				
SNP199	2H	73.75	0.14	0.0224	1.65	0.49	4.72	QTL4_PHT	*sdw3*	Gottwald et al. 2004
SNP200	2H	74.37	0.25	0.0162	1.79	0.54				
SNP254	2H	130.01	0.20	0.0117	1.93	0.56	-4.57	QTL5_PHT	QHt.StMo-2H.2	Hayes et al. 1993
SNP256	2H	131.77	0.28	0.0175	1.76	0.5				
SNP295	3H	36.49	0.13	0.0124	1.91	0.55	-5.57	QTL6_PHT	QHt.HaMo-3H	Hayes et al. 1993; Marquez-Cedillo et al. 2001
SNP303	3H	43.23	0.23	0.0083	2.08	0.61				
SNP304	3H	46.31	0.35	0.0141	1.85	0.54				
SNP312	3H	52.50	0.25	0.0002	3.55	1.15	-5.80	QTL7_PHT	*uzu*	Grain genes database
SNP313	3H	52.50	0.42	0.0220	1.66	0.48				
SNP404	3H	126.27	0.39	0.0129	1.89	0.55	4.11	QTL8_PHT	*sdw1/denso*	Jia et al. 2011; Yin et al. 1999
SNP406	3H	127.10	0.37	0.0061	2.21	0.65				
SNP427	3H	155.09	0.06	0.0120	1.92	0.56	-2.68	QTL9_PHT		
SNP429	3H	162.15	0.06	0.0053	2.28	0.75				
SNP519	4H	80.79	0.18	0.0306	1.51	0.4	3.70	QTL10_PHT	QHei.pil-4H.5	Pillen et al. 2003
SNP575	5H	50.27	0.31	0.0028	2.55	0.8	5.30	QTL11_PHT		
SNP588	5H	51.30	0.41	0.0159	1.80	0.5				
SNP623	5H	85.93	0.16	0.0164	1.79	0.51	4.95	QTL12_PHT		
SNP643	5H	110.26	0.10	0.0133	1.88	0.41	-5.58	QTL13_PHT	HT	Yin et al. 1999
SNP654	5H	132.63	0.44	0.0235	1.63	0.46	-4.15	QTL14_PHT	QHei.pil-5H.1	Pillen et al. 2003
SNP722	6H	12.54	0.42	0.0229	1.64	0.41	-5.11	QTL15_PHT		
SNP724	6H	16.97	0.30	0.0033	2.48	0.8				
SNP757	6H	55.36	0.09	0.0060	2.22	0.64	-8.54	QTL16_PHT		
SNP766	6H	55.36	0.32	0.0092	2.04	0.6				
SNP831	6H	124.85	0.28	0.0038	2.42	0.74	-4.89	QTL17_PHT		
SNP882	7H	73.75	0.45	0.0210	1.68	0.46	-4.23	QTL18_PHT	HT	Yin et al. 1999
SNP947	7H	144.45	0.41	0.0301	1.52	0.43	-3.77	QTL19_PHT		

Table 2.3.9 GWAS results for trait thousand grain weight. Significant markers associated for thousand grain weight with K-model

Marker	Chr	Position	MAF	P-value	-log10(P)	R^2 (%)	Effect	QTL	Reference QTL	Literature
SNP48	1H	55.49	0.47	0.0288	1.56	0.34	-2.19	QTL1_TGW		
SNP56	1H	61.53	0.26	0.0128	1.92	0.39	2.59	QTL2_TGW	*vrs 3*	Pourkheirandish et al. 2007
SNP62	1H	66.70	0.25	0.0019	2.77	0.70				
SNP68	1H	72.43	0.09	0.0263	1.60	0.37				
SNP76	1H	87.62	0.21	0.0225	1.67	0.38	2.29	QTL3_TGW		
SNP78	1H	88.23	0.26	0.0019	2.79	0.69				
SNP81	1H	92.04	0.28	0.0208	1.70	0.38				
SNP137	2H	38.03	0.35	0.0259	1.61	0.40	-1.20	QTL4_TGW		
SNP171	2H	63.53	0.10	0.006	2.26	0.50	2.27	QTL5_TGW	QTgw.pil-2H.2	Pillen et al. 2003; Marquez-Cedillo et al. 2001
SNP174	2H	63.53	0.41	0.029	1.55	0.35				
SNP175	2H	63.53	0.10	0.006	2.26	0.50				
SNP210	2H	82.75	0.36	0.0081	2.13	0.51	1.33	QTL6_TGW	*vrs1*	Pourkheirandish et al. 2007
SNP215	2H	86.63	0.32	0.0267	1.59	0.38				
SNP245	2H	117.91	0.42	0.0084	2.11	0.51	-1.66	QTL7_TGW		
SNP262	2H	139.65	0.31	0.0091	2.07	0.49	-1.62	QTL8_TGW		
SNP305	3H	47.09	0.16	0.0225	1.67	0.39	3.01	QTL9_TGW		
SNP385	3H	98.49	0.37	0.0131	1.91	0.45	1.94	QTL10_TGW	*vrs4*	Pourkheirandish et al. 2007
SNP395	3H	111.42	0.37	0.0302	1.54	0.36	-1.35	QTL11_TGW	QTgw.S42-2H.a	von Korff et al. 2006
SNP458	4H	26.19	0.34	0.0224	1.67	0.40	1.75	QTL12_TGW	*int-c*	Pourkheirandish et al. 2007; Ramsay et al. 2011
SNP460	4H	26.66	0.26	0.0034	2.52	0.63				
SNP467	4H	40.36	0.33	0.0007	3.21	0.74	2.52	QTL13_TGW	QTgw.pil-4H.3	Pillen et al. 2003
SNP469	4H	40.36	0.17	0.0006	3.28	0.82				
SNP643	5H	110.26	0.10	0.0312	1.52	0.28	-3.00	QTL14_TGW	QTgw.pil-5H.2	Pillen et al. 2003
SNP663	5H	142.2	0.16	0.0004	3.45	0.87	4.47	QTL15_TGW	QGwe.TaER-5H.2	von Korff et al. 2008
SNP664	5H	142.2	0.16	0.0002	3.79	1.00				
SNP666	5H	142.2	0.15	0.0012	3.00	0.74				
SNP709	5H	187.38	0.28	0.0082	2.12	0.50	2.02	QTL16_TGW	QTgw.pil-5H.4	Pillen et al. 2003
SNP739	6H	43.83	0.08	0.016	1.82	0.43	-1.91	QTL17_TGW		
SNP740	6H	44.77	0.42	0.0041	2.43	0.6				
SNP741	6H	44.77	0.41	0.0064	2.23	0.55				
SNP770	6H	55.94	0.31	0.003	2.58	0.54				
SNP851	7H	34.82	0.43	0.0056	2.29	0.52	-1.88	QTL18_TGW	QGwe.HaTR-7H.1	Szücs et al. 2009
SNP854	7H	37.55	0.36	0.0277	1.58	0.32				
SNP919	7H	88.65	0.13	0.0164	1.81	0.43	3.01	QTL19_TGW		
SNP934	7H	129.91	0.24	0.0048	2.36	0.59	1.84	QTL20_TGW		
SNP944	7H	143.68	0.12	0.0315	1.52	0.27	-1.37	QTL21_TGW	QTw.HaTR-7H.1	Pillen et al. 2003

Table 2.3.10 GWAS results for trait starch content. Significant markers associated for starch content with K-model

SNP	Chr	Position	MAF	P-values	-log10(P)	r^2 (%)	Effect	QTL	Reference QTL	Literature
SNP20	1H	43.28	0.099	0.0076	2.12	0.3	-0.915	QTL1_SC		
SNP22	1H	47.47	0.340	0.0045	2.35	0.34				
SNP36	1H	51.23	0.396	0.0299	1.52	0.2	-0.78	QTL2_SC		
SNP47	1H	55.49	0.495	0.0190	1.72	0.22				
SNP53	1H	60.19	0.309	0.0105	1.98	0.28	1.34	QTL3_SC		
SNP56	1H	61.53	0.264	0.0059	2.23	0.32				
SNP66	1H	69.53	0.474	0.0148	1.83	0.25				
SNP92	1H	101.45	0.288	0.0009	3.04	0.43	-0.70	QTl4_SC		
SNP108	1H	126.01	0.108	0.0236	1.63	0.32	-0.76	QTL5_SC		
SNP136	2H	33.74	0.349	0.0093	2.03	0.28	-0.63	QTL6_SC	Qsch2a	Abdel-Haleeem et al. 2010
SNP174	2H	63.53	0.406	0.0142	1.85	0.25	-1.14	QTL7_SC		
SNP176	2H	63.53	0.184	0.0315	1.50	0.19				
SNP180	2H	64.24	0.225	0.0066	2.18	0.31				
SNP181	2H	64.24	0.209	0.0137	1.86	0.26				
SNP192	2H	71.12	0.474	0.0259	1.59	0.21	-0.68	QTl8_SC	QStr.StMo-2H	Grain genes
SNP222	2H	90.10	0.485	0.0277	1.56	0.22	-1.05	QTL9_SC	Qsch2a	Abdel-Haleeem et al. 2010
SNP311	3H	51.73	0.214	0.0227	1.64	0.22	-1.15	QTL10_SC		
SNP334	3H	55.57	0.373	0.0067	2.17	0.31				
SNP358	3H	72.26	0.358	0.0087	2.06	0.29	0.96	QTL11_SC		
SNP507	4H	65.05	0.491	0.0160	1.80	0.23	-0.55	QTL12_SC		
SNP539	4H	111.68	0.175	0.0048	2.32	0.33	1.18	QTL13_SC		
SNP543	4H	123.29	0.256	0.0039	2.41	0.35	1.10	QTL14_SC		
SNP599	5H	58.70	0.351	0.0272	1.57	0.21	0.67	QTL15_SC	QStr.StMo-5H	Grain genes
SNP612	5H	65.49	0.445	0.0135	1.87	0.26				
SNP643	5H	110.26	0.104	0.0080	2.10	0.27	-1.79	QTL16_SC		
SNP725	6H	22.35	0.469	0.0117	1.93	0.27	0.73	QTL17_SC		
SNP727	6H	24.36	0.433	0.0244	1.61	0.22				
SNP795	6H	71.08	0.392	0.0282	1.55	0.2	-0.57	QTL18_SC	QStr.StMo-6H	Grain genes
SNP823	6H	112.32	0.299	0.0252	1.60	0.21	-0.81	QTL19_SC		
SNP836	7H	0	0.199	0.0175	1.76	0.24	-0.74	QTL 20_SC		
SNP844	7H	12.42	0.096	0.0003	3.50	0.65	-1.72	QTL21_SC	*waxy*	Grain genes
SNP893	7H	78.22	0.127	0.0040	2.40	0.53	0.81	QTL22_SC		
SNP918	7H	87.97	0.297	0.0296	1.53	0.17	0.38	QTL23_SC	Qsch7a	Abdel-Haleeem et al. 2010
SNP930	7H	121.09	0.074	0.0083	2.08	0.24	-1.74	QTL24_SC		
SNP951	7H	149.03	0.24	0.0168	1.77	0.27	-0.76	QTL25_SC		

Table 2.3.11 GWAS results for trait crude protein content. Significant markers associated for crude protein content with K-model

SNP	Chr	Position	MAF	P-Value	-log10 (P)	R^2 (%)	Effect	QTL	Reference QTL	Literature
SNP47	1H	55.49	0.50	0.0044	2.36	0.78	0.74	QTl 1_CPC		
SNP97	1H	114.84	0.25	0.0139	1.86	0.56	-0.85	QTL 2_CPC		
SNP136	2H	33.74	0.35	0.0310	1.51	0.45	0.39	QTL 3_CPC	QPc.nab-2H.1;Qcp2a	Marquez-Cedillo et al. 2001; Abdel-Haleem et al. 2010
SNP170	2H	63.53	0.32	0.0190	1.72	0.54	0.72	QTL 4_CPC	QGpc.StMo-2H.2	Szücs et al. 2009; Marquez-Cedillo et al. 2001
SNP173	2H	63.53	0.33	0.0115	1.94	0.62				
SNP174	2H	63.53	0.41	0.0055	2.26	0.74				
SNP177	2H	63.53	0.32	0.0116	1.94	0.61				
SNP200	2H	74.37	0.25	0.0071	2.15	0.72	-0.56	QTL5_CPC		
SNP205	2H	78.03	0.40	0.0296	1.53	0.47				
SNP226	2H	96.82	0.23	0.0056	2.25	0.66	-0.90	QTL6_CPC	QPc.nab-2H.1	Marquez-Cedillo et al. 2001
SNP227	2H	96.82	0.19	0.0160	1.80	0.55				
SNP244	2H	116.49	0.24	0.0242	1.62	0.48	-0.47	QTL7_CPC	QGpc.HaMo-2H.2	Szücs et al. 2009
SNP272	2H	147.94	0.26	0.0103	1.99	0.62	-0.52	QTL8_CPC		
SNP305	3H	47.09	0.16	0.0020	2.70	0.86		QTL9_CPC		
SNP322	3H	55.57	0.18	0.0093	2.03	0.64	-1.46			
SNP357	3H	72.26	0.32	0.0159	1.80	0.51	-0.47	QTL10_CPC		
SNP401	3H	122.14	0.24	0.0062	2.21	0.68	0.60	QTL11_CPC	Qcp3a	Abdel-Haleem et al. 2010
SNP409	3H	130.82	0.41	0.0146	1.84	0.54				
SNP518	4H	79.58	0.45	0.0006	3.22	1.09	-0.75	QTL12_CPC	QGpc.HaTR-4H.2	Mather et al. 1997
SNP531	4H	97.06	0.11	0.0281	1.55	0.45	0.79	QTL13_CPC	QGpc.StMo-4H	Hayes et al. (1993) ;
SNP534	4H	101.62	0.16	0.0025	2.60	0.87				
SNP537	4H	108.70	0.21	0.0016	2.80	0.92				
SNP616	5H	74.78	0.51	0.0107	1.97	0.54	-0.81	QTL14_CPC	QGpc.HaMo-5H	Szücs et al. 2009; Marquez-Cedillo et al. 2001
SNP623	5H	85.93	0.16	0.0082	2.09	0.61				
SNP643	5H	110.26	0.10	0.0055	2.26	0.73	1.52	QTL15_CPC	QGpc.DiMo-5H.2	Oziel et al. (1996)
SNP699	5H	171.66	0.11	0.0219	1.66	0.53		QTL16_CPC		
SNP807	6H	83.89	0.25	0.0020	2.70	0.91	0.67	QTL17_CPC		
SNP844	7H	12.42	0.10	0.0214	1.67	0.51	0.79	QTL18_CPC		
SNP855	7H	38.32	0.35	0.0019	2.72	0.91	-0.86	QTL19_CPC		
SNP860	7H	46.19	0.26	0.0314	1.50	0.45				
SNP871	7H	61.32	0.24	0.0285	1.55	0.46	-0.50	QTL20_CPC	QPc.nab-7H	Marquez-Cedillo et al. 2001
SNP904	7H	80.94	0.22	0.0036	2.44	0.85	-0.68	QTL21_CPC	QGpc.HaTR-7H	Mather et al. 1997
SNP925	7H	112.46	0.37	0.0210	1.68	0.49	0.41	QTL22_CPC		
SNP930	7H	121.09	0.07	0.00001	4.73	1.57	1.54	QTL23_CPC		

2.4 Discussion

In the present study we describe the application of whole genome association mapping in a panel of diverse spring barley genotypes for agronomic traits. For each of the analyzed trait 19 to 25 QTL were detected. A substantial portion of the derived QTL locations are congruent with previously identified QTL in various biparental mapping populations (**Tables 2.3.7 to 2.3.11**). GWAS are strongly influenced by the quality of the phenotypic data (Rafalski, 2010). In the present study, heritabilities for all traits exceeded 0.9 and phenotypic means reflected a broad variation in the panel. The observed differences for two-rowed and six-rowed groups were expected due to their different breeding histories and the pleiotropic effects of spike morphology (**Supp Table 2.3**). Phenotypic variation observed for all traits is higher in the six-rowed group than in the two-rowed group, which is in accordance with the higher genetic diversity of this subgroup (**Table 2.3.4**). A more detailed analysis of population structure revealed six subgroups, which were mostly defined by spike morphology and geographical origin, both of which are known to impinge on the expression of agronomic traits.

2.4.1 Genetic diversity and population structure

Arguably an association mapping panel should suffice both phenotypic and molecular diversity for the outcome of reliable association results. Owing to the availability of a large number of mapped SNP markers that can be interrogated in a multiparallel manner, high marker coverage amounting to 1 marker per 1.18 cM was achieved. The average PIC (0.30) and Gene diversity (0.33) values observed in this panel of accessions are comparable with the results in previous studies using bi-allelic markers. PIC values differed among chromosomes and among different germplasm subgroups (**Tables 2.3.3 & 2.3.4**). Among all chromosomes, the highest average PIC value (0.33) was detected for chromosome 6H - which corresponds to the observations made by Rostoks et al. (Rostoks et al., 2006) in a set of European barley cultivars. The population structure in the panel was detected by implementing various approaches (STRUCTURE, PCoA and NJ-dendrogram) and found similar results. Several previous studies (Hamblin et al., 2010; Malysheva-Otto et al., 2006; Rostoks et al., 2006; Zhang et al., 2009) have shown that growth habit, spike morphology and geographical origin are the major factors that mirror population structure in barley. Since the present study has been restricted to spring barley, spike morphology and geographical origin were the fundamental determinants for population substructuring (G1 to G6) (**Fig. 2.3.3**). The 55 landrace accessions included in this panel were distributed among all groups. The subgroups G1, G2 and G3 are mainly six-rowed barleys and the subgroups G4, G5 and G6 include mainly two-rowed barleys. Two-rowed barleys in the panel are more closely related to each other and less diverse than the six-rowed barleys, which is in contrast to the findings of Zhang et al. (Zhang et al., 2009) for Canadian germplasm. While in the panel two-rowed barleys even outnumbered the six-rowed accessions, the reason for their limited diversity might be that the majority originated from Europe. The geographical distribution of the accessions has a major influence on the diversity of alleles sampled in the population. In Europe, two-rowed barley is mainly grown as raw material for malt production. Malting quality is a quantitative trait. The use of a limited number of principal progenitors in the corresponding breeding programs has resulted in the reduction of genetic diversity and in the concomitant formation of a distinct subpopulation as it is seen in our present panel (Melchinger et al., 1994).

2.4.2 LD configuration and consequences

The resolution of LD mapping depends on the extent of LD across the genome and the rate of LD decay with genetic distance (Remington et al., 2001; Stracke et al., 2007). Genome-wide LD studies for barley have been previously reported in various populations using different molecular markers such as AFLP, SSR and DArT (Kraakman et al., 2004; Malysheva-Otto et al., 2006; Mather et al., 1997; Zhang et al., 2009), with few studies, however, relying on more than 1000 markers. In panel of spring barley accessions of worldwide origin, intra-chromosomal whole genome LD decays below the critical r^2-value (0.2) within a genetic distance of 5 cM. It needs to be kept in mind that this is an average value, which summarizes substantial intra-chromosomal LD variation. The extent of intra-chromosomal LD for different chromosomes in the current panel ranges from 5-10 cM with varying patterns along each chromosome (**Supp Fig. 2.5**). Previous studies found various levels of LD decay in different barley populations (Caldwell et al., 2006; Stracke et al., 2007; Waugh et al., 2009) and among different chromosomes (Rostoks et al., 2006). The LD decay was more rapid in the study of Comadran et al. (Comadran et al., 2009) probably due to the inclusion of landraces in the collection. Caldwell et al. (Caldwell et al., 2006) also showed that LD decays more rapidly in barley landraces compared to elite barley cultivars. Less extensive LD beyond 10 cM has been found in our panel, as the majority of significant LD values above the basal level (33.7%) are due to tightly linked markers. Significant inter-locus LD values of unlinked markers (4%) may be the result of population structure (**Table 2.3.6**). Some closely linked markers were found to be in complete Linkage Equilibrium (LE), while some distantly linked markers exhibited high LD values. This reflects the dynamic variation of LD patterns along the chromosomes as it has been shown in this panel at the sequence level for several transcription factors (Haseneyer et al., 2010a). As to the individual subgroups, the portion of significant r^2-values above the basal level (0.2) is higher within six-rowed than in two-rowed groups indicating high LD in these groups. Interestingly, LD in all subgroups extended beyond 30 cM except for G5 where LD extended to about 20-25 cM (**Fig. 2.3.5**). This is most likely because of the larger population size of G5 compared to the other subgroups. The extensive LD observed in the subgroups is probably due to their decreased population size and a concomitant increase in relatedness.

2.4.3 Genome-wide association mapping

Despite the advantages of GWAS to pinpoint genetic polymorphisms underlying agronomic traits, this approach may suffer from an inflation of false positives due to population structure (Kang et al., 2008; Lander and Schork, 1994; Zhang et al., 2010). Several statistical models to correct for the effect of population structure have been proposed and tested in previous studies (Kang et al., 2008;

Price et al., 2006; Stich and Melchinger, 2009). Since a considerable amount of structure was detected in the current panel, linear models were used to control for population structure and to reduce the false positive associations. Similar to the previous studies of comparing GWAS models in allogamous and autogamous species (Kang et al., 2008; Stich et al., 2008), our results suggest that K, QK and PK-models performed better than others (**Fig. 2.3.7**). Moreover, for the K-model computational time is faster and no additional steps like identifying appropriate population structure (Q-matrix) in the panel are required. Since in an exploratory analysis mostly consistent results were obtained for all three approaches, the K-model was employed in the complete analysis of all traits to avoid redundancy of data. Still it should be kept in mind that correcting for population structure not only reduces the frequency of false positives but also may entail false negatives in situations where a character state is strongly correlated with population structure (Cockram et al., 2008).

In order to confirm the efficiency and resolution of the panel for association mapping using the range of available markers, all 918 SNPs were re-mapped using the K-model. From 918 SNPs, 783 were re-mapped within 10 cM of their original positions. Only 14% of the markers mapped beyond 10 cM. Among the successfully re-mapped markers more than 95% markers are within 5 cM distance from the original map position indicating the mapping resolution of the panel (**Fig. 2.3.6**). Rostoks et al. (Rostoks et al., 2006) has used the same approach to evaluate their barley collection for GWAS with a subset of markers and successfully mapped 80% of the markers.

To demonstrate the suitability of the panel and the model for GWAS, we first analyzed spike morphology (row type) (**Fig. 2.3.8**). This trait can be easily scored and is important from the agronomic and the domestication point of view. The genetic basis of row type is already well known and several loci have been mapped and genes have been cloned (Pourkheirandish & Komatsuda 2007). For this trait 34 marker-trait associations were detected (Fig. 2.3.8). The identified marker-trait associations for row type are concurrent with all previously identified major loci - *vrs1* (Komatsuda et al., 2007), *vrs2, vrs3, vrs4* and *int-c* (Ramsay et al., 2011; Waugh et al., 2009). Additional, less significant associations detected for row type could not be associated to any known major locus, and need to be further explored. These results for row type act as a proof of concept for GWAS in current spring barley panel and reflect the efficiency of GWAS for high resolution QTL mapping in inbreeding species. Some of the row type QTL overlapped with associated regions for other traits, especially with the traits TGW, SC and CPC (**Supp**. **Fig. 2.7**). As expected, two-rowed barley has higher TGW than the six-rowed types, as the number of sink organs (kernels) in two-rowed spikes is smaller than in six-rowed spikes. While the effect of spike architecture on TGW is clearly pleiotropic, its influence on SC and CPC is the result of breeding history and end use quality. In case of malting barley, varieties are generally bred for high starch and low protein content. In Europe two-rowed barley is preferred for malt production while six-rowed barley is primarily used as feed and is characterized by high protein content (Hayes and Szucs, 2006). As a result, the two-rowed types in our panel have higher starch content and lower protein content than six-rowed types (**Supp Table 2.3**).

Heading date (HD) reflects the adaptation of a plant to its environment and is a complex trait affected by numerous QTL both in outbreeding (Buckler et al., 2009) and in inbreeding species (Wang et al., 2010a). Many SNP markers were found to be associated with the trait HD (**Fig. 2.3.9a**) and we report a total of 34 significant SNPs defining 19 QTL. Some of these QTL hit genomic regions that were previously reported to harbor major genes including *HvFT3, PpdH1, HvFT4*, *eps2*, *HvGI*, *HvCO3*, *HvFT1* and *HvCO1* (**Table 2.3.7**). In a previous study using the same panel, fragments from three flowering time candidate genes were re-sequenced and SNPs within the gene *PpdH1* revealed the largest effects on HD (Stracke et al., 2009). In the present GWAS, SNPs located in the vicinity (ca. 2 cM) of *PpdH1* showed significant associations with HD (**Table 2.3.7**).

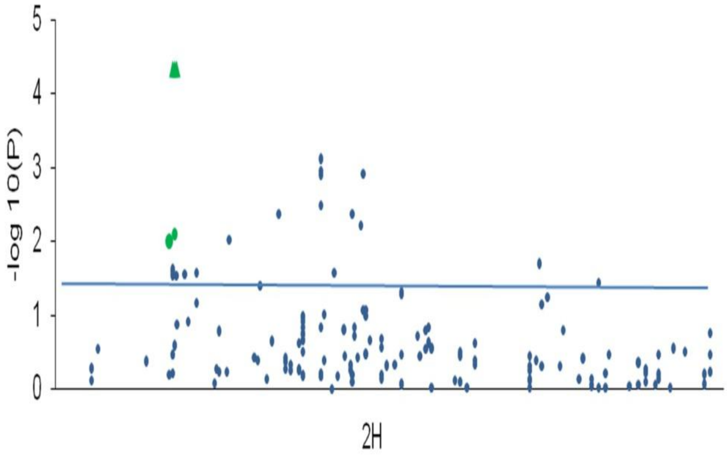


Fig. 2.4.1 Association analysis for the trait HD for chromosome 2H with SNPs from IPK-OPA and the re-sequenced *PpdH1* fragment. Blue circles represent the IPK-OPA SNPs on chromosome 2H, Green circles represent the IPK-OPA SNPs that are significantly associated with HD and are in vicinity of the *PpdH1* gene, green triangles represent SNPs from the re-sequenced *PpdH1* fragment

By further including all *PpdH1* SNPs from Stracke et al. (Stracke et al., 2009) into our GWAS, these SNPs revealed the highest association of all markers used (**Fig. 2.4.1**). These findings lend strength to the hypothesis that a further increase in marker coverage will either lead to the detection of additional associations or improve the significance of existing QTLs.

For the trait PHT we found 19 putative QTL regions located on chromosomes 2H, 3H, 4H, 5H, 6H and 7H comprising 32 marker trait associations. Semi-dwarf and dwarf cultivars have been developed worldwide to reduce lodging and to improve the harvest index. Different genes/alleles have been deployed in different geographic regions: the GA sensitive *sdw1* dwarfing gene has been deployed in America and Australia, while its allelic form, termed *denso*, is frequently seen in European two-rowed germplasm. The recessive *uzu* allele is found in Japanese, Chinese and Korean cultivars (Jia et al., 2011; Zhang et al., 2007). Many QTL for PHT coincide with previously mapped QTL and genes (**Table 2.3.8**). The QTL4_PHT on chromosome 2H coincides with the mapping position of *sdw3* which plays a major role in gibberellins-insensitive dwarfing barley (Gottwald et al., 2004). The dwarfing gene *denso/sdw1* maps to the same genomic region as QTL8_PHT located on the long arm of chromosome 3H (Jia et al., 2011). The QTL7_PHT is about 10 cM distant from the *uzu* locus based on the consensus map presented in grain genes.

Thousand grain weight (TGW) is one of the major yield components having direct effect on the final yield. Altogether 21 QTL were found for this trait and some of them are in vicinity of row type genes. Some of the QTL were further confirmed by previously mapped QTL in same genomic regions (**Table 2.3.9**). QTL14_TGW on 5HL is observed to effect other traits like PHT, SC and CPC. As outlined above, starch and protein content of the grain are major determinants of the end use quality. Several of the 25 QTL detected for starch content coincided with the previously identified QTL (**Table 2.3.10**). These include QTL for related traits like acid detergent fiber (ADF) content, starch granule size and granule shape (Abdel-Haleem et al., 2010). QTL21_SC on 7H is located in the region of the *waxy* locus known to encode *granule-bound starch synthase I* (*GBSS I*), which catalyses the synthesis of amylose (Kleinhofs, 1997; Rohde et al., 1988). For the grain crude protein content 23 QTL were identified, located on all seven chromosomes. Eleven of these QTL regions co-localize with previously mapped QTL, while 12 QTL are novel (**Table 2.3.11**). Interestingly, the majority of QTL for traits SC and CPC are located on chromosome 7H. Some of the QTL identified for SC coincide with QTL for CPC e.g. chromosomes 1H (55 cM), 2H (33.74 cM), 3H (55 cM), 5H (110 cM) and 7H (12 cM and 121 cM) (**Table 2.3.10 & 2.3.11**). The coincidence of the QTL for these two traits can be expected due to their negative correlation (**Table 2.3.2**). On the other hand, we cannot rule out that some of the shared QTL are the result of linkage of underlying genes.

2.4.4 GWAS reveals small effects only

Even the best associations observed in the present study showed only modest R^2 values (percentage of genetic trait variation explained) for the corresponding SNPs, implying low variance predicted by each SNP. This is exemplified by the QTL ‘Qsch7a’, which in a biparental QTL mapping study explained 47% of variation for SC (Abdel-Haleem et al., 2010). In the present study, ‘QTL23_SC’ located at the same genomic region as ‘Qsch7a’ explains only 0.2% of the variation. Many GWAS in humans have reported low R^2 values and the rest of the unexplained variation is termed as ‘unexplained missing heritability’ (Manolio et al., 2009). Roy et al. (Roy et al., 2010), among others, reported R^2-values to range from 0.2% to 3.95% in GWAS for plants, which corresponds well with our present results. In a consorted study for the trait "body height", an impressive number of 40 genotypic variants have been identified under a stringent threshold. Together these were only able to explain around 5% of the variation in humans body height (Maher, 2008; Visscher, 2008). Possible explanations for this "missing heritability" include i) insufficient marker coverage, in cases where the causal polymorphism is not in perfect LD with the genotyped SNP reduces the power to detect associations and the variation explained by such a SNP marker. This has been demonstrated in the present study for the effect of the *PpdH1* gene on HD; ii) rare alleles (MAF < 5%) with a major effect have been dropped from the analysis and will go undetected in cases where they are associated; iii) the expression of a character or trait depends on a large number of genes/QTL with small individual effects which escape statistical detection; iv) inadequacy of the statistical approaches available to detect epistatic interactions in GWAS and v) biased estimates of R^2 for individual SNPs due to the level of population stratification in the panel (Frazer et al., 2009; Gibson, 2010; Hall et al., 2010; Maher, 2008; Manolio et al., 2009). Although the above mentioned reasons were mainly discussed in the context of GWAS in humans, they also pertain to GWAS in plants and other organisms. In addition to the above mentioned reasons, the statistical model employed for the analysis will affect the variation explained by the SNPs. As the stringency and threshold of the models increases, the power of detecting small effect SNPs will be reduced. We observed that in the case of using stringent models for GWAS the larger portion of the trait variation is explained by the model itself and less variation is left to be explained by genetic effects. For the trait HD the K-model, explained nearly 70% of the variation of the trait. Reducing the stringency of the model would increase the variation explained by the marker, but at the same time would result in more false positives. Especially in inbreeding crops like barley, it is difficult to preclude completely the effect of relationship among genotypes by applying simpler models. Hence, GWAS in highly structured populations of inbreeding crops such as barley will depend on the careful optimization of the model regarding sensitivity vs. selectivity.

2.5 Conclusions

Overall, these results provide new details on the chances and pitfalls of GWAS in structured populations of inbreeding crops like barley. Results from the present study provide an insight into the genetic architecture of important agronomic traits for barley (HD, PHT, TGW, SC and CPC). In total, 107 QTL were identified for these traits. Some genomic regions harbor QTL for more than one trait and, based on map comparisons, 50 QTL have been found to concur with previously mapped QTL. For all traits together, 57 novel QTL have been detected. To mitigate the shortcomings of GWAS in inbreeding crops, future association studies might implement novel strategies such as joint linkage and LD mapping which were already successfully applied in various species (Blott et al., 2003; Brachi et al., 2010; Buckler et al., 2009; Mott and Flint, 2002). Furthermore, to fine map and “mendelize” selected QTL, staggered patterns of LD decay observed for different genepools of barley (cultivars, landraces, wild barley) may be exploited in combination with biparental mapping and marker saturation strategies exploiting the ever increasing body of genomic sequence information (Mayer et al., 2011; Waugh et al., 2009). The feasibility of such an approach was recently demonstrated by identifying a candidate gene for the *ANTHOCYANINLESS 2* locus using a combination of association mapping followed by a segregation analysis in a biparental population and a BAC contig analysis (Cockram et al., 2010).

CHAPTER THREE: Effects of marker density on QTL detection by genome wide association studies in a worldwide spring barley collection

3.1 Introduction

In this study the results and effects of using different marker densities for QTL detection by genome wide association studies (GWAS) were compared. It is hypothesized that if the LD decay is assumed to be within 100 kb, then the optimum SNP requirement for uniform coverage of the complete barley genome will be around 57,000 SNPs (Semagn et al., 2010). Such high marker densities are not yet available for barley and genotyping by sequencing methods are in a nascent stage for barley (Elshire et al., 2011). Until recently, BOPA (Barley Oligonucleotide Pool Assay) SNPs and DArT (Diversity Array Technology) markers were extensively used for GWAS in barley (Comadran et al., 2009; Comadran et al., 2011; Massman et al., 2011). Apparently, in a large genome like barley this marker coverage is inadequate for capturing all the loci in GWAS and also to directly identify the genes underlying detected QTL. With 918 SNPs in our association panel an approximate coverage of 1 SNP per 1.18 cM of genetic distance was achieved. This coverage is not uniform with several large gaps without markers and also in certain regions of barley genome 1 cM genetic distance can correspond to a large physical distance. Hence large marker coverage is required for efficiently covering all the loci. Based on the efforts of an international consortium an assay consisting of 7864 SNPs has been developed recently using the iSelect technology from Illumina (Comadran et al. unpublished). The spring barley collection (224) was genotyped using this iSelect assay. A mixed linear model (MLM) with kinship matrix (K) as a random effect, that accounts for the relatedness or kinship between the individuals was used to perform association analysis (Yu et al., 2006). Several studies reported the accuracy of randomly selected background markers for estimating population structure (Pritchard et al., 2000b; Zhu et al., 2008). The accurate estimate of K from molecular markers would result in a better fit of the model in explaining phenotypic variation with genetic relatedness. Deciding the optimum set of markers required for accurate estimation of K is still under discussion. The effect of background marker densities on robustness of the estimated kinship matrix and its effects on the outcome of GWAS in plant populations were not investigated previously. Thus we evaluated and compared the results of GWAS using three different kinship matrices generated using different marker sets. Moreover, the effect of marker density on GWAS was investigated by comparing the results obtained with the OPA assay based on 918 informative SNP markers (Pasam et al. 2012, Chapter 2 of this thesis) with those form the iSelect assay, relying on 6467 informative SNPs.

3.2 Materials and Methods

3.2.1 Association panel

The panel is described in chapter 2 in section **2.2.1,** and also described in detail by Haseneyer et al. (2010). The association mapping panel consists of 224 spring barley accessions selected from the Barley Core Collection (BCC) (Knüpffer and van Hintum, 2003) and the barley Genebank collection maintained at the IPK Genebank Gatersleben, Germany. However, only 212 accessions were considered for GWAS and comparison of matrices. The remaining 12 accessions that performed badly when assayed with IPK-OPA were excluded from all analyses.

3.2.2 Phenotypic evaluation

Previous phenotypic data have been used to estimate the marker trait associations between the markers and the traits row type (RT), heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and crude protein content (CPC). The phenotypic data is discussed in chapter 2 in section **2.2.2.**

3.2.3 Genotyping

The panel was genotyped using the iSelect SNP assay from Illumina. The assay is a designed bead chip that permits multiplexing of thousands of SNP markers. In barley, this chip has been mainly developed from the RNAseq data of the 10 diverse barley cultivars including high quality SNPs from previous BOPA arrays (Close et al., 2009) (Comadran et al. unpublished). Most of the SNPs from IPK-OPA are also included in this iSelect assay. Genetic map positions were obtained primarily from the genetic map of Morex x Barke RILs (Comadran et al. unpublished) and for the markers that were not mapped in the Morex x Barke RIL map, LD mapping was used. The map positions of BOPA1 and BOPA2 markers from the consensus genetic map (Close et al., 2009) were used for reference comparisons of LD mapped SNPs. The barley panel was genotyped by a service provider (TraitGenetics GmbH Gatersleben, Germany). After filtering and evaluating the SNPs, 7864 successfully called SNPs were used on the assay. In current panel, 864 SNPs that did not have any single data point across all the accessions were considered unsuccessful and excluded. The iSelect genotyping in the present collection resulted in 7000 successful SNPs with an average of 98% data points per each SNP in the panel (Supp Table 3.1).

3.2.4 Association analysis

The basic analysis including the implementation of association analysis is described in section 2.2.6. MLM with kinship matrix (K-Model) was used for accounting population structure and relatedness to avoid spurious association in GWAS. The analysis was performed using TASSEL v.2.1 (Bradbury et al., 2007) (www.maizegenetics.net). As the K-model was found to be performing better than other models in current panel (Chapter 2, Pasam et al. 2012), only the K-model was used for association analysis with iSelect SNPs. The K-model was tested with three different kinship matrices fitted as random effect and the results were compared. Three kinship matrices were generated using TASSEL i) kinship matrix using 6467 iSelect SNP markers that were successful in this collection (K_1); ii) kinship matrix with 918 IPK-OPA markers (chapter 2, K_2) and iii) kinship matrix generated by 362 selected high PIC value markers that are evenly distributed across the genome (K_3). The results from GWAS for row type and heading date were compared with the kinship matrix from three marker sets. For other traits results are discussed only for marker set K_3.

Several SNPs (240 to 304 SNPs) were associated with each of the traits at significance threshold of 0.03 [-log10 (*P*) =1.5]. The threshold 0.03 was used as a midway approach for stringent FDR (False Discovery Rate) correction and the more liberal approach proposed by Chan et al. (Chan et al., 2010). According to the liberal approach, the bottom 0.1 percentile distribution of P-values is considered significant. The lower 0.1 percentile distribution of P-values using the K_1 model for all traits ranged from 0.05 to 0.07, which is too permissive. Hence P-value =0.03 was used as a standard significance threshold, which was used in the previous analysis with IPK_OPA markers (Chapter 2, Pasam et al. 2012). Additionally, a conservative approach was used to compute corrections for multiple testing using the *q*-false discovery rate (*q*FDR) method (Storey, 2002). The *q*FDR method is an alternative to the FDR method proposed by Benjamini and Hochberg (Benjamini and Hochberg, 1995). All P-values from each trait were imported into the QVALUE R-package (Storey and Tibshirani, 2003) and analyzed using settings of FDR at 0.05 to obtain the *q*-values. SNPs with *q*-value < 0.05 were considered significantly associated above the threshold level. However, this approach is considered too conservative (Storey et al., 2004) and only few SNPs from the analysis crossed the significance threshold. Hence, we consider the earlier mentioned approach and report all SNPs below the significant P-value of 0.03 for further evaluation and discussion.

3.2 Results

Only 212 accessions were used for the analysis in order to compare the results with previous findings from GWAS using IPK-OPA (Chapter 2, Pasam et al. 2012). Out of 7864 SNP, 11% of SNPs did not work in our panel due to various reasons. Out of the successful 7000 SNPs, 533 SNPs had MAF below 0.05 and hence were excluded from the final analysis. The distribution of MAF for the SNPs in current panel is shown in **Fig. 3.3.1**. The Polymorphism Information Content (PIC) values for these SNPs ranged from 0 to 0.375. SNPs with PIC value zero (207 SNPs) are apparently monomorphic SNPs. Most of the SNPs amounting up to 60% of the total were in the PIC range of 0.3 to 0.375, indicating that the iSelect assay SNPs are very informative in current panel (**Fig. 3.3.2**). Finally, 6467 SNPs were used for GWAS. We present results from 5474 SNPs that were genetically mapped. The major portion of SNPs were mapped using Morex x Barke RIL population (3872 SNPs), 1465 SNPs were mapped by LD mapping using barley association mapping panel (Comadran et al unpublished) and 137 SNPs from BOPA were assigned map positions based on consensus map information of Close et al. (2009). Unmapped SNPs were excluded from further analysis.

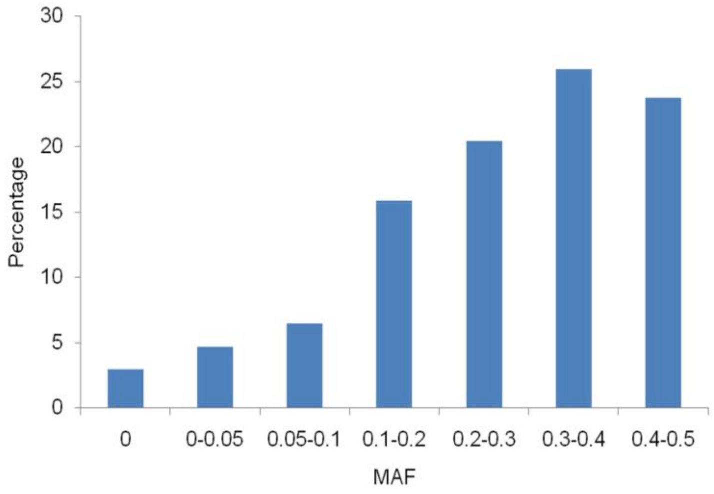


Fig 3.3.1 Minor allele frequency (MAF) distribution of 7000 SNP markers in the spring barley panel of 212 accessions. The range of MAF is indicated on the X-axis, and the percentage of markers falling in that range is represented on Y-axis

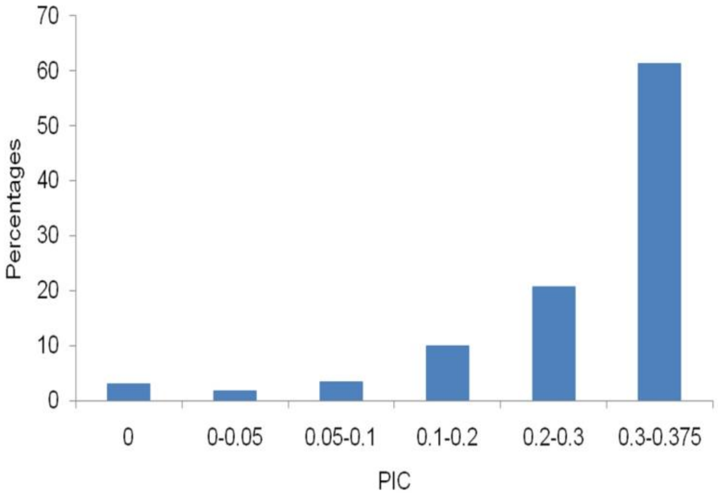


Fig 3.3.2 Distribution of Polymorphic information content (PIC) values of the 7000 SNP markers. The range is shown on X-axis, and the percentages of markers falling in that range are represented on Y-axis

3.3.1 Comparison of different kinship matrices

The heat plots of kinship coefficient matrices among 212 genotypes generated using different marker sets are presented in **Supp Fig 3.1.** Heat plots of the three matrices showed similar patterns and Mantel test correlations between the matrices were highly significant (**Table 3.3.1**). The correlation between K_1 and K_3 is 0.955, where as the correlation between K_2 and K_3 is 0.855. The percentage of SNPs falling into each kinship category slightly differed for three marker sets

(**Fig.3.3.3**). The impact of different kinships for GWAS varied for different traits. Use of different kinship matrices for GWAS resulted in change of significance of the marker trait associations and also altered the genetic variation explained by the marker. Nevertheless, for marker set K_3, the phenotypic variance (R^2) explained by the model was lowest for all traits (**Table 3.3.2**). The explained variance is a measure that indicates, how well the model explained the genetic variance of the phenotype. An increase in R^2 indicates that the kinship matrix captures more of the variance, leaving less for the SNP to explain.

Table 3.3.1 Mantel correlations between three different kinship matrices. The kinship matrices are generated from 7000 iSelect markers (K_1), 918 SNPs from IPK-OPA (K_2), and uniformly distributed 362 SNPs (K_3). All correlations are highly significant

	K1 (all SNPs)	**K2 (918 SNPs)**	**K3 (362 SNPs)**
K1 (all SNPs)	1	0.001	0.001
K2 (918 SNPs)	0.901	1	0.001
K3 (362 SNPs)	0.955	0.885	1

Table 3.3.2 Average R^2 explained by MLM with kinship for each trait using three different kinships. The higher the R^2 value, higher is the phenotype variation explained by the model

	K_1 (all SNPs)	**K_2 (918 SNPs)**	**K_3 (362 SNPs)**
HD	0.82	0.809	0.669
PHT	0.83	0.82	0.713
TGW	0.808	0.817	0.685
SC	0.91	0.91	0.801
CPC	0.879	0.768	0.685

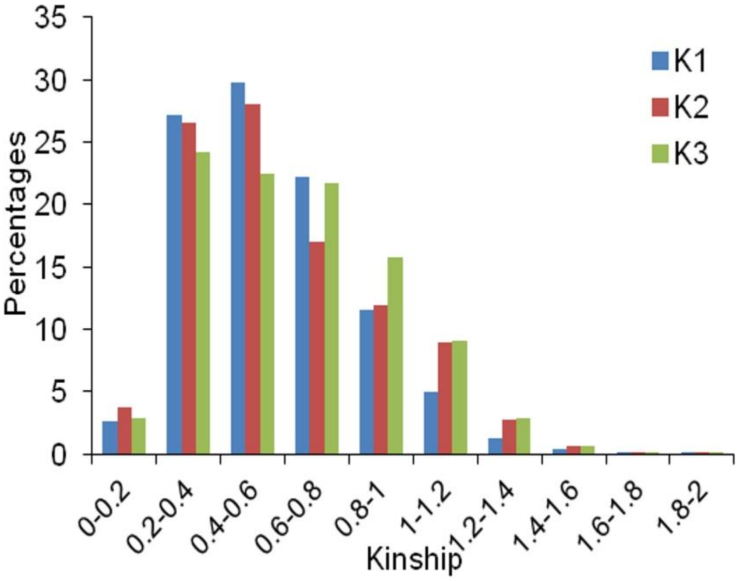


Fig 3.3.4 Distribution of pair-wise kinship estimates for 212 barley genotypes for different marker sets: K_1 with all iSelect markers; K_2 with 918 OPA markers; K_3 with 362 selected iSelect markers. The kinship estimates are grouped into different ranges and the percentage of estimates in each range are shown on the y-axis

3.3.2 GWAS scans and comparison with previous results

The results of GWAS for the traits row type, heading date, plant height, thousand grain weight, starch content and crude protein content are given in **Fig. 3.3.4 - 3.3.9**. The graphs comparing GWAS scans with IPK-OPA markers and with iSelect markers using different kinships (K_1, K_2, K_3) are presented for the traits row type and HD. GWAS scans for other traits are provided in supplementary figures **(Sup Fig 3.2-3.5**). For the traits PHT, TGW, SC and CPC the graphs from iSelect using the K_3 model are presented under **Fig 3.3.6, 3.3.7, 3.3.8 and 3.3.9**. The model using the K_3 matrix appears to be appropriate for all traits based on two grounds: i) the K_3 matrix captures the same amount of structure as the one using 6467 SNPs (K_1), which is evident from the Mantel test and the similarity of matrix heat plots (**Table 3.3.2**) and ii) MLM when used with K_3 explained less heritability than other K-models (**Supp Table 3.1**), leaving more heritability to be explained by the SNPs, in turn resulting in lower p-values and higher SNP effects. The explained variance by the model was lowest for the K_3 model for all traits studied (**Table 3.3.2**). Notwithstanding this, great overlap of significant SNPs with the other kinship models was observed. Hence only the results for the K_3 model are presented and discussed further.

It is not surprising that most of the QTL detected in GWA scans with OPA markers (Chapter 2) were also detected with GWA scans using iSelect markers. However, the QTL positions from the two studies cannot be matched exactly, as the consensus maps used for mapping the SNPs are different in both cases. Using common SNPs present in both assays a qualitative comparison of QTL positions is possible though. GWAS with iSelect resulted in multi-fold higher significant SNP number than the SNPs detected with IPK-OPA analysis for all the traits. More SNPs were found to be associating with a given trait and at some loci the peak markers displayed much higher significances.

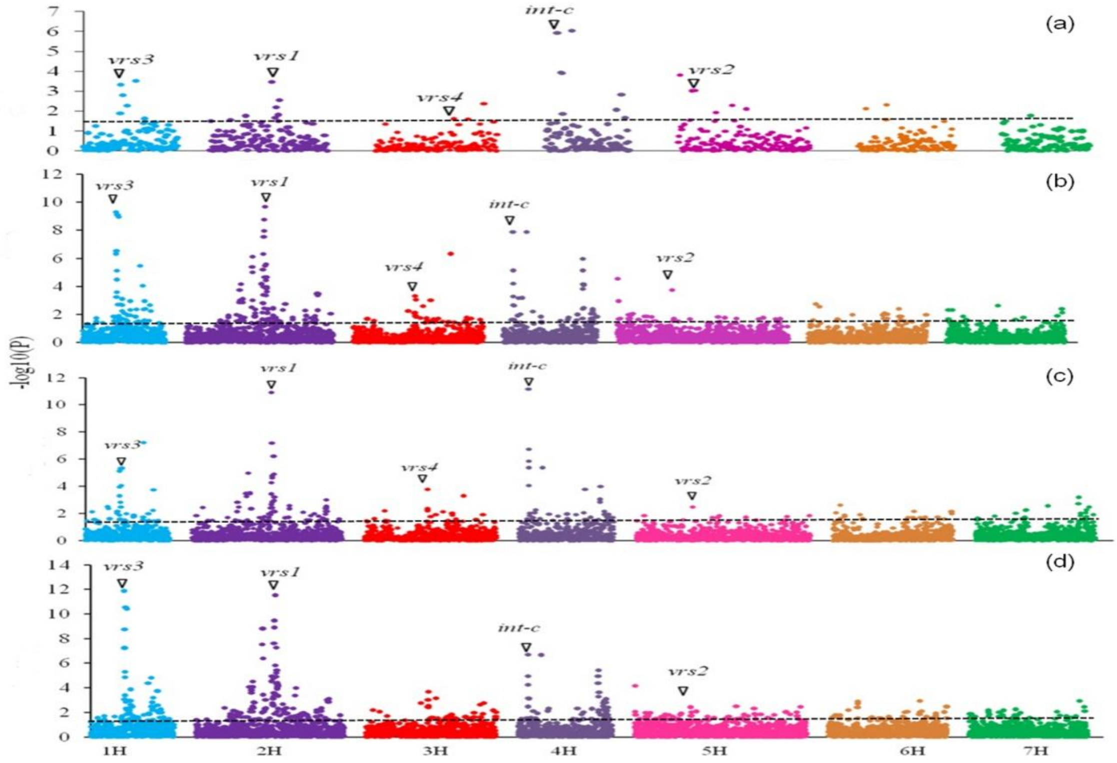


Fig 3.3.4 GWAS scans for row type using the Kinship (K) model. The Y-axis represents – log10 (*P*) values for maker trait associations. The X-axis corresponds to map positions. Individual barley chromosomes are denoted 1H through 7H. (a) GWAS with 918 SNPs (IPK-OPA) (b) GWAS with 5474 SNPs (iSelect) using kinship based on all SNPs that were employed in association analysis (K_1) (c) GWAS with 5474 SNPs using kinship from 918 SNPs (K_2) and (d) GWAS with 5574 SNPs using kinship from 362 SNPs (K_3)

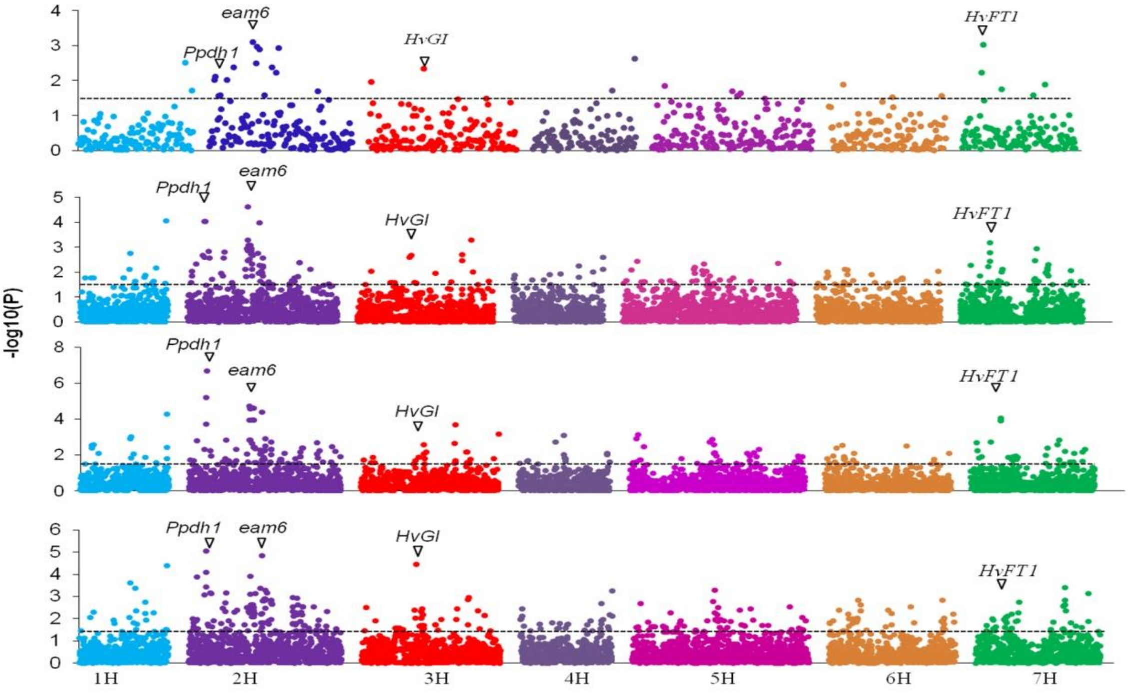


Fig 3.3.5 GWAS scans for heading date (HD) using the K-model. (a) GWAS with 918 SNPs (IPK-OPA) (b) GWAS with 5474 SNPs (iSelect) using K_1 (c) GWAS with 5474 SNPs using K_2 and (d) GWAS with 5574 SNPs using K_3

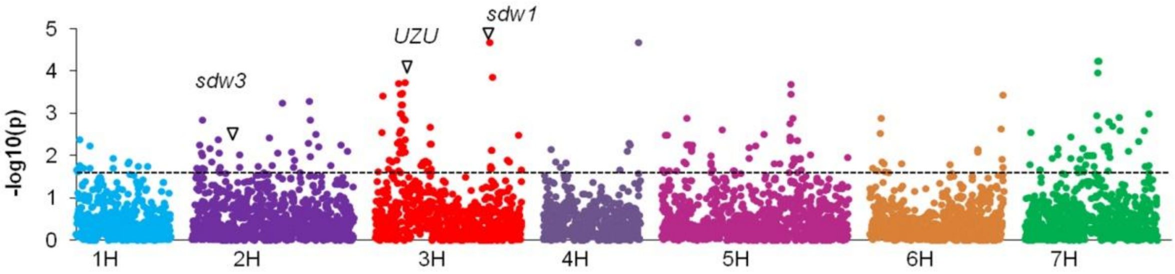


Fig 3.3.6 GWAS scans for plant height (PHT) with K-model using Kinship generated from equally spaced 362 SNPs (K_3). The Y-axis represents –log10 (P) values and the X-axis corresponds to the map position of the marker

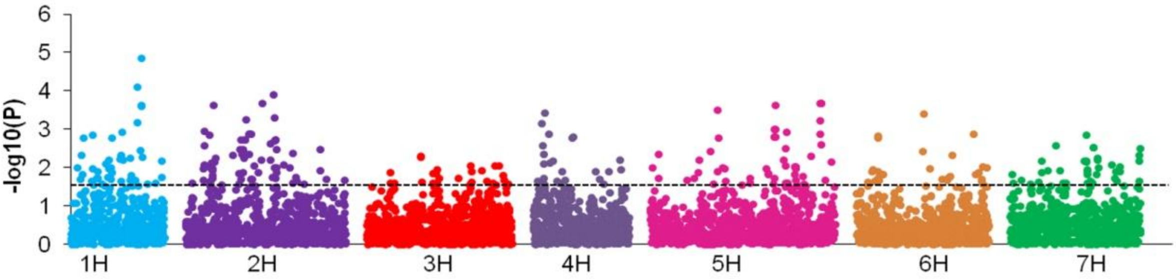


Fig 3.3.7 GWAS scans for thousand grain weight (TGW) with K-model using Kinship generated from equally spaced 362 SNPs (K_3)

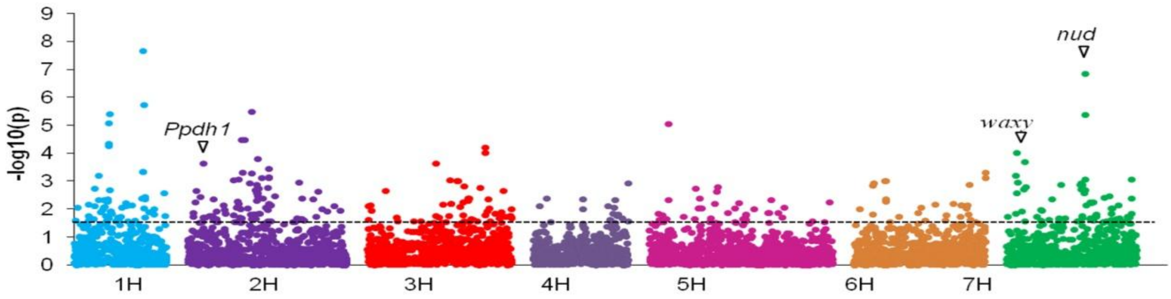


Fig 3.3.8 GWAS scans for starch content (SC) with K-model using Kinship generated from equally spaced 362 SNPs (K_3).

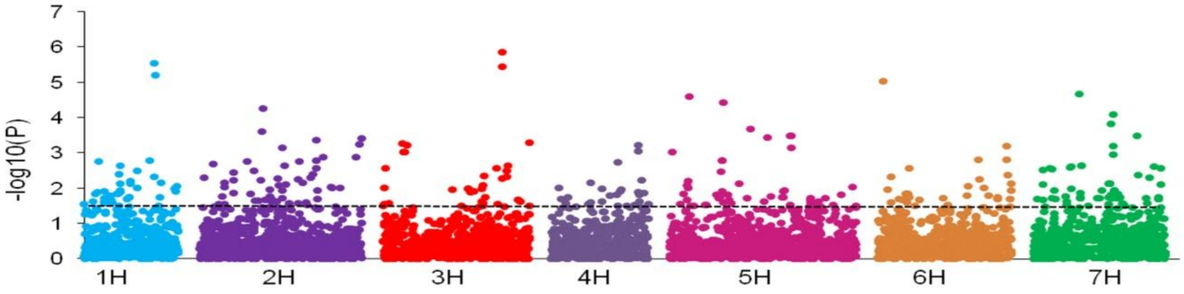


Fig 3.3.9 GWAS scans for crude protein content (CPC) with K-model using Kinship generated from equally spaced 362 SNPs (K_3).

Heading date			
Chromosome	SNPs[†]	QTL	Q-value‡
1H	15	7	1
2H	98	6	2
3H	30	6	1
4H	23	5	0
5H	45	7	0
6H	30	6	0
7H	28	6	0
Total	269	43	4
unmapped	70	-	1

Table 3.2.2 GWAS results for heading date (HD). Number of significant markers (P-value < 0.03) and number of probable QTL regions are given for each chromosome. The number of associations crossing the stringent *q*FDR threshold is given in column Q-value

Plant height (PHT)			
Chromosome	SNPs[†]	QTL	Q-value‡
1H	15	5	0
2H	36	7	0
3H	56	6	1
4H	9	5	1
5H	49	8	0
6H	15	6	0
7H	60	7	3
Total	240	44	5
unmapped	57	-	2

Table 3.2.3 GWAS results for plant height (PHT). Number of significant markers and number of probable QTL regions are given for each chromosome

Thousand grain weight			
Chromosome	SNPs[†]	QTL	Q-value‡
1H	40	5	0
2H	71	7	0
3H	20	6	0
4H	22	6	0
5H	47	7	0
6H	30	6	0
7H	36	8	0
Total	266	45	0
unmapped	66	-	0

Table 3.2.4 GWAS results for thousand grain weight (TGW). Number of significant markers and number of probable QTL regions are given for each chromosome

Starch content			
Chromosome	SNPs[†]	QTL	Q-value‡
1H	51	7	6
2H	82	7	5
3H	44	9	3
4H	13	6	0
5H	33	7	1
6H	25	5	0
7H	56	9	4
Total	304	50	19
unmapped	77	-	12

Table 3.2.4 GWAS results for starch content (SC). Number of significant markers and number of probable QTL regions are given for each chromosome

Crude protein content			
Chromosome	SNPs[†]	QTL	Q-value‡
1H	35	6	2
2H	48	9	2
3H	26	8	2
4H	19	5	0
5H	39	9	3
6H	29	8	1
7H	49	8	3
Total	245	53	13
unmapped	82	-	15

Table 3.2.4 GWAS results for crude protein content (CPC). Number of significant markers and number of probable QTL regions are given for each chromosome

[†] SNPs with significance < 0.03 (-log10(p) >1.5) ‡ Number of markers with significant qFDR

The large number of SNPs associating with the traits is expected because of their complex nature as most analyzed traits are controlled by numerous loci, and also due to the presence of more SNPs on the assay that are linked to the causal gene. The *P*-value distributions for different traits are provided in S**upp Fig 3.6**. The significant marker trait associations (*P*-value < 0.03) pertaining to each trait were grouped into probable QTL regions based on LD (**Table 3.3.3-3.3.7**). Within a range of 5-10 cM significant SNPs were grouped and accounted as a single QTL region. Map positions of many QTL regions were congruent with previously described QTL. In addition novel QTL regions were identified.

Row type

Before correction for multiple testing, 297 SNPs were found to be significantly associated (*P*-value < 0.03) with row type (**Supp Table 3.2**). The associated SNPs are distributed across all chromosomes with maximum number on 2H (104) **(Fig 3.3.4***)*. The major loci *vrs1*, *vrs3*, *vrs4* and *int-c* (Komatsuda et al., 2007; Pourkheirandish and Komatsuda, 2007; Ramsay et al., 2011; Waugh et al., 2009) are prominent with many significant SNPs in the corresponding regions. The regions in vicinity to the loci *vrs1*, *vrs3*, *vrs4* and *int-c* showed 35, 29, 10 and 7 SNPs significantly associating to the trait, indicating the important role of these genes in row type determination and underlying pathways. Even after *q*FDR testing, 59 SNPs were found to be highly significant at 0.05 *q*- value.

Heading date (HD)

A total of 269 SNPs were significantly associated (*P*-value < 0.03) to the trait heading date, accounting to 43 QTL (**Supp Table 3.3**). Chromosome 2H showed large number of SNPs (98) associating to this trait (**Fig 3.3.5**). This provides a suggestive evidence of the presence of either several loci affecting HD on 2H or the presence of major loci affecting HD that are in linkage with other genes on this chromosome. The regions corresponding to *Ppd-H1* (Laurie et al., 1995) and *eam6* (Comadran et al., 2011) on 2H have several SNPs associated to the trait HD. Similarly, strong associations were detetcted in vicinity of *HvFT1* on chromosome 7H (Faure et al., 2007) and *HvCO1* region on 7H (Griffiths et al., 2003; Wang et al., 2010a). Only four SNPs, one SNP from the *Ppd-H1* region, one from *eam6* region, and one from 3H and from 1HL surpassed the *q*FDR threshold (**Table 3.3.3**). All QTL regions detected using the IPK-OPA marker set were also identified using iSelect markers by K_3 model (**Fig 3.3.5**). From the unmapped markers, 70 SNPs were significantly associated with the trait HD.

Plant height (PHT)

Plant height (PHT) showed significant associations (*P*-value < 0.03) with 240 SNPs and among them five SNPs surpassed the *q*FDR threshold (**Table 3.3.4**). Significantly associated SNPs were detetcted on all chromosomes, with 60 SNPs on 7H and 56 SNPs on 3H. The genomic region on 3HL close to centromere corresponds probably to the *uzu* gene (Saisho et al., 2004) and on the distal end of 3HL to *sdw1*(Jia et al., 2011). After grouping the significant SNPs based on LD, a total of 44 probable QTL regions were identified. This is twice the number of QTL detected using the IPK-OPA markers (**Fig 3.3.6**). However, only three of these QTL passed the stringent *q*FDR threshold level (**Supp Table 3.4**). Among the unmapped markers, 57 SNPs were significantly associated with PHT.

Thousand grain weight (TGW)

For the trait thousand grain weight (TGW) 266 SNPs were significantly associated (*P*-value < 0.03) across all chromosomes (**Table 3.3.5**). None of these SNPs surpassed the *q*FDR threshold level pointing at the high stringency of this threshold. Significant SNPs were grouped into 45 QTL regions, which again exceeds the number of QTL detected using IPK-OPA. SNPs from the genomic regions at *Ppd-H1*, *eam6*, *vrs1*, *vrs3*, and *int-c* were significantly associated to TGW (**Supp Table 3.5**). Chromosomes 2H (70) and 5H (47) harbor more SNPs associated to TGW than the remaining chromosomes (**Figure 3.3.7**). Sixty-six unmapped SNPs were significantly associated with TGW.

Starch content (SC)

Maximum number of significant marker associations (304) were detected for the trait starch content (SC) in our analysis (**Table 3.3.6**) (**Supp Table 3.6**). These SNPs were grouped into 50 QTL regions. Surprisingly, 19 of these SNPs surpassed the significant *q*FDR threshold and they fall under 9 QTL regions. SNPs from regions corresponding to *vrs3*, *Ppd-H1*, and *waxy* loci showed significant associations. Most associated SNPs were found on 2H (82), followed by SNPs on 7H (56) (**Figure 3.3.8**). From the unmapped markers, 77 SNPs were significantly associated with SC and among them 12 SNPs were highly significant at *q*FDR threshold.

Crude protein content (CPC)

For crude protein content (CPC), 245 SNPs were significantly associated (*P*-value < 0.03) which grouped into 53 QTL regions (**Table 3.3.7**). However, only 13 SNPs corresponding to eight QTL regions surpassed the *q*FDR threshold level (**Supp Table 3.7**). Interestingly, 7H harbors the largest number of SNPs significantly associated for CPC (**Figure 3.3.9**). From the unmapped SNPs, 82 were significantly associated with CPC and among them 15 SNPs crossed the *q*FDR threshold limit.

3.3 Discussion

In recent years there has been a surge in GWAS using different marker systems with different marker coverage in barley and various other crops (Atwell et al., 2010; Comadran et al., 2009; Roy et al., 2010). Bearing in mind the fact that marker coverage in barley is not as extensive as in model plants or in completely sequenced genomes, it is difficult to analyze the trait associations to fine resolution of candidate genes in many barley GWA studies. Till recently, marker coverage achievable with the OPA SNP markers is at maximum 3000 markers across the whole genome of barley (Close et al., 2009). SNP assays with millions of markers and genotyping by sequencing (GBS) methods have already paved their way into the association mapping studies in humans and plants (Atwell et al., 2010; Huang et al., 2010; Tian et al., 2011). Recently, a 9K SNP chip (iSelect) became available for barley that can now be used for GWAS, population diversity studies, genomic selection and other plant breeding approaches (Comadran et al. unpublished). From the iSelect chip the successfully mapped SNPs that showed MAF >0.05 (5474 SNPs) were used for GWA mapping. A subset of the markers included in the iSelect array was used in a pilot study using the IPK OPA array. A total of 790 SNPs are commonly present both on iSelect assay (6467 successfully mapped SNPs) and IPK OPA assay (918 SNPs). The effects of using different marker sets for calculating kinship and their effects on GWA studies were compared. This study demonstrates the advantages of increased marker density on the number of QTL detected and on the significance of the detected QTL. We detected a multifold increase in significantly associated SNPs to the trait of interest, compared to GWAS with previously used OPA markers.

3.4.1 Comparison of different kinship matrices

Despite the vast increase in the number of GWA studies, the effects of using different sets of markers on statistical models that use kinship generated by various markers has not been well studied. Most of the studies comparing different statistical models for GWAS concluded that the mixed linear models proposed by Yu et al (2006) fit best for association mapping (Kang et al., 2008; Stich et al., 2008). The use of MLM with STRUCTURE Q-matrix and kinship matrix was shown to be highly successful, along with other models replacing the Q-matrix with principal components or only using kinship matrix (Kang et al., 2008; Price et al., 2006; Stich et al., 2008). In a previous study we demonstrated that the K-model performed best in the current panel (Pasam et al. 2012, Chapter 2). Therefore, the next issue we explored was the choice of kinship matrix to be used for MLM. The robustness of predicting population structure using different background markers has been previously reported (Falush et al., 2003; Kaeuffer et al., 2007), but studies reporting the effect of using different background markers on kinship estimates are still limiting. Here we compared different kinship matrices K_1 (all iSelect markers), K_2 (918 SNPs) and K_3 (362 SNPs) generated by various SNP marker sets. The matrices were observed to be highly correlated with significant Mantel correlations (**Table 3.3.1**). The correlation was higher between the K_1 and K_3 although the marker number used for kinship estimation differed significantly. Uniformly spaced markers with high PIC values were carefully selected to estimate the K_3 matrix. This matrix captured nearly the same amount of relatedness as the K_1 matrix estimated from all 6467 markers. For population structure estimation it is suggested to use uniformly distributed markers, to avoid biased estimation due to the correlated allele frequencies of markers that are in close LD (Falush et al., 2003; Kaeuffer et al., 2007). The same applies for kinship matrix estimation, as kinship is estimated based on allelic frequencies. The present SNP arrays reflect an approximate marker density between 1.18 (IPK-OPA) and 4.01 markers/recombination unit (iSelect SNPs mapped using Morex x Barke RILs). As a consequence, the large number of markers mapped closely at same genomic position and are in strong LD (see Supp Fig 2.5, chapter 2). For instance, on iSelect assay 58 SNPs are mapped at 51cM on chromosome 3H among which many are in high LD. Such groups of SNPs possibly cause biased estimation of kinship, which could be avoided by selecting a set of equally distributed markers based on LD decay in the panel. However, the increasing number of markers might lead to biased calculation of kinship estimates due to dependency among the markers (Browning, 2008), which could be avoided by pruning the marker number. Besides, the optimum requirement of markers for kinship estimation is still a frequent and important question. From simulation studies in maize it was reported that 1000 random markers is optimum requirement to estimate kinship and further increase of markers might lead to biased estimation due to marker dependency (Yu et al., 2009). In a previous study in barley, 384 SNPs instead of 1536 SNPs were suggested to be ideal for population structure estimation, as increasing the number of markers did not provide any additional information (Moragues et al., 2010). We found this also applies to the kinship estimation in our studies. The kinship with 362 SNPs has illustrated similar relatedness as the kinship with 6467 markers (**Table 3.3.1**). However, detailed examination of marker dependency and its effect on kinship can be assessed by extensive simulation studies to provide a realistic solution, which is beyond the scope of this study.

3.4.2 Comparison of kinship effects on GWAS

In spite of using different kinship matrices (K_1, K_2, and K_3), the GWAS results showed considerable overlap (**Fig. 3.3.4**). Highly associated SNPs were found to be significant through all analyses using different matrices, but the significance level of associations differed. Likewise, the variation explained by the SNPs also differed for the three different models. Examining **Table 3.3.2** provides an overview of the variation explained by the model using different kinships for each trait. It was observed that for all traits the model with K_3 explained less variance than the other two models, leaving more variation to be explained by the markers. This is good in one way that it prevents false negatives due to over correction and at same time controls for false negatives by capturing the population relatedness to an optimum extent. Hence in present analysis using iSelect markers, the K_3 model for QTL detection was used. For instance, 160, 180 and 269 SNPs were associating significantly for the trait HD using kinships K_1, K_2, and K_3 for population correction. Similar trend was observed for the number of significant SNPs for each of the trait with use of different kinship matrices. The requirement of background markers for assessing relatedness in a population depends on the species, population size, LD pattern in the population, marker type, and available marker density and distribution across the genome. In summary, researchers have to make rational choices about the marker sets used to estimate kinship and assess the tradeoff between the false positives and false negatives while balancing the power to detect QTL in a devised study.

3.4.3 GWAS with iSelect markers

As expected, the number of SNPs significantly associated with the trait of interest and the number of QTL detected increased with iSelect markers compared to the results obtained with IPK-OPA markers (**Table 3.3.3 to 3.3.7**). We considered a P-value of 0.03 as a significance threshold similar to previous studies (Chapter 2), but additionally applied a threshold to account from multiple testing (*q*FDR). Apparently, increasing the marker density resulted in detecting more QTL for all traits and with much higher significances. The increase in significance of SNP associations is due to an increased number of SNPs that are in close linkage with the causal gene. This was demonstrated for *Ppd-H1,* a gene involved in the regulation of heading date (**Chapter 2**). The variance explained by individual associated marker also increased for each trait when compared to previous results. However, the variances are still much less when compared to variances observed in bi-parental QTL mapping (**Supp Table 3.2 to 3.7**). The marker coverage of 7000 SNPs across 5.1 Gb barley genome (very roughly 1 SNP/760 Kb), though is the highest coverage achieved in barley till now, it is still exiguous in comparison to other studies. Previous studies suggested that most SNPs associated with the trait would be located very close to the causative genetic variant (Myles et al., 2009). Regardless, for some of the traits the detected significant associations were in concordance with previously detected genes/QTL. Some SNP loci were very close to known genes in barley like *Ppd-H1* on 2H for HD and *sdw1* on 3H for PHT. In any GWA scan, the genomic region associated with the trait will show either a single marker associating or multiple markers associating in form of a smooth rise and fall peak. The latter pattern confers more confidence to the association results, although single marker associations cannot be ruled out completely. The number of SNPs associated with the trait and with different significances in a given region is the consequence of LD decay in that region (Wang et al., 2010b).

Barley spike morphology is one of the determining factors of population structure and several genes have been proposed to influence this trait (Pourkheirandish and Komatsuda, 2007). This trait was investigated by simple scoring of two-rowed and six-rowed phenotype. Surprisingly, 297 SNPs were associated with row type and 59 SNPs crossed the *q*FDR threshold. Highly significant associations were detcted for row type and some of the loci detected were having pleiotropic effects on other agronomic traits. The candidate gene for *int-c* locus is a maize domestication gene *Teosinte branched 1* (TB1), which has an orthologue gene in rice at loci Loc_OS03g49880 ('TCP transcription factor') (Ramsay et al., 2011). Exploration of syntenic conservation between barley and rice for the significant SNPs (11_20606) from our GWAS revealed that the *int-c* homologue gene 'TCP transcription factor' is only thirteen gene models away in rice from the associated SNPs.

Likewise, the SNP markers associating with row type on 2H at 80 cM correspond to the *vrs1* gene. The *HvHOx2* gene is the candidate for *vrs1* locus (Komatsuda et al., 2007) and its homologue in rice is *Oshox14* (Loc_OS07g39320). The significantly associated SNPs 12_30896 and 12_30897 are syntenic to rice loci 'Loc_OS07g39320' indicating the preciseness of the GWAS approach. However, the other significantly associating SNPs in the same region in barley are syntenic to rice chromosome 4, suggesting breakdown of the micro-collinearity between rice and barley in this region. Substantial sequence collinearity was established between barley 2H chromosome and rice 4th and 7th chromosomes (Close et al., 2009) and the breakdown of the rice-barley micro-collinearity was reported by Pourkheirandish et al. (Pourkheirandish et al., 2007). This suggests the limitations in synteny based gene cloning, and emphasizes caution before deducing the predicted genes based on syntenic relationships. Significant associations detected on 1H at 50 cM corresponds to the *vrs3* region, for which the candidate gene is not yet known. Furthermore, we observed other associating SNPs that are in concurrent positions to known loci like *vrs4* on 3H and *vrs2* on 5H (Pourkheirandish and Komatsuda, 2007). Apart from them novel genomic regions associating for row type were found which were not detected with IPK OPA markers and they need to be explored further (**Supp Table 3.2**).

For HD, several SNPs were found to be significantly associating with the trait (**Table 3.3.3**). HD is a quantitative trait of complex genetic architecture with many pathways and genes underlying the phenotype variation (Buckler et al., 2009). In barley, several genetic and molecular studies tried to unravel the involved pathways and genes for flowering (Clark, 1967; Dunford et al., 2005; Griffiths et al., 2003; Wang et al., 2010a). In our studies, 43 probable QTL regions were detected under optimum threshold (*P* =0.03), but only four regions surpassed the stringent threshold (1H-132cM; 2H-18 cM; 2H-63cM; 3H-55cM). The 1HL region corresponds to *eam8* region (Sameri et al., 2011), the region on 2HS corresponds to *Ppd-H1* (Wang et al., 2010a), whereas the 2HL region to *eam6* (Comadran et al., 2011). The flanking markers of these gene/QTL regions were compared with the positions of associated SNPs using consensus map and common marker positions on these maps. In case of cloned genes, he syntenic conservation information of rice and Brachypodium was used to validate the associations and see how close the associated SNPs are to the gene (Mayer et al., 2011). The *Praematurum-a* (*Mat-a*) that is allelic to *eam8* was cloned and mapped to the long arm of 1H (Zakhrabekova et al., 2012) is mapped to the same region.

The region on 3H 55 cM (SNP marker SCRI_RS_168173) has an effect of 2.35 days on heading date which is the highest effect observed in our analysis. This region is close to the *HvGI* gene (Dunford et al., 2005), but the syntenic conservation between rice and barley has shown *HvGI* is located very distant from the significantly associated SNPs. Furthermore, one of the SNPs from gene *HvGI* (BK-08) included in iSelect assay was not significantly associated to HD and also mapped 10 cM away from the significantly associated SNP. Hence, it is not yet clear if the association on 3HS corresponds to *HvGI* gene or whether it is a novel QTL. Though the remaining associated SNPs did not cross the *q*FDR, there is evidence of co-location of some important flowering time genes with these SNP positions. For instance, SNPs associated to 1H at 97 cM correspond to *Ppd-H2* (Sameri et al., 2011) and SNPs on 7H corresponds to *HvFT1* gene (Wang et al., 2010a). SNPs included in iSelect assay from *Ppd-H1* gene were in group of unmapped SNPs and showed highly significant association with the trait HD.

For PHT. 240 significant SNP associations were detcted among which 5 SNPs were crossing the *q*FDR threshold. It is not surprising that we found nearly 44 QTL regions that associate to the plant height in our analysis. In barley nearly 30 types of dwarfs and semi-dwarfs have been found including breviaristatum-*ari*, brachytic-*br*, curly dwarf*cud*, denso dwarf-*denso*, erectoides-*ert*, lazy dwarf-*lzd*, many noded dwarf-*mnd*, many noded dwarf-*mnd*, 'semidwarf-*sdw*, 'single node dwarf-*sid*, 'slender dwarf-*sld*, uzu or semibrachytic-*uzu* and vegetative dwarf-*dwf* (Zhang and Zhang, 2003). Semi-dwarfing genes were preferred over dwarfing genes and have been extensively used in barley breeding to reduce plant height and improve resistance to lodging. Most of the European and American semi-dwarf cultivars possess the *denso*/*sdw1* gene which is mapped to 3HL. The candidate gene for *denso*/*sdw1* gene is *Hv20ox*$_2$ in barley, which is homologue of rice *sd1* gene encoding gibberellin (GA) 20-oxidase enzyme (Jia et al., 2011). SNPs associated to PHT on 3HL at 108 cM (12_31525, SCRI_RS_120973, SCRI_RS_121052, SCRI_RS_103215, SCRI_RS_165334) are in a concurrent position to the *sdw1* region. Furthermore, syntenic comparisons with rice revealed that the most significantly associated marker is only 18 gene models away from the *sd1* loci in rice. Among the significantly associated markers, SNP SCRI_RS_103215 is only one gene away from *sd1* (*OsGA20ox2*) loci in rice. Different alleles of *sdw1* [*sdw1-a* (Jotun), *sdw1-c* (*denso*) and *sdw1-d* (Diamant)] were used in barley, especially in feed barley in Europe and USA.

Few European barley cultivars are reported to carry dwarfing allele of gene *ari-e* also known as *GPert* (Zhang and Zhang, 2003). The dwarfing mutant of the Golden Promise cultivar (*GPert* or *ari-e.GP*) is allelic to the breviaristatum mutant allele *ari-e* which was mapped to 5H short arm (Lundqvist and Franckowiak, 2003; Thomas et al., 1984). The significant associations were observed in the region of 44-59 cM on 5H, which on comparison with consensus map are in concurrent position to *ari-e* locus. Though we do not have further evidence to confirm this, there are no other genes effecting PHT reported in this region so far. Moreover, several European cultivars were included in our association panel which allows us to expect this gene to show up, as *ari-e.GP* is reported to be more prominent in European barley.

Likewise, on chromosome 3HL, close to the centromere, the GA insensitive semi-dwarf gene *uzu* has been mapped. Semi-dwarf varieties with *uzu* gene are distributed exclusively in East Asia. Barley *HvBRI-1* (*Brassinosteroid Insensitive-1*) gene is a homolog of rice and Arabidopsis *BRI1* and was identified as candidate gene for *uzu* gene (Chono et al., 2003; Gruszka et al., 2011). SNPs associated to PHT in our analysis close to the centromere at 50-52 cM are close to *uzu* gene position when compared on consensus maps using Graingenes database. As current panel consists of barley accessions of worldwide origin, including from Japan, China and Korea, where the *uzu* gene is the major semi dwarf gene exploited in breeding, we expected significant associations in this genomic region for plant height. Similarly, the Gibberellic acid (GA) insensitive dwarfing gene *sd3* has been fine mapped in the centromere region of 2H by exploiting the syntenic relationship between 2H and long arm of rice chromosome 7 (Börner et al., 1999; Gottwald et al., 2004). In line with these findings, significant associations were detected on 2H in a window between 55-67 cM which is close to the centromere and corresponds well to the region where *sdw3* has been located on the consensus map. As confirming evidence, two of the significantly associated SNPs (SCRI_RS_136233 and SCRI_RS_103572) in this region are collinear to rice chromosome 7 and are very close to the position of a set of projected candidate genes identified for *sdw3* in rice by using syntenic conservation models between the genomes (Vu et al., 2010).

Using OPA SNPs we previously showed significant associations on 2H at 73 cM and assigned this region to *sdw3* (Pasam et al. 2012, **Table 2.3.8**). The difference in the observed positions observed is due to inherent inaccuracies within the consensus map that was used as a reference for locating IPK-OPA markers. By using a single segregating population Morex x Barke (Comadran et al. unpublished) to map the majority of the SNPs present on the iSelect array SNPs positions could be estimated more accurately. Moreover, re-examination of the associated markers of the IPK-OPA panel has confirmed that they were in common positions to those from the iSelect panel confirming the congruency of both approaches. Recently, a novel gene *btwd1* for PHT was discovered in Chinese barley landraces which maps to the 7HL close to the centromere (Ren et al., 2010). The location of *btwd1* is similar to the significant associations we found in our analysis on chromosome

7H at 70 cM. Another QTL detected towards the distal end of 7HL corresponds to QTL-PH7 described by Yu et al. (Yu et al., 2010). However, during the comparisons with consensus markers we observed a difference of 10 cM for these regions. The difference might be because of difference in marker positions or probably they are two different loci. In addition to those known QTL, several novel loci for plant height were observed, which should be further investigated to confirm their effects and utility in plant breeding.

For thousand grain weight (TGW) we observed 266 significant SNPs above the *P*-value of 0.03, but none of them surpassed the *q*FDR threshold. Several of the associated genomic regions correspond to known row type and heading date loci and QTL providing circumstantial evidence for their pleiotropic effects. For instance, *vrs3*, *vrs1* and *int-c* regions showed significant marker trait associations for TGW. The profound impact of row type on grain characters is apparent due to the source-sink differences and due to the pleiotropic nature of some of these loci (Pourkheirandish and Komatsuda, 2007; Saisho et al., 2009). Separate breeding histories and end-usability preferences have further accentuated the row type differences for various traits especially for the grain traits. Also genes affecting heading date showed significant associations to TGW, indicating the importance of flowering time on final yield. The flowering time genes are important for local adaptation and are known to impinge on major yield components (Cockram et al., 2007). In barley, flowering time genes were reported to affect TGW and other yield component traits in several studies (Pourkheirandish and Komatsuda, 2007; Wang et al., 2010a). The SNPs in genomic regions of *Ppd-H2* (1H, 100 cM), *Ppd-H1* (2H, 20cM) and *eam6* (2H, 60-70 cM) are significantly associated with TGW. Numerous QTL were reported in the past for TGW and some of them are in concurrent position to the present detected QTL (Chapter 2, Table 2.3.9) and besides there are several new QTL detected in our GWA scan.

We observed maximum number of SNPs (304 SNPs) associated for starch content (SC), in our GWA scans. In cereals, a chain of enzymes encoded by multitude of genes are involved in starch biosynthesis including *ADP-glucose pyrophosphorylase*, *starch synthase*, starch branching enzymes and starch debranching enzymes. These enzymes are encoded by different classes of genes and are known to effect the composition and content of grain starch at various levels (James et al., 2003; Li et al., 2011). Not surprisingly, the SNPs closer to *Ppd-H2*, *Ppd-H1* and *eam6* genes were significantly associated to SC in our analysis. SNPs at around 50 cM on IH are significantly associated to SC, which is close to the position of *vrs3*. However, the *amo1* mutant locus, which is believed to play a role in starch accumulation, was mapped to the same region and *starch synthase IIIa* (*ssIIIa*) gene is suggested to be the involved candidate gene (Li et al., 2011). As the associations are LD based, it is difficult to clearly assign the SNPs to either of the two loci mapped to the same region. Similarly, several SNPs on 3H associated significantly to SC. Numerous QTL and functional genes related to carbohydrate accumulation, starch degradation and protein content have been mapped to 3H in previous studies (Hayes et al., 2003). We also observed several SNPs on 7H associated to SC. The SNPs associating on 7HS at 9 cM (SNPs SCRI_RS_137983, SCRI_RS_132017, 11_20245 and SCRI_RS_152931) correspond to the *waxy* locus. The *waxy* gene encodes the *granule-bound starch synthase I* (*GBSS-1*), the key enzyme for amylase synthesis in cereal endosperm (Ma et al., 2010). By exploiting the syntenic conservation between rice and barley (Mayer et al., 2011) the significantly associated SNPs were found to be syntenic and only few gene models far from the *GBSS-1* encoding locus in rice. The significant peak close to the centromere of 7H (73 cM) in our GWA scan is close to the *nud* locus position (Taketa et al., 2006). Regardless, *sex6* mutant locus which is suggested to be *starch synthase IIa* (*ssIIa*) is also mapped to the same position on 7H (Morell et al., 2003). The association peak on 7H at position 73 cM in our GWAS can correspond to any one of these loci. Twelve unmapped SNPs (**Table 3.3.6**) associated to SC and can provide more information once they are mapped.

For the trait crude protein content (CPC), 245 SNPs were significantly associated which were grouped to 53 QTL regions. The traits SC and CPC are highly correlated to each other as a consequence many genomic regions were commonly associated with both traits. The regions corresponding to *Ppd-H2,* and *eam6* regions were associated to the CPC. The SNPs from *waxy* locus and *nud* locus region on 7H were also highly associated to CPC. Several *hordein* genes were mapped to chromosome 1H and many malting quality genes mapped across all the chromosomes in different studies were shown to effect the grain protein content (Hayes et al., 2003). Some of the detected genomic regions in our GWAS are concurrent with these QTL/genes and the remaining are novel regions which need to be further investigated. The QTL detected in the GWAS can be ascertained by confirming the QTL either by biparental mapping studies or by using a different LD mapping panel.

3.4 Conclusion

High number of marker trait associations were detected using iSelect assay markers due to the increased marker density compared to the OPA markers. The significance of the associations increased in many cases with addition of new markers to the region. As expected, most of the QTL regions detected using OPA markers were confirmed with GWAS using iSelect markers. The comparison of different kinship matrices and their effects on the results of GWAS has revealed that the kinship generated from uniformly distributed high PIC value markers is optimum for GWAS. The diversity captured by the kinship generated using evenly spaced 362 SNPs (K_3) is similar to the diversity captured by the whole iSelect marker set and K_3 further resulted in less false negatives. The phenotype variation explained by the model was less when K_3 was used compared to other kinships. In GWAS, for most of the traits we detected associations closely linked to major candidate genes affecting the trait. It was possible to predict candidate genes underlying a QTL in few cases by using the genome models exploiting the syntenic conservation between rice, *Brachypodium* and barley. However, micro-synteny studies reveal various rearrangements which complicate the synteny comparison based gene cloning strategies (Delseny, 2004; Pourkheirandish et al., 2007). Hence it demands extreme prudence in predicting the candidate genes solely based on the information from a synteny based model genome. The resolution of the panel and the marker density is sufficient for detecting QTL, but for further fine mapping or gene identification the present resolution is still limiting in many cases. To attempt for further fine mapping of the detected QTL the resolution of the panel could be improved by increasing population size and denser marker coverage can be advantageous. In recent times there are examples in rice and other crops where genes have been identified by GWAS using genotyping by sequencing approach for dense marker coverage (Huang et al., 2010). The tiered pattern of LD among genepools can be exploited to develop association mapping population of higher resolution in barley (Waugh et al., 2009). Therefore, we developed a large population of spring landraces for increased genetic resolution that could help in fine mapping of traits.

CHAPTER FOUR: Analysis of genetic diversity and population structure in spring barley landraces and pertinence for association mapping

4.1 Introduction

The apparent dearth of large scale genetic and diversity studies in barley landraces of varied origins resulted in the limited use of these genetic resources in plant breeding programs. Knowledge of genetic diversity in landraces will help exploiting the underlying natural variation for better understanding of the genetic basis of their environmental adaptation. This deeper understanding of genetic diversity serves as a prerequisite for effective utilization of landraces in future breeding strategies to achieve long term gains in agriculture. Association mapping is one of the recent methods proposed and presumed to champion the use of natural genetic diversity. The extent of LD decreases gradually from modern cultivars to landraces to wild genepools, and in case of allelic diversity a reverse trend of increased allelic diversity from cultivars to wild is observed. This varying trend of LD in different genepools stored in genebanks provides an opportunity for establishing populations with high genetic resolution using landraces and wild populations (Caldwell et al., 2006; Waugh et al., 2009). The study of genetic diversity and evaluation of population structure is foremost and decisive step in association studies to determine the appropriate statistical strategies and asses the power and usability of the mapping population.

In this study, a set of 1491 landraces originating from 41 countries was selected from a large collection of spring barley landraces stored in the Genebank at IPK (Leibniz Institute of Plant Genetics and Crop Plant Research) based on the nomenclature, morphological and agronomic descriptions available from passport data. The collection comprises two-rowed, six-rowed, naked and hulled barleys from 5º N to 62.5º N and 16º W to 71º E. The landrace collection was evaluated with 45 SSR markers to assess the genetic diversity, population structure and genetic differentiation among the collection to provide useful information for their efficient utilization. The accessions were from different climatic zones with regions of annual mean temperature ranging from 0ºC to 25ºC, and with annual precipitation of 30 mm to 1952 mm. The present study attempts to look at the genetic diversity and the population structure of a large collection of landraces originating from various eco-geographical and climatic regions using SSR markers in a global perspective. The characterization of these landraces will provide an interesting insight into the population structure which can be deployed in devising association mapping studies using this germplasm.

4.2 Materials and methods

4.2.1 Plant material

A representative subset of spring barley landraces originating from temperate and tropical zones of the northern hemisphere comprising of different climatic regions was selected (oceanic, Mediterranean, humid subtropical, continental, arid and semi arid climates). The material consists of 1491 landraces originating from 41 countries, based on the passport data available from the European Barley Database (EBDB, http://barley.ipk-gatersleben.de/ebdb.php3). Apart from considering geographical origins (or collection sites), morphological characters (viz., two-rowed type, six-rowed type, naked barley, hulled barley, hooded barley, colored seed and colorless seed) (Mansfeld, 1950) were used as additional criteria to select the accessions. Seeds were provided by IPK Genebank, Gatersleben Germany. The genebank at IPK practices the splitting of variable landrace accessions and stores them as morphologically distinct lines to counteract the possible loss of rare alleles in the population (Hamilton et al., 2002; Lehmann and Mansfeld, 1957). For this reason, in our collection, each landrace accession might correspond to a single representative of the original landrace population collected. The information about collection sites, scientific nomenclature and morphological character descriptions of each accession are provided in **Supp Table 4.1**. The landrace distribution according to the county of origin and row type is presented in **Table 4.2.1**.

Table 4.2.1 Distribution of landraces according to countries of origin, caryopsis type and row type

		Hulled		Naked	
Country of origin	No. of accessions	Two-rowed	Six-rowed	Two-rowed	Six-rowed
Afghanistan	107	9	87	1	10
Albania	22	5	10	2	5
Algeria	5	2	3	0	0
Armenia	4	4	0	0	0
Austria	56	36	20	0	0
Azerbaijan	1	0	1	0	0
Bulgaria	15	0	14	0	1
Croatia	3	2	1	0	0
Czech	17	7	1	7	2
Denmark	2	0	0	0	2
Egypt	5	1	4	0	0
Ethiopia	299	54	46	95	104
Finnland	3	3	0	0	0
France	9	2	6	0	1
Georgia	80	47	33	0	0
Germany	37	26	7	3	1
Greece	70	19	50	0	1
Hungary	3	2	0	0	1
Iran	84	44	31	9	0
Iraq	37	14	22	1	0
Italy	42	9	28	0	5
Kazakhstan	1	0	0	1	0
Latvia	1	1	0	0	0
Libya	123	13	107	0	3
Lithuania	1	1	0	0	0
Macedonia	1	0	1	0	0
Morocco	50	1	48	0	1
Netherlands	1	1	0	0	0
Norway	1	1	0	0	0
Poland	58	40	8	10	0
Romania	10	5	3	1	1
Russia	23	5	3	12	3
Slovakia	149	146	3	0	0
Spain	34	0	34	0	0
Sweden	3	2	1	0	0
Switzerland	10	8	1	0	1
Syria	6	5	1	0	0
Tunisia	4	1	3	0	0
Turkey	99	48	50	1	0
Ukraine	6	2	3	0	1
Yugoslavia*	9	4	5	0	0
Total	1491	570	635	143	143

* Former Yugoslavia (accessions from Serbia, Bosnia and Herzegovina)

4.2.2 Ecogeographic data

For all accessions, latitude and longitude coordinates were inferred using the original collection site names when the precise collection information is documented. For the accessions wherein the exact collection location is not documented, a broader source location (nearby city or province or state or country capital) were considered to infer geographic coordinates. The location name searches were performed with Google maps (http://maps.google.com/maps) and global Gazetteer version 2.2 (http://www.fallingrain.com/world/index.html). For each collection site ecogeographic data were collected. The climatic parameters like annual mean temperature (AMT), annual precipitation (APT), mean diurnal temperature range (MDR), maximum temperature of warmest month (MTW) and others were projected with software DIVA-GIS 7.4 (Hijmans et al., 2005b) using WORLDCLIM database (Hijmans et al., 2005a). Data from 10 arc min spatial resolution grid (approx 18.6 Km grid) was used in this study.

4.2.3 Molecular genetic studies

DNA isolation

Four plants per accession were grown in the greenhouse at IPK and were checked for heterogeneity. Leaves from one representative plant per accession were harvested at 3-week seedling stage for DNA extraction. DNA was extracted from dried leaves using BioRobot 9600 Work Station (Qiagen) with MagAttract 96 DNA plant core kit (Qiagen, Germany). Quality and concentration of the DNA was checked on 1% agarose gels followed by normalization to uniform concentration (50ng/µl).

SSR evaluation

Forty five fluorescence-labeled SSR markers were selected based on the barley genetic map (Thiel et al., 2003; Varshney et al., 2007a). The markers are uniformly distributed over all seven barley chromosomes (**Table 4.2.2, Supp Fig 4.1)**. Primers were labeled with HEX, FAM and TAMRA dyes allowing multiplexing of primers pairs into 15 multiplexes with three primer pairs per multiplex (M1 to M15). PCR reactions were performed following PCR profile described by Haseneyer et al. (Haseneyer et al., 2010b). Amplification products were separated on a capillary electrophoresis instrument MegaBACE 1000 capillary sequencer (Amersham Biosciences). Fragment sizes were analyzed and recorded using MegaBACE fragment profiler (Amersham Biosciences) software version 1.2. For each SSR, the allele sizes, peak intensities and stuttering of bands were carefully checked manually and low intensity ambiguous bands were given missing values. Six accessions with more than 50% missing values were excluded from some of the analyses.

4.2.4 Data analysis

4.2.4.1 Genetic structure analysis

Three of the 45 SSRs evaluated were either monomorphic or with ambiguous multiple amplifications and thus excluded from further analysis (**Table 4.2.2**). Genetic diversity and population structure were studied using 42 SSR markers. The genetic structure of the 1491 landraces was explored using model-based method based on multi-locus genotypic data using STRUCTURE 2.2 (Falush et al., 2003; Pritchard et al., 2000a). The approach uses Bayesian clustering analysis to assign individuals to clusters or groups (*k*) without prior knowledge of their population affinities and assumes loci in Hardy-Weinberg equilibrium. Individuals are assigned a probability (inferred ancestry) of belonging to a cluster or jointly to two or more clusters if their genotype is admixed. The STRUCTURE simulations were performed with the number of presumed populations from *k* = 1 to *k* = 20 (hypothetical number of groups) and 5 runs per *k* value. For each run, the initial burn-in period was set to 50000 followed by 100000 MCMC (Markov Chain Monte Carlo) iterations for accurate parameter estimation. All runs were based on admixture model assuming correlated allele frequency. The most probable number of groups was determined by plotting the estimated likelihood values [LnP(D)] obtained from STRUCTURE runs as a function *k*. The value of *k* at which the log likelihood data is maximum or reaches a plateau is considered as the best value to assign the individuals to clusters (Pritchard et al., 2000a). Initially the accessions were assigned to a particular group, when the membership coefficient (likelihood ratios of Q-matrix) of the accession is high in that group. This resulted in lot of admixtures even with individuals less than 50% membership coefficient being assigned to the groups. Therefore, a cut-off limit of 60% was considered to assign the individuals to a particular group. Accessions with greater than 60% membership coefficient (Q-matrix) to a particular group were assigned to that group. All remaining accessions that do not meet this criterion were not included to any group and considered as admixed. Graphical representation of the results was obtained using DISTRUCT 1.1. (Rosenberg, 2004). Principal Component Analysis (PCA) was performed using PAST 2.12 (Hammer et al., 2001) with 42 SSR markers. Relationships among accessions were further evaluated through Neighbor-Joining (NJ) cluster analysis using a shared allele distance matrix. Shared allele distance between the individuals was calculated using the formula given by Chakraborty and Jin (Chakraborty and Jin, 1993).

$$D_{SA} = 1 - \left(\frac{a}{2n}\right)$$

Where *a* is the number of common alleles to individuals *i* and *j*, and *n* the number of loci studied. Bootstrap support of the branches was estimated from 1000 bootstrap replicates and NJ tree was constructed on PAST (Hammer et al., 2001). The individuals in the dendrogram were colored according to STRUCTURE defined groups.

4.2.4.2 Genetic polymorphism and population differentiation

The polymorphism level at each locus, heterozygosity (H_e), number of alleles per locus, frequency of each allele and gene diversity was estimated for all loci across the total population using Powermarker 3.25. (Liu and Muse, 2005). Polymorphism Information Content (PIC) values are determined according to Botstein et al. (Botstein et al., 1980) using the formula:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{k} p_{i^2} - \Sigma_{i=1}^{k-1} \Sigma_{j=i+1}^{k} 2p_i^2 p_j^2$$

Where p_i and p_j are the frequencies of alleles *i* and *j* respectively.

Gene diversity (D_l) and is defined as the probability that two random chosen alleles from population are different at *l* th locus. The averaged gene diversity across all m loci is termed as D and is calculated using the formula (Weir, 1996):

$$D = 1 - \frac{1}{m}\sum_{l}\sum_{u}\tilde{p}_{lu}^{2}$$

Where p_{lu} is the frequency of the *u* th allele at the *l* th locus.

The genetic diversity index was calculated based on 42 SSR markers. The genetic distance matrix was calculated using shared allele distance approach (Chakraborty and Jin, 1993) based on allele frequencies at 42 loci. The Partition of genetic variation within and among populations was further assessed by Analysis of Molecular Variance (AMOVA) using ARLEQUIN 3.1 (Excoffier et al., 2005). Initially AMOVA was conducted between the two-rowed and six-rowed type barley followed by naked and hulled barley types to see the differentiation among these groups. To further investigate the molecular variance, AMOVA was conducted among the groups inferred by STRUCTURE analysis. Genetic differentiation among the groups was calculated based on unbiased F_{st} estimators (Weir and Cockerham, 1984). Pair wise population comparisons using Fixation statistics (F_{st}) were produced among all groups using FSTAT 2.9.3 (Goudet, 1995). Allelic richness, gene diversity (GD) and the number of alleles in each group were estimated using FSTAT. Allelic richness values for the subpopulations are calculated based on rarefaction procedure to account for varying sample size. This approach uses the frequency distribution of alleles at a locus to estimate the number of alleles that would occur in smaller samples of individuals (Leberg, 2002).

In all cases statistical significance was determined by performing 1000 permutations. NJ clustering of the groups was constructed based on the overall genetic distances between the groups. PCA based on genetic distances was performed separately for each group to further investigate the structuring and relationship among the accessions. The accessions were allocated to the core groups using M strategy employed in the software MSTRAT. The M strategy examines all possible core groups and singles out the core groups that maximize the number of observed alleles at the marker loci (Gouesnard et al., 2001).

4.2.4.3 Spatial genetic structure, analysis of geographic and climatic variables

Geographic ground distances in kilometers between the accessions were calculated based on latitude and longitude coordinates. Similar distance matrices among individuals for latitude and climatic variables like AMT, APT, MDR and MTW were calculated with PASSaGE 2.1 (Rosenberg and Anderson, 2011). Mantel tests were conducted between the genetic distance (shared allele distance) and other distance matrices to verify the relationship between these matrices. However, a simple Mantel test apart from indicating the relationship between the distance matrices will not provide any further information. In order to explore more details about the correlation between the matrices we generated Mantel correlograms which are completely analogous to autocorrelation function (Escudero et al., 2003). In a Mantel correlogram, one of the distance matrixes is partitioned into a subset of discrete distance classes. Then the Mantel statistics will be calculated for all the pairs that fall in these classes separately. Mantel correlograms allowed to assess the overall correlation between the matrices and to determine the significance level of correlation at the level of each distance class. Mantel correlograms were constructed using PASSaGE 2.1. Distribution of accessions according to geographic, population sub-groups and climatic factors were visualized using DIVA-GIS.

4.3 Results

4.3.1 Distribution of landraces

In total we surveyed a collection of 1491 landraces from the IPK Genebank that were originating from 43 countries ranging from 5º N to 62.5º N and 16º E to 71º W. The collection sites for the landraces are represented on the map (see **Fig. 4.3.1**). The collection comprises of 712 two-rowed and 779 six-rowed barleys. Among all accessions, 20 percent (299) were of Ethiopian origin (5º N to 14.5º N). The collection comprises of 286 hulless barley (naked barley), among them 199 were from Ethiopia. Only few Landraces from Syria (only six accessions) and Jordan were included in our studies as barley landraces from this region were extensively investigated by Russell et al. (Russell et al., 2011). Apart from other considerations, the numbers of landraces included in the study were in proportion to the number of landraces present from that country in the IPK genebank barley landrace collection.

4.3.2 Allelic variability, level of polymorphism and overall genetic diversity

Out of the forty five SSRs analyzed for the collection of 1491 accessions, two markers were monomorphic (GBM1043 & GBM 1036) and one marker amplified multiple fragments (GBM1326). These three markers were thus excluded from further analysis. All markers were uniformly distributed across the 7 chromosomes (**Supp Fig. 4.1**). The level of missing data across all loci is very low (1.79%). For the remaining 42 SSRs, 372 alleles with fragment size ranging from 90 to 360 bp were detcted. The number of alleles per locus ranged from 3 (GBM 1363) to 22 (GBM 1007, GBM 1015) with an average of 8.85 alleles per locus. The major allele frequency for each locus is in the range of 0.212 to 0.974 with a mean value of 0.512. PIC values ranged from 0.049 (GBM 1404) to 0.839 (GBM 1256) with an average of 0.548. The majority (75%) of the markers showed PIC values in the range of 0.4 to 0.839. The average allelic richness amounted to 5.743, ranging from 2.195 (GBM 1363) to 15.183 (GBM 1015). Very low heterozygosity (He) was detected in the collection and ranged between of 0 to 0.050 with an average value of 0.011 (See **Table 4.2.2**). Three SSR loci detected no heterozygosity, while 4 loci showed $H_e < 0.0001$, 22 loci showed H_e between 0.001 and 0.01, and 14 loci showed H_e between 0.01 and 0.05. Average gene diversity (GD) value of 0.603 was obtained, indicating a high level of genetic variation among these accessions. An allele is considered to be rare if the allelic frequency is less than 1% in the total population. Among all the SSRs, only 30 loci (71%) showed rare alleles. A total of 152 rare alleles were detected amounting to 41% of the total alleles discovered in the whole collection.

Table 4.2.2 Diversity statistics in the landrace collection. Marker name, multiplex (Mplex) number, SSR motif, number of alleles per locus (allele number), percentage of missing data per marker, heterozygosity and Polymorphic Information Content values (PIC) across 45 SSR loci

Marker	Mplex	Dye	SSR Motif	Linkage group	Position	Fragment range	Allele number	Missing (%)	Hetero-zygosity	PIC
GBM1404	M1	FAM	TATG	6H	129.76	220-290	8	1.95	0.0062	0.0499
GBM1363	M1	HEX	(AGG)7	5H	120.68	110-120	3	1.54	0.0068	0.3776
GBM1461	M1	TAMRA	NA	1H	135.94	190-240	20	8.32	0.0022	0.8108
GBM1033	M2	FAM	(AT)9	7H	67.13	270-290	8	2.21	0.0000	0.5830
GBM1110	M2	HEX	(AAG)6	3H	60.27	210-245	10	1.34	0.0394	0.5051
GBM1326	M2	TAMRA	(CTT)8	7H	31.24			Excluded		
GBM1013	M3	FAM	(CTG)9	1H	67.50	160-175	5	1.07	0.0075	0.3538
GBM1015	M3	HEX	ACAT	4H	115.94	190-275	22	1.54	0.0204	0.8217
GBM1176	M3	TAMRA	AT	5H	18.59	280-295	7	4.36	0.0000	0.6430
GBM1043	M4	FAM	AAC	3H	90.39			Excluded		
GBM1031	M4	HEX	AG	3H	50.26	280-295	6	0.80	0.0014	0.6090
GBM1212	M4	TAMRA	(AGG)5	6H	55.10	100-111	5	0.87	0.0014	0.4857
GBM1064	M5	FAM	AGGG	5H	157.60	280-300	8	1.41	0.0000	0.4078
GBM1035	M5	HEX	CT	2H	29.46	270-285	5	1.41	0.0020	0.7076
GBM1003	M5	TAMRA	CTT	4H	79.53	185-220	11	1.68	0.0075	0.5446
GBM1334	M6	FAM	(GGC)8	1H	70.69	120-140	5	2.48	0.0014	0.3162
GBM1036	M6	HEX	CT	2H	156.39			Excluded		
GBM1020	M6	TAMRA	AC	4H	64.05	240-250	4	2.28	0.0007	0.3810
GBM1413	M7	FAM	(TCATA)6	3H	49.69	150-175	6	0.27	0.0303	0.5738
GBM1047	M7	HEX	AGC	2H	129.66	205-222	6	0.13	0.0027	0.5792
GBM1029	M7	TAMRA	AG	1H	60.42	220-230	5	0.27	0.0007	0.3767
GBM1021	M8	FAM	AC	6H	40.17	250-280	15	1.41	0.0129	0.7317
GBM1060	M8	HEX	GGT	7H	8.78	200-220	6	1.27	0.0007	0.4343
GBM1075	M8	TAMRA	GT	6H	50.08	290-305	5	1.14	0.0007	0.4877
GBM1007	M9	FAM	AC	1H	26.45	180-230	22	1.07	0.0298	0.6938
GBM1483	M9	HEX	(GCG)7	5H	80.64	150-180	6	1.74	0.0150	0.3901
GBM1256	M9	TAMRA	(GA)8	6H	75.40	340-360	9	1.81	0.0034	0.8396
GBM1516	M10	FAM	CT	7H	81.21	90-115	10	0.87	0.0419	0.6599
GBM1221	M10	HEX	(AC)10	4H	14.65	105-135	12	1.01	0.0027	0.7304
GBM1405	M10	TAMRA	(CGCA)5	3H	86.33	270-290	4	0.87	0.0047	0.6853

Marker	Mplex	Dye	SSR Motif	Linkage group	Position	Fragment range	Allele number	Missing (%)	Hetero- zygosity	PIC
GBM1063	M11	FAM	(ACAT)7	6H	63.49	195-220	7	0.60	0.0337	0.6864
GBM1280	M11	HEX	CTT	3H	3.83	270-295	6	0.60	0.0054	0.5970
GBM1323	M11	TAMRA	(GCC)8	4H	28.96	110-135	7	0.80	0.0020	0.5159
GBM1464	M12	FAM	AT	7H	53.53	130-220	15	1.01	0.0102	0.7254
GBM1501	M12	HEX	(TAGA)6	4H	0.00	250-290	12	1.01	0.0264	0.4830
GBM1002	M12	TAMRA	CCT	1H	101.50	250-355	12	0.80	0.0041	0.3353
GBM1026	M13	FAM	AC	5H	53.08	210-220	5	2.55	0.0014	0.3948
GBM1459	M13	HEX	(AC)7	2H	64.35	150-175	10	2.55	0.0489	0.6670
GBM1419	M13	TAMRA	CTCAT	7H	95.75	90-130	8	1.88	0.0164	0.5025
GBM1018	M14	FAM	(CCG)6	4H	132.69	250-285	7	2.15	0.0034	0.3950
GBM1061	M14	HEX	(GGT)6	1H	130.75	320-350	10	6.30	0.0508	0.6318
GBM1208	M14	TAMRA	(AG)6	2H	102.85	130-160	10	2.48	0.0034	0.5373
GBM1054	M15	FAM	CCG	5H	132.16	255-275	9	2.28	0.0034	0.5866
GBM1008	M15	HEX	(AAC)10	6H	95.37	150-180	10	2.01	0.0397	0.6467
GBM1218	M15	HEX	GA	2H	72.45	130-150	11	2.88	0.0069	0.5498
Mean							8.85	1.79	0.01	0.54

4.3.3 Population structure

STRUCTURE runs were performed for *k* =1-20 based on the distribution of 372 different alleles at 42 SSR loci among 1491 accessions. The value of log likelihood of the model increased for *k* values from 1 to 20 and the LnP(D) value seems to reach a plateau at *k*=10 with slow rate of change of LnP(D) values subsequently (**Fig 4.3.2**). When the landrace collection was divided into two clusters (K=2), 92% of the landraces (threshold > 60) were assigned to one or the other group. The primary division was into Ethiopian landraces (red) and non Ethiopian landraces (green). Very few of the Iranian, Iraq and Afghanistan landraces also grouped along with the Ethiopian landraces. The non-Ethiopian group consisted of both two-rowed and six-rowed barley landraces (**Supp Fig 4.2**). When the landraces were divided into three clusters (*k* =3) 89% of the landraces were assigned to one of the groups and the proportions assigned to each group are asymmetric. The major groups observed were: 1. Ethiopian landraces and two-rowed barley from Europe (green). 2. Majority of six-rowed barley (red) 3. Two-rowed and six-rowed barley from Iran, Iraq, Afghanistan and Georgia (blue) (**Supp Fig 4.2**). When the landraces were inferred into four distinct clusters based on STRUCTURE Q-profiles (*k* =4), 92% of the landraces were assigned to one of the groups. The major division of the landrace collection into groups was: 1. Ethiopian landraces (green); 2. Majority of six-rowed barley from Europe, Mediterranean regions and North Africa (blue); 3. Majority of two-rowed barley (yellow); 4. two-rowed and six-rowed barley from Iran, Iraq, Afghanistan and Georgia (red) (**Supp Fig 4.2**). Even though at *k* =4 the landraces were distributed according to their spike morphology and geographical origins, we still observed more structuring in respective PCAs (not shown) for each of these groups.

For *k* =5, 6, 7, 8 and 9, 89%, 89%, 87%, 84% and 84% of the landraces were distributed to one or another group, respectively. The observed structuring patterns are based on spike morphology, naked or hulled seed trait and geographical origins, but still these groups are not distinct enough. At *k* =10 the rate of change of log likelihood is nearly a plateau and also the distribution of the landraces was apt with 87% landraces assigned to the 10 clusters (**Fig. 4.3.3**). The proportions of individuals assigned to each group are asymmetric and the distinctness of the 10 groups is optimum. Thus we decided to focus on k = 10 for our further analysis. A total of 197 accessions with less than 60% membership coefficient were considered as admixed. To find the key determinants in the inferred structure of these 10 groups, we compared the clustering with morphological descriptions and geographical origins of the landraces. In this study, geographical origin, spike morphology and hulled or naked seed trait are found to be the major players in delineating the collection into groups. The groups were named as Group 1 (G1) up to Group 10 (G10) and the constitution of each group is given in **Table 4.3.1**.

Group 1 (G1) consists of 200 landraces and six (3%) of them were considered admixed. The remaining 194 are naked barleys, notably most of them from Ethiopia (5.07-14.03ºN). The admixed lines were not included in to the group for further analysis. Nine accessions from outside of Ethiopia were included in this group.

Group 2 (G2) consists of 77 landraces and 11 (14%) of them were considered admixed and remaining 64 are six-rowed barley. The six-rowed hulled barleys are majorly from Libya (31), Iraq (17) and Iran (15) with in latitude range of 23.31-37.65ºN. One two-rowed barley from Libya and one six-rowed barley from Greece were exceptionally included into this group.

Group 3 (G3) with 258 landraces was the second largest group, and 32 (12 %) of them were considered admixed. The remaining 218 landraces are six-rowed barley. Four two-rowed accessions from Libya and one each from Tunisia, Algeria, Ethiopia and Greece were also included into this group. The majority of the six-rowed accessions are from Libya (61), Morocco (48), Spain (30), Italy (21), Greece (28) and Turkey (10). This group includes the landraces mostly from North Africa and from Mediterranean regions (26.33-41.59ºN).

Group 4 (G4) consists of 66 landraces with ten (15 %) considered as admixed and the remaining 56 landraces are all from Georgia (40.18-45.37º N). Interestingly both two-rowed and six-rowed barleys originated from Georgia were included in this single group.

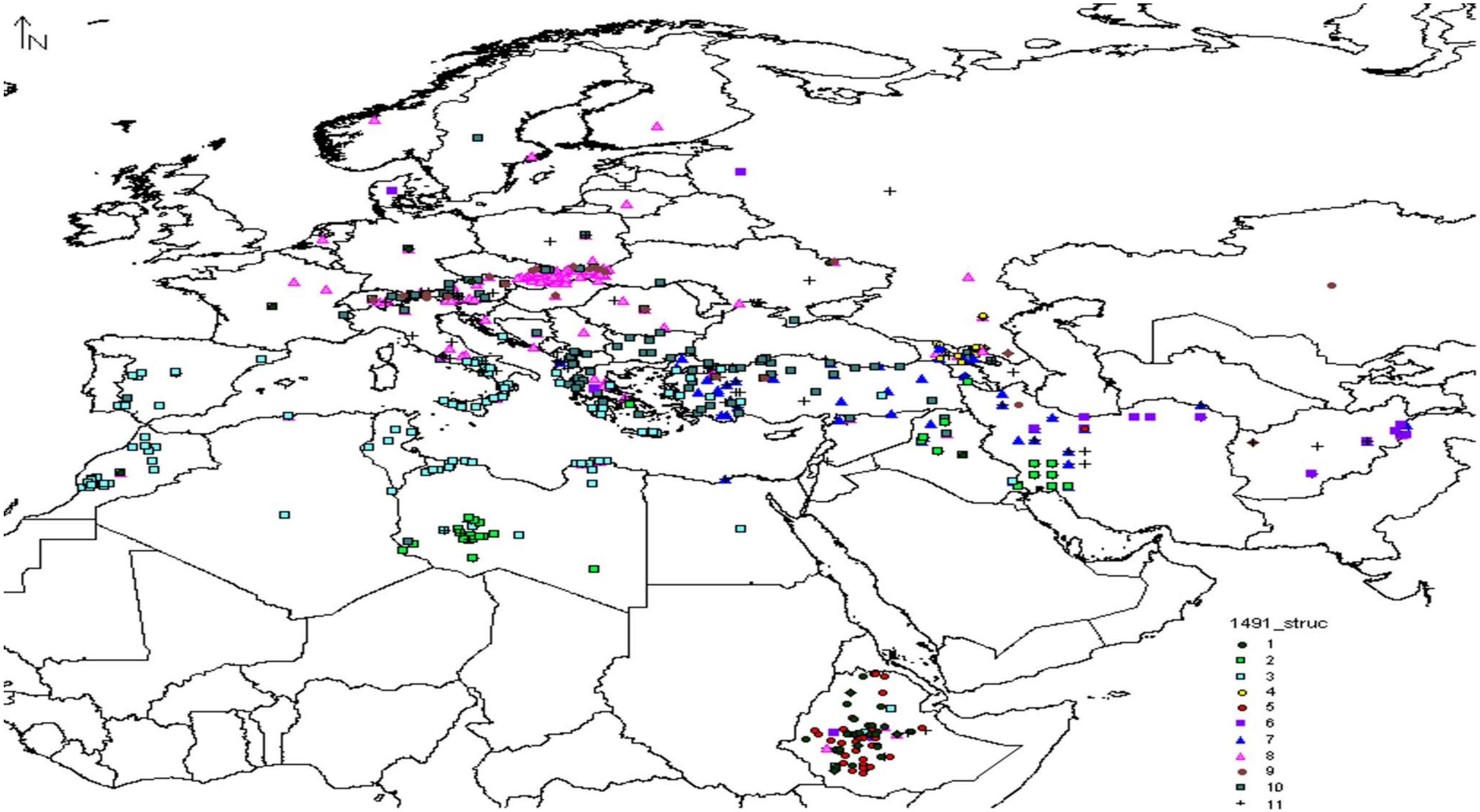


Fig. 4.3.1 Geographical distribution of accessions inferred to each group. Genotypes are plotted according to latitude and longitude of collection sites. Individual accession is represented by a symbol; there is lot of overlapping of the symbols due to same or closer reference data points. The two-rowed type groups are represented with triangles, the six-rowed type with squares and the Ethiopian, Georgian and naked barley are represented by circles. The cross symbol represents admixed accessions. The color and number corresponds to the STRUCTURE inferred groups (1-G1 to 10-G10 and 11-admixed)

Group 5 (G5) consists of 89 landraces with six (7%) considered as admixed. The remaining83 landraces were all from Ethiopia. This group consists of hulled Ethiopian barleys of tow-rowed (43) and six-rowed (40) type.

Group 6 (G6) consists of 97 landraces including 17 (17.5%) admixed accessions. Among the 80 accessions only a single accession is two-rowed type. Nine accessions are six-rowed naked barley from Afghanistan (5), Denmark (2), Greece (1) and Russia (1). Seventy of the accessions are six-rowed hulled barley from Afghanistan (61) and Iran (6) (31.61-36.61º N).

Group 7 (G7) constitutes 134 landraces and 36 (27%) of them were admixed. The remaining 98 accessions are majorly two-rowed barley. Majority of the accessions are from Iran (34), Turkey (29), Iraq (11), Afghanistan (9) and Georgia (7). The percentage of admixtures excluded is high for this group compared to all other groups (between 30-41ºN).

Group 8 (G8) was the largest group constituting 23% (336) of total landraces. Thirty nine (12%) landraces of them were considered admixed and remaining 297 are majorly two-rowed hulled barley. Three hulled six-rowed barley and four naked barley were included into this group. In total 290 hulled two-rowed type barley landraces from Slovakia, Poland, Germany, Sweden, Finland and Norway are included in this group (between 41-61ºN).

Group 9 (G9) was the smallest group with 57 landraces. Two landraces were considered admixed and the remaining 55 accessions are naked barleys. Interestingly, most of the naked barleys of non-Ethiopian origin were included into this group.

Group 10 (G10) consists of 177 landraces, 38 (21%) of them are admixed. The remaining 139 accessions are majorly six-rowed barley. Exceptionally, one naked six row barley and five hulled two-rowed barley from Austria were included into this group. In total 133 hulled six-rowed landraces from Austria, Bulgaria, Greece and Turkey were included (majorly between 40-60º N).

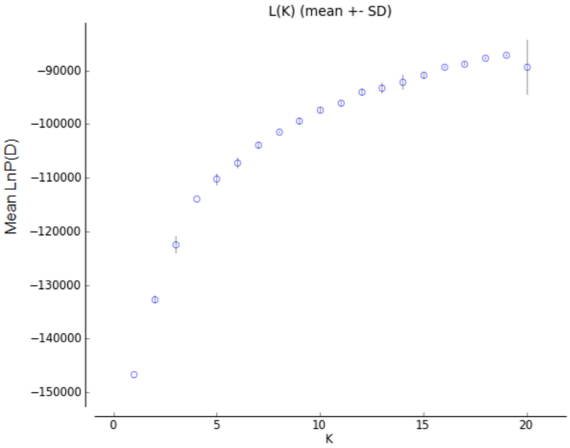


Fig. 4.3.2 STRUCTURE analysis of 1491 barley landraces**.** Log probability data (LnP(D)) plotted as function of *k* (number of clusters). LnP(D) values are the mean values of five replications. The plateau of the graph at *k*=10 indicates the minimum number of subgroups possible in the panel

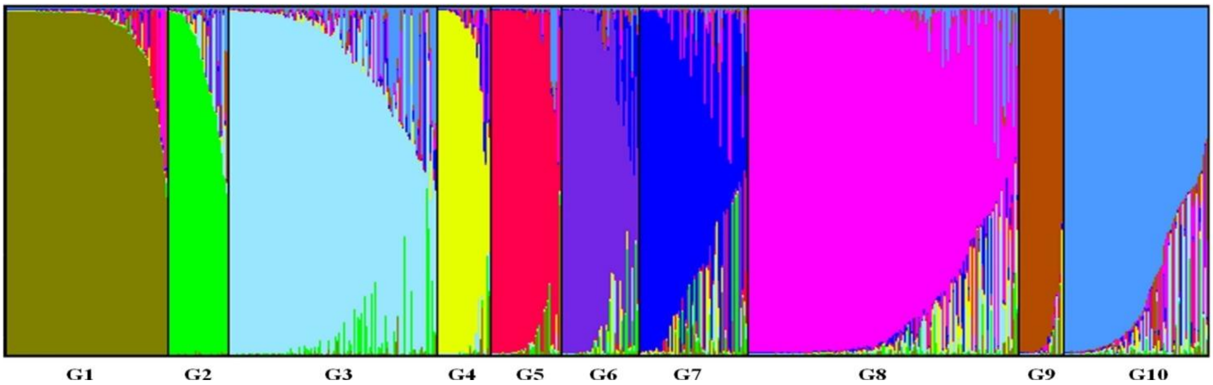


Fig. 4.3.3 Inferred population structure of 1491 spring landraces is represented by a vertical bar which is partitioned into 10 groups (*k*=10). The genotypes are ordered according to the membership coefficient (Q) values that are identified by different colors. Each color represents one group and the length of the color shows the accessions estimated membership in that group.

Table 4.3.1 Distribution of the landraces into ten groups based on STRUCTURE analysis. Total accessions represent the number of accessions that have the highest membership coefficient from that group. “Admixed” corresponds to accessions with less than 60% membership coefficient to a particular group. “Assigned” corresponds to accessions with coefficient above the threshold of 60%

				Assigned			
				Hulled		Naked	
Group	Total	Admixed (%)	Assigned	Two-rowed	Six-rowed	Two-rowed	Six-rowed
G1	200	06 (03.0)	194	0	0	88	106
G2	77	11 (14.3)	66	1	64	0	1
G3	258	32 (12.4)	226	8	218	0	0
G4	66	10 (15.2)	56	33	23	0	0
G5	89	06 (06.7)	83	43	40	0	0
G6	97	17 (17.5)	80	1	70	0	9
G7	134	36 (26.9)	98	92	6	0	0
G8	336	39 (11.6)	297	290	3	4	0
G9	57	02 (03.5)	55	0	0	43	12
G10	177	38 (21.5)	139	5	133	0	1
Total	1491	197	1294	473	557	135	129

To further place the STRUCTURE inferred groups in a cogent scenario, two other approaches to detect population structure were also evaluated. The second method employed to investigate the population structure in the landrace collection was hierarchical clustering. The NJ tree generated with bootstrap support values (based on 1000 bootstraps) for landrace collection is shown in **Fig. 4.3.4**. The accessions names were represented in different colors according to their classification from STRUCTURE (*k*=10) (**Fig. 4.3.4**). All admixed accession names were colored in black. The STRUCTURE inferred groups showed good agreement with cluster analysis as most of the accessions from a group were clustered together. The main exception was some of the accessions from G3 were seen clustering with accessions from G10. We have also generated a NJ tree from overall genetic distances matrix between each of these groups to see the relationship and distance among the groups (**Fig. 4.3.5**). The results were interesting as the groups G3 and G10, G6 and G2, G7 and G4, G1 and G5 were clustering together. These results were further confirmed by the Pair-wise F_{st} comparisons from population differentiation tests.

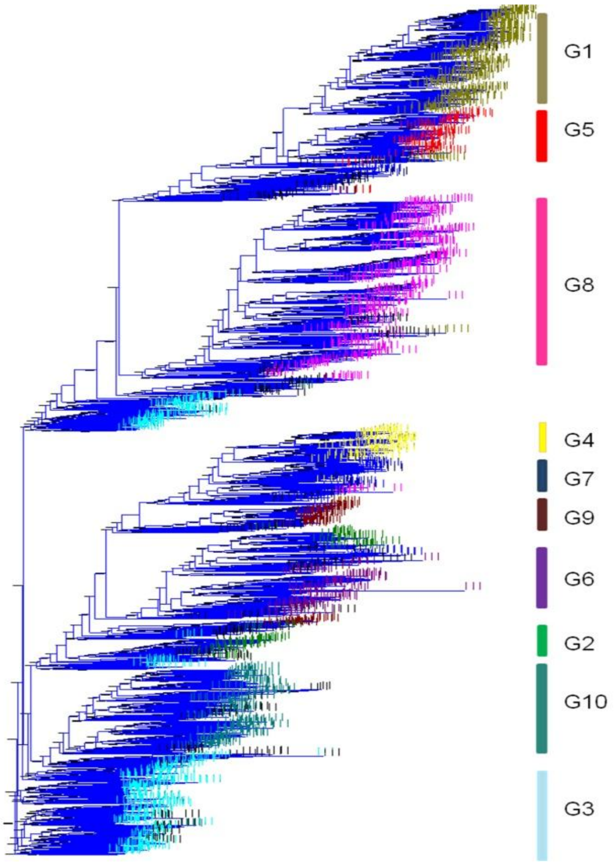


Fig. 4.3.4 NJ clustering Dendrogram of 1491 accessions based on the shared allele similarity matrix obtained from 42 SSR markers

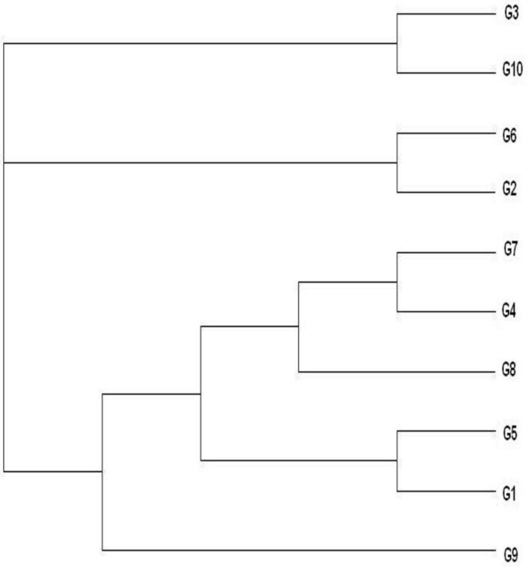


Fig. 4.3.5 Cluster analysis of different STRUCTURE inferred groups. STRUCTURE inferred groups are clustered based on their genetic similarity matrix

PCA was used as third approach to detect population structure and visualize the relationship of different groups. PCA (**Fig. 4.3.6**) was visualized for the first two components with accessions represented in different colors according to their classification from STRUCTURE (*k*=10) (**Figure 4.3.3**). From PCA, it is visualized that the discreteness between some of the groups is less pronounced than in the other two methods. Principal component 1 (PC1) and Principal component 2 (PC2) explained 11.6% and 8.97% variation respectively. PC1 primarily separates the Ethiopian accessions from the rest of the accessions and also separates the two-rowed barley from six-rowed barley. Interestingly, groups G2, G6, G7, and G9 are spread across the axis without distinct structuring and the groups G3 and G10 are overlapping. This apparently indicates the presence of greater diversity and genetic structure in these groups. We further tried to elucidate the relationships among the accessions within each group by PCA (**Fig. 4.3.7**). For each of the groups explained variation by PC1 and PC2 together ranged from 39% to 84% (G1-52.3, G2-84.5, G3-39.3, G4-56.7, G5-57.9, G6-41.2, G7-58.4, G8-45.9, G9-83.1 and G10-46). For the groups G2 and G9, PC1 and PC2 explained the highest variation indicating further distinct separation of accessions in these groups. PCA for G1 did not show any distinct separation, but a subtle structuring of two-rowed naked and six-rowed naked was observed (**Fig. 4.3.7a**). PCA for G2 showed a clear separation of Libyan and non-Libyan accessions on PC1. Further, PC2 separates Iraq and Iranian landraces into distinct groups (**Fig. 4.3.7b**). PCA for G3 did not show any pattern of sub structuring, besides the accessions were distributed across the axis (**Fig. 4.3.7c**). PCA for G4 showed a more distinct separation of Georgian two-rowed and six-rowed type barley (**Fig. 4.3.7d**). PCA for G5, G6, G7 and G8 did not show any further distinct structuring (**Fig. 4.3.7e to Fig. 4.3.7h**). PCA for G9 showed further structuring of the group into two-rowed naked barley and six-rowed naked barley accessions. G10 PCA did not indicate any further structuring among the accessions (**Fig. 4.3.7a to 7j**).

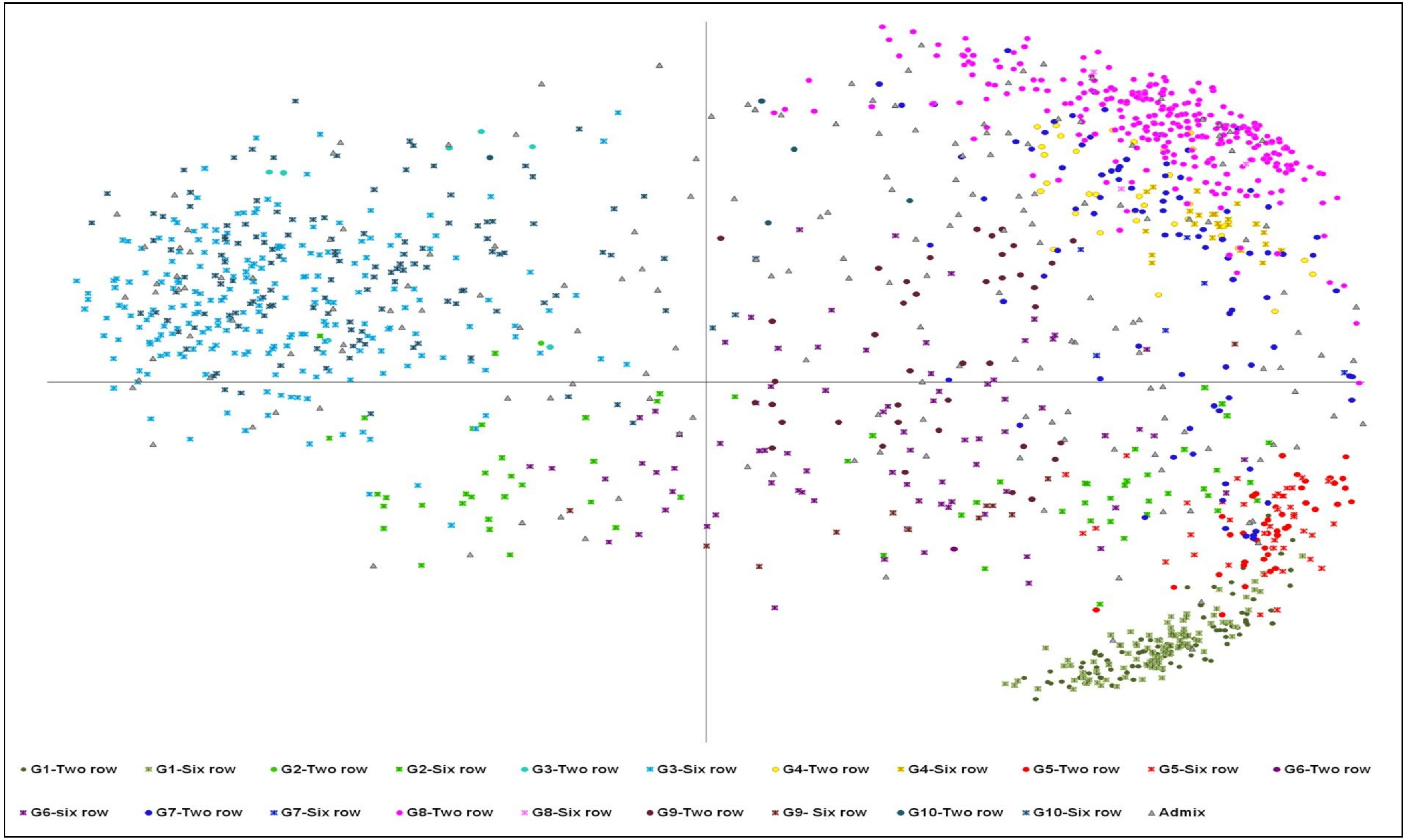


Fig. 4.3.6 Scatter plot of 1491 landraces from Principal Component Analysis calculated from 42 SSR data. The different colors correspond to different STRUCTURE inferred groups

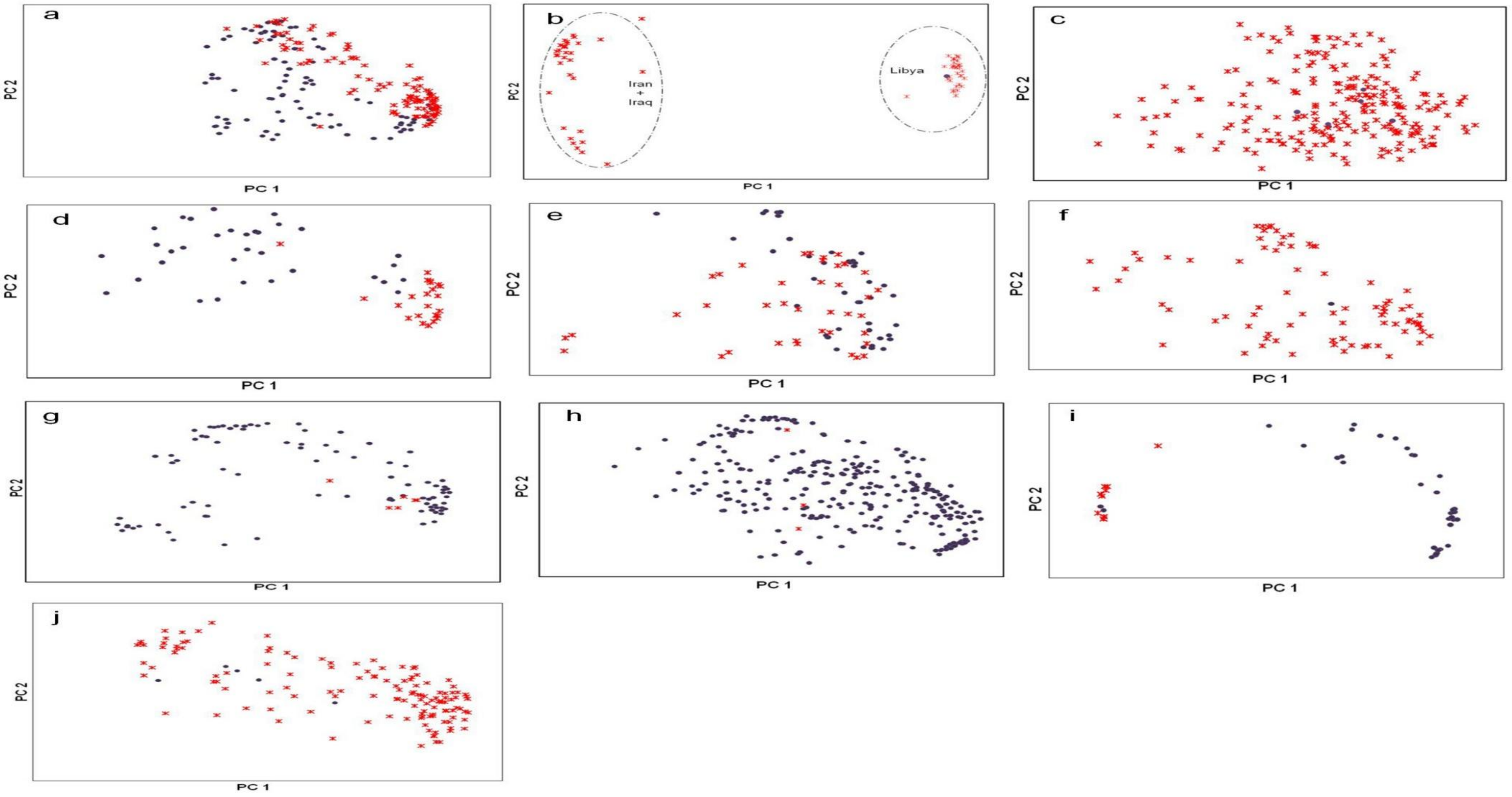


Fig. 4.3.7: Scatter plot of each STRUCTURE inferred group from Principal Component Analysis. The two-rowed barley are represented by blue colour circle and six-rowed barley are represented by red colour cross symbols. The PCA was performed independently for each group with 42 SSR data. Each plot represents a single group: (a) G1 (b) G2 (c) G3 (d) (e) G4 (f) G5 (e) G6 (f) G7 (g G8 (h) G9 (i) G10

4.3.4 Genetic variation and Population differentiation

Analyses of molecular variance (AMOVA) were performed to quantify the differentiation between different groups. Initially AMOVA was performed between the two-rowed and six-rowed barley, between naked and hulled barley, between Ethiopian and non-Ethiopian origin barley, between hulled and non-Ethiopian naked barely and between the STRUCTURE inferred groups (**Table. 4.3.2**). All sources of variation indicate highly significant level of genetic differentiation between groups and within groups. The percentage of genetic variation among the groups ranged from 8.75% to 36.63 and the percentage of genetic variation within the groups ranged from 63.37% to 91.24%. Analysis of AMOVA for two-rowed and six-rowed groups showed moderate differentiation (percentage variation=8.755) among the groups, and with maximum variation present within the groups (percentage variation =91.245). For hulled and naked barley groups the differentiation between the groups was moderate (percentage variation=15.754), but surprisingly higher than the differentiation between row type groups. The differentiation observed was high between Ethiopian and non-Ethiopian origin barley (percentage variation=19.610). The differentiation observed was very high among STRUCTURE inferred groups (percentage variation=36.622) with only 63% of explained variation within the groups. As fixation indices (F_{st}) measures the amount of differentiation between the population groups (Wright, 1951), Pairwise F_{st} comparisons among the STRUCTURE inferred groups were computed. Overall F_{st} among all clusters is 0.366 with F_{st} for each locus ranging from 0.090 to 0.651. Pairwise comparison on the basis of F_{st} values can be interpreted as standardized population distances between two populations. The pairwise F_{st} comparisons among the ten groups ranged from 0.187 between G4 and G7 to 0.544 between G1 and G4 (**Table 4.3.3**).

Average alleles per locus were higher in group G3, followed by groups G10, G8, G7, G6, G2, G1, G5, G4 and G9. The G3 group has twice the number of alleles compared to the number of alleles present in G9 group. Allelic richness was calculated to account for the differences in individual group sizes. A similar trend as observed for allele number was echoed for allelic richness among the 10 groups. Group G3 had the highest allele richness (4.91) followed by other groups, and G9 had the lowest allele richness (3.23). Gene diversity values were computed for all loci among the 10 groups. Gene diversity values were high for G7 (0.488) and lowest for G5 (0.256), G1 (0.279) and G9 (0.295). The number of unique rare alleles found are higher for group G10 (17) followed by G7 (13) and G8 (12) (**Table 4.3.4**).

Table 4.3.2 Analysis of Molecular Variance (AMOVA) Summary of partitioning of genetic variation among different groups of the landrace collection

	Total	Among populations	Within populations	P-value	F_{st}
Two-rowed & six-rowed barley groups					0.088
Variance components	13.261	1.161	12.100	0.000	
Percentage variation	100	8.755	91.245	0.000	
Hulled & naked barley groups					0.158
Variance components	14.218	2.240	11.978	0.000	
Percentage variation	100	15.754	84.246	0.000	
Hulled & non-Ethiopian naked barley					0.123
Variance components	14.147	1.753	12.394	0.000	
Percentage variation	100	12.328	87.672	0.000	
Ethiopian origin & non-Ethiopian barley					0.195
Variance components	14.392	2.822	11.570	0.000	
Percentage variation	100	19.610	80.390	0.000	
Ethiopian origin hulled & total collection					0.170
Variance components	13.735	2.534	12.384	0.000	
Percentage variation	100	16.987	83.013	0.000	
Structure inferred groups^{+}					0.366
Variance components	13.249	4.852	8.397	0.000	
Percentage variation	100	36.622	63.378	0.000	

*all the P-values are highly significant. $^{+}$ G1 to G10 groups inferred from structure analysis

Table 4.3.3 Pair wise comparison of F_{st} values between the STRUCTURE inferred groups. The P-values of significances computed after 1000 permutations are represented above diagonal and the F_{st} values are presented below the diagonal

Groups	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
G2	0.406		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
G3	0.429	0.273		0.001	0.001	0.001	0.001	0.001	0.001	0.001
G4	0.544	0.403	0.396		0.001	0.001	0.001	0.001	0.001	0.001
G5	0.331	0.394	0.423	0.494		0.001	0.001	0.001	0.001	0.001
G6	0.439	0.257	0.324	0.383	0.437		0.001	0.001	0.001	0.001
G7	0.386	0.260	0.291	0.187	0.345	0.251		0.001	0.001	0.001
G8	0.440	0.338	0.325	0.329	0.387	0.355	0.212		0.001	0.001
G9	0.476	0.324	0.349	0.406	0.495	0.315	0.284	0.371		0.001
G10	0.481	0.329	0.234	0.420	0.442	0.333	0.295	0.354	0.352	

Table 4.3.4 Diversity and summary statistics for the ten STRUCTURE inferred groups

Group	Sample size	Major Allele Frquency	Allele No	Mean Allele No	Availab ility	Gene diversity	Hetero-zygosity	Allele richness	Group specific rare alleles
G1	194	0.791	174	4.143	0.987	0.279	0.012	3.25	9
G2	66	0.625	189	4.500	0.984	0.474	0.016	4.44	4
G3	226	0.646	252	6.000	0.980	0.465	0.015	4.92	10
G4	56	0.776	140	3.333	0.987	0.295	0.005	3.32	1
G5	83	0.815	146	3.476	0.985	0.256	0.012	3.31	2
G6	80	0.632	206	4.857	0.984	0.475	0.010	4.69	10
G7	97	0.613	213	5.071	0.988	0.488	0.011	4.77	13
G8	295	0.713	218	5.190	0.988	0.393	0.008	3.95	12
G9	56	0.718	136	3.238	0.983	0.361	0.008	3.23	3
G10	137	0.665	220	5.238	0.981	0.440	0.011	4.67	17

*alleles < 1% frequency in the whole collection are considered rare alleles

4.3.5 Eco-geography and spatial genetics

In spatial genetic analysis significant relationship was found between the genetic distances of accessions and the eco-geographical parameters of their site of origin. The molecular genetic distance between the accessions is significantly correlated to their geographical distance followed by latitude differences. The Mantel test between the genetic distance matrix (shared allele distance matrix) and geographical distance matrix displayed a significant Mantel correlation of 0.357 ($P < 0.0001$). The correlation between genetic distance and longitude differences ($r=0.305$, $P < 0.0001$) was high. The correlation between genetic distance and latitude differences ($r=0.193$, $P < 0.0001$) was 2-fold lower than the correlation between genetic and geographical distances. However, the correlation of climatic parameters to the genetic distance has displayed low correlation. Annual mean temperature (AMT) has shown significant positive correlation with the genetic distance matrix. Though the correlation was lower it is highly significant ($r=0.189$, $P < 0.0001$). Similarly, mean diurnal temperature range (MDR) and also maximum temperature of warmest month (MTW) displayed low positive correlation with high significances (for MDR $r=0.177$, $P < 0.001$; for MTW $r=0.158$, $P < 0.001$). Significant but low correlation for annual precipitation (APT) with genetic distance was detected ($r=0.100$, $p < 0.001$). The same pattern was observed for the spatial Mantel correlograms, wherein different distance classes displayed different Mantel correlations of high to low significant values (**Fig 4.3.8a- 4.3.8g**).

In the Mantel correlogram between genetic distance and geographic distance matrix, the matrix is subdivided into 20 discrete distance classes and the correlation between the matrixes in these distance classes were evaluated. The r-value (Mantel correlation) is higher in the class with geographically closer individuals than in the other classes (**Fig 4.3.8a**). Within a distance class of 300 km between the accessions, the correlation was found to be high (r=0.523). The r-values and their significances also declined with the increasing distance classes (**Supp Table 4.2a**). Spatial Mantel correlograms between the genetic distance and longitude also displayed similar pattern for the different longitude classes (**Fig 4.3.8b**). The matrix is subdivided into 9 discrete longitude difference classes and Mantel correlations within these groups were estimated (**Supp Table 4.2b**). The first class with a difference of 10º longitude showed high and significant correlation (r=0.304). Mantel correlograms between the genetic distance and latitude also displayed similar pattern for the different latitude classes (**Fig 4.3.8c; Supp Table 4.2c**).For AMT the spatial correlogram was evaluated based on seven classes and only four classes showed significant r values and the remaining classes displayed very low significances (**Fig 4.3.8d; Supp Table 4.2d**). Spatial Mantel correlogram for MDR and MTW also showed similar pattern (**Fig 4.3.8e & 4.3.8f; Supp Table 4.2e & 4.2f)**. Mantel correlogram between genetic distance and annual precipitation (APT) was assayed based on 10 classes. The pattern of Mantel correlations within in these classes is similar to the trend observed for annual temperature; correlations in all the classes were significant but displayed very low r-values (**Fig 4.3.8g, Supp Table 4.2g**).

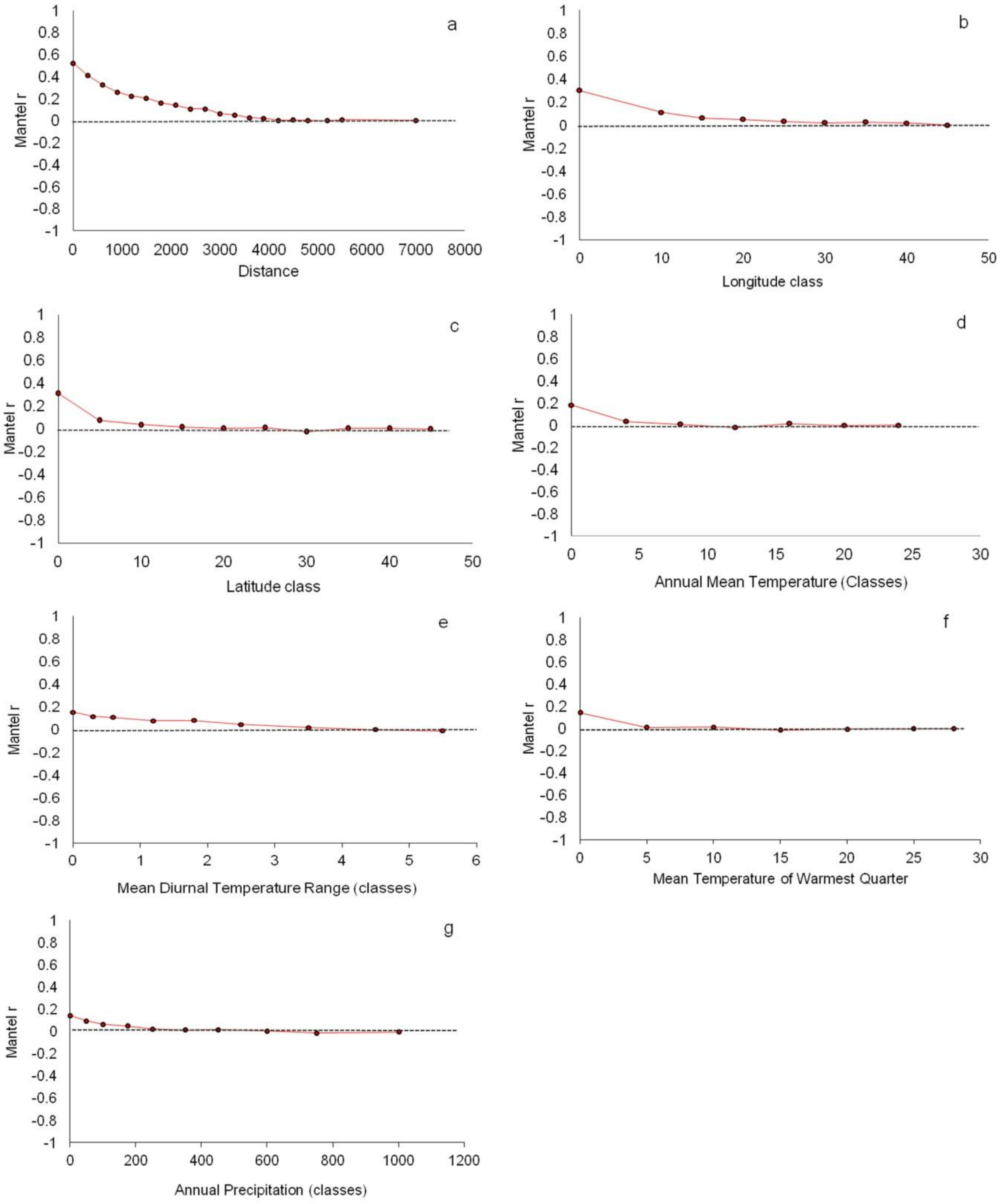


Fig. 4.3.8: Spatial genetic patterns observed among the 1491 landraces. Correlogram showing spatial genetic autocorrelation between: (a) genetic distance and geographical distance. (b) genetic distance and longitude (c) genetic distance and latitude (d) genetic distance and annual mean temperature (e) genetic distance and mean diurnal range (f) genetic distance and mean temperature of warmest quarter (g) genetic distance and annual precipitation.

4.4 Discussion

Knowledge on genetic diversity and population structure is of great importance for ongoing crop improvement efforts. Emerging genome wide association studies (GWAS) are considered as an alternate approach for QTL detection in comparison to traditional QTL mapping . Genetic diversity and population structure information are imminent prerequisites for GWAS (D’hoop et al., 2010). Several studies provided an insight into genetic diversity of barley cultivars, landraces and wild collections (Hubner et al., 2009; Malysheva-Otto et al., 2006). However, in case of barley landraces, studies were mostly confined to a limited number of accessions sampled from specific geographical regions (Leino and Hagenblad, 2010; Pandey et al., 2006; Russell et al., 2011; Yahiaoui et al., 2008). Due to the worldwide distribution of barley, evaluation of genetic diversity among the germplasm from a large geographical area encompassing different countries can provide a better understanding of the diversity, patterns of distribution, inter-regional seed exchange and admixtures, evolution and domestication. The primary aim of this study was to explore the diversity and population structure in barley landraces of wide range of origins and investigate their genetic distinctness in relation to their distribution and eco-geographical relationships. Subsequently, the genetic diversity of landrace collection was compared with already reported cultivar panel (Chapter 2 & Chapter 3) to assess the usefulness of landrace collection for association mapping studies. This information can help in further efficient utilization of this landrace collection for developing core groups, for selecting germplasm, for allele mining approaches and for GWAS. In this study landraces from 41 countries encompassing varied climatic zones were considered (**Table 4.2.1**). Our results furnish the presence of strong population structure in the collection, and indicate significant relationships between the genetic distinctness and patterns of eco-geographical distribution.

4.4.1 Genetic diversity, population structure and geographical distribution

4.4.1.1 Genetic diversity

A collection of 1491 barley landraces was characterized with 45 SSR loci distributed across seven chromosomes. Three-hundred-seventy-two alleles were detected using 42 SSR markers with an average of 8.85 alleles per SSR locus. These figures are comparable to the results from the studies of Varshney et al. (Varshney et al., 2010) using 223 cultivars and wild barleys of worldwide origin (average allele number =7.7; PIC=0.6). The observed average allele number (AN) is higher than the allele number reported in Eritrean landraces (AN=7.6) by Backes et al. (Backes et al., 2009) and in Himalayan landraces (5.54) by Pandey et al (Pandey et al., 2006). Higher average allele number values were reported for a worldwide collection of cultivars (AN=16.7) (Malysheva-Otto et al., 2006) and for Syrian and Jordanian landraces (AN=11.6) (Russell et al., 2003). The average number of alleles per locus depends on the existing genetic diversity, population size and apparently on the selected marker set. Comparison of diversity statistics among the same markers in different populations will provide a more sensible conclusion.

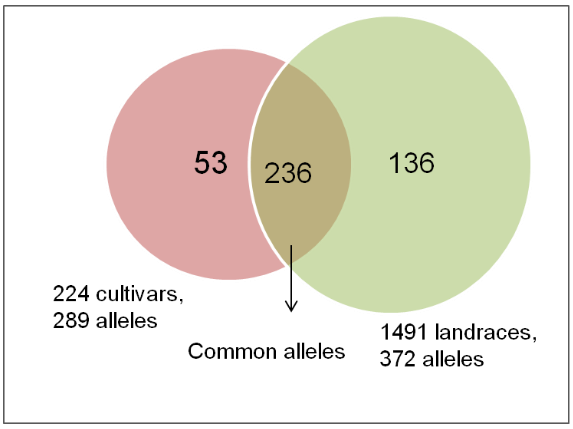


Fig. 4.4.1 Comparision of allele richness among 224 spring barley cultivar collection and 1491 spring barley landrace collection using 42 SSR markers

The same set of SSRs were used to study genetic diversity in a worldwide collection of 224 spring barleys (Haseneyer et al., 2010b). The genetic diversity of the 224 spring barley worldwide collection was compared with the landrace collection. Number of alleles detected across 42 SSR markers in the 224 collection was 289, while 372 alleles were detected in landraces. Among the total 425 alleles, only 236 alleles were common in both the collections. We detected 53 unique alleles for 224 collection and 136 alleles unique for landrace collection (**Fig 4.4.1**). The large number of unique alleles observed in this landrace collection when compared to the diverse worldwide collection of spring barley indicates the larger diversity and allelic richness present in the landrace collection. A total of 152 rare alleles were detected in the whole landrace collection. Rare alleles were mostly detected at SSR loci which displayed an above average number of polymorphic alleles. More than 40% of the detected alleles are rare alleles in our collection, which indicates their usefulness for plant breeding and genetic research purposes. However, these results require further studies for accurate estimation of allelic frequencies and their specific effects. Among the rare alleles detected, 81 alleles (53%) are group specific rare alleles (**Table 4.3.4**). The amount of admixed accessions (13%) observed during STRUCTURE analysis, indicates there is genetic exchange and gene flow between closely related groups and between two-rowed and between six-rowed barley groups from same geographical origin. The average PIC value for overall 42 SSR loci was 0.55 and is comparable to that observed by Haseneyer et al. (PIC=0.54; 2009). The average gene diversity (GD) of 0.603 was observed which is high compared to other studies (Backes et al., 2009; Castillo et al., 2010). As anticipated, very little heterozygosity was observed in the whole collection.

4.4.1.2 Population structure

In order to ensure the reliability of the inferences made regarding structure in the collection, various statistical approaches were employed to determine the population structure. All the approaches have given nearly similar results. STRUCTURE analysis divided the collection into k=10 groups (**Fig. 4.3.3**). As noted in previous studies, our observations showed that Ethiopian barley are different in comparison to all the other accessions and always grouped separately (Negassa, 1985; Tanto Hadado et al., 2009). The different evolutionary and domestication history of barleys from Ethiopia offers a plausible explanation for the observed differences (Orabi et al., 2007; Saisho and Purugganan, 2007). Most of the Ethiopian barley were divided into two groups constituting naked barley (G1) and hulled barley (G5). The row type in barley is also an important determinant of the population structure (Haseneyer et al., 2010b; Malysheva-Otto et al., 2006; Pourkheirandish and Komatsuda, 2007). In our studies, except four groups all the other groups constituted either two-rowed or six-rowed barley. The groups with naked barley (G1 & G9), Georgian (G4) and Ethiopian hulled barley (G5) constitute both two-rowed and six-rowed barleys and were majorly structured by their geographical origin. The collection showed clear geographical structuring (**Fig.4.3.1**) with the differentiation associated with geographical distance and latitudinal differences.

The STRUCTURE results were further ascertained by the results from hierarchical cluster analysis (**Fig. 4.3.5**) and PCA (**Fig. 4.3.6**). The cluster analysis results were in accord to the STRUCTURE inferred groups. The results from PCA showed overlapping of some of the groups and represented a blurred distinction among the groups. Especially the groups G3 and G10, G2, G6, and G9, G4 and G7 show greater overlap and are distributed across the axes. The large dispersal of these groups along the axes, suggests gene flow between the overlapping groups. This is evident from the higher percentage of admixed lines present in these groups. Initially, when the accessions were assigned to the groups without any threshold limit of membership coefficient, groups G7 (26%), G10 (21.5%), G6 (17.5%) and G4 (15.2%) constituted more admixed accessions (**Table 4.3.1**). The average membership coefficient values over all accessions in a group were also low for the groups G7, G6, G3, G2 and G10, ascertaining the presence of admixes. However, PCA does not classify accessions into discrete populations in all cases, especially when admixed accessions and accessions of various geographical origins with a constant gene flow are included in the collection (Patterson et al., 2006). Considering the collection size and large geographical range of this study, groups defined by our analysis should not be considered as populations. The groups for k=10, was optimum number of groups detected in the collection, and there is possibility of further structuring within these groups. These groups still contain a regional level of genetic differentiation comprising more closely related sub-groups or populations. In order to illustrate the substructuring within the groups, individual PCA for each group were investigated (**Fig. 4.3.7a to 4.3.7j**). Interestingly, the PCA plots for some of the groups showed no further structure. Some groups showed subtle structuring and some other groups a distinct structuring within the groups. For instance, group G1 with all the naked Ethiopian barleys suggests subdivision into two-rowed and six-rowed barley. But the distinction is not clear and the accessions are dispersed along the axes indicating strong gene flow among these row type sub-groups in Ethiopia. The group G2 displayed higher level of distinctness between the subgroups. The primary axis separated Libyan landraces from Iranian and Iraqi landraces. The Iranian and Iraq landraces are dispersed along the secondary axis suggesting the gene flow between these two groups and also indicates the higher diversity among these accessions compared to Libyan subgroup (**Fig. 4.3.7b)**. Group G4 shows two distinct sub-groups of two-rowed and six-rowed accessions. Group G9 with all non-Ethiopian naked barley accessions showed two distinct sub-groups (**Fig. 4.3.7i)** dividing into two-rowed naked and six-rowed naked barley. The remaining PCA plot does not reveal any further sub-structuring patterns in the groups.

4.4.1.3 Population differentiation and geographical distribution

To further explore the genetic diversity and relationships among the groups and within the group, various diversity statistics were assessed for each of the STRUCTURE inferred groups (**Table 4.3.4**). The GD over 42 loci was high for the groups G7 (0.48), G6 (0.475) followed by G2 (0.47), G3 (0.46) and G10 (0.44). These were the groups with high number of admixed accessions, indicating the relationship between GD and the number of admixed accessions in a group. The observed GD for these individual groups was comparable to the GD values of various collections from previous studies (Backes et al., 2009; Castillo et al., 2010). Interestingly, lower GD values were observed for Ethiopian barley groups G1 (0.27) and G5 (0.25) followed by Georgian landraces G4 (0.295). Previous studies also showed that lower diversity was revealed by nuclear SSR markers in Ethiopian landraces due to the selection pressure during their independent domestication period. Chloroplast SSR markers revealed greater diversity with same population as they are less influenced by selection pressure (Orabi et al., 2007). Group G4 consists of accessions from a region of narrow geographical range in Georgia. The number of alleles detected and group specific rare alleles observed were also low for this group (**Table 4.3.4**). G8 showed a low GD when compared to its sample size, apparently because most of the accessions are two-rowed barleys from Slovakia (145) and remaining are from Europe. The two-rowed barley from Europe has shown low diversity in previous studies (Chapter2) (Liu et al., 2000; Pasam et al., 2012). The number of alleles detected was high for G3 (252), G10 (220) and G8 (218).

The number of alleles detected in a group is biased and depends on the sample size of the group. Therefore, allelic richness values based on rarefaction method were detected for each group and compared. Allelic richness values were high for G3, G7, G6, G2 and G10 groups. Interestingly, though the sample size was small for groups G6 and G7 an impressive number of alleles 206 and 213 were detected. Group specific rare alleles detected for the groups G6 (10) and G7 (13) were also high, emphasizing the presence of large diversity. The accessions in these groups are majorly from Iran, Iraq and Afghanistan. The high diversity observed in these accessions may be due to: 1. Eco-geographical diversity of the origins of sampled accessions (**Fig 4.3.1**) 2. Seed exchange and gene flow between the accessions in this region both from Fertile Crescent areas and from Asian regions like Tibet, Nepal and China. The theories about multiple barley domestication sites apart from Fertile Crescent have been proposed in past studies. Morrell and Clegg (Morrell and Clegg, 2007) proposed a secondary domestication site of barley somewhere 1500-3000 km east of Fertile Crescent and indicated greater allelic differences between cultivated Western barleys and Eastern barleys. The differences between eastern and western origin barley were also indicated in various other barley diversity studies (Haseneyer et al., 2010b; Saisho and Purugganan, 2007). The land locked regions between these centers of diversity, might had a free seed exchange and gene flow from both directions that have accumulated greater allelic diversity. These regions (Iran, Iraq, and Afghanistan) are covered by the ancient land route of the silk road that connects ancient Anatolia (ancient Turkey) and China. Probably the exchange of seed materials and introduction of new alleles might have been facilitated by this land route in both directions (Taketa et al., 2004). As anticipated, barley from Eastern regions grouped separately from the accessions from Western geographical regions, with some exceptions though. The six-rowed landraces from south of Libya sampled from oasis were grouped together with the landraces from Iraq and Iran (Group G2). And also few landraces from Turkey were found to be grouped with the landraces from Iran, Iraq and Afghanistan regions (Group G7). The ancient trade routes and the migrating populations between these regions due to various socio-political reasons might have resulted in the exchange of germplasm.

Analysis of molecular variance (AMOVA) among and within groups was calculated based on various setups that included two-rowed vs. six-rowed barley, Ethiopian vs. non-Ethiopian barley, naked vs. hulled barley, hulled barley vs. naked non-Ethiopian, and STRUCTURE inferred groups (**Table 4.3.4**). Due to different domestication and breeding histories of both two-rowed and six-rowed barleys, row type is one of the primary determinants in structuring of barley (Pourkheirandish and Komatsuda, 2007). The division of the accessions based on row type into two groups explained minimum variation (8.75%) between the groups. It indicates the presence of large variation and structure within these groups. Further structuring in the row type groups is evident from the geographical structuring we see within these large groups. Due to the large geographical range of our collection, besides row type, geographical distances also play an important role and affect the allelic distribution and frequencies.

AMOVA for hulled and naked barley groups explained relatively higher variation (15.75%) between the groups. This large explained variation might be of two probable reasons. Firstly, the naked barleys were selected for food purposes by farmers and are mostly distributed towards East Asian regions (Pandey et al., 2006) and Ethiopia. Majority of naked barley is adapted to part of the distribution range only. Naked barley is believed to be of monophyletic origin and later on migrated to other parts of the world (Taketa et al., 2004). The domestication and selection process of naked barley was independent of other hulled barley types with exceptional introgressions between the two types resulting in the formation of distinct groups and less gene flow. Similar distinct groups of naked and hulled barley were observed by Strelchenko et al. (Strelchenko et al., 1999). Secondly, most of the naked barley accessions in our collection are from Ethiopia (199) which can result in the biased estimation of variation. Therefore, we estimated the molecular variation between the non-Ethiopian naked barley (87) and all hulled barleys. The explained variation between these groups was still high (12.33%), confirming the distinctness between the naked and hulled barley types (**Table 4.3.4**). Group G9 representing the non-Ethiopian naked barleys is distinct from all other groups in the cluster analysis of groups (**Fig. 4.3.5**). AMOVA between the Ethiopian origin barley and non-Ethiopian origin barley displayed a higher level of variation between the groups (19.61%). In order to avoid the biased estimation due to large number of naked barley in Ethiopian collection, AMOVA was conducted between total hulled barley and Ethiopian origin hulled barley. Variation explained between hulled barley and Ethiopian origin hulled barley was still higher (16.99%). This is also evident from the distinct structuring of the Ethiopian lines in PCA and also in cluster analysis of groups (**Fig. 4.3.5**). As discussed above and in previous studies the distinctness of Ethiopian barley is well proclaimed (Orabi et al., 2007; Saisho and Purugganan, 2007). The explained variation between the STRUCTURE inferred groups was substantially higher (36.62%), indicating the presence of strong population differentiation. This larger variation between groups showed they are significantly distinct. From the distinctness observed between the STRUCTURE inferred groups, it can be concluded that the for structure *k*=10 the groups are significantly distinct.

Further interesting feature of distinctness and relationship between these groups was portrayed by pair wise comparisons of F_{st} values (**Table 4.3.3**). Fixation indices (F_{st}) measure the amount of differentiation among subpopulations derived from the subdivision of an original population. F_{st} values range from 0 for non-differentiation to 1 for complete differentiation between an original population and its subpopulations (Wright, 1951) and values above 0.25 indicate great genetic differentiation. Fixation statistics (F_{st}) for the whole collection when STRUCTURE inferred groups were compared was high (0.36). Large F_{st} values ($\geq$0.4) were found between G1 and all the other groups except with G5. Apparently, as both groups G1 and G5 represent Ethiopian barleys they are much closer. Interestingly the F_{st} between G1 (Ethiopian naked) and G9 (non-Ethiopian naked) was high (0.476) indicating greater differentiation between these groups. This suggests that the Ethiopian naked barleys are distinct from the other naked barleys. Lowest F_{st} (0.187) was observed between group G4 and G7, probably because of the closer proximity of the geographical origins of these two groups. F_{st} value (0.251) was observed to be low between the groups G6 and G7. Although, the groups G6 and G7 represent different row type barleys, they originate from the same geographical regions (Afghanistan, Iran, Iraq and Turkey) hence the groups showed less distinctness.

4.4.2 Eco-geographical factors and spatial genetics

Knowledge about the available eco-geographical variables for each of the accession will help in determining the most relevant climatic variables influencing the genetic differentiation. This expert knowledge will help in selecting diverse accessions to investigate environmental effects on various agronomic and physiological traits. As our accessions are selected from diverse climatic regimes, the impact of the micro and macro environment on the adaptation behavior of the accessions can be addressed from a global perspective. The distribution of our accessions projected over the layers of climatic data like annual mean temperature (AMT), and annual precipitation (APT) is visualized in **Supp Fig 4.3 and 4.4**. Though altitude was considered as an important factor influencing the genetic distribution (Tanto Hadado et al., 2010), this variable is not considered in our studies. For prediction of accurate altitudes from Geographic Information Systems (GIS) the required exact geographic coordinates, were not available for some of the accessions. The 10 arc min grid (18.6 km) used to predict environmental variables is fairly acceptable for climatic factors (Hijmans et al., 2005a) but it would be too risky to predict altitude at this resolution. Hence altitude cline of variation in the genetic differentiation and adaptation behavior was not investigated. In our studies the genetic distance was found to be significantly associated with the geographical distance (0.357), longitude (0.305), latitude (0.193) and other climatic factors. Especially for accessions within closer distance classes, the correlation to genetic distance was high (0.523) and as the distance class increased the correlation declined and their significance reduced (**Fig. 4.3.8a**). Similar pattern was observed for correlations of genetic distance with longitude and latitude respectively (**Fig. 4.3.8b, Fig. 4.3.8c)**. These results suggest the presence of stronger local adaptation in accessions as the correlation was high in accessions of closer distance classes. Significant Mantel correlation between geographic distance and genetic distances were reported in previous studies for Ethiopian barley by Tanto Hadado et al. (Tanto Hadado et al., 2010) and for flax by Uysal et al. (Uysal et al., 2010). The latitudinal cline of the genetic variation suggests the photoperiod adaptation of the accessions to the various latitudinal regimes. The Mantel correlations for genetic distance and the climatic factors like AMT (**Fig. 4.3.8d**), MDR (**Fig. 4.3.8e**), MTW (**Fig. 4.3.8f**) and APT (**Fig. 4.3.8g**) were low but were still significant. However, considering the large geographical area of origins of the collection and their diverse climatic profiles, it would be ambivalent to draw any strong interpretations from these correlograms. A further investigation with accessions from a limited geographical area in detail is required to derive at a conclusive interpretation of these results. Regardless, it is obvious from the correlograms that the climatic factors showed a lower but still significant association with genetic distances. This points the impact of the climatic factors on the adaptation of the barley germplasm. Even from the genetic structuring of our collection the effect of geographical and climatic conditions on the genetic distribution is evident. Six-rowed barley is majorly grouped into G3, G10, G2 and G6. The G3 group represents the six-rowed type barleys from Mediterranean regions of northern Libya, Morocco, Iberian Peninsula, Greece and south of Italy. The six-rowed barley from Northern latitude regions like Austria, Bulgaria, Sweden and Northern Turkey are included into group G10. Impact of climatic factors on the genetic distribution was explored in previous studies (Yahiaoui et al., 2008) and temperature was determined as the major climatic factor followed by annual precipitation that affected the distribution of alleles in Spanish barley landraces. Also in our studies AMT showed a higher correlation (0.189) with genetic distance than the other climatic factors did.

4.5 Conclusion and prospects

Our results revealed the accessible allelic diversity present in the large collection of landraces, the inherent population structure and distribution in relation to the domestication and eco-geographical factors. The statistical analysis revealed high allelic richness and diversity present in the collection and its significant correlation to eco-geographical factors. The association between environmental data and the genetic diversity patterns in landraces provides an interesting scenario for understanding barley distribution and adaptation to the local environments.

Beyond providing insights into the diversity, our data will allow to construct core groups based on maximizing allelic diversity approaches. The core groups can be constructed for various purposes by maintaining the allelic richness and giving weightage to the relevant factors or traits. As an example, we generated association mapping panels of different sizes (200, 400, 600 and 800) for heat stress studies by using molecular data for maximizing allelic diversity and giving additional weightage to AMT and MTW climatic factors for selection of the lines. Core group construction strategy of M-STRAT was used for selecting the lines (Gouesnard et al., 2001). The panels comprised lines that were well distributed across all temperature regimes and also the allele diversity was high for all panels (**Table 4.4.1**). This example demonstrates that eco-geographical data can be used to predict the agronomic (Endresen, 2010) and adaptive traits of the accessions and will aid in selection of the lines for diverse small sized association mapping panels. This collection can also be used for allele mining, as already our results indicate high allelic diversity in the collection. It is however, important to note that the presence of rare alleles in the association mapping panel might result in false associations and hence are excluded from the analysis (Comadran et al., 2009; Myles et al., 2009). This concern could be overcome to an extent by careful trading between selectivity and sensitivity of the analysis by fixing an optimal threshold for allelic frequencies that could be used in association studies.

Table 4.4.1 Comparison of the diversity statistics for different sizes of core groups generated for heat adaptation from 1491 accessions using marker data and climatic variables.

	Group size					
Accession number	200	400	600	800	1000	1491
Allele number	8.167	8.619	8.714	8.857	8.857	8.857
GD	0.613	0.608	0.608	0.606	0.606	0.606
PIC	0.561	0.554	0.553	0.551	0.551	0.551
MAF	0.506	0.510	0.506	0.508	0.509	0.512

The presence of rare alleles and group specific rare alleles in the collection suggest their potential to provide useful alleles for further plant breeding efforts. Population structure is one of the major limiting factors for association mapping as it results in more false positive associations (Zhu et al., 2008). We assessed population structure by different approaches and detected strong genetic structure in the landracecollection, indicating the need to account for structure in association mapping studies. Several statistical models have been developed for controlling the population structure caused spurious associations (Kang et al., 2008; Yu et al., 2006). It is anticipated that the staggering patterns of linkage disequilibrium exhibited in different genepools (Caldwell et al., 2006; Tenaillon et al., 2001) in combination with the phenotype and ever increasing marker resources, would enable identification of the genes and QTL underlying complex agronomic and adaptation traits (Waugh et al., 2009). In this context the utilization of this landrace collection for association mapping studies can aid in fine mapping of genes, as LD decay is presumed to be fast in landrace collections. Using a stringent statistical model to correct for population structure, our collection can be efficiently used to detect meaningful marker trait associations useful for marker assisted breeding approaches. Here in this studies LD decay was not calculated using the 42 SSR marker data. The marker density is too low and would result in overestimation of LD hence we avoided LD studies using this data.

CHAPTER FIVE: Summarized discussion and outlook

5.1 Summarized discussion

The findings described within this thesis indicate the potential of LD mapping for exploiting genetic diversity and variation present in spring barley for localizing QTL that could be used for breeding purposes. In the thesis, the genetic and phenotypic diversity of one association mapping panel comprising worldwide spring barley cultivars (224) was investigated. One of the objectives was to gain insight into the complex genetics underlying the phenotypic variation of agronomic traits by localizing the corresponding QTL. Besides detecting several QTL for each of the traits, patterns and extent of LD in the panel was also investigated. The recent development of high throughput genotyping methods will have a significant impact on the fundamental and applied research in crop species like barley . The thesis investigates the advantages of increased marker density in whole genome association mapping approaches. Furthermore, in anticipation of fine mapping of QTL using staggering pattern of LD decay in different genepools, we established a spring barley landrace association mapping panel. The landrace collection was studied for genetic diversity, genetic relationships and population structure using 42 SSR markers. This thesis presents work both in QTL detection by association mapping methods and large scale diversity studies in barley.

Chapter 2 is a QTL detection study using spring barley cultivar panel with IPK OPA SNPs (918 successful SNPs) by a GWAS approach. The success of the association mapping depends on the choice of the mapping panel (Myles et al., 2009). Therefore, the spring barley association panel was studied extensively for population structure and linkage disequilibrium patterns using SNPs across the whole genome. In first place, the spring association panel (224) constituted accessions that were carefully selected from a barley core collection (BCC) and then complemented with additional accessions from the entire distribution range and maximize diversity. Strong population structure was detected in this population, as the accessions clustered into six groups based on their spike morphology and geographical origins (**Fig.2.2.1**). As construed from several previous studies, population structuring is a general feature hitherto seen in most of the plant populations (Comadran et al., 2011; Zhu et al., 2008). To abstain from the spurious associations caused by LD due to structure, the aspect of population structure need to be understood and adjusted by adequate statistical corrections.

Candidate gene association studies were reported successfully using this collection (Haseneyer et al., 2010a; Haseneyer et al., 2008; Stracke et al., 2009). We now performed genome wide association studies for five major agronomic traits (HD, PHT, TGW, SC and CPC) previously analyzed in the field. As a starting point for GWAS, it is important to gain good knowledge about patterns and extent of LD in the panel to design and conduct unbiased association mapping (Mather et al., 2007). LD was observed to decay below a critical level within a map distance of 5-10 cM in our panel. The extent and distribution of LD varied across the genome and also across the sub-groups. LD extent was observed to be larger in the sub-groups than in the whole population. In the two-rowed and six-rowed barley groups LD decayed within 10-15 cM while in the subgroups LD extended beyond 20-25 cM. The LD detection power is biased by the number of markers used, as average LD decay with few markers ranged from 10 cM using AFLPs (Kraakman et al., 2004) to 50 cM with SSRs in another worldwide barley collection (Malysheva-Otto et al., 2006). When denser marker coverage was used, LD decayed below 5 cM for a germplasm collection from particular region (Rostoks et al., 2006; Zhang et al., 2009). The subgroups of these collections showed extended LD (Rostoks et al., 2006), emphasizing that the LD depends on recombination and selection pressure in the population besides marker number and population structure. Estimates of average LD across the genome are often used to predict the required number of markers and accuracy of the GWAS (Comadran et al., 2009). These estimates do not take into account the dynamic and extremely variable pattern of LD across the genome (Hamblin et al., 2010). The patterns of LD observed here indicate that the panel can be used for GWAS with the available modest marker coverage (918 IPK-OPA SNPs) to detect QTL. However, the resolution of mapping can be increased by filling the marker gaps with increased marker coverage.

The accuracy of GWAS was evaluated by mapping all 918 SNPs using LD by a GWAS approach. Interestingly, more than 85% of the markers mapped within 0-10 cM of the original map position and 80% of SNPs mapped within 5 cM. This suggests that resolution of the panel is approximately between 0-10 cM. For further fine resolution mapping, either the maker density needs to be increased or larger population with faster LD decay across the genome can be used. The same panel was investigated for GWAS results sing an increased marker density (iSelect) in chapter 3. The impact of the population type on LD is obvious and it is known fact that LD decays rapidly in wild and landraces and slowly in cultivated varieties (Caldwell et al., 2006; Gouesnard et al., 2001). Thus, we established a large panel of spring barley landraces to increase the resolution and power of QTL detection. The details of the panel and landrace diversity are discussed in chapter 4.

In an attempt to obtain a statistical model which best fits for GWAS in our panel, we evaluated different General Linear Models (GLM) and Mixed Linear Models (MLM) proposed for correcting population structure in structured populations. Several statistical models were presented in the past using Q-matrix, PCA and Kinship matrix for correcting structure in GWAS (Kang et al., 2008; Yu et al., 2006). The models QK (with Q-matrix and kinship), PK (with principal components and kinship matrix) and K (with kinship matrix) performed best in our panel and showed a good fit for P-value distribution (**Fig. 2.3.7**). Henceforth, only the K model which showed best fit and is time saving for our further analysis is used.

Another aspect that was discussed is the number of markers needed for estimation of the kinship matrix. Several studies recommended to use randomly distributed markers for population structure estimation (Falush et al., 2003), but studies on marker requirement for kinship estimation are limiting (Yu et al., 2009). For association analysis with IPK-OPA markers, the whole set of SNPs were used for kinship estimation rather than a selected subset of markers. The genome coverage by IPK-OPA SNPs is only modest and gaps without marker coverage existed along the genome. Hence, randomly selected markers would further bias the kinship estimation due to large unequal gaps. Therefore, the whole set of markers was used estimate kinship rather than selecting a subset. Nevertheless, with the use of iSelect assay, the SNP number increased multifold and consequently the marker coverage was improved. The use of all markers (7000 SNPs) to estimate kinship would result in biasness due to the over representation of certain regions by more markers. Besides, the use of kinship generated from all markers for GWAS resulted in an overkill causing many false negatives. It was reported that for population structure estimation in barley, 384 randomly selected markers are optimum requirement (Moragues et al., 2010). We devised a comparison study using different marker sets (n=6467, n=918, n=362) for Kinship estimation (K_1, K_2, K_3) and their impact on the results of GWAS. The 362 markers were carefully selected at equidistant across all seven chromosomes to avoid any possible biasness. The K_3 matrix successfully captured similar diversity as captured by K_1 and also controlled spurious associations effectively.

Using IPK-OPA SNPs, a total of 205 marker trait associations (MTA) were detected for the traits row type (RT), heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and crude protein content (CPC). These SNPs were grouped into QTL based on LD. Several of these QTL regions were concurrent to the previously reported QTL regions for the respective traits. For the trait RT significant associations were observed in the regions of *vrs3*, *vrs1*, and *int-c* locus. Similarly for HD, significant associations were observed in the regions of *Ppd-H1*, *HvFT1*, *HvCO1* and *eam6*. However, the observed significances and the variance explained by the markers were not high. Low variance explained by the marker in GWAS is universal and can be attributed to several causes. When we included SNPs from *Ppd-H1* gene into the analysis, the SNPs from gene showed higher association to HD than any other SNP included in the assay (**Fig 2.3.10**.). This confirms that the chance of detecting association with a SNP increases with its proximity to the causual SNP. These findings emphasized the need of further improvement of genome coverage for accuracy and power of QTL detection.

Consequently, the current spring barley panel was genotyped using the newly established iSelect assay (Comadran et al. in prep), which yielded 7000 successful SNPs. After removing SNPs with MAF less than 5%, 6467 SNPs were for GWAS. It was anticipated that the improved genome coverage will help in increasing the power of QTL detection and for further fine mapping the QTL (Yu et al., 2011). We observed multifold increase in the number of SNPs associating with the trait, and some of these markers associated with a very high significance. When compared to the rice syntenic loci, some of the significant SNPs were only few gene models away from the candidate gene. Several SNPs associated to the traits in a certain regions is mainly due to the LD extent in the population. Hence, these SNPs were grouped into probable QTL regions and some of these regions were confirmed by QTL reported in past from linkage mapping studies. QTL detected by GWAS and also confirmed from other studies are promising candidates for further studies by fine mapping using traditional mapping approaches or using joint linkage mapping approaches (Brachi et al., 2010; Buckler et al., 2009) or by using different association mapping panels with high resolution. Generally, association panel with LD decay at shorter distances and with good genome coverage can be suitable for high resolution mapping (Myles et al., 2009; Waugh et al., 2009).

As noted previously, LD decays rapidly in wild and landraces and slowly in cultivated varieties (Caldwell et al., 2006; Gouesnard et al., 2001). The role of different selection pressures and domestication bottlenecks resulted in the varied pattern of LD in the crop genepools. Hence a landrace population with fast decaying LD and sufficient marker coverage can be a choice for further fine mapping of QTL. Besides, such a collection can be exploited to mine new alleles that may be successfully used in crop improvement. Our goal was to select spring barley landraces from varying eco-geographical regions and to establish a diverse well representative association mapping panel. The barley from East Asian regions were not included in our present landrace collection. Collection sites extend from 5º N to 62.5º N and 16º W to 71º E. The accessions (1491) were selected based on the nomenclature and morphological descriptions available from the genebank database. The study of the diversity and population structure is the primary step for assessing the feasibility of using the collection for different crop improvement purposes. Molecular analysis using 42 SSRs has shown considerable genetic variation in the landraces. Using SSRs, a total of 372 alleles were detected and among them 152 are rare alleles (allele frequency < 1%). The collection is diverse with an average gene diversity of 0.60% and with average allelic richness of 5.74. The collection showed strong population structure with 10 subgroups (K=10). Valuation of genetic diversity among the germplasm from a large area encompassing different countries provides an understanding of the diversity, patterns of distribution, inter-regional seed exchange and admixtures in global perspective.

The association between environmental data and genetic diversity in landraces provided an interesting scenario for understanding barley distribution and adaptation to local environments. This eco-geographical data can be used to predict the agronomic (Endreson 2010) and adaptive traits of the accessions and will aid in selection of the lines for diverse smaller sized association mapping panels. We generated different core groups for heat adaptation studies (with annual mean temperature (AMT) and annual precipitation (APT) as weighted variables) from the whole collection and compared the genetic diversity among these groups. Interestingly, we found similar allelic richness for core groups above the size of 800 accessions. Smaller core groups showed lower number of alleles per loci (**Table 4.3.1**). We simulated groups ranging from size n =1, n =49 … to n =1491 with a difference of 50 accessions, and measured the diversity over 5 replications using MSTRAT (Hamilton et al., 2002). The average diversity scores were plotted with respective group size and the optimum size of the panel for capturing all alleles present in the panel was determined to be n =745 (**Fig.5.1**). For the group size n =745, the diversity score was similar to the diversity score of the whole collection and as the number of accessions decreased the score also declined. We also compared the random sampling and MSTRAT sampling approach. The allele maximizing strategy of MSTRAT performed better in capturing the alleles rather than random selection of the accessions. Based on these results it can be concluded that association mapping panel developed from this collection with size anywhere between n=650 to n=745 would capture the maximum diversity of whole collection and could be used for GWAS with proper statistical corrections for population structure.

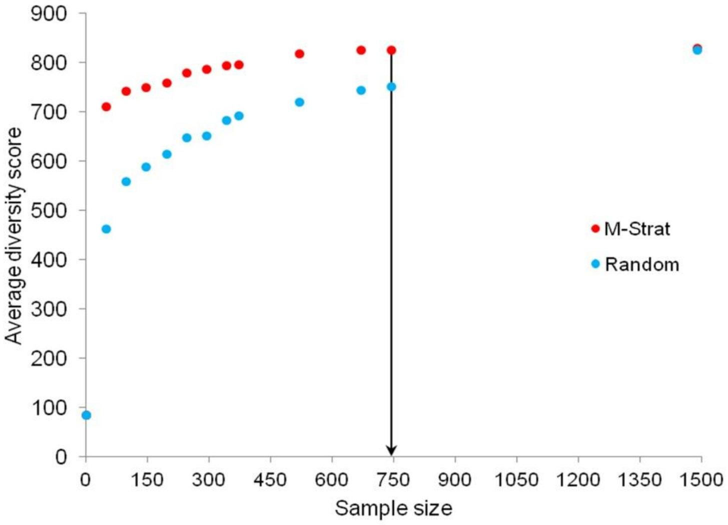


Fig. 5.1 Sampling efficiency based on MSTRAT strategy and random sampling to capture the diversity in the collection. Diversity score calculated based on allele richness plotted against the size of each core group. Red circles indicate score of the core collection by MSTRAT and blue circles indicate score by of the random selected accessions

GWAS is a successful approach for QTL detection in barley. However, it is necessary to validate the numerous small effect QTL detected in GWAS either using different association populations or using biparental populations (Atwell et al., 2010). GWAS in synergy with linkage mapping studies can effectively validate the QTL and identify the genes. Moreover, the emerging resources like nested association mapping (NAM) populations and multiparent advanced generation intercross (MAGIC) populations established with designed genetic structure from diverse parents are evolving as ideal resources for QTL validation and gene identification in plants (Yan et al., 2011). The power of association studies is determined by the size of population, the trait, density of markers used, LD and the population structure in the population and the statistical approaches used (Myles et al., 2009). Increasing the number of accessions in the association mapping population can have substantial effect on the power of QTL detection. Therefore, accessions form the landrace collection after efficient phenotyping and genotyping can be used for GWAS to increase the power of QTL detection.

5.2 Prospects

The findings within this thesis indicate the usefulness of GWAS in detecting QTL for economic traits. The worldwide barley collection with improved marker coverage could be used to detect QTL with increased resolution. The concerns caused by inherent population structure in barley populations for GWAS can be overcome by appropriate statistical methodology. The new wave of next generation sequencing technologies along with the anticipated barley genome reference sequence in near future (Schulte et al., 2009), will allow genotyping by re-sequencing in large collections of barley. This can provide extensive genome coverage without biasness and more power to localize the QTL and underlying candidate genes. The markers associated with these QTL can be translated into diagnostic markers after sufficient validation and used for marker assisted selection in future. Further validation of these markers can be done either by traditional biparental crosses or different association panels. The landrace collection assessed for genetic diversity in the current studies can be used for fine resolution mapping of these QTL. Both the worldwide collection and landrace collection can be exploited to mine new alleles for agronomic and adaptive traits. Especially the landrace collection can prove to be a trove of useful alleles, as we discovered many rare alleles in this collection. Several studies in barley have reported the detection of useful alleles from landraces (Piffanelli et al., 2004; Wang et al., 2010b). This material can be useful for targeted candidate gene re-sequencing for allele mining purposes. For these reasons we consider the landrace collection as a valuable resource for future scientific research and crop improvement. Furthermore, in hindsight this thesis has provided a valuable landrace collection for future research purposes.

6 Summary

Genome-wide association studies (GWAS) based on Linkage Disequilibrium (LD) provides a promising tool for detecting and fine mapping of Quantitative trait loci (QTL) underlying complex traits. The LD between the genotyped marker and the causal gene allows the detection of QTL depending on the extent of LD in the association mapping panel.

One of the major objectives of this thesis was to identify QTL for agronomic traits in a diverse spring barley panel using GWAS approach. The association panel comprising of 224 worldwide spring barleys was used for GWAS, as well as for LD and diversity studies. The Phenotypic traits row type (RT), heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and crude protein content (CPC) were investigated. The panel was initially genotyped using a customized DNA marker assay (IPK OPA assay) which yielded 918 successful SNPs with approximate genome coverage of one SNP per 1.18 cM. Average LD was observed to decay below a critical level (r^2-value= 0.2) within a map distance of 5-10 cM. Different statistical models were tested to control spurious LD caused by population structure and to calculate the P-value of marker-trait associations. The mixed linear model (MLM) with kinship to control spurious LD effects, performed best in this panel. Using MLM with kinship (K-model), a total of 171 significant marker trait associations were detected, which delineated into 107 QTL regions. Across all traits these were grouped into 57 novel QTL and 50 QTL that are congruent with previously mapped QTL positions.

These results demonstrate that the described diverse barley panel can be efficiently used for GWAS of various quantitative traits, provided that population structure is appropriately taken into account. The observed significant marker trait associations provide a refined insight into the genetic architecture of important agronomic traits in barley. However, individual QTL accounted only for a small portion of phenotypic variation. The fact that the combined SNP effects fall short of explaining the complete phenotypic variance may support the hypothesis that the expression of a quantitative trait is caused by a large number of very small effects that escape detection.

Consequently the current spring barley panel was genotyped using a newly established iSelect assay, which yielded 7000 successful SNPs. Finally 6467 SNPs were used for GWAS, after the SNPs with minor allele frequency less than 5% were excluded. Multifold increase in the number of SNPs associating with the trait was observed. The significance of the associations also increased in many cases with new markers in the region. The effects of use of different kinship matrices on the GWAS results were compared. The kinship matrix generated using evenly distributed 362 SNPs excelled in performance when used with K-model. GWA scans with iSelect SNPs showed 297, 269, 240, 266, 304 and 245 SNPs associating with the traits RT, HD, PHT, TGW, SC and CPC respectively. In GWAS with iSelect SNPs, for most of the traits we detected associations closely linked to major candidate genes affecting the trait. It is possible to predict candidate genes underlying a QTL in few cases by using the genome models exploiting the syntenic conservation between rice, *Brachypodium* and barley.

The variance explained by individual associated marker also increased for each trait when compared to GWAS results with SNPs from IPK OPA assay. However, the variances are still much less when compared to those observed in bi-parental QTL mapping. This study demonstrates the advantages of an increased marker density on the number of QTL detected and on QTL significances. The resolution of the panel and the marker density are sufficient for detecting QTL, but for further fine mapping or gene identification the present resolution is still limiting in many cases. Therefore, a large population of spring landraces was developed for increased genetic resolution that can be used in fine mapping of traits

Landraces offer important genetic resources for cultivated barley, which has narrow genetic background due to intensive breeding. Besides, LD decays faster in landraces than in the cultivar collection. This different LD patterns can be exploited for high resolution mapping of the QTL using GWAS in cultivar and landrace collections. Therefore, it is pragmatic to study the genetics of the traits in landrace collections for precise mapping, and also before they are utilized for crop improvement. Hence we investigated a large collection of barley landraces (1491 accessions) for genetic diversity and population structure to establish spring barley landrace association mapping population. The collection comprises two-rowed, six-rowed, naked and hulled barleys from 41 countries. The landrace collection was evaluated with 45 SSR markers to assess the genetic diversity, population structure and genetic differentiation among the collection. A total of 372 alleles among which 152 are rare alleles (allele frequency < 1%) were detected. The collection was diverse with an average gene diversity of 0.603 and average allelic richness over all loci was 5.74. The landraces were differentiated into subgroups majorly based on row type and their geographical origins. The genetic distance between the accessions was significantly correlated with the geographical, latitudinal, longitudinal distances and also with other eco-geographical parameters.

Beyond providing insights into the diversity, our data allow to construct core groups based on maximizing allelic diversity approaches. Different core groups were generated for heat adaptation studies from the whole collection and compared the genetic diversity among these groups. Core groups above the size of 800 accessions showed similar allelic richness equal to the whole collection. Further small core groups showed declining allelic richness with the sample size. Simulating populations of different sizes from this landrace collection and comparing their genetic diversity revealed that the population size n=745 captures the diversity present in the whole collection. This collection can also be used for allele mining strategies to discover new alleles, as already our results indicate high allelic diversity in the collection. For further research, the landrace collection can be used for fine mapping of these detected QTL. The markers corresponding to the QTL detected in this study can be verified in other populations and then efficiently used for barley crop improvement.

7 Zusammenfassung

Genomweite Assoziationsstudien (GWAS) basierend auf dem Kopplungsungleichgewicht (LD) stellen ein vielversprechendes Verfahren für die Erfassung und Feinkartierung von quantitativen Merkmalen dar. Eines der Ziele dieser Studie war die Identifizierung von *Quantitative Trait Loci* (QTL) für bedeutende agronomische Eigenschaften in einer diversen Sommergerstekollektion. Hierzu wurde eine Kollektion, welche 224 weltweit verbreitete Sommergerstensorten umfasst, im Hinblick auf die Merkmale Zeiligkeit (ZT, zweizeilig/sechszeilig), Blühzeitpunkt (BZ), Pflanzenhöhe (PH), Tausendkorngewicht (TKG), Stärkegehalt (SG) und Rohproteingehalt (REG) untersucht. Die Kollektion wurde zunächst mit dem „IPK OPA Assay“ im Hinblick auf Einzelnukleotidpolymorphismen (Single Nucleotide Polymorphism, SNPs) genotypisiert. Daraus resultierten 918 informative SNPs mit einer Genomabdeckung von einem SNP pro 1,18 centiMorgan (cM). Innerhalb einer Kartierungsentfernung von 5-10 cM sank das durchschnittliche LD unter die kritische Schwelle von $r^2 = 0,2$. Um das Auftreten falscher Marker-Merkmal Assoziationen bedingt durch den Einfluss der Populationsstruktur zu minimieren, wurden unterschiedliche statistische Modelle verglichen. Gemischte lineare Modelle (Mixed Linear Models, MLM), in denen die genetische Distanz in Form der Verwandtschaftsmatrix Kinship (K) verwendet wurden, zeigten die besten Ergebnisse. Unter Nutzung eines entsprechenden Modells konnten insgesamt 171 signifikante Assoziationen für die o.a. Merkmale gefunden werden, welche zusammen 107 QTL ergaben. Hierunter befanden sich 57 neue, bisher nicht beschriebene QTLs und 50 QTLs, welche mit bereits kartierten QTL Positionen übereinstimmen.

Diese Ergebnisse zeigen, dass die diverse Gerstenkollektion effektiv für GWAS für quantitative Eigenschaften genutzt werden kann, vorausgesetzt, dass die Populationsstruktur berücksichtigt wird. Die beobachteten signifikanten Marker-Merkmal-Zusammenhänge liefern einen präzisen Einblick in die genetische Struktur von wichtigen agronomischen Eigenschaften in Gerste. Allerdings erfassen die individuellen QTLs nur einen kleinen Teil der phänotypischen Varianz. Der Fakt, dass die kombinierten SNP-Effekte nicht ausreichen um die komplette phänotypische Varianz zu erklären, unterstützt die Hypothese, dass die Expression eines quantitativen Merkmals durch ein

Vielzahl von sehr kleinen Effekten verursacht wird, welche mit GWAS nicht detektiert werden können.

Im nächsten Schritt wurde die Sommergerstenkollektion mit dem neu entwickelten „iSelect Assay“ genotypisiert. Insgesamt konnten 7000 SNPs erfolgreich detektiert werden. Letztendlich wurden 6467 SNPs für GWAS verwendet. Hierfür wurden nur SNPs mit einer Allelfrequenz über 5% berücksichtigt. Es wurde ein deutlicher Anstieg von Marker-Merkmal Assoziationen beobachtet. Die Signifikanz der Assoziation erhöhte sich in vielen Fällen. Die Effekte der Nutzung unterschiedlicher Verwandtschaftsmatrizen wurden ebenfalls verglichen. Die Kinship matrix basierend auf 362 gleichmäßig verteilten SNPs (K3) erzielte das beste Ergebnis. Insgesamt konnten 297, 269, 240, 266, 304 und 245 signifikant assoziierte SNPs in Zusammenhang mit den Eigenschaften ZT, BZ, PH, TKG, SG und REG detektiert werden. In den GWAS mit den iSelect SNPs wurden für die meisten Eigenschaften signifikante Assoziationen detektiert, welche nah an bekannten Kandidatengenen für diese Merkmale liegen. Mit Hilfe von Genmodellen und syntänen Zusammenhängen zwischen Reis, *Brachypodium* und Gerste, ist es möglich zugrundeliegende Kandidatengene eines QTL näher einzugrenzen. Die erklärte Varianz eines individuellen, assoziierten Markers erhöhte sich ebenfalls für jede Eigenschaft im Vergleich zu den „IPK OPA Assay“ Ergebnissen. Dennoch sind die erklärten Varianzen vergleichsweise gering verglichen mit den Varianzen, die in bi-parentalen QTL Kartierungen beobachtet werden. Diese Studie zeigte die Vorteile einer erhöhten Markerdichte in Bezug auf die Anzahl der detektierten QTLs und deren Signifikanz auf. Die Auflösung der Kollektion und die Markerdichte sind ausreichend, um QTLs zu detektieren. Für weitere Feinkartierungen oder Genidentifikationen könnte die derzeitige Auflösung jedoch nicht ausreichen. Es wurde daher eine große Sommergerste-Landrassen-Population aufgebaut, die eine höhere genetische Auflösung aufweisen sollte.

Gerste-Landrassen bilden eine wichtige genetische Ressource für die Verbesserung von Hochleistungssorten, welche aufgrund intensiver Züchtung nur einen beschränkten genetischen Hintergrund aufweisen. Weiterhin fällt das LD in Landrassen schneller ab als in Kultursorten. Die präzise Erfassung der Ausdehnung des LD in Landrassen ist eine Grundvoraussetzung, bevor sie für eine Sortenverbesserung genutzt werden können. Es konnte eine große Sommergersten-Landrassen-Kollektion (1491 Akzessionen) für zukünftige Assoziationskartierungen etabliert werden. Die Kollektion umfasst zweizeilige, sechszeilige, nackt- und bedecktsamige Gersten aus 41 Ländern. Die Landrassenkollektion wurde bisher mit 45 SSR Markern untersucht um die genetische Diversität, Populationsstruktur und die genetische Differenzierung innerhalb der Kollektion zu beurteilen. Es wurden 372 Allele detektiert, darunter befanden sich 152 seltene Allele (Allelfrequenz < 1%). Die Kollektion erwies sich als divers. Dies spiegelt sich in der durchschnittlichen Gendiversität von 0,603 und einer durchschnittlichen Allelhäufigkeit über alle Loci von 5,74 wider. Die Untergruppen der Landrassen bedingten sich durch die Zeiligkeit und ihre geographische Herkunft. Die genetische Distanz zwischen Akzessionen war signifikant korreliert mit geographischen und öko-geographischen Parametern.

Zusätzlich zum Einblick in die Diversität, erlauben die Daten die Erstellung von *Core groups* basierend auf maximierten Alleldiversitäten. Dies wurde am Beispiel von verschiedenen *Core groups* für Studien zur Hitzetoleranz demonstriert, indem die genetische Diversität zwischen den Gruppen verglichen wurde. Es fanden sich gleiche Allelhäufigkeiten in *Core groups* bis zu einer Mindestgröße von 800 Akzessionen. Kleine *Core groups* (<800 Akzessionen)zeigten dagegen eine niedrigere Anzahl an Allelen per Locus. Die Simulation unterschiedlicher Populationsgrößen der Landrassen-Kollektion ergab, dass eine Populationsgröße von n=745 die Diversität der gesamten Kollektion (n=1491) beinhalten kann. Optimierte *Core groups* können daher für *allele mining* Studien genutzt werden um neue Allele zu identifizieren. Zusätzlich dazu kann die Landrassenkollektion für die Feinkartierung von detektierten QTLs verwendet werden.

8 References

Abdel-Ghani, A.H., H.K. Parzies, A. Omary, and H.H. Geiger, 2004: Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. Theoretical and Applied Genetics **109**, 588-595.

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

Annonymus, 2008: Global strategy for the ex situ conservation and use of barley germplasm, Global Crop Diversity Trust, Rome, Italy.

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

Atwell, S., Y.S. Huang, B.J. Vilhjalmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Muliyati, X. Zhang, M.A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. de Meaux, J.R. Ecker, N. Faure, J.M. Kniskern, J.D. Jones, T. Michael, A. Nemri, F. Roux, D.E. Salt, C. Tang, M. Todesco, M.B. Traw, D. Weigel, P. Marjoram, J.O. Borevitz, J. Bergelson, and M. Nordborg, 2010: Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature **465**, 627-31.

Backes, G., J. Orabi, A. Wolday, A. Yahyaoui, and A. Jahoor, 2009: High genetic diversity revealed in barley (*Hordeum vulgare*) collected from small-scale farmer’s fields in Eritrea. Genetic Resources and Crop Evolution **56**, 85-97.

Badr, A., K. Müller, R. Schäfer-Pregl, H. El Rabey, S. Effgen, H.H. Ibrahim, C. Pozzi, W. Rohde, and F. Salamini, 2000: On the origin and domestication history of Barley (Hordeum vulgare). Molecular Biology and Evolution **17**, 499-510.

Baik, B.-K., and S.E. Ullrich, 2008: Barley for food: Characteristics, improvement, and renewed interest. Journal of Cereal Science **48**, 233-242.

Barrett, J.C., B. Fry, J. Maller, and M.J. Daly, 2005: Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21**, 263-5.

Benjamini, Y., and Y. Hochberg, 1995: Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological **57**, 289-300.

Bennett, M.D., and J.B. Smith, 1976: Nuclear DNA amounts in Angiosperms. Philosophical Transactions of the Royal Society of London. B, Biological Sciences **274**, 227-274.

Bernardo, R., 2008: Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. **48**, 1649-1664.

Bhullar, N., Z. Zhang, T. Wicker, and B. Keller, 2010: Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene Pm3: a large scale allele mining project. BMC Plant Biology **10**, 88.

Blott, S., J.-J. Kim, S. Moisio, A. Schmidt-Kuntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir, and W. Coppieters, 2003: Molecular dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine substitution in the transmembrane domain of the Bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics **163**, 253-266.

Börner, A., V. Korzun, S. Malyshev, V. Ivandic, and A. Graner, 1999: Molecular mapping of two dwarfing genes differing in their GA response on chromosome 2H of barley. Theoretical and Applied Genetics **99**, 670-675.

Bothmer, R.V., N. Jacobsen, C. Baden, C. Jørgensen, and I. Linde-Laursen, 1995: An ecogeographical study of the genus Hordeum. Systematic and ecogeographic studies on crop genepools 7. 2nd ed. International Plant Genetic Resource Institute, Rome.

Bothmer, R.V., K. Sato, T. Komatsuda, S. Yasuda, and G. Fischbeck, 2003: Chapter 2 The domestication of cultivated barley, In: Bothmer, R.V., Hintum T. V., H. Knüpffer and K. Sato, (eds.) Developments in Plant Genetics and Breeding, 9-27, Vol. Volume 7. Elsevier.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis, 1980: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics **32**, 314-331.

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

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler, 2007: TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23**, 2633-5.

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

Brown, A., 1992: Genetic variation and resources in cultivated barley, In: L. Munch, (ed.) Barley Genetics VI, pp. 669-682, Vol. Volume II. Munksgaard International Publishers, Copenhagen.

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

Browning, S.R., 2008: Estimation of Pairwise Identity by Descent From Dense Genetic Marker Data in a Population Sample of Haplotypes. Genetics **178**, 2123-2132.

Brücher, H., and E. Åberg, 1950: Die primitiv-gersten des Hochlands von Tibet, Ihre bedeutung für die züchtung und das verständnis des ursprungs und der Klassifizierung der gersten Royal Agricultural College, Uppsala.

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J.C. Glaubitz, M.M. Goodman, C. Harjes, K. Guill, D.E. Kroon, S. Larsson, N.K. Lepak, H. Li, S.E. Mitchell, G. Pressoir, J.A. Peiffer, M.O. Rosas, T.R. Rocheford, M.C. Romay, S. Romero, S. Salvo, H. Sanchez Villeda, H.S. da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M.D. McMullen, 2009: The genetic architecture of maize flowering time. Science **325**, 714-8.

Caldwell, K.S., J. Russell, P. Langridge, and W. Powell, 2006: Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, Hordeum vulgare. Genetics **172**, 557-67.

Castillo, A., G. Dorado, C. Feuillet, P. Sourdille, and P. Hernandez, 2010: Genetic structure and ecogeographical adaptation in wild barley (*Hordeum chilense*) as revealed by microsatellite markers. BMC Plant Biology **10**, 266.

Ceccarelli, S., and S. Grando, 1996: Drought as a challenge for the plant breeder. Plant Growth Regulation **20**, 149-155.

Chakraborty, R., and L. Jin, 1993: Determination of Relatedness between Individuals Using DNA-Fingerprinting. Human Biology **65**, 875-895.

Chan, E.K., H.C. Rowe, and D.J. Kliebenstein, 2010: Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. Genetics **185**, 991-1007.

Chono, M., I. Honda, H. Zeniya, K. Yoneyama, D. Saisho, K. Takeda, S. Takatsuto, T. Hoshino, and Y. Watanabe, 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.

Clark, H.H., 1967: The Origin and early History of the cultivated Barleys. The agricultural history review **15**, 1-18.

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

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

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

Cockram, J., J. White, D.L. Zuluaga, D. Smith, J. Comadran, M. Macaulay, Z.W. Luo, M.J. Kearsey, P. Werner, D. Harrap, C. Tapsell, H. Liu, P.E. Hedley, N. Stein, D. Schulte, B. Steuernagel, D.F. Marshall, W.T.B. Thomas, L. Ramsay, I. Mackay, D.J. Balding, R. Waugh, D.M. O'Sullivan, and A. Consortium, 2010: Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proceedings of the National Academy of Sciences of the United States of America **107**, 21611-21616.

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

Comadran, J., W.T. Thomas, F.A. van Eeuwijk, S. Ceccarelli, S. Grando, A.M. Stanca, N. Pecchioni, T. Akar, A. Al-Yassin, A. Benbelkacem, H. Ouabbou, J. Bort, I. Romagosa, C.A. Hackett, and J.R. Russell, 2009: Patterns of genetic diversity and linkage disequilibrium in a highly structured Hordeum vulgare association-mapping population for the Mediterranean basin. Theoretical and Applied Genetics **119**, 175-87.

Comadran, J., J.R. Russell, A. Booth, A. Pswarayi, S. Ceccarelli, S. Grando, A.M. Stanca, N. Pecchioni, T. Akar, A. Al-Yassin, A. Benbelkacem, H. Ouabbou, J. Bort, F.A. van Eeuwijk, W.T.B. Thomas, and I. Romagosa, 2011: Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. Theoretical and Applied Genetics **122**, 1363-1373.

D’hoop, B., M. Paulo, K. Kowitwanich, M. Sengers, R. Visser, H. van Eck, and F. van Eeuwijk, 2010: Population structure and linkage disequilibrium unravelled in tetraploid potato. Theoretical and Applied Genetics **121**, 1151-1170.

Delseny, M., 2004: Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. Current Opinion in Plant Biology **7**, 126-131.

Diamond, J., 2002: Evolution, consequences and future of plant and animal domestication. Nature **418**, 700-707.

Dunford, R.P., S. Griffiths, V. Christodoulou, and D.A. Laurie, 2005: Characterisation of a barley (*Hordeum vulgare* L.) homologue of the *Arabidopsis* flowering time regulator *GIGANTEA*. Theoretical and Applied Genetics **110**, 925-931.

Ehrenreich, I.M., Y. Hanzawa, L. Chou, J.L. Roe, P.X. Kover, and M.D. Purugganan, 2009: Candidate Gene Association Mapping of Arabidopsis Flowering Time. Genetics **183**, 325-335.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell, 2011: A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE **6**.

Endresen, D.T.F., 2010: Predictive Association between Trait Data and Ecogeographic Data for Nordic Barley Landraces. Crop Science **50**, 2418-2430.

Escudero, A., J.M. Iriondo, and M.E. Torres, 2003: Spatial analysis of genetic diversity as a tool for plant conservation. Biological Conservation **113**, 351-365.

Excoffier, L., G. Laval, and S. Schneider, 2005: Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics **1**, 47 - 50.

Falush, D., M. Stephens, and J.K. Pritchard, 2003: Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics **164**, 1567-1587.

Fan, J.B., K.L. Gunderson, M. Bibikova, J.M. Yeakley, J. Chen, E. Wickham Garcia, L.L. Lebruska, M. Laurent, R. Shen, and D. Barker, 2006: Illumina Universal Bead Arrays, In: K. Alan and O. Brian, (eds.) Methods in Enzymology, 57-73, Vol. Volume 410. Academic Press.

Fan, J.B., A. Oliphant, R. Shen, B.G. Kermani, F. Garcia, K.L. Gunderson, M. Hansen, F. Steemers, S.L. Butler, P. Deloukas, L. Galver, S. Hunt, C. McBride, M. Bibikova, T. Rubano, J. Chen, E. Wickham, D. Doucet, W. Chang, D. Campbell, B. Zhang, S. Kruglyak, D. Bentley, J. Haas, P. Rigault, L. Zhou, J. Stuelpnagel, and M.S. Chee, 2003: Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology **68**, 69-78.

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

Feuillet, C., P. Langridge, and R. Waugh, 2008: Cereal breeding takes a walk on the wild side. Trends in genetics : TIG **24**, 24-32.

Fischbeck, G., 2003: Chapter 3 Diversification through breeding, In: Bothmer, R.V., Hintum T. V., H. Knüpffer and K. Sato, (eds.) Developments in Plant Genetics and Breeding, 121-141, Vol. Volume 7. Elsevier.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler, 2003: Structure of linkage disequilibrium in plants. Annu Rev Plant Biol **54**, 357-74.

Flint-Garcia, S.A., A.C. Thuillet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler, 2005: Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J **44**, 1054-64.

Frary, A., T.C. Nesbitt, A. Frary, S. Grandillo, E.v.d. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert, and S.D. Tanksley, 2000: fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. Science **289**, 85-88.

Frazer, K.A., S.S. Murray, N.J. Schork, and E.J. Topol, 2009: Human genetic variation and its contribution to complex traits. Nature Reviews Genetics **10**, 241-51.

Fuller, D.Q., 2007: Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World. Annals of Botany **100**, 903-924.

Gibson, G., 2010: Hints of hidden heritability in GWAS. Nature Genetics **42**, 558-60.

Gong, X., S. Westcott, C. Li, G. Yan, R. Lance, and D. Sun, 2009: Comparative analysis of genetic diversity between Qinghai-Tibetan wild and Chinese landrace barley. Genome **52**, 849-861.

Gottwald, S., N. Stein, A. Borner, T. Sasaki, and A. Graner, 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.

Goudet, J., 1995: FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. Journal of Heredity **86**, 485-486.

Gouesnard, B., T.M. Bataillon, G. Decoux, C. Rozale, D.J. Schoen, and J.L. David, 2001: MSTRAT: An Algorithm for Building Germ Plasm Core Collections by Maximizing Allelic or Phenotypic Richness. Journal of Heredity **92**, 93-94.

Grando, S., and H. Gómez Macpherson, 2005: Food barley : importance, uses and local knowledge ICARDA, Aleppo, Syria.

Graner, A., W.F. Ludwig, and A.E. Melchinger, 1994: Relationships among European barley germplasm. 2.Comparision of RFLP and pedigree data Crop Science **34**, 1199-1205.

Graner, A., A. Kilian, and A. Kleinhofs, 2010: Barley Genome Organization, Mapping, and Synteny Barley, 63-84. Wiley-Blackwell.

Graner, A., s. Bjørnstad, T. Konishi, and F. Ordon, 2003: Chapter 7 Molecular diversity of the barley genome, In: Bothmer, R.V., Hintum T. V., H. Knüpffer and K. Sato, (eds.) Developments in Plant Genetics and Breeding, 121-141, Vol. Volume 7. Elsevier.

Graner, A., H. Siedler, A. Jahoor, R.G. Herrmann, and G. Wenzel, 1990: Assessment of the degree and the type of restriction fragment length polymorphism in barley (*Hordeum vulgare*). Theoretical and Applied Genetics **80**, 826-832.

Graner, A., A. Jahoor, J. Schondelmaier, H. Siedler, K. Pillen, G. Fischbeck, G. Wenzel, and R.G. Herrmann, 1991: Construction of an RFLP map of barley. Theoretical and Applied Genetics **83**, 250-256.

Griffiths, S., R.P. Dunford, G. Coupland, and D.A. Laurie, 2003: The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant Physiology **131**, 1855-67.

Gruszka, D., I. Szarejko, and M. Maluszynski, 2011: New allele of *HvBRI1* gene encoding brassinosteroid receptor in barley. Journal of Applied Genetics **52**, 257-268.

Gupta, P.K., and R.K. Varshney, 2000: The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica **113**, 163-185.

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

Hall, D., C. Tegstrom, and P.K. Ingvarsson, 2010: Using association mapping to dissect the genetic basis of complex traits in plants. Brief Funct Genomics **9**, 157-65.

Hamblin, M.T., E.S. Buckler, and J.-L. Jannink, 2011: Population genetics of genomics-based crop improvement methods. Trends in Genetics **27**, 98-106.

Hamblin, M.T., T.J. Close, P.R. Bhat, S. Chao, J.G. Kling, K.J. Abraham, T. Blake, W.S. Brooks, B. Cooper, C.A. Griffey, P.M. Hayes, D.J. Hole, R.D. Horsley, D.E. Obert, K.P. Smith, S.E. Ullrich, G.J. Muehlbauer, and J.-L. Jannink, 2010: Population Structure and Linkage Disequilibrium in U.S. Barley Germplasm: Implications for Association Mapping. Crop Sci. **50**, 556-566.

Hamilton, S.N.R., J.M.M. Engels, T.J.L. van Hintum, B.a. Koo, and M. Smale, 2002: Accession management. Combining or splitting accessions as a tool to improve germplasm management efficiency IPGRI Technical Bulletin No. 5, International Plant Genetic Resources Institute, Rome, Italy.

Hammer, Ø., D.A.T. Harper, and P.D. Ryan, 2001: Past paleontological statistics software package for educaton and data anlysis.

Harlan, J.R., and J.M.J. de Wet, 1971: Toward a Rational Classification of Cultivated Plants. Taxon **20**, 509-517.

Haseneyer, G., S. Stracke, H.P. Piepho, S. Sauer, H.H. Geiger, and A. Graner, 2010a: DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. Bmc Plant Biology **10**.

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

Haseneyer, G., C. Ravel, M. Dardevet, F. Balfourier, P. Sourdille, G. Charmet, D. Brunel, S. Sauer, H.H. Geiger, A. Graner, and S. Stracke, 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. Theoretical and Applied Genetics **117**, 321-331.

Hastbacka, J., A. Delachapelle, I. Kaitila, P. Sistonen, A. Weaver, and E. Lander, 1992: Linkage Disequilibrium Mapping in Isolated Founder Populations - Diastrophic Dysplasia in Finland. Nature Genetics **2**, 204-211.

Hayes, P., and P. Szucs, 2006: Disequilibrium and association in barley: thinking outside the glass. Proceedings of the National Academy of Sciences U S A **103**, 18385-6.

Hayes, P.M., A. Castro, L. Marquez-Cedillo, A. Corey, C. Henson, B.L. Jones, J. Kling, D. Mather, I. Matus, C. Rossi, and K. Sato, 2003: Chapter 10 Genetic diversity for quantitatively inherited agronomic and malting quality traits, In: Bothmer, R.V., Hintum T. V., H. Knüpffer and K. Sato, (eds.) Developments in Plant Genetics and Breeding, 201-226, Vol. Volume 7. Elsevier.

Heun, M., A.E. Kennedy, J.A. Anderson, N.L.V. Lapitan, M.E. Sorrells, and S.D. Tanksley, 1991: Construction of a restriction fragment length polymorphism map for barley (Hordeum vulgare). Genome **34**, 437-447.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis, 2005a: Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**, 1965-1978.

Hijmans, R.J., L. Guarino, P. Mathur, A. Jarvis, E. Rojas, M.a. Cruz, and I. Barrantes, 2005b: DIVA-GIS Manual.

Hill, W.G., and A. Robertson, 1968: Linkage disequilibrium in finite populations. Theoretical and Applied Genetics **38**, 226-231.

Hirschhorn, J.N., and M.J. Daly, 2005: Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics **6**, 95-108.

Hoisington, D., M. Khairallah, T. Reeves, J.-M. Ribaut, B. Skovmand, S. Taba, and M. Warburton, 1999: Plant genetic resources: What can they contribute toward increased crop productivity? Proceedings of the National Academy of Sciences **96**, 5937-5943.

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

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

James B, H., 2007: Genetic architecture of complex traits in plants. Current Opinion in Plant Biology **10**, 156-161.

James, M.G., K. Denyer, and A.M. Myers, 2003: Starch synthesis in the cereal endosperm. Current Opinion in Plant Biology **6**, 215-222.

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

Jones, H., P. Civan, J. Cockram, F. Leigh, L. Smith, M. Jones, M. Charles, J.-L. Molina-Cano, W. Powell, G. Jones, and T. Brown, 2011: Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces. BMC Evolutionary Biology **11**, 320.

Kaeuffer, R., D. Reale, D.W. Coltman, and D. Pontier, 2007: Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. Heredity **99**, 374-380.

Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin, 2008: Efficient control of population structure in model organism association mapping. Genetics **178**, 1709-23.

Kilian, B., 2006: Haplotype structure at seven barley genes : relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. MGG : Molecular genetics and genomics **276**, 230-241.

Kilian, B., H. Özkan, C. Pozzi, and F. Salamini, 2009: Domestication of the Triticeae in the Fertile Crescent, In: G. J. Muehlbauer and C. Feuillet, (eds.) Genetics and Genomics of the Triticeae, 81-119, Vol. 7. Springer New York.

Kilian, B., H. Özkan, J. Kohl, A. von Haeseler, F. Barale, O. Deusch, A. Brandolini, C. Yucel, W. Martin, and F. Salamini, 2006: Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. Molecular Genetics and Genomics **276**, 230-241.

Kleinhofs, A., 1997: Integrating barley RFLP and classical marker maps Barley Genet News letter 105-112, Vol. 27.

Kleinhofs, A., A. Kilian, M.A. Saghai Maroof, R.M. Biyashev, P. Hayes, F.Q. Chen, N. Lapitan, A. Fenwick, T.K. Blake, V. Kanazin, E. Ananiev, L. Dahleen, D. Kudrna, J. Bollinger, S.J. Knapp, B. Liu, M. Sorrells, M. Heun, J.D. Franckowiak, D. Hoffman, R. Skadsen, and B.J. Steffenson, 1993: A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theoretical and Applied Genetics **86**, 705-712.

Knüpffer, H., and T. van Hintum, 2003: Chapter 13 Summarised diversity—the Barley Core Collection, In: Bothmer, R.V., Hintum T. V., H. Knüpffer and K. Sato, (eds.) Developments in Plant Genetics and Breeding, 259-267, Vol. Volume 7. Elsevier.

Komatsuda, T., M. Pourkheirandish, C. He, P. Azhaguvel, H. Kanamori, D. Perovic, N. Stein, A. Graner, T. Wicker, A. Tagiri, U. Lundqvist, T. Fujimura, M. Matsuoka, T. Matsumoto, and M. Yano, 2007: Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proceedings of the National Academy of Sciences U S A **104**, 1424-9.

Kraakman, A.T., R.E. Niks, P.M. Van den Berg, P. Stam, and F.A. Van Eeuwijk, 2004: Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics **168**, 435-46.

Lai, J., R. Li, X. Xu, W. Jin, M. Xu, H. Zhao, Z. Xiang, W. Song, K. Ying, M. Zhang, Y. Jiao, P. Ni, J. Zhang, D. Li, X. Guo, K. Ye, M. Jian, B. Wang, H. Zheng, H. Liang, X. Zhang, S. Wang, S. Chen, J. Li, Y. Fu, N.M. Springer, H. Yang, J. Wang, J. Dai, P.S. Schnable, and J. Wang, 2010: Genome-wide patterns of genetic variation among elite maize inbred lines. Nature Genetics **42**, 1027-1030.

Lander, E.S., and N.J. Schork, 1994: Genetic dissection of complex traits. Science **265**, 2037-48.

Laurie, D.A., N. Pratchett, J.W. Snape, and J.H. Bezant, 1995: RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter x spring barley (*Hordeum vulgare* L.) cross. Genome **38**, 575-85.

Leberg, P.L., 2002: Estimating allelic richness: Effects of sample size and bottlenecks. Molecular Ecology **11**, 2445-2449.

Lehmann, C.O., and R. Mansfeld, 1957: Zur Technik der Sortimentserhaltung. Genetic Resources and Crop Evolution **5**, 108-138.

Leino, M.W., and J. Hagenblad, 2010: Nineteenth Century Seeds Reveal the Population Genetics of Landrace Barley (*Hordeum vulgare*). Molecular Biology and Evolution **27**, 964-973.

Li, J.Z., X.Q. Huang, F. Heinrichs, M.W. Ganal, and M.S. Röder, 2006: Analysis of QTLs for yield components, agronomic traits, and disease resistance in an advanced backcross population of spring barley. Genome **49**, 454-466.

Li, Y.H., C. Zhang, Z.S. Gao, M.J.M. Smulders, Z.L. Ma, Z.X. Liu, H.Y. Nan, R.Z. Chang, and L.J. Qiu, 2009: Development of SNP markers and haplotype analysis of the candidate gene for rhg1, which confers resistance to soybean cyst nematode in soybean. Molecular Breeding **24**, 63-76.

Li, Z., D. Li, X. Du, H. Wang, O. Larroque, C.L.D. Jenkins, S.A. Jobling, and M.K. Morell, 2011: The barley amo1 locus is tightly linked to the starch synthase IIIa gene and negatively regulates expression of granule-bound starch synthetic genes. Journal of Experimental Botany **62**, 5217-5231.

Liu, F., R. von Bothmer, and B. Salomon, 2000: Genetic diversity in European accessions of the Barley Core Collection as detected by isozyme electrophoresis. Genetic Resources and Crop Evolution **47**, 571-581.

Liu, K., and S.V. Muse, 2005: PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics **21**, 2128-2129.

Lundqvist, U., and J.D. Franckowiak, 2003: Chapter 5 Diversity of barley mutants, In: T. v. H. H. K. Roland von Bothmer and S. Kazuhiro, (eds.) Developments in Plant Genetics and Breeding, 77-96, Vol. Volume 7. Elsevier.

Ma, J., Q.-T. Jiang, Y.-M. Wei, L. Andre, Z.-X. Lu, G.-Y. Chen, Y.-X. Liu, and Y.-L. Zheng, 2010: Molecular characterization and comparative analysis of two *waxy* alleles in barley. Genes & Genomics **32**, 513-520.

Maccaferri, M., M.C. Sanguineti, E. Noli, and R. Tuberosa, 2005: Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Molecular Breeding **15**, 271-290.

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

Maher, B., 2008: Personal genomes: The case of the missing heritability. Nature **456**, 18-21.

Malysheva-Otto, L., M. Ganal, and M. Roder, 2006: Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genetics **7**, 6.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F. Mackay, S.A. McCarroll, and P.M. Visscher, 2009: Finding the missing heritability of complex diseases. Nature **461**, 747-53.

Mansfeld, R., 1950: Das morphologische System der Saatgerste *Hordeum vulgare* L. s. l. Theoretical and Applied Genetics **20**, 8-24.

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

Mather, D.E., N.A. Tinker, D.E. LaBerge, M. Edney, B.L. Jones, B.G. Rossnagel, W.G. Legge, K.G. Briggs, R.B. Irvine, D.E. Falk, and K.J. Kasha, 1997: Regions of the genome that affect grain and malt quality in a North American two-row barley cross. Crop Science **37**, 544-554.

Mather, K.A., A.L. Caicedo, N.R. Polato, K.M. Olsen, S. McCouch, and M.D. Purugganan, 2007: The extent of linkage disequilibrium in rice (*Oryza sativa* L.). Genetics **177**, 2223-32.

Matus, I.A., and P.M. Hayes, 2002: Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome **45**, 1095-1106.

Maxted, N., B. Ford-Lloyd, S. Jury, S. Kell, and M. Scholten, 2006: Towards a definition of a crop wild relative. Biodiversity and Conservation **15**, 2673-2685.

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

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

Melchinger, A.E., A. Graner, M. Singh, and M.M. Messmer, 1994: Relationships among European Barley Germplasm .1. Genetic Diversity among Winter and Spring Cultivars Revealed by RFLPs. Crop Science **34**, 1191-1199.

Molina-Cano, J.L., J.R. Russell, M.A. Moralejo, J.L. Escacena, G. Arias, and W. Powell, 2005: Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. Theoretical and Applied Genetics **110**, 613-619.

Moragues, M., J. Comadran, R. Waugh, I. Milne, A. Flavell, and J. Russell, 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, M.K., B. Kosar-Hashemi, M. Cmiel, M.S. Samuel, P. Chandler, S. Rahman, A. Buleon, I.L. Batey, and Z. Li, 2003: Barley sex6 mutants lack starch synthase IIa activity and contain a starch with novel properties. The Plant Journal **34**, 173-185.

Morrell, P.L., and M.T. Clegg, 2007: Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proceedings of the National Academy of Sciences **104**, 3289-3294.

Morrell, P.L., D.M. Toleno, K.E. Lundy, and M.T. Clegg, 2005: Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proceedings of the National Academy of Sciences of the United States of America **102**, 2442-2447.

Mott, R., and J. Flint, 2002: Simultaneous Detection and Fine Mapping of Quantitative Trait Loci in Mice Using Heterogeneous Stocks. Genetics **160**, 1609-1618.

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, and E.S. Buckler, 2009: Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell **21**, 2194-202.

Naumann, C., and R. Bassler, 2004: VDLUFA-Methodenbuch III: Die Chemische Untersuchung von Futtermitteln, 5. Ergänzungslieferung, Darmstadt, VDLUFA.

Negassa, M., 1985: Patterns of phenotypic diversity in an Ethiopian barley collection, and the Arussi-Bale Highland as a center of origin of barley. Hereditas **102**, 139-150.

Nei, M., 1972: Genetic Distance between Populations. American Naturalist **106**, 283.

Nevo, E., 2006: Genome evolution of wild cereal diversity and prospects for crop improvement. Plant Genetic Resources **4**, 36-46.

Nevo, E., A. Ordentlich, A. Beiles, and I. Ràskin, 1992: Genetic divergence of heat production within and between the wild progenitors of wheat and barley: evolutionary and agronomical implications. Theoretical and Applied Genetics **84**, 958-962.

Newman, C.W., and R.K. Newman, 2006: A Brief History of Barley Foods. CEREAL FOODS WORLD **51**, 4-7.

Orabi, J., G. Backes, A. Wolday, A. Yahyaoui, and A. Jahoor, 2007: The Horn of Africa as a centre of barley diversification and a potential domestication site. Theoretical and Applied Genetics **114**, 1117-1127.

Ordon, F., K. Werner, B. Pellio, A. Schiemann, W. Friedt, and A. Graner, 2003: Molecular breeding for resistance to soil-borne viruses (BaMMV, BaYMV, BaYMV-2) of barley (*Hordeum vulgare* L.). Journal of Plant Disease and Protection **110**.

Palaisa, K.A., M. Morgante, M. Williams, and A. Rafalski, 2003: Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell **15**, 1795-1806.

Pandey, M., C. Wagner, W. Friedt, and F. Ordon, 2006: Genetic relatedness and population differentiation of Himalayan hulless barley (*Hordeum vulgare* L.) landraces inferred with SSRs. Theoretical and Applied Genetics **113**, 715-729.

Parzies, H.K., W. Spoor, and R.A. Ennos, 2000: Genetic diversity of barley landrace accessions (*Hordeum vulgare ssp. vulgare*) conserved for different lengths of time in ex situ gene banks. Heredity **84**, 476-486.

Pasam, R.K., R. Sharma, M. Malosetti, F.A. van Eeuwijk, G. Haseneyer, B. Kilian, and A. Graner, 2012: Genome-wide association studies for Agronomical Traits in a world wide Spring Barley Collection. BMC Plant Biology **12**, 16.

Patterson, N., A.L. Price, and D. Reich, 2006: Population Structure and Eigenanalysis. PLoS Genet **2**, e190.

Payne, R.W., Murray, D.A., Harding, S.A., Baird, D.B. & Soutar, D.M. , 2006: GenStat for Windows (9th Edition) Introduction. VSN International, Hemel Hempstead.

Perrier, X., and J.P. Jacquemound-Collet, 2006: DARwin Software.

Pickering, R.A., G.M. Timmerman, M.G. Cromey, and G. Melz, 1994: Characterisation of progeny from backcrosses of triploid hybrids between *Hordeum vulgare L.* (2x) and *H. bulbosum* L (4x) to *H. vulgare*. Theoretical and Applied Genetics **88**, 460-464.

Pickering, R.A., A.M. Hill, M. Michel, and G.M. Timmerman-Vaughan, 1995: The transfer of a powdery mildew resistance gene from *Hordeum bulbosum* L to barley (*H. vulgare L.*) chromosome 2 (2I). Theoretical and Applied Genetics **91**, 1288-1292.

Piffanelli, P., L. Ramsay, R. Waugh, A. Benabdelmouna, A. D'Hont, K. Hollricher, J.H. Jorgensen, P. Schulze-Lefert, and R. Panstruga, 2004: A barley cultivation-associated polymorphism conveys resistance to powdery mildew. Nature **430**, 887-891.

Pillen, K., A. Zacharias, and J. Léon, 2003: Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics **107**, 340-352.

Pillen, K., A. Zacharias, and J. Léon, 2004: Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare ssp. spontaneum*. Theoretical and Applied Genetics **108**, 1591-1601.

Pillen, K., A. Binder, B. Kreuzkam, L. Ramsay, R. Waugh, J. Förster, and J. Léon, 2000: Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theoretical and Applied Genetics **101**, 652-660.

Pins, J.J., and H. Kaur, 2006: A Review of the Effects of Barley beta-Glucan on Cardiovascular and Diabetic Risk. CEREAL FOODS WORLD **51**, 8-11.

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

Pourkheirandish, M., T. Wicker, N. Stein, T. Fujimura, and T. Komatsuda, 2007: Analysis of the barley chromosome 2 region containing the six-rowed spike gene vrs1 reveals a breakdown of rice-barley micro collinearity by a transposition. Theoretical and Applied Genetics **114**, 1357-65.

Powell, W., W.T.B. Thomas, E. Baird, P. Lawrence, A. Booth, B. Harrower, J.W. McNicol, and R. Waugh, 1997: Analysis of quantitative traits in barley by the use of Amplified Fragment Length Polymorphisms. Heredity **79**, 48-59.

Prada, D., 2009: Molecular population genetics and agronomic alleles in seed banks: searching for a needle in a haystack? Journal of Experimental Botany **60**, 2541-2552.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich, 2006: Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics **38**, 904-9.

Pritchard, J., M. Stephens, and P. Donnelly, 2000a: Inference of population structure using multilocus genotype data. Genetics **155**, 945 - 959.

Pritchard, J.K., M. Stephens, N.A. Rosenberg, and P. Donnelly, 2000b: Association mapping in structured populations. American Journal of Human Genetics **67**, 170-181.

Pswarayi, A., F.A. Van Eeuwijk, S. Ceccarelli, S. Grando, J. Comadran, J.R. Russell, E. Francia, N. Pecchioni, O. Li Destri, T. Akar, A. Al-Yassin, A. Benbelkacem, W. Choumane, M. Karrou, H. Ouabbou, J. Bort, J.L. Araus, J.L. Molina-Cano, W.T.B. Thomas, and I. Romagosa, 2008: Barley adaptation and improvement in the Mediterranean basin. Plant Breeding **127**, 554-560.

Qi, X., P. Stam, and P. Lindhout, 1996: Comparison and integration of four barley genetic maps. Genome **39**, 379-394.

Rae, S., M. Macaulay, L. Ramsay, F. Leigh, D. Matthews, D. O'Sullivan, P. Donini, P. Morris, W. Powell, D. Marshall, R. Waugh, and W. Thomas, 2007: Molecular barley breeding. Euphytica **158**, 295-303.

Rafalski, J.A., 2010: Association genetics in crop improvement. Current Opinion in Plant Biology **13**, 174-180.

Ramsay, L., M. Macaulay, S. degli Ivanissevich, K. MacLean, L. Cardle, J. Fuller, K.J. Edwards, S. Tuvesson, M. Morgante, A. Massari, E. Maestri, N. Marmiroli, T. Sjakste, M. Ganal, W. Powell, and R. Waugh, 2000: A Simple Sequence Repeat-Based Linkage Map of Barley. Genetics **156**, 1997-2005.

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

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S.t. Buckler, 2001: Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences U S A **98**, 11479-84.

Ren, X.F., D.F. Sun, W.W. Guan, G.L. Sun, and C.D. Li, 2010: Inheritance and identification of molecular markers associated with a novel dwarfing gene in barley. BMC Genetics **11**.

Rode, J., J. Ahlemeyer, W. Friedt, and F. Ordon, 2011: Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulgare L.*). Molecular Breeding, 1-13.

Rohde, W., D. Becker, and F. Salamini, 1988: Structural analysis of the waxy locus from Hordeum vulgare. Nucleic Acids Res **16**, 7185-6.

Rosenberg, M., and C. Anderson, 2011: PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. Methods in Ecology and Evolution **2**, 229-232.

Rosenberg, N., 2004: distruct: a program for the graphical display of population structure. Molecular Ecology Notes **4**, 137-138.

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

Roy, J.K., K.P. Smith, G.J. Muehlbauer, S. Chao, T.J. Close, and B.J. Steffenson, 2010: Association mapping of spot blotch resistance in wild barley. Mol Breed **26**, 243-256.

Russell, J., I.K. Dawson, A.J. Flavell, B. Steffenson, E. Weltzien, A. Booth, S. Ceccarelli, S. Grando, and R. Waugh, 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, J.R., J.D. Fuller, M. Macaulay, B.G. Hatz, A. Jahoor, W. Powell, and R. Waugh, 1997: Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theoretical and Applied Genetics **95**, 714-722.

Russell, J.R., A. Booth, J.D. Fuller, M. Baum, S. Ceccarelli, S. Grando, and W. Powell, 2003: Patterns of polymorphism detected in the chloroplast and nuclear genomes of barley landraces sampled from Syria and Jordan. Theoretical and Applied Genetics **107**, 413-421.

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

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

Saisho, D., K.-i. Tanno, M. Chono, I. Honda, H. Kitano, and K. Takeda, 2004: Spontaneous Brassinolide-insensitive Barley Mutants *uzu* Adapted to East Asia. Breeding Science **54**, 409-416.

Salamini, F., 2002: Genetics and geography of wild cereal domestication in the Near East. Nature reviews. Genetics **3**, 429-441.

Salvi, S., and R. Tuberosa, 2005: To clone or not to clone plant QTLs: present and future challenges. Trends in Plant Science **10**, 297-304.

Sameri, M., M. Pourkheirandish, G.X. Chen, T. Tonooka, and T. Komatsuda, 2011: Detection of photoperiod responsive and non-responsive flowering time QTL in barley. Breeding Science **61**, 183-188.

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

Sax, K., 1923: the association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics **8**, 552-560.

Schmierer, D.A., N. Kandemir, D.A. Kudrna, B.L. Jones, S.E. Ullrich, and A. Kleinhofs, 2005: Molecular marker-assisted selection for enhanced yield in malting barley. Molecular Breeding **14**, 463-473.

Schulte, D., T.J. Close, A. Graner, P. Langridge, T. Matsumoto, G. Muehlbauer, K. Sato, A.H. Schulman, R. Waugh, R.P. Wise, and N. Stein, 2009: The International Barley Sequencing Consortium—At the Threshold of Efficient Access to the Barley Genome. Plant Physiology **149**, 142-147.

Semagn, K., Å. Bjornstad, and Y. Xu, 2010: The genetic dissection of quantitative traits in crops. Electronic Journal of Biotechnology **13**.

Silvar, C., A. Casas, E. Igartua, L. Ponce-Molina, M. Gracia, G. Schweizer, M. Herz, K. Flath, R. Waugh, D. Kopahnke, and F. Ordon, 2011: Resistance to powdery mildew in Spanish barley landraces is controlled by different sets of quantitative trait loci. Theoretical and Applied Genetics, 1-10.

Singh, A., S. Reimer, C.J. Pozniak, F.R. Clarke, J.M. Clarke, R.E. Knox, and A.K. Singh, 2009: Allelic variation at Psy1-A1 and association with yellow pigment in durum wheat grain. Theoretical and Applied Genetics **118**, 1539-1548.

Smith, A., B. Cullis, and A. Gilmour, 2001: The Analysis of Crop Variety Evaluation Data in Australia. Australian & New Zealand Journal of Statistics **43**, 129-145.

Smith, E.G., 1927: The Beginning of Agriculture. Nature **119**, 81-82.

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

Stich, B., and A.E. Melchinger, 2009: Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics **10**.

Stich, B., J. Mohring, H.P. Piepho, M. Heckenberger, E.S. Buckler, and A.E. Melchinger, 2008: Comparison of mixed-model approaches for association mapping. Genetics **178**, 1745-1754.

Storey, J.D., 2002: A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64**, 479-498.

Storey, J.D., and R. Tibshirani, 2003: Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences **100**, 9440-9445.

Storey, J.D., J.E. Taylor, and D. Siegmund, 2004: Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **66**, 187-205.

Stracke, S., T. Presterl, N. Stein, D. Perovic, F. Ordon, and A. Graner, 2007: Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding Bymovirus resistance in barley. Genetics **175**, 805-17.

Stracke, S., G. Haseneyer, J.B. Veyrieras, H.H. Geiger, S. Sauer, A. Graner, and H.P. Piepho, 2009: Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theoretical and Applied Genetics **118**, 259-73.

Strelchenko, P., O. Kovalyova, and K. Okuno, 1999: Genetic differentiation and geographical distribution of barley germplasm based on RAPD markers. Genetic Resources and Crop Evolution **46**, 193-205.

Sutton, T., U. Baumann, J. Hayes, N.C. Collins, B.-J. Shi, T. Schnurbusch, A. Hay, G. Mayo, M. Pallotta, M. Tester, and P. Langridge, 2007: Boron-Toxicity Tolerance in Barley Arising from Efflux Transporter Amplification. Science **318**, 1446-1449.

Taketa, S., T. Awayama, S. Amano, Y. Sakurai, and M. Ichii, 2006: High-resolution mapping of the nud locus controlling the naked caryopsis in barley. Plant Breeding **125**, 337-342.

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

Tanksley, S.D., 1993: Mapping Polygenes. Annual Review of Genetics **27**, 205-233.

Tanksley, S.D., and S.R. McCouch, 1997: Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. Science **277**, 1063-1066.

Tanto Hadado, T., D. Rau, E. Bitocchi, and R. Papa, 2009: Genetic diversity of barley (*Hordeum vulgare* L.) landraces from the central highlands of Ethiopia: comparison between the *Belg* and *Meher* growing seasons using morphological traits. Genetic Resources and Crop Evolution **56**, 1131-1148.

Tanto Hadado, T., D. Rau, E. Bitocchi, and R. Papa, 2010: Adaptation and diversity along an altitudinal gradient in Ethiopian barley (*Hordeum vulgare* L.) landraces revealed by molecular analysis. BMC Plant Biology **10**, 121.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley, and B.S. Gaut, 2001: Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). Proceedings of the National Academy of Sciences **98**, 9161-9166.

Thiel, Michalek, Varshney, and Graner, 2003: Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare L.*). Theoretical and Applied Genetics **106**, 411-422.

Thomas, W.T.B., W. Powell, and W. Wood, 1984: The chromosomal location of the dwarfing gene present in the spring barley variety golden promise. Heredity **53**, 177-183.

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler, 2001: Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics **28**, 286-289.

Tian, F., P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T.R. Rocheford, M.D. McMullen, J.B. Holland, and E.S. Buckler, 2011: Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics **43**, 159-162.

Ullrich, S.E., 2010: Significance, Adaptation, Production, and Trade of Barley Barley, 3-13. Wiley-Blackwell.

Uysal, H., Y.-B. Fu, O. Kurt, G. Peterson, A. Diederichsen, and P. Kusters, 2010: Genetic diversity of cultivated flax (*Linum usitatissimum* L.) and its wild progenitor pale flax (*Linum bienne* Mill.) as revealed by ISSR markers. Genetic Resources and Crop Evolution **57**, 1109-1119.

Varshney, R., M. Baum, P. Guo, S. Grando, S. Ceccarelli, and A. Graner, 2010: Features of SNP and SSR diversity in a set of ICARDA barley germplasm collection. Molecular Breeding **26**, 229-242.

Varshney, R., T. Marcel, L. Ramsay, J. Russell, M. Röder, N. Stein, R. Waugh, P. Langridge, R. Niks, and A. Graner, 2007a: A high density barley microsatellite consensus map with 775 SSR loci. Theoretical and Applied Genetics **114**, 1091-1103.

Varshney, R.K., P. Langridge, and A. Graner, 2007b: Application of Genomics to Molecular Breeding of Wheat and Barley, In: C. H. Jeffery, (ed.) Advances in Genetics, 121-155, Vol. Volume 58. Academic Press.

Varshney, R.K., K. Chabane, P.S. Hendre, R.K. Aggarwal, and A. Graner, 2007c: Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Science **173**, 638-649.

Vavilov, N.I., 1940: The new systematic of cultivated plants, In: J. Huxley, (ed.) The New Systematics, 549-566. Clareon Press, Oxford.

Visscher, P.M., 2008: Sizing up human height variation. Nature Genetics **40**, 489-90.

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

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

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

von Korff, M., H. Wang, J. Léon, and K. Pillen, 2008: AB-QTL analysis in spring barley: III. Identification of exotic alleles for the improvement of malting quality in spring barley (*H. vulgare ssp. spontaneum*). Molecular Breeding **21**, 81-93.

Vu, G., T. Wicker, J. Buchmann, P. Chandler, T. Matsumoto, A. Graner, and N. Stein, 2010: Fine mapping and syntenic integration of the semi-dwarfing gene *sdw3* of barley. Functional & Integrative Genomics **10**, 509-521.

Wang, G., I. Schmalenbach, M. von Korff, J. Leon, B. Kilian, J. Rode, and K. Pillen, 2010a: 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. Theoretical and Applied Genetics **120**, 1559-74.

Wang, H., K. Smith, E. Combs, T. Blake, R. Horsley, and G. Muehlbauer, 2012: Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. Theoretical and Applied Genetics **124**, 111-124.

Wang, J., J. Yang, D. McNeil, and M. Zhou, 2010b: Identification and molecular mapping of a dwarfing gene in barley (*Hordeum vulgare* L.) and its correlation with other agronomic traits. Euphytica **175**, 331-342.

Wang, L., R.K. Newman, C.W. Newman, L.L. Jackson, and P.J. Hofer, 1993: Tocotrienol and fatty acid composition of barley oil and their effects on lipid metabolism. Plant Foods for Human Nutrition (Formerly Qualitas Plantarum) **43**, 9-17.

Waugh, R., J.L. Jannink, G.J. Muehlbauer, and L. Ramsay, 2009: The emergence of whole genome association scans in barley. Curr Opin Plant Biol **12**, 218-22.

Weir, B.S., 1996: Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates, Sunderland, Massachusetts.

Weir, B.S., and C.C. Cockerham, 1984: Estimating F-Statistics for the analysis of population structure. Evolution **38**, 1358-1370.

Wenzl, P., J. Carling, D. Kudrna, D. Jaccoud, E. Huttner, A. Kleinhofs, and A. Kilian, 2004: Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proceedings of the National Academy of Sciences of the United States of America **101**, 9915-9920.

Wenzl, P., H. Li, J. Carling, M. Zhou, H. Raman, E. Paul, P. Hearnden, C. Maier, L. Xia, V. Caig, J. Ovesná, M. Cakir, D. Poulsen, J. Wang, R. Raman, K. Smith, G. Muehlbauer, K. Chalmers, A. Kleinhofs, E. Huttner, and A. Kilian, 2006: A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics **7**, 206.

Wright, S., 1951: The Genetical Structure of Populations. Annals of Eugenics **15**, 323-354.

Xu, Y., and J.H. Crouch, 2008: Marker-Assisted Selection in Plant Breeding: From Publications to Practice. Crop Sci. **48**, 391-407.

Yahiaoui, S., E. Igartua, M. Moralejo, L. Ramsay, J. Molina-Cano, F. Ciudad, J. Lasa, M. Gracia, and A. Casas, 2008: Patterns of genetic and eco-geographical diversity in Spanish barleys. TAG Theoretical and Applied Genetics **116**, 271-282.

Yan, J., M. Warburton, and J. Crouch, 2011: Association Mapping for Enhancing Maize (*Zea mays* L.) Genetic Improvement. Crop Sci. **51**, 433-449.

Yu, G., R. Horsley, B. Zhang, and J. Franckowiak, 2010: A new semi-dwarfing gene identified by molecular mapping of quantitative trait loci in barley. Theoretical and Applied Genetics **120**, 853-861.

Yu, H., W. Xie, J. Wang, Y. Xing, C. Xu, X. Li, J. Xiao, and Q. Zhang, 2011: Gains in QTL Detection Using an Ultra-High Density SNP Map Based on Population Sequencing Relative to Traditional RFLP/SSR Markers. PLoS ONE **6**, e17595.

Yu, J., Z. Zhang, C. Zhu, D.A. Tabanao, G. Pressoir, M.R. Tuinstra, S. Kresovich, R.J. Todhunter, and E.S. Buckler, 2009: Simulation Appraisal of the Adequacy of Number of Background Markers for Relationship Estimation in Association Mapping. Plant Gen. **2**, 63-77.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler, 2006: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics **38**, 203-8.

Zakhrabekova, S., S.P. Gough, I. Braumann, A.H. Müller, J. Lundqvist, K. Ahmann, C. Dockter, I. Matyszczak, M. Kurowska, A. Druka, R. Waugh, A. Graner, N. Stein, B. Steuernagel, U. Lundqvist, and M. Hansson, 2012: Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. Proceedings of the National Academy of Sciences.

Zhang, J., and W. Zhang, 2003: Tracing sources of dwarfing genes in barley breeding in China. Euphytica **131**, 285-293.

Zhang, J., Z. Li, and C.H. Zhang, 2007: Analysis of dwarfing genes in Zhepi 1 and Aizao 3: Two dwarfing gene donors in barley breeding in China. Canadian Journal of Plant Science **87**, 93-96.

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

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

Zhao, K.Y., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C.L. Tang, C. Toomajian, H.G. Zheng, C. Dean, P. Marjoram, and M. Nordborg, 2007: An Arabidopsis example of association mapping in structured samples. Plos Genetics **3**.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu, 2008: Status and Prospects of Association Mapping in Plants. The Plant Genome Journal **1**.

Zohary, D., and M. Hopf, 2000: Domestication of plants in the old world the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley. Oxford University Press, Oxford.

9 Supplementary Material

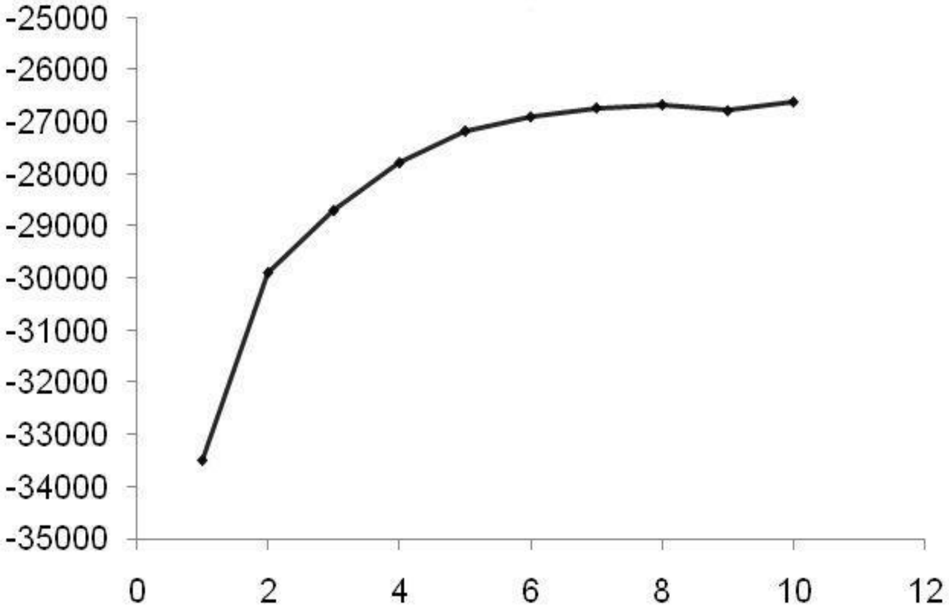
Chapter 1:

Supplementary Figures

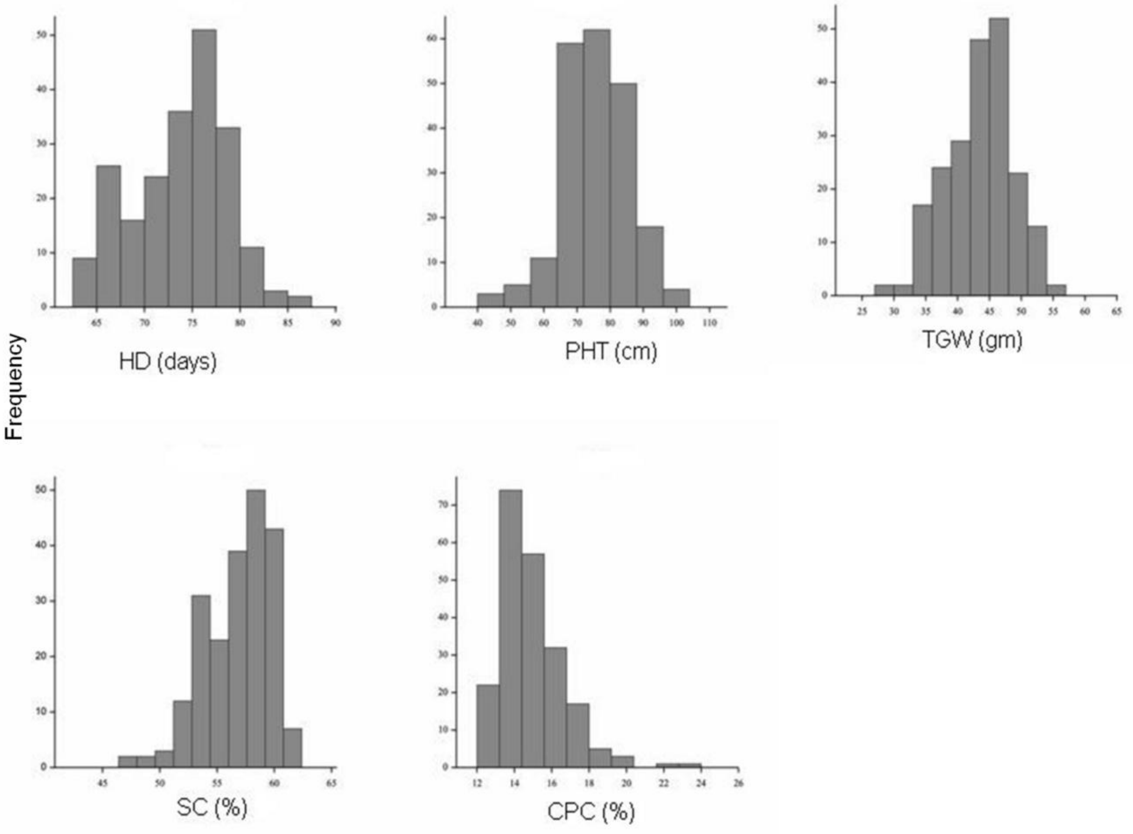
Supplementary Tables

Chapter2:

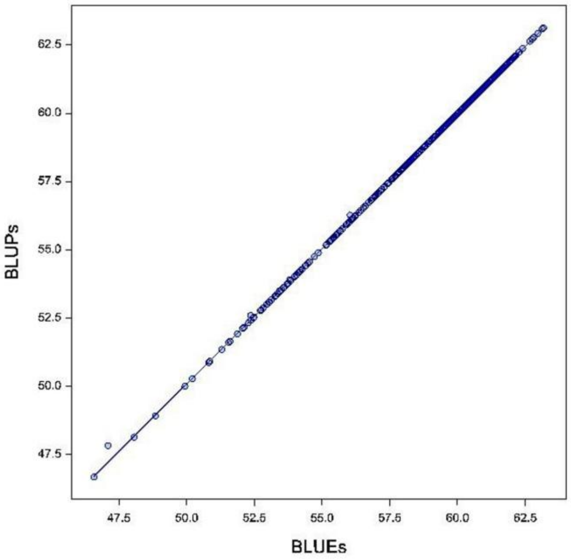
Supplementary Figures



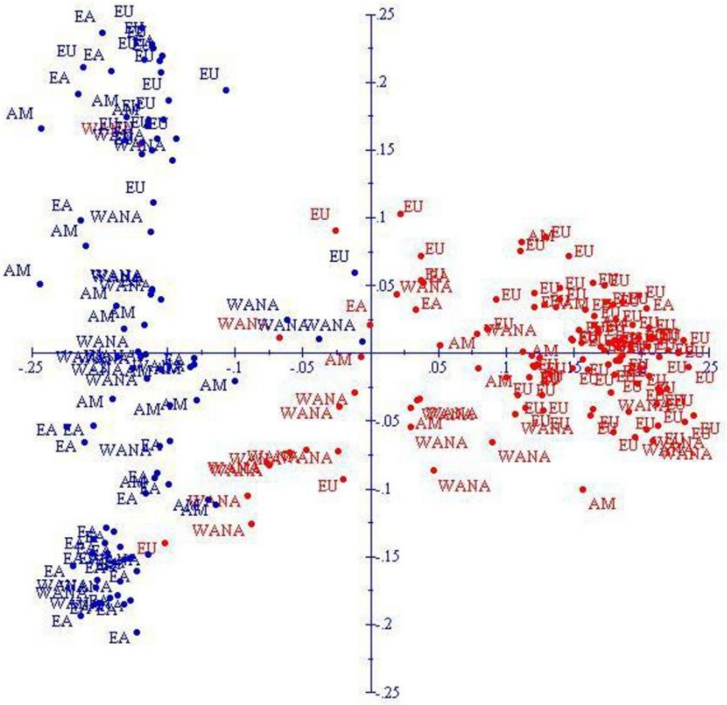
Supplementary Fig.2.1 STRUCTURE results using DArT markers. Log probability data (LnP(D)) as function of *k* (number of clusters) from the STRUCTURE run using 1088 DArT markers with the same association panel. The plateau of the graph at *K*=6 indicates the minimum number of subgroups possible in the panel



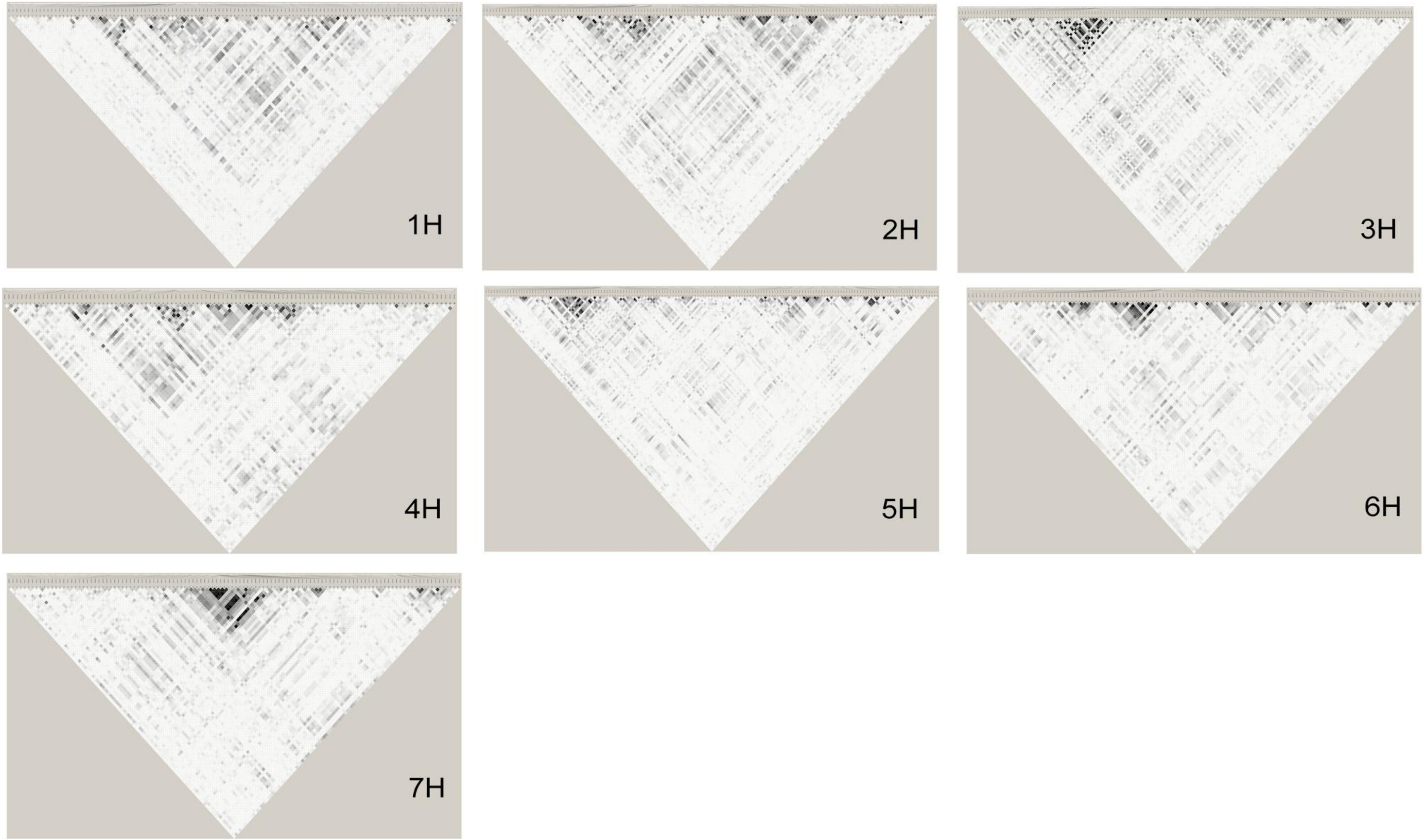
Supplementary Fig.2.2 Phenotypic distribution of 224 spring barley accessions for the traits heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and protein content (CPC)



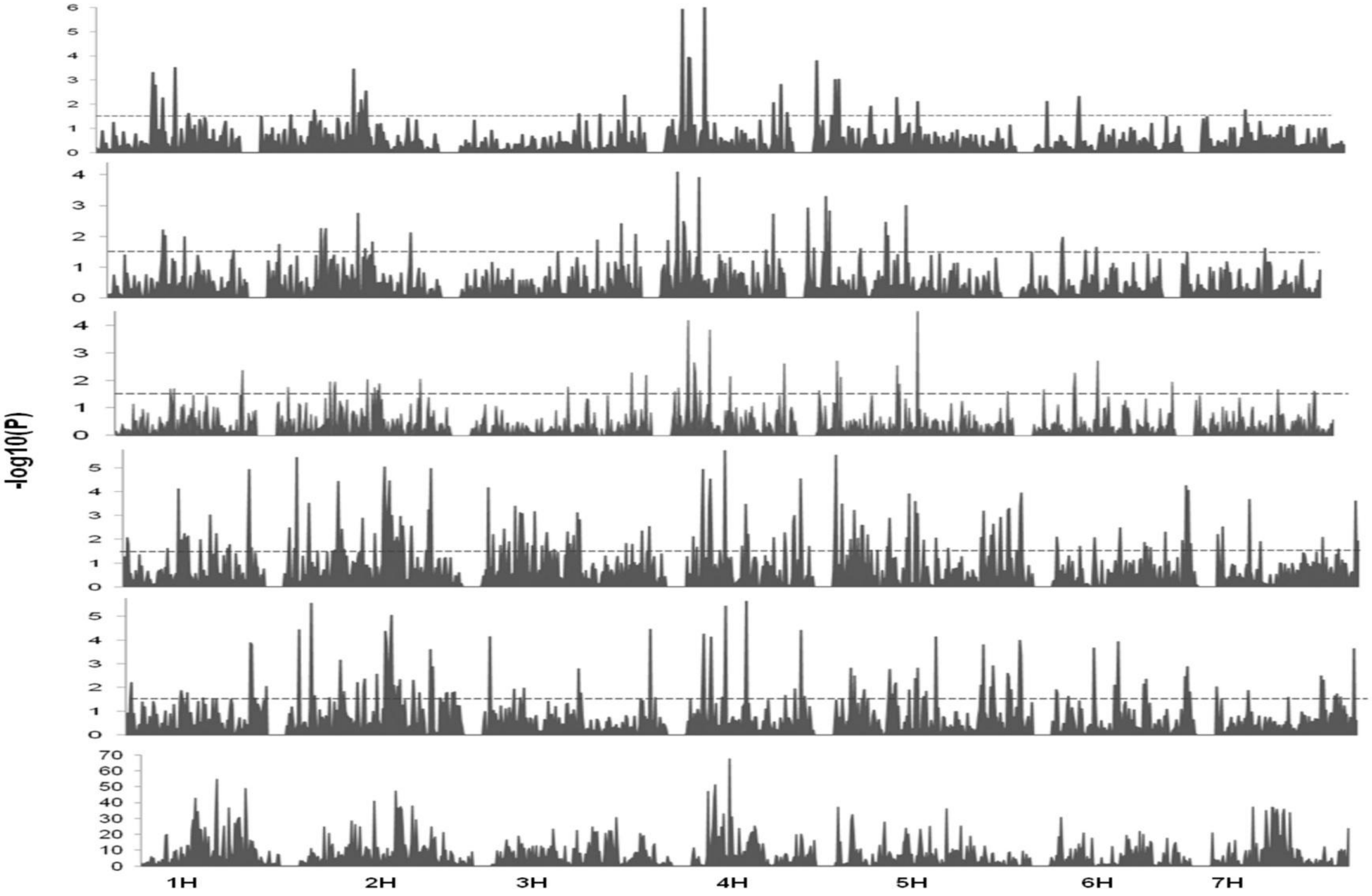
Supplementary Fig.2.3 Comparison of BLUPs and BLUEs for starch content. The graph implies that there is not much difference between the BLUPs and BLUEs in our experiment



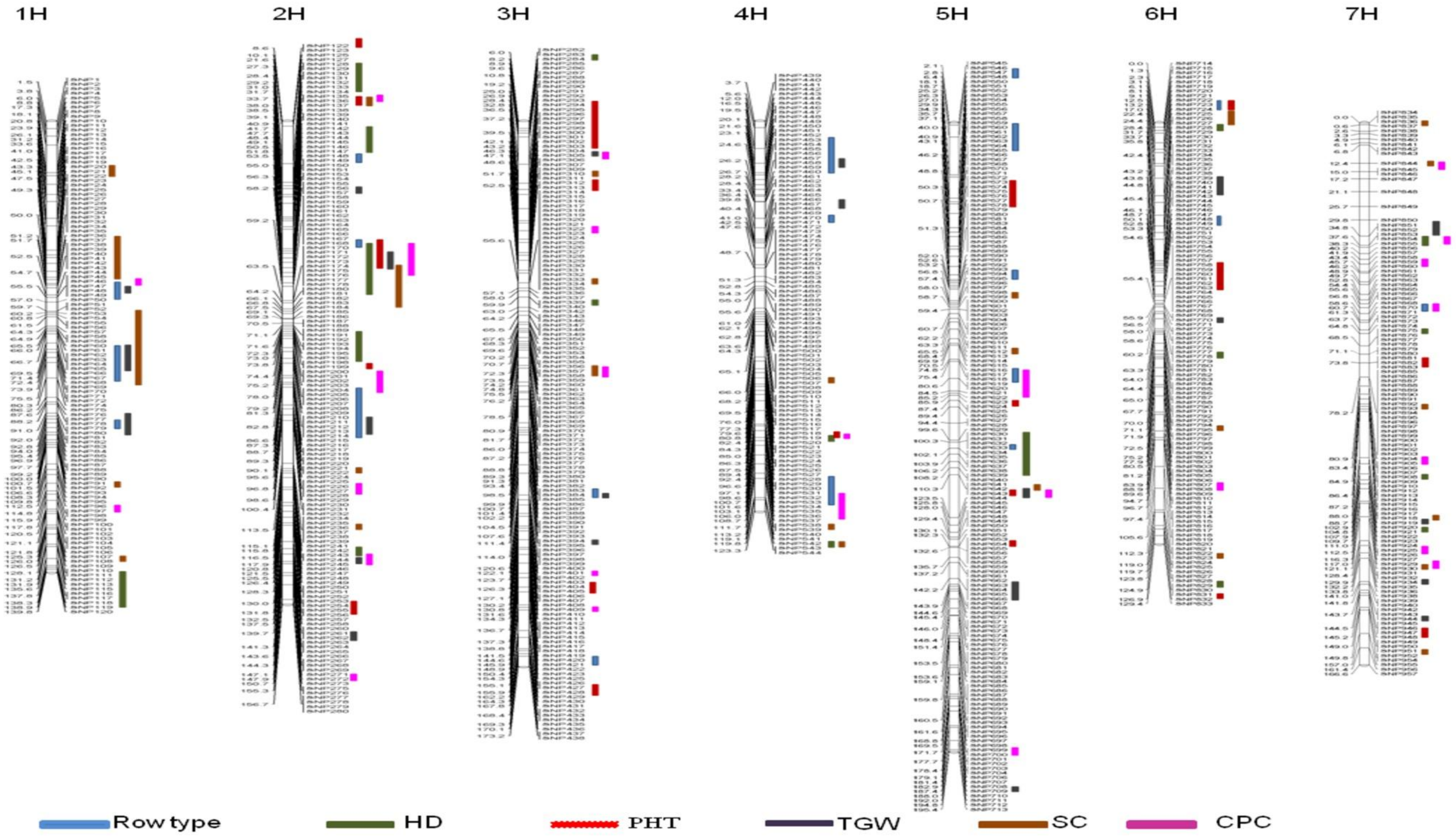
Supplementary Fig.2.4 Principal Co-ordinate analysis (PCoA) of the panel based on the first two components derived using 918 SNPs. The primary axis tends to separate into subgroups based on their spike morphology character (blue: six-rowed barley; red: two-rowed barley). Further clustering is based on origin of the accessions



Supplementary Fig.2.5: LD plots for each chromosome in barley. The color of squares illustrate the strength of pairwise r^2 values on a black and white scale, where black indicates perfect LD ($r^2 = 1.00$) while white indicates perfect equilibrium ($r^2 = 0$). Failed and monomorphic SNPs as well as SNPs with MAF < 0.05 are not considered.



Supplementary Fig.2.6 GWAS whole genome scans for row type using different association models (naive, P, Q, QK, PK and K)



Supplementary Fig.2.7 GWAS for all traits. Localization of QTL and candidate genes for the traits row type (RT), heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and crude protein content (CPC) on the genetic map with 918 SNP markers

Supplementary Tables

Supplementary Table 2.1 Information of 957 mapped SNP markers from the IPK customized OPA that were successful in our panel (**Attached CD**)

Supplementary Table 2.2 Details of the 212 accessions used for GWAS. Name of the accession, row type, number of successful markers, Structure group, and region of origin and country of origin (**Attached CD**)

Supplementary Table 2.3 Phenotypic variation among two-rowed and six-rowed groups. Estimation of means, standard deviation (SD), variation (VAR), standard error variation (SEVAR) and coefficient of variance (CV%) for each trait among two-rowed and six-rowed groups

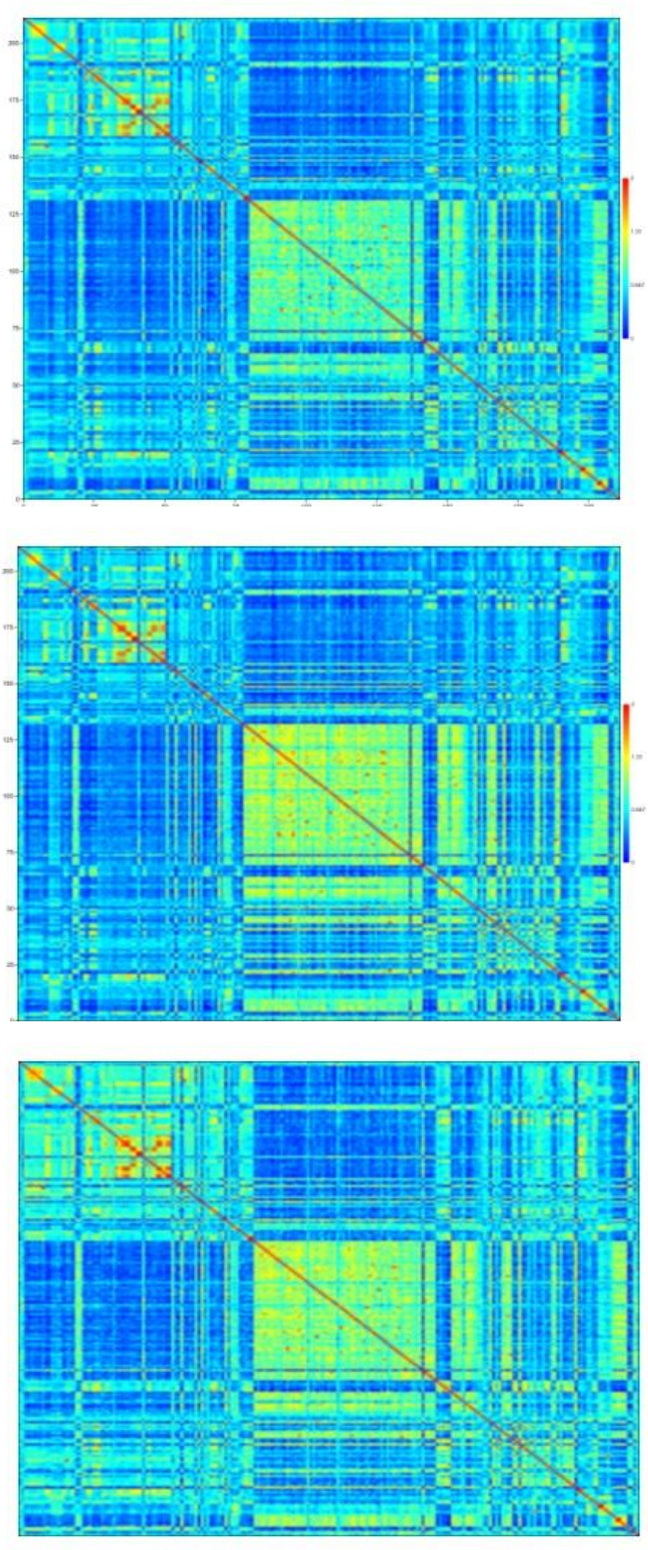
	HD		PHT		TGW		SC		CPC	
	2-rowed	6-rowed	2-rowed	6-rowed	2-rowed	6-rowed	2-rowed	6-rowed	2-rowed	6-rowed
MEAN	75.25	71.54	75.08	75.92	46.20	40.03	58.29	54.82	14.40	15.65
SD	4.45	4.80	9.01	11.83	3.41	5.25	1.92	2.85	1.21	1.97
VAR	19.79	23.01	81.24	140.01	11.60	27.56	3.70	8.10	1.47	3.90
SEVAR	2.76	2.82	11.91	22.79	1.44	3.82	0.51	1.29	0.25	0.87
% CV	5.91	6.71	12.01	15.59	7.37	13.11	3.30	5.19	8.44	12.61

Supplementary Table 2.4 Trait distribution in the whole population and subgroups. Estimation of means, SD, variation (VAR), standard error variation (SEVAR) and coefficient of variance (CV%) among all six subgroups in the panel

	Groups	Genotypes	Mean	SD	VAR	SEVAR	% CV
HD (days after sowing)	Total	212	73.68	4.94	24.40	1.93	6.70
	1	23	70.86	4.69	21.95	3.65	6.61
	2	31	70.42	5.16	26.60	5.85	7.32
	3	31	73.57	3.95	15.60	4.57	5.37
	4	24	75.41	5.22	27.24	9.08	6.92
	5	79	76.13	3.42	11.69	2.79	4.49
	6	23	70.83	5.18	26.88	6.98	7.32
PHT (cm)	Total	212	75.43	10.28	105.82	11.92	13.63
	1	24	73.50	7.56	57.22	19.85	10.29
	2	31	69.07	9.80	96.08	31.60	14.19
	3	31	85.95	8.02	64.36	15.21	9.33
	4	24	83.92	9.47	89.65	36.00	11.28
	5	79	74.17	7.35	54.00	13.18	9.91
	6	23	67.38	9.06	82.07	22.19	13.44
TGW (g)	Total	212	43.58	5.25	27.60	2.57	12.05
	1	24	44.05	5.78	33.45	7.77	13.13
	2	31	38.10	4.15	17.21	5.15	10.89
	3	31	38.89	4.79	22.91	7.00	12.31
	4	24	45.19	3.87	14.95	4.41	8.56
	5	79	46.07	3.08	9.48	1.42	6.68
	6	23	46.54	4.24	17.96	5.10	9.11
SC (%)	Total	212	56.82	2.91	8.47	0.90	5.12
	1	24	54.64	3.04	9.27	3.19	5.57
	2	31	53.64	2.84	8.07	1.82	5.30
	3	31	56.52	1.82	3.30	0.79	3.21
	4	24	57.18	1.70	2.89	0.96	2.97
	5	79	59.19	1.32	1.75	0.38	2.23
	6	23	55.26	2.11	4.47	0.82	3.83
CPC (%)	Total	212	14.93	1.69	2.87	0.47	11.36
	1	24	15.43	1.91	3.66	1.81	12.39
	2	31	17.02	1.93	3.71	1.35	11.32
	3	31	14.47	1.19	1.41	0.37	8.20
	4	24	15.17	1.21	1.46	0.60	7.97
	5	79	13.89	0.93	0.86	0.28	6.68
	6	23	15.54	1.06	1.13	0.22	6.84

Chapter 3:

Supplementary Figures

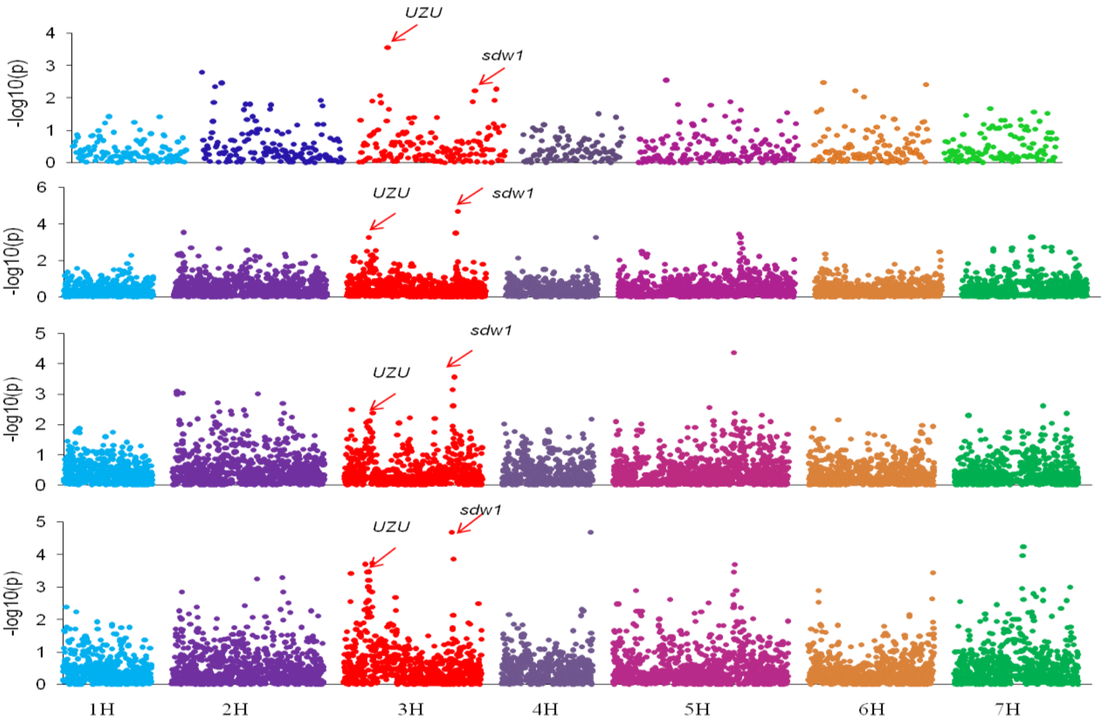


(a) Kinship matrix generated from 7000 iSelect markers (K_1)

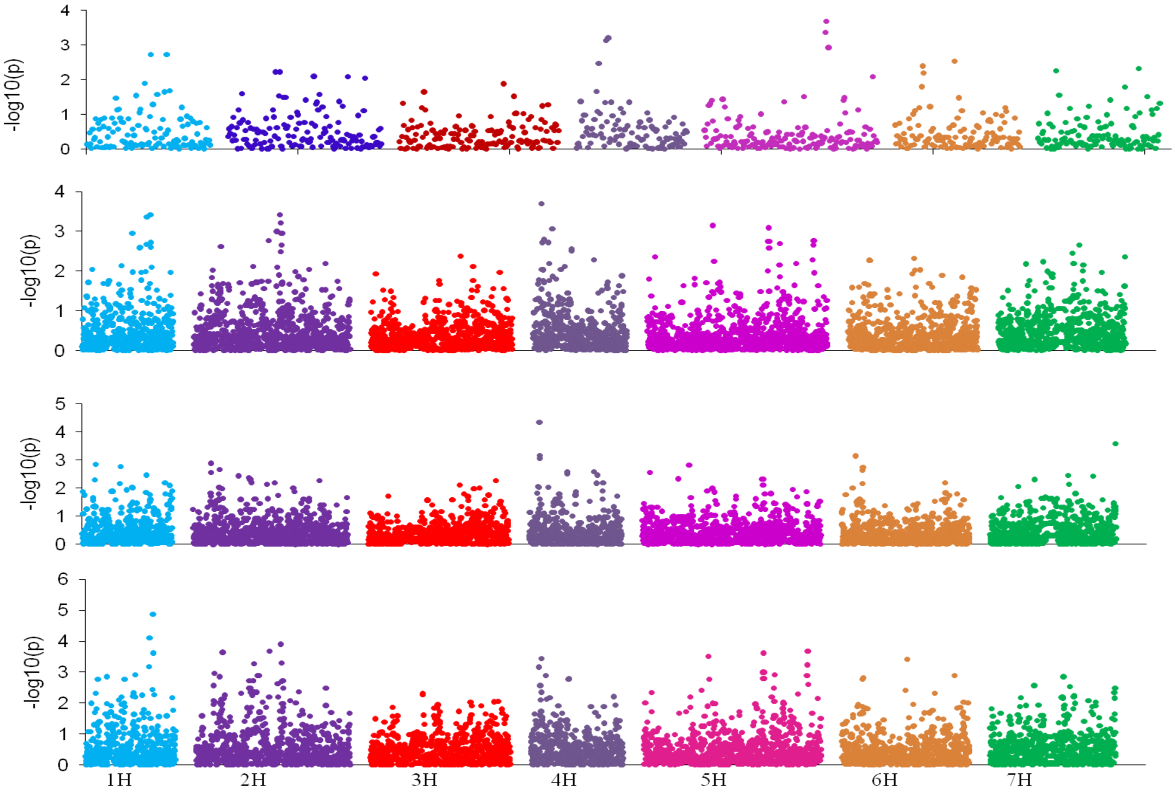
(b) Kinship matrix generated from 918 SNPs from IPK-OPA (K_2)

(c) Kinship matrix generated from uniformly distributed 362 SNPs (K_3)

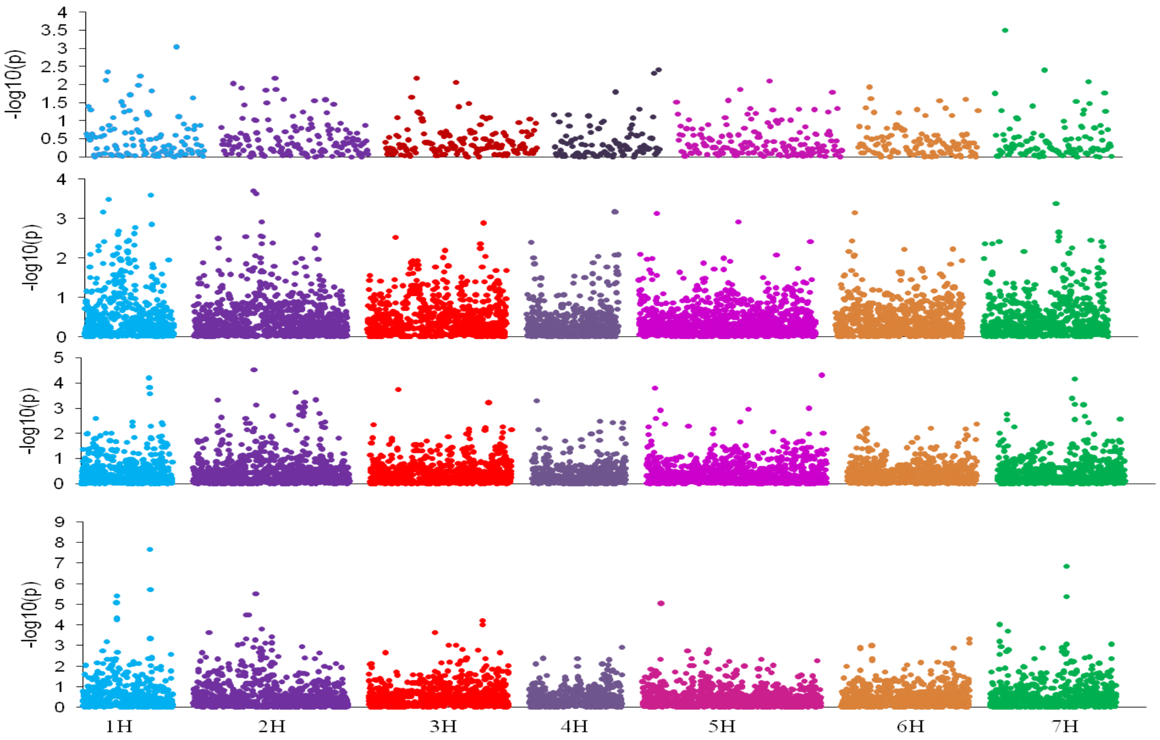
Supplementary Fig 3.1 Heat plots of different kinship matrices. The figure shows heat plots of Kinship developed from (a) all iSelect markers (K_1), (b) with 918 SNPs from IPK-OPA (K_2), and (c) with selected 362 SNPs (K_3)



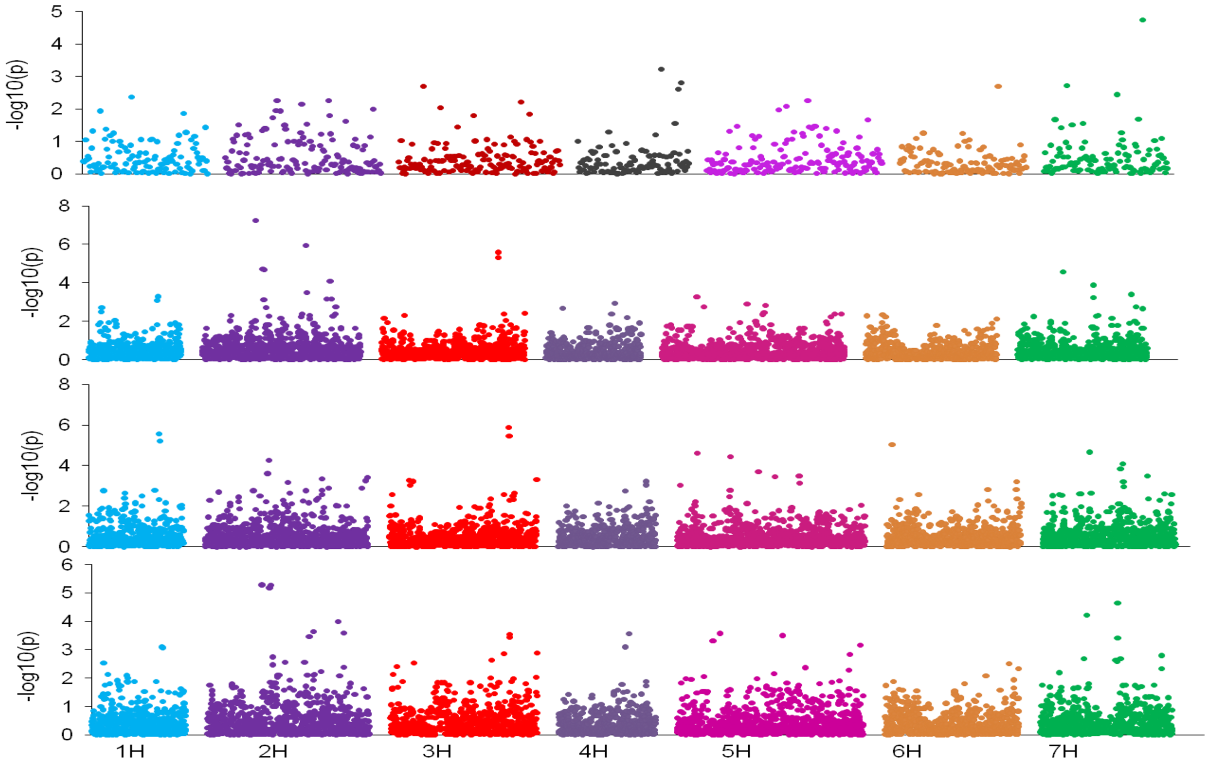
Supplementary Fig. 3.2 GWAS scans for plant height (PHT) using K-model. (a) GWAS with 918 SNPs (IPK-OPA) (b) GWAS with 5474 SNPs (iSelect) using K_1 (c) GWAS with 5474 SNPs using K_2 (d) WGA with 5474 SNPs using kinship from K_3.



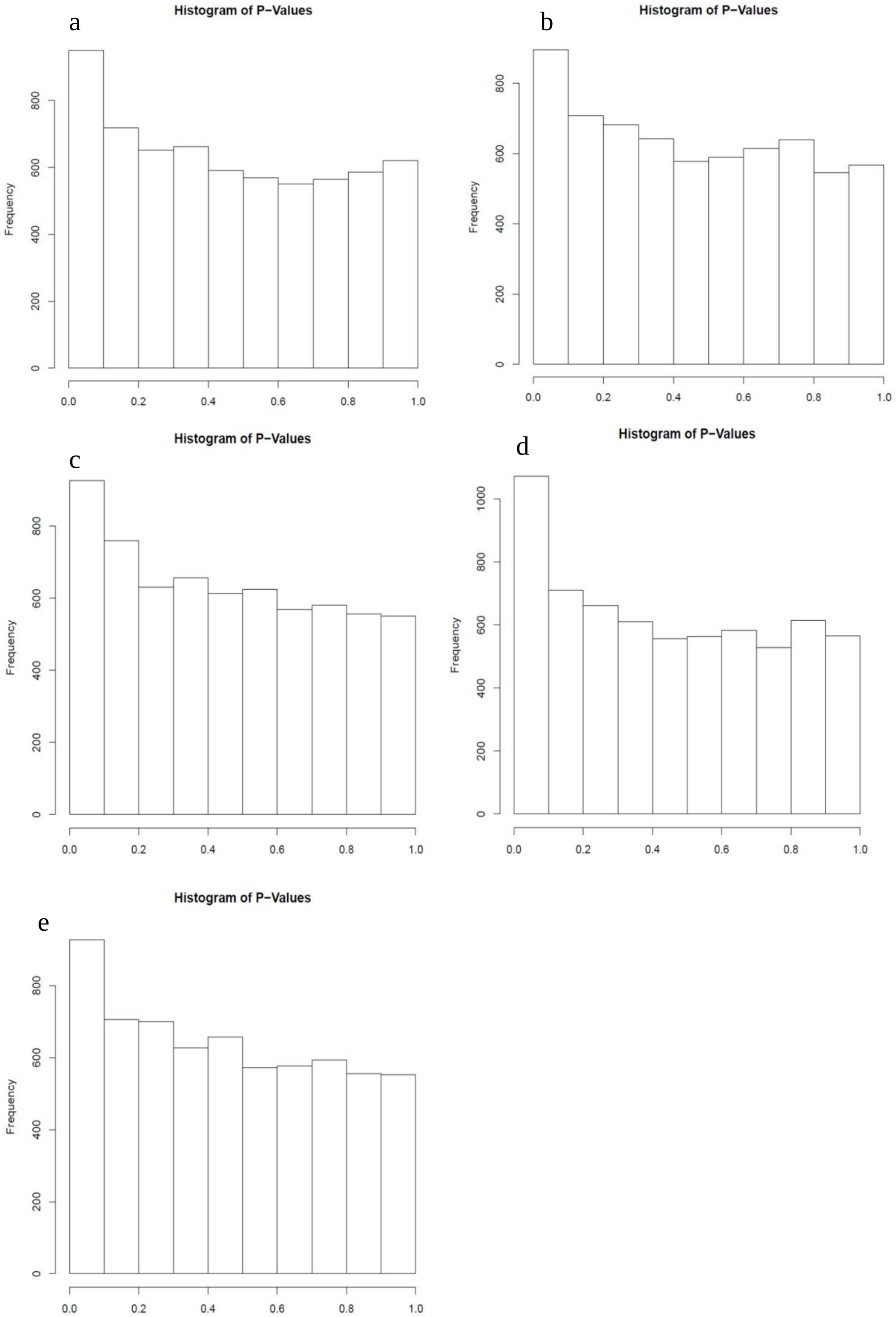
Supplementary Fig 3.3 GWAS scans for thousand grain weight (TGW) using K-model. (a) GWAS with 918 SNPs (IPK-OPA) (b) GWAS with 5474 SNPs (iSelect) using K_1 (c) GWAS with 5474 SNPs using K_2 (d) GWAS with 5474 SNPs using kinship from K_3.



Supp Fig. 3.4 GWAS scans for starch content (SC) using K-model. (a) GWAS with 918 SNPs (IPK-OPA) (b) GWAS with 5474 SNPs (iSelect) using K_1 (c) GWAS with 5474 SNPs using K_2 (d) GWAS with 5474 SNPs using kinship from K_3.



Supp Fig. 3.5 GWAS scans for protein content (CPC) using K-model. (a) GWAS with 918 SNPs (IPK-OPA) (b) GWAS with 5474 SNPs (iSelect) using K_1 (c) GWAS with 5474 SNPs using K_2 (d) GWAS with 5474 SNPs using kinship from K_3.



Supp Fig. 3.6. Distribution of *P*-values from the GWA analysis of each trait with iSelect markers using K-model. a) HD b) PHT c) TGW d) SC and e) CPC

List of Supplementary Tables

Supp Table. 3.1 List of all iSelect SNPs and their diversity statistics. Successful SNPs, their allele frequency, PIC and Gene diversity values. ‘F’ represents failed SNPs and ‘S’ represents successful SNPs (**Attached CD**)

Supp Table. 3.2 GWAS results for trait row type using iSelect SNPs with K-model**.** Significant SNPs from iSelect associated to trait row type, corresponding map position, *P*-value of association, variance explained by marker (R^2), effect of the significant marker. The markers with map ‘MxB’ are the mapped using Morex X Barke RIL population and their map positions correspond to MxB map. The markers with map LD are mapped by linkage disequilibrium mapping approach using bin map positions (Comadran et al unmapped). Column *q*FDR shows the SNPs crossing the FDR threshold (**Attached CD**)

Supp Table. 3.3 GWAS results for trait heading date using iSelect SNPs with K-model. Significant SNPs from iSelect associated to trait heading date, corresponding map position, *P*-value of association, variance explained by marker (R^2), effect of the significant marker (**Attached CD**)

Supp Table. 3.4 GWAS results for trait plant height using iSelect SNPs with K-model. Significant SNPs from iSelect associated to trait plant height, corresponding map position, *P*-value of association, variance explained by marker (R^2), effect of the significant marker (**Attached CD**)

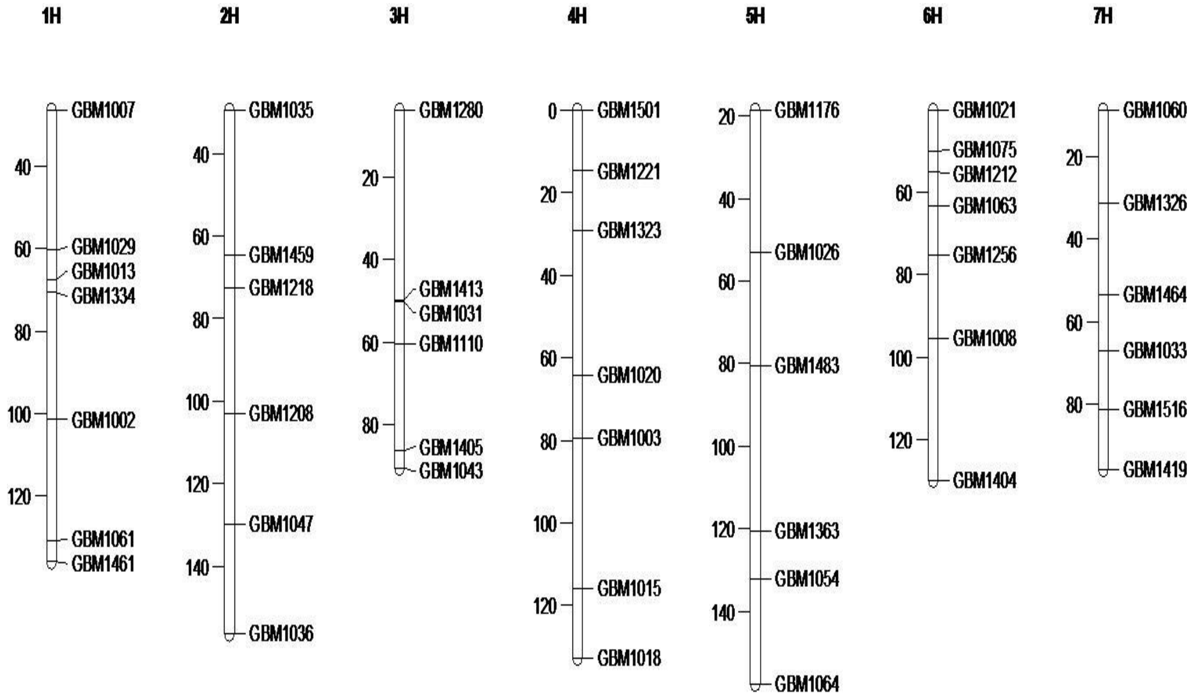
Supp Table. 3.5 GWAS results for trait thousand grain weight using iSelect SNPs with K-model. Significant SNPs from iSelect associated to trait thousand grain weight, corresponding map position, *P*-value of association, variance explained by marker (R^2), effect of the significant marker (**Attached CD**)

Supp Table. 3.6 GWAS results for trait starch content using iSelect SNPs with K-model. Significant SNPs from iSelect associated to trait starch content, corresponding map position, *P*-value of association, variance explained by marker (R^2), effect of the significant marker (**Attached CD**)

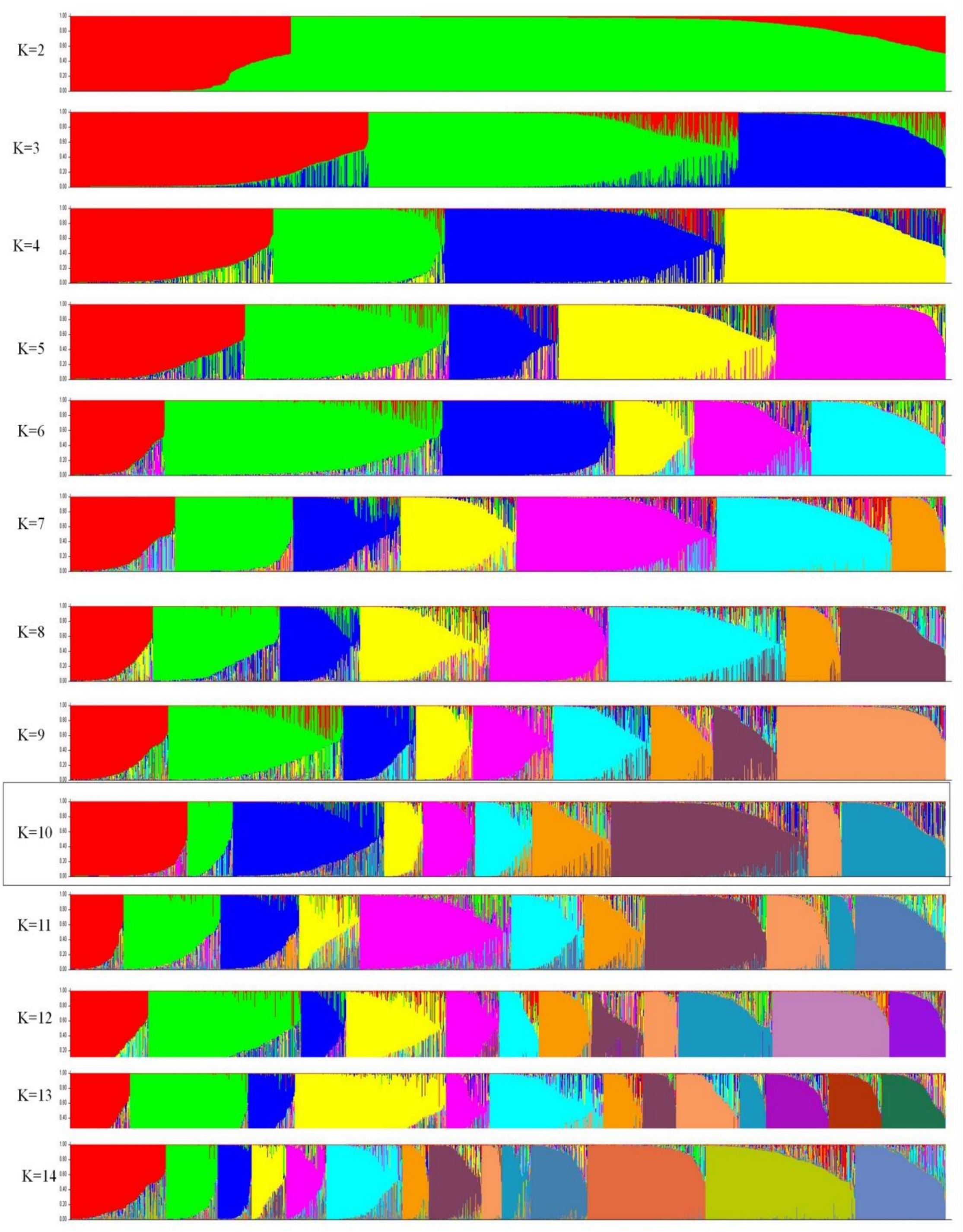
Supp Table. 3.7 GWAS results for trait crude protein content using iSelect SNPs with K-model. Significant SNPs from iSelect associated to trait crude protein content, corresponding map position, *P*-value of association, variance explained by marker (R^2), effect of the significant marker (**Attached CD**)

Chapter 4:

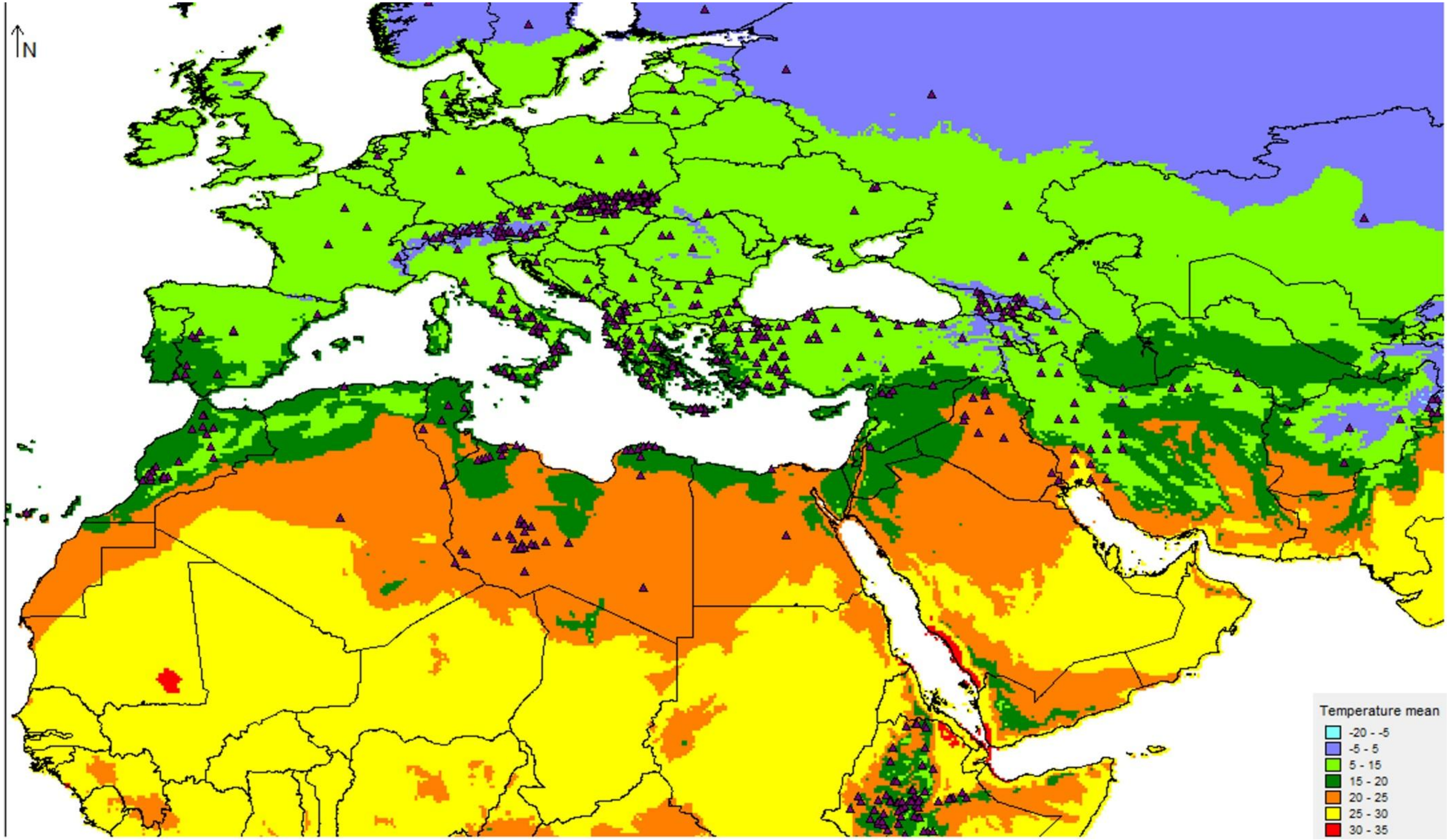
Supplementary Figures



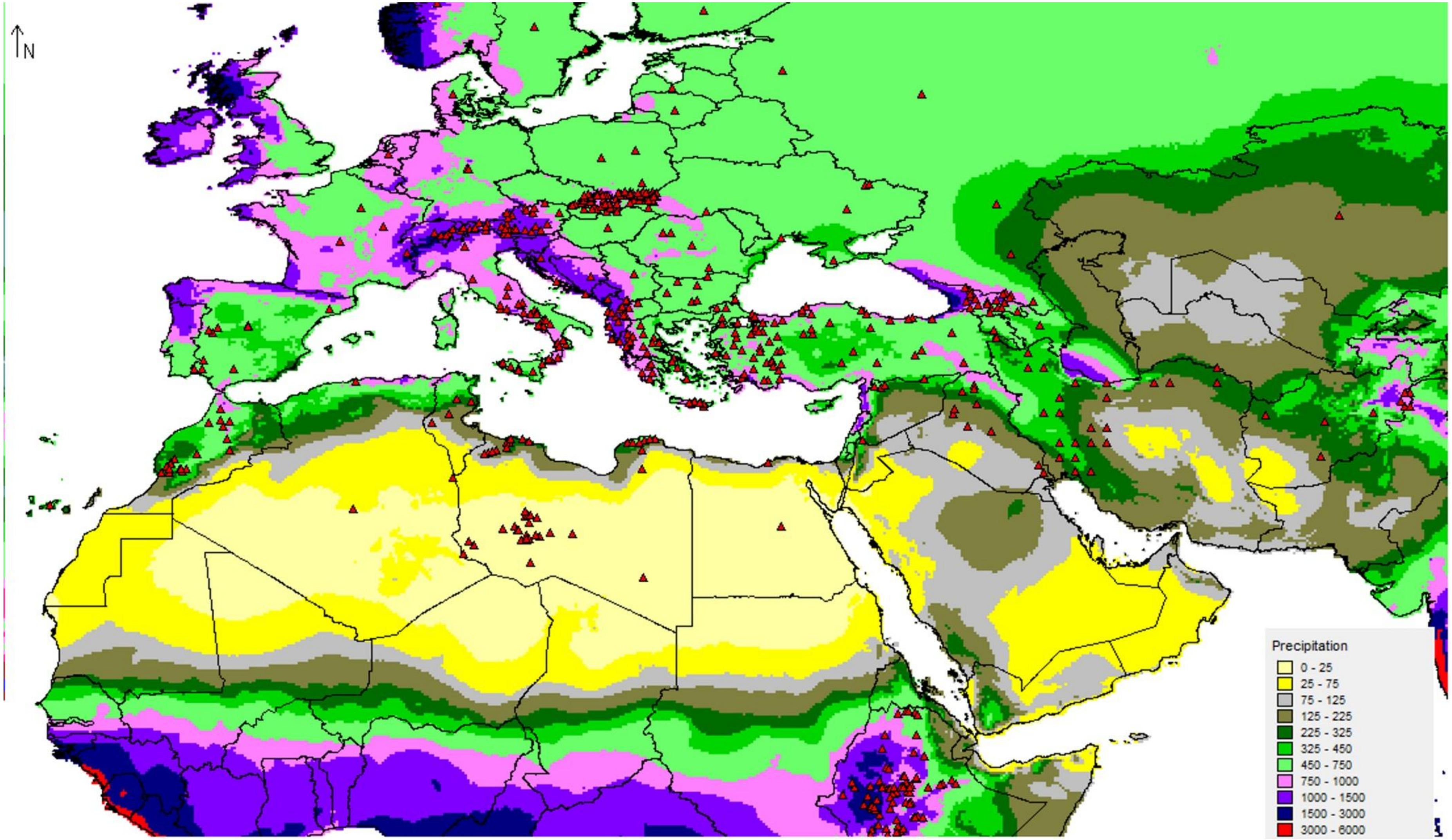
Supplementary Fig. 4.1: Distribution of the 42 SSR markers across the seven linkage groups



Supplementary Fig. 4.2: STRUCTURE results for *k*=2 to 12 using 42 SSR data in 1491 barley landraces. Population structure plots inferred for different number of proposed groups



Supplemenatry Fig. 4.3: Geographical distribution of 1491 landraces over various temperature regimes. The annual mean temperature data is projected over the landrace collection sites. The landrace collection is spread across all the extreme temperature regimes. Triangle symbols represent each landrace.



Supplementary Fig. 4.4: Geographical distribution of 1491 landraces over various precipitation regimes. The annual precipitation data is projected over the landrace collection sites. Triangle symbols represent each landrace

Supplementary Tables

Supplementary Table 4.1 List of all accessions in landrace collection. Accessions collection sites, Geo-references and botanical nomenclature. The STRUCTURE inferred group for each of the accession is included (**Attached CD**)

Supp Table 4.2: Mantel correlogram tables. Class indicates different distance classes. Min and Max are the lower and upper boundary values of each class. "Pairs" is number of pairs for which correlation is calculated within each distance class. "Mantel r" is the mantle correlation for each class and P-value is the significance of mantel correlation for each class. Mantel correlogram tables between: (a) genetic distance and geographic distance (b) genetic distance and longitude difference matrix (c) genetic distance and latitude difference matrix (d) genetic distance and annual mean temperature (e) genetic distance and mean diurnal range (f) genetic distance and temperature of warmest quarter (g) genetic distance and annual precipitation.

Supplementary Table 4.2a Mantel correlogram between genetic distance and geographic distance

Class	Min	Max	Pairs	Mantel r	P-value
1	0	300	73503	0.52327	0.002
2	300	600	51623	0.4131	0.002
3	600	900	54160	0.32875	0.002
4	900	1200	55200	0.25814	0.002
5	1200	1500	52360	0.2235	0.002
6	1500	1800	48593	0.20506	0.002
7	1800	2100	73270	0.16337	0.002
8	2100	2400	59780	0.14068	0.002
9	2400	2700	67085	0.10775	0.002
10	2700	3000	55063	0.10537	0.002
11	3000	3300	74808	0.06373	0.002
12	3300	3600	56716	0.05144	0.002
13	3600	3900	64816	0.02578	0.002
14	3900	4200	43833	0.02011	0.002
15	4200	4500	69709	0.00342	0.03393
16	4500	4800	53184	0.0064	0.002
17	4800	5200	86547	0.00311	0.03992
18	5200	5500	32814	0.001	0.17166
19	5500	7000	32051	0.00969	0.002
20	7000	8000	4249	0.0019	0.002

Supplementary Table 4.2b. Mantel correlogram between genetic distance and longitude difference matrix

Class	Min	Max	Pairs	Mantel r	P-value
1	0	10	372933	0.30142	0.000
2	10	15	99706	0.11062	0.000
3	15	20	133628	0.06165	0.000
4	20	25	131885	0.04863	0.000
5	25	30	124302	0.03342	0.000
6	30	35	78835	0.02128	0.000
7	35	40	28578	0.02702	0.000
8	40	45	25651	0.01875	0.000
9	45	50	41363	0.00026	0.000

Supplementary Table 4.2c. Mantel correlogram between genetic distance and latitude difference matrix

Class	Min	Max	Pairs	Mantel r	P-value
1	0	5	297005	0.30807	0.001
2	5	10	203938	0.07606	0.001
3	10	15	150791	0.03572	0.001
4	15	20	92041	0.01649	0.001
5	20	25	73157	0.002	0.045
6	25	30	90996	0.01077	0.041
7	30	35	77656	-0.02792	0.011
8	35	40	58811	0.00204	0.096
9	40	45	61506	0.00296	0.026
10	45	65	4894	-0.00026	0.473

Supplementary Table 4.2d. Mantel correlogram between genetic distance and annual mean temperature (AMT) difference matrix

Class	Min	Max	Pairs	Mantel r	P-value
1	0	4	367291	0.18153	0.001
2	4	8	300352	0.0327	0.001
3	8	12	225124	0.00794	0.001
4	12	16	129637	-0.02379	0.001
5	16	20	62951	0.01253	0.001
6	20	24	20489	-0.00525	0.001
7	24	28	4951	-0.00109	0.001

Supplementary Table 4.2e. Mantel correlogram between genetic distance and mean diurnal range (MDR) difference matrix

Class	Min	Max	Pairs	Mantel r	P-value
1	0	0.3	92152	0.15405	0.001
2	0.3	0.6	76932	0.11576	0.001
3	0.6	1.2	127131	0.10915	0.001
4	1.2	1.8	118454	0.07773	0.001
5	1.8	2.5	122936	0.08152	0.001
6	2.5	3.5	126397	0.04573	0.001
7	3.5	4.5	133784	0.01896	0.001
8	4.5	5.5	113506	0.00025	0.876
9	5.5	6.5	106984	-0.00941	0.081

Supplementary Table 4.2f. Mantel correlogram between genetic distance and mean temperature of warmest quarter (MTW) difference matrix

Class	Min	Max	Pairs	Mantel r	P-value
1	-5	5	536095	0.14177	0.001
2	5	10	332327	0.01248	0.001
3	10	15	164292	0.01382	0.001
4	15	20	63174	-0.01488	0.001
5	20	25	13556	-0.00465	0.041
6	25	28	1121	-0.00023	0.002
7	28	30	230	-0.00008	0.008

Supplementary Table 4.2g. Mantel correlogram between genetic distance and annual mean precipitation (APT) difference matrix

Class	Min	Max	Pairs	Mantel r	P-value
1	0	50	90136	0.14122	0.001
2	50	100	84094	0.09324	0.001
3	100	175	120210	0.06312	0.001
4	175	250	121581	0.04899	0.001
5	250	350	145333	0.02449	0.001
6	350	450	124786	0.01413	0.001
7	450	600	152274	0.01497	0.001
8	600	750	107351	0.00268	0.018
9	750	1000	103732	-0.0134	0.001
10	1000	1500	54523	-0.0047	0.001

Chapter 5:

Supplementary Figures

Supplementary Tables

Acknowledgments

My personal experience of working with barley in field, greenhouse and laboratory is vacillating; no doubt a great learning experience. During my PhD, I treasured moments filled with delight, excitement, cheerfulness, frustration, apathy, confidence, success and contentment. I will not be overstating, if I assert my success at all junctures of PhD in the last four years was possible only because of the kind and helpful people around me.

Foremost, I would like to express my deep and sincere gratitude to Prof. Dr. Andreas Graner for giving me this opportunity to work in the GABI -GENOBAR project. My work in this project was conceived and initiated by Prof. Graner after the results of GABI-Genoplante. This gave me a platform for easy takeoff initially, and hence would like to thank all the GABI-Genoplante partners for providing the initial phenotype data. This thesis would not be possible without the support, help, advice and encouragement of my principal supervisor Prof. Dr. Andreas Graner. His unsurpassed knowledge of barley, his enthusiasm in science, his efforts to explain the things and his insightful criticism has motivated, inspired and helped me to complete my PhD. I would like to thank Dr. Benjamin Kilian for his good advices, support, help and contributions. Dr. Kilian critically read and improved my thesis and all my manuscripts. I acknowledge him for all his suggestions throughout the work and thesis writing which are invaluable. I thankfully acknowledge the help and suggestions provided by Prof. Dr. Klaus Pillen, Head of the chair of Plant breeding MLU, Halle. Statistical support provided by Prof. Fred van Eeuwijk and Dr. Marcos Malosetti (University of Wageningen) was instrumental in completing and validating my results. I acknowledge Fred and Marcos for all their support and lengthy discussions.

I am grateful to my colleague and dear friend Rajiv Sharma for all his information, help, suggestions and assistance in work, analysis and writing. Rajiv was always there to discuss any subject starting from science to politics and it was a great fun to work along with him. I sincerely wish him all the best for his future scientific career. I would like to thank Dr. Nils stein, Dr. Kerstin Neumann, Dr. Stephan Weise and other colleagues for their interactive and fruitful scientific discussions and inputs. I also wish to thank all the Genome Diversity group scientific and technical staff during the years I worked at IPK. I would like to mention and thank Kathrin baake, Kerstin Wolf, Peter Schreiber, Jürgen Marlow and other field technicians for their help in the lab, greenhouse and field experiments. Special thanks to Naser for all the coffee break discussions which helped to vent off the work stress. I am also fortunate enough to come across many good friends here at IPK and thank all of them for their support, encouragement and happy times spent together. Especially I am grateful to dear friends Harsha, Rajesh, Geetha and Shailendra for making me feel at home here in Gatersleben.

I owe a debt of special thanks to Dr. Heiko Parzies (1959-2011) for inspiring me to choose scientific career. I am grateful for all his support, encouragement and advices during master thesis and later on his continued encouragement till 2011. I always honor him and his memory for being a great human with generous helping nature and as a person with great enthusiasm in science and its applications to humankind.

I also wish to acknowledge all the collaborating partners of GABI GENOBAR consortium for their contributions to the discussions and project. I am grateful to the project principal investigator Dr. Frank Ordon, for his interesting and stimulating comments and feedback on the work during the course of the project. I would like to thank German ministry of education and research (BMBF) for funding my work through GABI program.

Above all, I am grateful to almighty god for giving me health, strength, opportunity and fortitude to pursue my dreams. My parents, brothers and family are always there for me and motivated me with their unequivocal support in all possible ways. My brothers Krishna and Deevenaiah always stood by me and provided the emotional support and encouragement needed. My thanks wouldn’t be sufficient and I will always be indebted to them for lifetime. Special thanks to my wife kavitha for her patience, understanding and personal support at all the time. Special mention for my daughter Abhishikta is rightful, due to the joyous bliss and abiding hope she brought into my life.

This Thesis is dedicated to my sweet daughter ABHISHIKTA who came into my life on 30th September 2011

Curriculum vitae

Raj Kishore Pasam

Correnstrasse 3,	E-mail: pasam@ipk-gatersleben.de
06466 Gatersleben,	Phone (Off): +49-(0)39482-5455
Germany.	Phone (Mob): +49-(0)17662010405

Gender	**Nationality**	**Date of birth**
Male	Indian	August 15, 1982

Academic Qualifications:

June 2008 to present: PhD, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany.

- Thesis: "*Whole genome association studies in a collection of worldwide spring barley*"

April 2006 to June 2008: University of Hohenheim, Stuttgart, Germany

- Master of Sciences in Organic Food Chain Management (OGPA 3.0/4.0)
  Specialization in Plant Breeding and Genetics. (Faculty of Agriculture Sciences)

- Master's thesis: "*Evaluation of photoperiod response and characterization of population structure of pearl millet germplasm from west and central Africa*"

2000 to June 2004: Acharya N.G. Ranga Agricultural University, Hyderabad, India.

- 2000 to 2004 Bachelor of Agriculture Sciences (B.Sc.Ag)
- Major Academic highlights: Both basic and practical courses in Agronomy, Entomology, Plant Pathology, Plant Breeding and Population Genetics, Soil Science, Horticulture, Plant Physiology, Economics and Statistics (OGPA 3.32/4.0).

Research Interests:

My research interests lies in the field quantitative genetics and functional genomics, Specifically: (Interested in gene/ QTL mapping, cloning and the challenges beyond cloning)

- Application of Linkage mapping and whole genome association approaches for fine mapping and gene discovery
- Investigate the genetics underlying the complex nature of physiological traits
- Investigating physiological and genetic parameters underlying abiotic and biotic stress responses
- Origin, maintenance and utilization of natural diversity (both genetic and phenotypic)
- Developing molecular tools and genomic resources that could enable in trait dissection

List of Publications in connection to the thesis

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

2. **Pasam, R.K.,** R. Sharma, M. Malosetti, F.A. van Eeuwijk, G. Haseneyer, B. Kilian, and A. Graner, 2012: Genome-wide association studies for Agronomical Traits in a world wide Spring Barley Collection. BMC Plant Biology **12**, 16

3. **Pasam R.K.**, Sharma R, Kilian B, Andreas Graner: "Analysis of genetic diversity and population structure in spring barley landraces and pertinence for association mapping" (ready to submit).

4. **Pasam RK**, Sharma R, Kilian B, Andreas Graner: "Genome wide association a in spring barley cultivar collection using >7000 SNP markers" (in preparation)

Posters and Talks in connection to the thesis

1. **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.
2. **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
3. **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
4. Rajiv Sharma, **Raj K Pasam**, Benjamin Kilian, Andreas Graner: Genome wide association studies for plant architecture traits in different barley genepools. Genomics Based Breeding Conference, Giessen, 26-28.10.2010
5. 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
6. **Raj K Pasam**, Rajiv Sharma, Grit Haseneyer, Benjamin Kilian, Andreas Graner: Whole Genome Association Mapping in Spring Barley. GPZ-Tagung in Freising, 15. - 17. March 2010 (Talk)
7. **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

Gatersleben, 10. May, 2012 **(Raj Kishore Pasam)**

ERKLÄRUNG

Hiermit versichere ich an Eides Statt, dass ich die eingereichte Dissertation „*Development of stratified barley populations for association mapping studies*" selbstständig angefertigt und diese nicht bereits für eine Promotion oder andere Zwecke an einer anderen Universität eingereicht habe. Weiterhin versichere ich, dass ich die zur Erstellung der Dissertationsschrift verwendeten Arbeiten und Hilfsmittel genau und vollständig angegeben habe.

Des Weiteren erkläre ich, dass keine Strafverfahren gegen mich anhängig sind.

Halle/Saale, 10. May 2012

Raj Kishore Pasam