The University of Milan

UNIVERSITÀ DEGLI STUDI DI MILANO

Graduate School of Molecular Sciences and Plant, Food and Environmental Biotechnology

Department of Agricultural and Environmental Sciences
Production, Landscape, Agroenergy (*DiSAA*)

PhD School in Plant Biology and Crop Production
Disciplinary sector: Genetics (AGR07)

Association genetics approaches for the identification of genes associated to barley agronomic traits in a Mediterranean environment

PhD program coordinator: Prof. Piero Bianco

Supervisor: Dr. Laura Rossini

Co-supervisors: Dr. Luigi Cattivelli, and Dr. Alessandra Stella

PhD candidate: NADER Mohammed Mohammed ElBadry

December 15th, 2012

Dedication

I would like to dedicate my work to the

Spirit of my Dad

Love of my wife

Care of my mother

Encouragement of my siblings

Loyalty of my friends

Acknowledgement

I would like to express my gratitude, appreciation and thanks to my main supervisor Laura Rossini, head of Plant Genomics group, Parco Tecnologico Padano (PTP), Lodi and Department of Plant Production, Università degli Studi di Milano for her guidance, dedication and valuable pieces of advice in my research project. This work would not have been completed without her wise mentor. I am also indebted to Laura for her guidance and support when first I arrived to Italy.

I would like to express my great appreciation for my co-supervisors Dr. Luigi Cattivelli, CRA, Centro di Ricerca per la Genomica e la Postgenomica Animale e Vegetale, Fiorenzuola d'Arda and Dr. Alessandra Stella, CNR, Consiglio Nazionale delle Ricerche and head of Bioinformatics group at PTP. Their mutual guidance and support along the way of my research that made this work to be well revised and presented.

I am deeply indebted to my colleague Elahe Tavakol for her technical and scientific support. I am also indebted to Agostino Fricano for his dedication and guidance in this research project with special concern in the AFLP analysis part. Special thanks to Alessandro Tondelli for his valuable pieces of advice. Many thanks to my friends and colleagues Ahmed ElSayed, Bouabid Badaoui, and Cassia da Silva for their dedication, mutual technical and scientific help and finally proof reading of my thesis. Alberto Vecchietti for his co-supervise while Laura's leave. Raoul Pirona and Stefano Ciannamea for their support in various parts along my work. Andrea Alberini who guided me in primer design. Tiziana Fusca helped me a lot in barley threshing. Special appreciation for Elena Audia and Iban Eduardo Munoz who offered me a lot of guidance getting me to settle down when I first arrived to Lodi. Thanks again for all of them for all their technical and scientific cooperation whether in the direct advice for my research work. Thanks for the department lab manager Cinzia Colombi, PTP for facilitating the lab coordination during my work.

Thanks to the Italy-Egypt FIRB project awarded to Parco Tecnologico Padano and funded this project, Prof. Francesco Salamini. Genomics platform group where I carried out my research and benefited from being of such scientific park. Thanks to EXBARDIV consortium that granted us access to their data. Special thanks to Emmanuelle and Danielle and workers in the greenhouse at Tavazzano for facilitating all cultivation conditions to us.

Especially, I would like to express my special thanks to my wife Eman whose patient love enabled me to complete this work.

Acronyms

Abbreviation	Description
AFLP	Amplified Fragment Length Polymorphism
SNP	Single Nucleotide Polymorphism
PCR	Polymerase Chain Reaction
GWAS	Genome Wide Association Study
DNA	Deoxyribonucleic Acid
QTL	Quantitative Trait Locus
LD	Linkage Disequilibrium
AM	Association Mapping
PCA	Principal Component Analysis
PCoA	Principle Coordinate Analysis
TASSEL	Trait Analysis by aSSociation, Evolution and Linkage
DT	Drought Tolerance
CG	Candidate Gene
EST	Expressed Sequence Tag
InDel	Insertion/Deletion
FD	Flowering Date
PH	Plant Height
GY	Grain Yield
GLM	General Linear Model
MLM	Mixed Linear Model
MAF	Minor Allele Frequency
EXBARDIV	"Genomics-assisted exploitation of barley diversity" international research consortium
MCMC	Markov chain Monte Carlo
NCBI	National Center for Biotechnology Information

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

Abstract

Barley (*Hordeum vulgare*) is one of the four most economically important cereal crops worldwide. It serves as a major animal feed crop, with smaller amounts used for malting and in health food. Among environmental factors influencing barley production, drought is recognized as the most common in the Mediterranean area and this problem is expected to worsen with on-going climate changes. Breeding for drought resistance is therefore an important objective to ensure stable crop yields. In this context, genetic and molecular dissection of drought tolerance is expected to lead to the identification of key genes/loci and favourable alleles through exploration of biodiversity.

Association genetics focuses on the identification of links between phenotypic traits and genetic markers with the aim to identify QTLs and locate the underlying genes in the genome. In general, different barley genotypes are expected to carry different loci at genomic regions of interest. The probability of chromosomal recombination between two loci is proportional to the physical distance between those loci.

General objective of this project was the identification of loci subtending major agronomic traits under different water regimes, using an association genetics approach. To this end, we considered a phenotypic panel consisting of 83 barley cultivars representing European diversity for drought tolerance. This germplasm collection included 2 and 6 rows winter and spring barleys, that were previously evaluated for plant height, flowering time and yield under irrigated and rainfed conditions over three successive years (Rizza *et al.*, 2004). At the beginning of this project, an initial objective was to evaluate the potential association between allelic variants of candidate genes (CGs) selected for their known roles in drought responses and phenotypic variation for barley grain yield under different watering regimes. Thus, we re-sequenced an initial set of 3 CGs -*HvCbf2, HvCbf4a* and *HvCbf7*- and we identified 4 Single Nucleotide Polymorphisms (SNPs) in the coding region of *HvCbf4a*. In parallel, we tried to assess the genetic diversity and structure of our barley panel by utilizing the low cost molecular marker approach of Amplified Fragment Length Polymorphism (AFLP). Incorporation of structure information in association analysis is important to prevent recovery of false associations. However, when we run structure analyses using our AFLP data, we failed to recover the classical

barley subdivisions reported in the literature indicating that these markers may not be reliable for our purposes. For this reason and considering the limited chances of detecting association with few CGs, we took advantage of the recently established iSELECT Infinium® Illumina 9k SNP platform (Comadran *et al.*, 2012) to carry out a Genome Wide Association Study (GWAS) on our barley panel. Filtering out the monomorphic and failed-genotyped markers resulted in the identification of a total of 4,661 SNPs distributed over the 7 barley chromosomes. Population stratification was investigated with a subset of 260 SNPs selected as highly informative using admixture model implemented in *Structure* software. Accordingly, three main subgroups were identified corresponding to winter-2rows, winter-6rows and spring-2rows barleys, respectively. Association between barley genotypic data and flowering date (FD) was processed using general linear model and afterward compared with mixed linear model.

Quantitative Trait Loci (QTLs) controlling grain yield (GY), flowering date (FD), and plant height (PH) were identified using the general linear model. Our results provide a starting point for the identification of potentially useful genes and markers for future applications in barley breeding schemes.

CHAPTER 1

INTRODUCTION AND LITRATURE REVIEW

Introduction

In the 21st century, cereals continue to constitute the most important crops with an annual output of more than 2 billion tons (FAOSTAT, 2011; http://www.fao.org). In today's worldwide production, barley (*Hordeum vulgare* L.) ranks fourth among cereals and is considered one of the top ten crop plants in the world (Akar *et al*., 2004). Barley is mostly used as feed grain, as a raw material for beer production and distilling, and to a smaller extent as food. Barley is one of the first agricultural species (Ullrich *et al*., 2011) having initially been domesticated in the fertile crescent of the Neolithic Near East over 10,000 years ago (Salamini *et al*., 2002). During domestication process, barley has gradually accumulated traits that facilitated agricultural production as a result of environmental selection, or deliberately as a result of targeted choice by humans (von Bothmer *et al* 2003). In the subsequent millennia, farmers continuously adapted local populations to their needs, leading to a great variety of landraces. About 100 years ago, these landraces formed the basis for the development of modern cultivars by cross breeding. During this time, grain yield was more than doubled with an estimated genetic contribution to this increase of about 30–50% (Schuster *et al*., 1997). However, to feed a growing world population, and to provide renewable resources to satisfy the soaring demand for energy, genomics-based technologies have to be efficiently implemented to study the genetic basis of plant performance and to isolate agronomically-important genes from the genetic diversity present in the gene pool of barley. Barley was among the first crops to be considered in genetic analyses (Von Tschermak, 1901), with induced mutations reported over 80 years ago. Since then, thousands of morphological and physiological mutants of barley have been identified worldwide providing a basis for investigating the mechanisms underlying plant growth, development and responses to environmental conditions (Lundqvist *et al*., 1997).

H. vulgare subsp. *spontaneum* (wild barley) has been a model plant to study links between genetic diversity and ecological niches, physiological-adapted-traits, and geographic variation. Significant genetic divergence and diversity were found in four microniches (Huang *et al.*, 2002). Action of natural selection upon specific genomic regions resulted in adaptive divergence and directional selection at particular loci. Various stress–related characters have been widely studied, such as salt (Pakniyat *et al.*, 2003), and drought tolerance (Chen *et al.*, 2002; Elberse *et al.*, 2003; Suprunova *et al.*, 2004).

A broad spectrum of resources has been developed during the last two decades to facilitate the systematic analysis of the barley genome (Sreenivasulu *et al*., 2008). These include a large number of mapped molecular markers, comprehensive EST collections, BAC libraries, mutant collections, DNA arrays, and enabling technologies such as the large scale production of doubled haploids and efficient transformation protocols (Gujaria *et al*., 2011). Advances made in barley genomics and recent progress in construction of anchored physical maps and sequencing of the barley gene space (Sreenivasulu *et al*., 2008; Schulte *et al*., 2009; http://barleygenome.org) will largely contribute to a comprehensive understanding of gene functions in the context of agronomical important phenotypes. Great efforts have been made for this through an international consortium also to develop a high-density genetic map, assess new sequencing technologies, and generate substantial datasets of genomic survey information (Schulte *et al*., 2009; Comadran *et al*., 2012; Kilian and Graner 2012). All this is becoming a milestone toward understanding grass genomics and systems biology.

An important factor in determining the past and future distributions of biodiversity and crop production is climate change. Such a phenomenon is not new, and species have traditionally responded to such change over evolutionary timescales. The key question today is how organisms will respond to the current apparently rapid rate of climate change. Drought is one of the major outcomes of global warming and it is considered one of the most serious abiotic stress factors that occur throughout the development of the plant, resulting in the modification of plant physiology and limitation of crop productivity. Plants have evolved a range of defense and escape mechanisms (Pennisi, 2008), mediated by multiple rather than by single genes.

Egyptian hieroglyphic scripts suggested that barley was more important than wheat for human food because of its tolerance against salt, when the irrigated lands of southern Mesopotamia began to salt up (Jacobsen and Adams, 1958). Globally barley production, area, and yield have been relatively stable this century, but have decreased by about 12% in overall production in the past 2 decades (Ullrich *et al*., 2011).

Improving the level of drought tolerance is an important objective in barley breeding programmes, as it would help to stabilize production in drought-prone environments such as the Mediterranean region. The application of genomics-based strategies offers the possibility to accelerate and focus genetic improvement. Such approaches take advantage of modern DNA marker technologies and knowledge of genes that are involved in stress responses. In barley, Quantitative Trait Loci (QTLs) underlying drought tolerance has been mapped to almost every chromosome (von Korff *et al*., 2008; Chen *et al*., 2010). However, little information has been gathered to date regarding the genomic location of drought-response genes, either expressed throughout plant development or at late reproductive stages influencing seed yield and quality. Barley is an ideal candidate plant for association mapping. It has a long history of recombination events and conserved linkage disequilibrium, and is highly autogamous (Caldwell *et al*., 2006). This means that fewer markers are required to survey the whole genome in comparison to out-breeding species such as maize (Remington *et al*., 2001). In recent years, association genetics has become a popular methodology for mapping purposes in crop plants with barley studies often leading the way (Waugh *et al*., 2009). As an example, several marker-trait associations were reported for yield and yield stability in a modern 2 row spring barley collection of 146 elite lines representing material that had been evaluated in official Danish variety trials from 1993 to 2000 (Kraakman *et al*., 2004).

Genomic tools can be used in association genetics to explore the genetic diversity within the germplasm of the species and to evaluate the possible association between selected candidate genes and genetic variation for drought tolerance.

Results will provide a deeper understanding of the genetic and molecular basis of drought responses in barley, potentially allowing the identification of underlying chromosomal regions, useful markers or alleles for future applications in barley breeding for yield stability in a changing climate.

1 Barley taxonomy and morphology

1.1 Taxonomy and major domestication traits

Barley is an annual diploid self-pollinating species with 2n = 14 chromosomes with a genome size of 5.1 Gbp (Klaus *et al*., 2012). Primitive landraces and the wild progenitor of barley (*H. spontaneum*) exhibit large variations in physiology, morphology and genetics, which might be used to improve cultivated barley (Nevo 1992; Forster *et al*., 2000). Barley belongs to the genus *Hordeum*, which forms part of the tribe Triticeae of the grass family Poaceae.

The Triticeae is composed of over 350 species. Cultivated barley (*Hordeum vulgare* L. spp. *vulgare*) is one of 32 *Hordeum* species (Pourkheirandish and Komatsuda, 2007; Forster *et al*., 2007) *Hordeum* is a monophyletic group with a common origin. All species, even cultivated barley, are thus related, some of them more distantly so *Hordeum* is evidently an ancient genus, splitting from the wheat species some 13 million years ago (Von Bothmer and Komatsuda, 2011).

In addition to barley, the Triticeae tribe includes other important small grain cereals, such as bread and durum wheats (*Triticum* spp.), rye (*Secale cereale*) and the wheat/rye hybrid crop, triticale (×*Triticosecale*), as well as several forage grass species (von Bothmer, 1992; von Bothmer, 1995). Molecular evidence has revealed considerable genomic co-linearity between barley, wheat, and rye. The relative simplicity of its genetics and ample genetical diversity make barley an ideal study organism. In addition, phytomeric models developed for barley development (Bossinger 1992 and Forster *et al*., 2000) have implications for related species and other monocots.

The ancestral form of barley, *Hordeum vulgare* subsp. *spontaneum* has been proven to be of interest as a gene source because of its agricultural traits of interest, such as stress tolerance, disease resistance, and various quality traits. Wild germplasm has been utilized in barley breeding. Interspecific hybridization and backcrosses have been performed with almost all wild species in combination with barley. Some of the wild, perennial *Hordeum* species are important as components in natural pastures used for foraging in central Asia and South America (Von Bothmer and Komatsuda, 2011).

The barley inflorescence is called spike, head or earThe spike axis is called rachis and bears three spikelets at each node. A spikelet is one of the flower clusters, the unit of inflorescence, consisting of two or more flowers and subtended by one or more glumes variously disposed around a common axis. In wild barley and two-rowed cultivars only the central spikelet is fertile, but not the lateral spikelets.

In six-rowed barley all three spikelets can produce grains (Kirby and Appleyard, 1987; Komatsuda *et al.*, 2007). Different mutations, dominant and recessive, result in the fertility of such lateral spikelets to produce six-row barleys (Zohary and Hopf, 2000). Recent genetic studies have revealed mutations in one gene, *vrs1*, are responsible for the transition from two-row to six-row barley (Komatsuda *et al*., 2006, Figure 1). Komatsuda *et al*. (2007) found that the expression of *Vrs1* was strictly localized in the lateral-spikelet primordia of immature spikes, suggesting that the wild-type VRS1 protein suppresses development of the lateral rows.

Winter barley varieties require a period of cold stimulus called vernalisation to begin floral development. Spring barleys do not require such vernalisation period. Flowering in many barley varieties responds to temperature as well as day length, so the development patterns can vary with latitude.

Two-rowed varieties have a higher number of tillers per plant and heavier seed than six-rowed varieties (Hayes *et al.*, 2003). Six-rowed varieties on the other hand, usually have more seeds per inflorescence. Thus the compensatory effects of yield components lead to similar levels of yield potential (Hayes *et al.*, 2003).

In wild barley, the spike rachis is brittle; spikelets tend to separate upon maturity to facilitate seed dispersal. In addition, this adaptive specialization ensures that the seeds will bypass stones to reach soil when they fall to the ground (Zohary, 1963). This feature is an evolutionary advantage offered by the two-rowed spikes in nature, and spontaneous six-rowed mutants are eliminated naturally and rapidly from wild barley populations because they lack this adaptation (Zohary, 1963). However, the domesticated barley spikes are non-shattering, making it much easier to harvest (Zohary and Hopf, 2000). According to Zohary and Hopf (2000) the tough rachis trait is caused by a mutation in one of two strongly linked genes known as Bt_1 and Bt_2; many cultivars possess both mutations. The non-shattering condition is recessive, so varieties of barley that exhibit this condition are homozygous for the mutant allele.

Figure 1: Barley spikes: Represent row types. 2r = 2 row barley spike, 6r = 6 row barley spike
Source: Wikipedia, by Xianmin.Chang

1.2 Morphology development and reproduction

Barley is an annual grass that stands 60-120 cm tall. It has two types of root systems, seminal and adventitious. The depth of the roots depends on the condition and texture of the soil, as well as on the temperature. The seminal rootlets of barley emerge when the seed germinates and form a fibrous branched mass of roots. At tillering stage, the adventitious root system arises from the crown, and this tends to be thicker and less branched. Under adverse conditions such as drought, the adventitious roots may not develop. In other conditions, the seminal roots stop functioning during the life of the plant. Different barley varieties can vary significantly in rooting system, and this can impact on their competitive ability (Briggs, 1978). If the grain is deeply planted a rhizomatous stem is formed, from which leaves form when it reaches the surface.

Beside the main stem or culm, barley generally has several lateral stems or tillers. Stems are erect and made up of hollow, cylindrical internodes, separated by the nodes, which bear the leaves (Gomez-Macpherson 2001). The ability of the barley plant to develop new tillers in response to favorable environmental conditions is a useful mechanism for adapting to changes during the growing season.

The apex of the main stem and each fertile tiller carry a spike. Near the soil surface, the part of the stem carrying the leaf bases swells to form the crown. It is from the crown that the adventitious roots and tillers develop (Briggs, 1978).

Barley leaves are strap-shaped with parallel veins and a prominent midrib, and are produced on alternate sides of the stem (Bossinger *et al.*, 1992). The leaf structure consists of the ligule, auricles, blade, and sheath. The sheath surrounds the stem completely (Figure 2). Ligules and auricles distinguish barley from other cereals since they are smooth and envelope the stem and also can be pigmented with anthocyanins (Gomez-Macpherson, 2001).

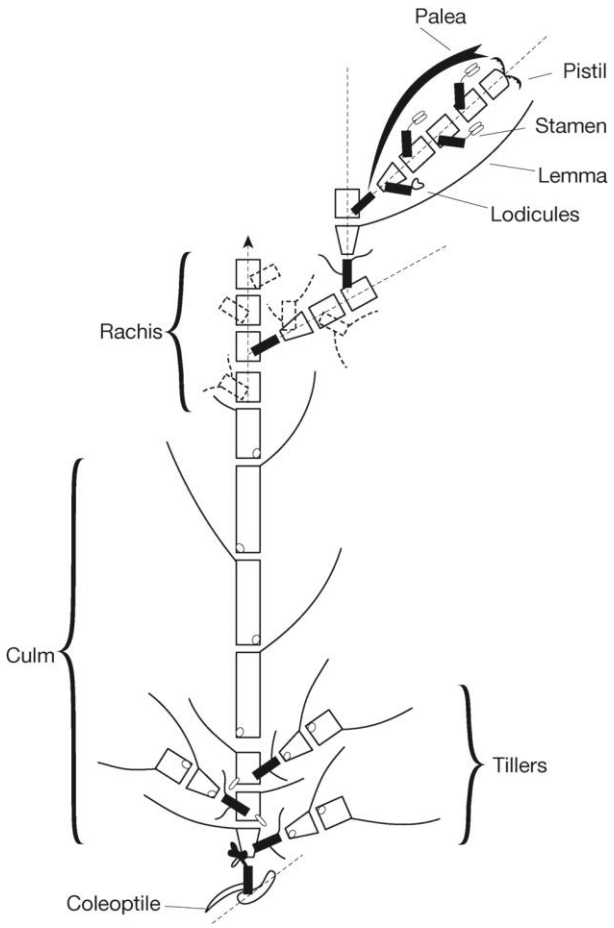


Figure 2: Schematic structure of the barley crop.
Source: CABI, PROSEA Foundation

Barley growth can be divided into a number of stages; germination, seedling development, tillering, stem elongation, heading (ear emergence), flowering and ripening (Figure 3). Tillers start to emerge at about the 3rd leaf stage. The number of tillers and duration of tillering vary according to variety and growth conditions (Briggs, 1978). In general, field grown barley plants typically produce 2-5 or more tillers (Briggs, 1978; Gomez-Macpherson, 2001). Most tillers initiate adventitious roots, although later appearing tillers often remain un-rooted and die prematurely (Anderson-Taylor and Marshall, 1983).

After a number of leaves have been initiated the stem apex gives rise to spikelet initials which form the inflorescence or spike. The first spikelets form at the base of the spike, which terminates with the formation of sterile florets. The spike is initially contained within the sheath of the flag leaf, which swells and is called the boot (Briggs, 1978). Spikelets consist of a floret and two subtending bracts called glumes. In most varieties the spike eventually becomes clear of the boot, and flowering generally occurs in the newly emerged spike. Flowering usually begins in the florets around the middle of the ear and spreads upwards and downwards, taking 1-4 days to complete. Ears on different tillers may mature at varying times (Briggs, 1978).

The pollen and ovules in each floret mature together in barley (Briggs, 1978). According to Riddle and Suneson (1944) pollen viability ranges from a few hours to at least 26 hours, while stigma are receptive and able to be fertilized for a period of 6-8 days following the first flower opening. Cereals can be either closed-flowering (cleistgamous) or open-flowering. Many winter barley varieties are open-flowering whereas spring barleys are usually closed-flowering (Nelson *et al*., 2001). In closed-flowering types, self-pollination occurs as anthers remain inside each floret. In open-flowering barley lodicules (tiny scales at the base of the ovary that represent the corolla in grass flowers) become turgid pushing the palea and lemma apart, so that the anthers may emerge (Briggs, 1978). In the latter case, pollen shedding starts before the spikelet opens and continues after it opens, thus out-crossing is possible (Turuspekov *et al*., 2005). Nevertheless, most pollen is shed before the spikelet opens, so that self-fertilization is usual (Briggs, 1978).

Floral traits such as high anther extrusion, large anthers and vigorous stigmas may increase the level of outcrossing in barley plants. Such traits are influenced by both genetic and environmental factors (Abdel-Ghani *et al*., 2005).

Few studies of barley pollen viability have been published. Earlier work suggests that barley pollen is extremely sensitive to drying and remains viable for only a few hours after dehiscence (Pope, 1944; Bennett *et al*., 1973; Gupta *et al*., 2000; Parzies *et al*., 2005).

All annual *Hordeum* species are mainly inbreeds, although none are obligate inbreeds (Von Bothmer, 1992). Cultivated barley and its wild progenitor both reproduce almost entirely by self-fertilization (~99%) (Wagner and Allard, 1991; Von Bothmer, 1992; Ellstrand, 2003), and gene flow in barley is low (Ritala *et al*., 2002).

1.3 Growth habit

Vernalization is a critical step in the transition from the vegetative to the reproductive stage (Saisho *et al.*, 2011). In general, wild barley requires vernalization and migration of cultivated barley outside its origin place was accelerated through mutations and recombination events to develop reduced vernalization requirement and photoperiod insensitivity (Salamini *et al.*, 2002). Cultivated barley may be either a winter or spring annual. Both winter and spring barleys are cultivated in mid-latitudinal regions including North Africa, Europe and Asia. The duration of the different developmental stages (Figure 3) varies widely. Winter varieties usually produce more tillers than spring varieties, during vegetative growth period over winter. Spring varieties do not have a typical rosette stage and so develop fewer tillers than winter varieties. However, growth rate also depends on environmental conditions, eg weather, water supply, soil fertility, and degree of competition with other plants, presence of pests and diseases, and time of planting. Initially growth is slow while the seedlings establish. Total time to maturity depends on variety, location and planting date (Thomas and Fukai, 1995).

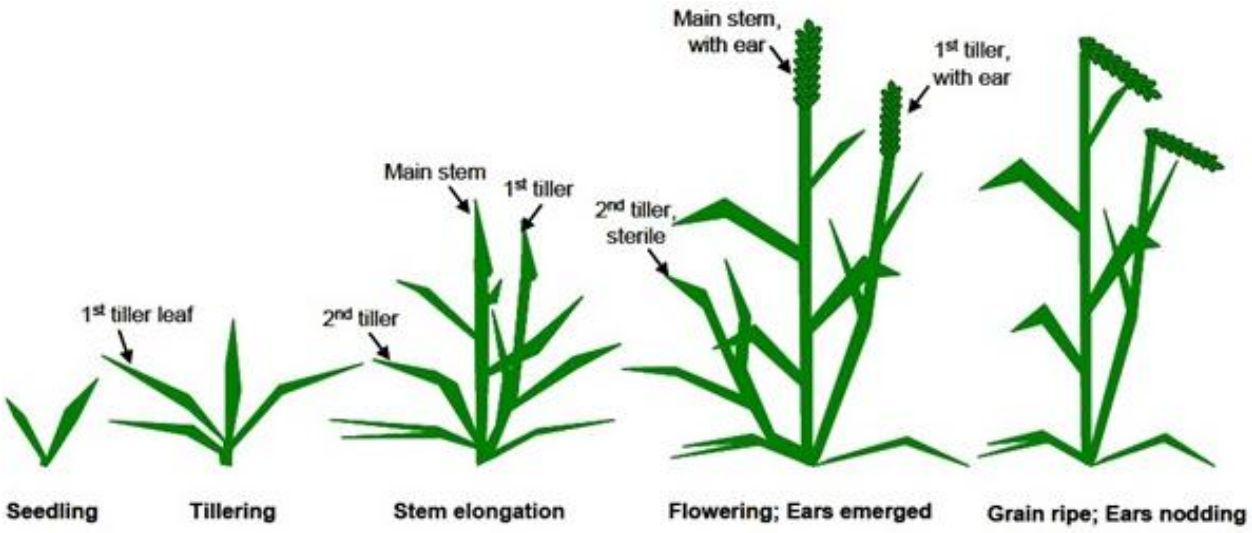


Figure 3: Schematic diagram of barley plants at different stages of development
Source: www.ogtr.gov.au

The emergence of the flag leaf is an important growth stage for timing the application of certain growth regulators. The mature leaves progressively senesce and gradually the whole plant dries out until full maturity, when the grain is ripe (Briggs, 1978).

Some older cultivars and landraces produce many tillers but develop few spikes, while most modern genotypes have a higher percentage of tillers that develop spikes (Gomez-Macpherson, 2001). In addition, winter varieties usually produce more tillers than spring varieties, during vegetative growth period over winter. Spring varieties do not have a typical rosette stage and so develop fewer tillers than winter varieties.

Both winter and spring barley are cultivated in mid-latitudinal regions including North Africa, Europe and Asia. The spread of barley cultivation lines was development by lacking vernalization requirement, leading to expand in areas where spring sowing is necessary to avoid winter injury (Pourkheirandish and Komatsuda, 2007).

According to Pourkheirandish and Komatsuda, 2007, the first domesticated barleys are likely to have had a winter growth habit.

1.4 Importance of barley in agriculture

1.4.1 History of domestication

Barley is one of the first crops to be domesticated 10,000 years ago by ancient farming communities in the Fertile Crescent region (Salamini *et al.*, 2002). Its scientific name, *Hordeum*, refers to Roman gladiators and means "barley eater". The most ancient remains uncovered from archaeological research that document collection of food staples from the natural flora date back about 17,000 years B.C. (Ohalo II, south shore of the Sea of Galilee (Klislev *et al*., 1992). The three of the so-called Neolithic founder crops in the development of agriculture are composed of wild barley (*Hordeum spontaneum*), wild emmer (*Triticum dicoccoides*), and wild einkorn (*Triticum boeoticum*) in pre-agricultural layers retrieved from Tell Abu Hureyra, Syria (Hillmann, 1975), and date back to 8,000 to 10,000 B.C. From around such date and later, remains of non-brittle rachis barley appear in an increasing number of excavations, in most cases together with tough rachis types of einkorn and emmer wheat. Compared to the time that has elapsed from the remains found at Ohalo II, a rapid spread of agriculture activities throughout the Fertile Crescent region is noted, as shown by Zohary and Hopf (1993).

Cultivated barley grains found in pits and pyramids in Egypt indicate that barley was cultivated there more than 5000 years ago. Ancient pictographs found for barley are dated about 3000 B.C. Many references to barley as well as beer are found in the earliest Egyptian and Sumerian writings. In the subsequent millennia, farmers continuously adapted local populations to their needs, leading to a great variety of landraces. About 100 years ago, these formed the basis for the development of modern cultivars by cross breeding. During this time, grain yield was doubled with an expected genetic involvement to this increase of about 30–50% (Schuster, 1997).

1.4.2 Cultivation and uses

In the 2011, barley worldwide production amounted to approximately 155.3 million metric tons, ranking it the fifth among cereals (FAOSTAT, 2011; http://www.fao.org). Barley serves as a major animal feed crop, with smaller amounts used for malting and in health food. The European Union

constitutes the biggest barley producer in the world (Figure 4). Barley grain is particularly high in soluble dietary fiber, which significantly reduces the risk of serious human diseases including type II diabetes, cardiovascular disease and colorectal cancers that afflict hundreds of millions of people worldwide (Collins *et al.*, 2010).

Barley has a wide range of climatic adaptation as it is characterized for relatively high drought tolerance. Cultivated barley is grown in a range of diverse environments that vary from sub-Arctic to sub-tropical, with greater concentration in temperate areas. Other than the cool highlands, barley is rarely grown in the tropics as it is not suited to warm humid climates (Nevo, 1992). Barley genotypes, in particular landraces and wild species, represent an important source of variation for adaptive traits that may contribute to increase yield and yield stability under drought conditions, and that could be introgressed into improved varieties. Traits that have been investigated include physiological/biochemical and developmental/morphological traits (Baum *et al.*, 2007). Yield performance under drought is a particularly complex phenomenon, and plants exhibit a diverse range of genetically complex mechanisms for drought resistance.

Barley breeding in Egypt started more than a century ago. Farmers relied on the own local varieties improved by selecting superior plants from existing landraces. This has enriched the Egyptian genetic resources, especially in relation to adaptation to abiotic stresses such as drought, salinity, and poor soil fertility. Barley breeders continued to cross local varieties and introductions to produce new varieties with better adaptation to low rainfall areas. For example, varieties Giza 119 and Giza 121, produced in 1973 and 1980, respectively, as selections from the cross Baladi 16 × Gem (Ullrich, 2011). In parallel, breeding in Morocco dates back to 1920 and was based on the improvement of local landraces and on the introduction of two – row foreign varieties. Selection was conducted mainly for disease resistance and yield (Ullrich, 2011).

The history of barley breeding in Tunisia is characteristic of most of North Africa as well as the Near East. However, the first serious attempt for barley breeding program was made in 1973 with crosses of early material for the semiarid environments (Ullrich, 2011).

In contrast to the Near East, six-rowed barley is representing the majority of landraces and improved varieties grown in North Africa (Ullrich, 2011).

Barley is an established crop because of its high demand for livestock feed and is a good source of source of protein. The consumption of barley has increased as a human food in some regions such as Egypt and Algeria (Grando and Gomez Macpherson 2005). In addition barley malt is used to produce beer, distilled alcohol, malt syrup, malted milk, and breakfast foods (Kling, 2004).

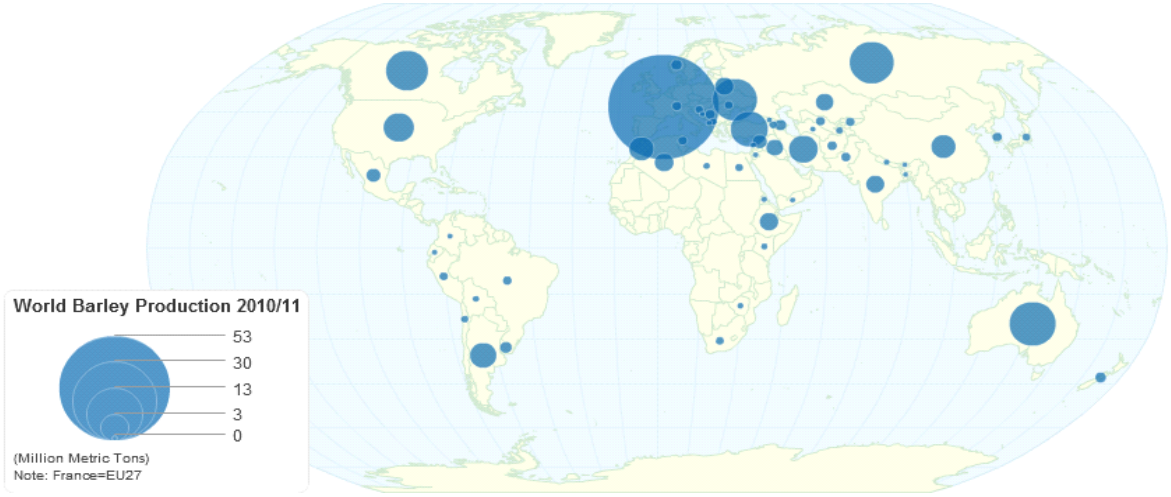


Figure 4: World Barley Production in Tons, FAOSTAT 2010

1.5 Barley as genetic system

In addition to its ecological and economic importance, barley has a long history of genetics research, making it a model for cereal crop genetics and breeding. Barley has one of the largest and most complex genomes of all economically important food crops (Wicker *et al.*, 2008). Barley has seven pairs of chromosomes currently designated according to their homoeologous relationships with other *Triticeae* species (Linde-Laursen 1997). As a result, barley chromosomes are now officially assigned as 7H, 2H, 3H, 4H, 1H, 6H, and 5H, respectively. Homology is well supported by biochemical, morphological, and molecular studies and the ability of barley chromosomes to substitute for the equivalent wheat chromosomes in substitution lines (Shepherd and Islam, 1992).

As indicated in section 1.1.2. (Morphology development and reproduction) barley lines are almost completely homozygous. F1-hybrids are produced by emasculation of the female parent and addition of pollen of the male parent one to three days later to the (bagged) female spike. The F1 can be developed into inbred lines by self-fertilization, or also by the production of doubled haploids (DH). Selfing is time consuming, as at least 7 or 8 cycles of selfing are necessary to reach homozygosity, alas in the later stages of this process many inadequate lines could be discarded. Hence, DHs are a fast road to homozygosity, and the selection would be possible after the DHs have been created. The most widespread techniques to obtain DHs are the bulbosum method (*Hurdeum bulbosum*) (Kasha and Kao 1970) and anther culture (Friedt and Foroughti-Wehr 1981).

Over the past century, genetic research focused on trait inheritance and mapping (Barley Genetics Newsletter, 1971 – 2010), induced mutagenesis (Nilan 1981; Gottwald *et al.*, 2008; Kurowska *et al.*, 2011), and recently on molecular and physical mapping and genomic (e.g. Hayes *et al.*, 1993; Yu *et al.*, 2000; Caldwell *et al.*, 2004; Close *et al.*, 2004; Druka *et al.*, 2006; Varshney *et al.*, 2007; Hamblin *et al.*, 2010).

1.6 Barley breeding

Traditional plant breeding has involved cross pollination between varieties of the same species as well as hybridization between different species. Wild barley represents an important genetic resource for cultivated barley that has a narrowed gene pool due to intensive breeding. For example, introduction of resistance to biotic and abiotic stresses from wild barley is recognized as an important target for barley breeding (Ellis *et al.*, 2000).

However, classic approaches for selection of the new varieties can take several years; also, traditional plant breeding is costly and time consuming, due to the necessity of evaluating high numbers of plants in field trials often considering several phenotypic traits. The development of DNA-based markers facilitated greatly the evaluation and selection process in plant breeding. These molecular tools have increased the speed and precision for achieving desired agronomic traits (Ullrich *et al.*, 2011).

2 Genomic tools

2.1 Expressed sequence tags (ESTs), physical map and genomic sequences

One of the most popular tools among barley molecular geneticists is HarvEST (http://harvest.ucr.edu). HarvEST is principally an EST database that was developed at the University of California, Riverside by Timothy Close, Steve Wanamaker, Mikeal Roose, and Matthew Lyon (Close *et al.*, 2009). HarvEST illustrates comparative genomics and gene function, the design of oligonucleotides, in support of activities such as microarray content design, functional annotation, as well as physical and genetic mapping. The "HarvEST:Barley" component of HarvEST has additional functions to support comparative genome mapping.

EST projects are primarily used to either complement the existing genome projects or to serve as alternatives for the purposes of gene discovery (Parkinson and Blaxter, 2009). The technology of EST sequencing offers a relatively inexpensive alternative to whole genome sequencing and has become a valuable resource for gene identification (Lindlöf, 2003).

Multinational collaboration, International Barley Sequencing Consortium (IBSC) has been established with the objective of obtaining the whole sequence of barley genome. (http://barleygenome.org; Schulte *et al.*, 2009). In many plant species, transcript contigs have been constructed by assembling all the EST data available in Plant Genome Data Base (www.plantgdb.org), with the rational of identifying a data set of unique mRNA sequences and maximizing the information obtained for both protein-coding and noncoding regions in these sequences (Duvick et *al.*, 2008). A large set of ESTs (501,620 from the *vulgare* subspecies and 24,161 from the *spontaneum* subspecies in the NCBI has been accumulated in the public domain (Matsumoto *et al.*, 2011).

Last month, an integrated and ordered physical, genetic and functional sequence resource that describes the barley gene-space in a structured whole-genome context has been made available to the community providing a powerful platform for trait dissection and breeding (http://barleygenome.org, The International Barley Genome Sequencing Consortium 2012).

2.2 Molecular markers, SNPs and SNP platforms

The increased availability of high throughput genotyping technology, the advances in DNA sequencing together with the development of statistical methodologies appropriate for genome-wide association mapping in presence of considerable population structure contributed to the increased interest for association mapping in plants. High-throughput platforms are now able to profile thousands of single nucleotide polymorphic (SNP) markers on large numbers of samples with high reliability and cost-efficiency (Tuberosa and Salvi, 2006). With these advancements, it is now possible to fast-track QTLs to potential candidate genes using approaches based upon association mapping and synteny conservation with sequenced cereal genomes (Waugh *et al*., 2009).

ESTs deriving from a range of barley genotypes have been extensively used to develop molecular markers, especially Single Nucleotide Polymorphisms (SNPs) (Rostoks et al., 2006, Stein *et al*., 2007, Close et al. 2009, Sato et al., 2009). A SNP is a DNA sequence variation occurring when a single nucleotide A, C, G or, T in the genome varies between paired chromosomes in an individual or between members of a species. SNPs may occur in the coding, non-coding and intergenic regions of the genome.

Gene-based SNP genotyping platforms using Illumina® technology have been used to construct barley consensus maps (Close *et al*., 2009, Munoz-Amatriain 2011) carry out association mapping analyses and map mutants that had been back-crossed to generate near-isogenic lines (NILs) (Druka *et al*., 2011). Raw datasets consisting of millions of data points can be generated in a single experiment using either Illumina® technologies that can be interrogated using various statistical algorithms

In barley, 4596 SNPs were initially arranged in three GoldenGate Pilot Oligonucleotide Pool Assays (POPA): based on preliminary results from these, two barley OPAs, BOPA1 and BOPA2, were used to develop a consensus genetic linkage map composed of 2943 SNPs from Steptoe × Morex, OWB and Morex × Barke doubled - haploid mapping populations (Close *et al*., 2009).

The Diversity Arrays Technology (DArT) is another high-throughput technique for genetic analyses (Jaccoud *et al*., 2001). DArT was developed to provide a practical and cost-effective whole-genome

fingerprinting tool (Jaccoud *et al.*, 2001). High throughput is achieved with instrumentation increasingly becoming standard. The technology allows rapid development of hundreds of markers distributed throughout the genome as well as inexpensive and fast routine genome scans. DArT was validated in several species including cereals such as barley (*H. vulgare* ssp. *vulgare* L.) (Wenzl et al. 2004), wheat (*Triticum aestivum* L.) (Akbari *et al.*, 2006) and sorghum (*Sorghum bicolor* (L.) Moench) (Mace *et al.*, 2008). DArT markers can be used to track phenotypic traits in breeding. The high throughput and low cost nature of the technology makes DArT more affordable for marker assisted selection. Such markers can be tracked though an introgression or crossing program, and used to supplement phenotyping to reduce potential miss-identification of a trait due to environmental effects (Lande and Thompson, 1990). However, it is unavoidable that there will be a small degree of incorrect ordering between very closely linked markers in some populations. This is due to the ordering of closely linked markers within component maps that are inherently difficult as a result of the limited resolution provided by the population sizes used for map construction (Alsop *et al.*, 2010). To this end, the utilization of a robust portion of genotyping data derived from four mapping populations in Illumina GoldenGate assay (Illumina Inc., San Diego, CA) has provided a new element of a high fidelity and dense consensus map produced entirely from transcribed gene SNPs (Close *et al.*, 2009).

In conclusion, the great improvement of technology made the use of SNP and InDel markers attractive for high-throughput use in marker-assisted breeding, EST mapping and the integration of genetic and physical maps.

2.3 Synteny

Comparative mapping of cereal genomes revealed extensive conservation of genome content and order co-linearity (synteny) despite significant differences in chromosome number and genome size (Feuillet and Keller 2002; Song *et al.*, 2002).

Barley has a high degree of synteny with other grass genomes. Rice diverged from barley and wheat about 50 million years ago and was initially proposed as a model also for Triticeae species (Paterson *et al.*, 2004). Later, *Brachypodium* emerged as a better model grass for Triticeae crops (Graner *et al.*,

2011). It has a small genome of ca. 350 Mb (Huo *et al.*, 2008), self-fertility, rapid generation time, simple growth requirements, and is easy to transform (Draper et *al.*, 2001; Vogel *et al*., 2006a; Vogel and Hill 2008; Garvin *et al*., 2008). Vogel *et al*. (2006b) and Huo *et al*. (2008) found a close relationship between *Brachypodium* and barley and wheat, and some more distant relationship with rice, maize, and sorghum.

Known genes in model plants can serve as a cloning vehicle for synteny-based gene isolation in the large genome species like barley. In a study by Jia *et al*. (2009), they used this strategy to target the barley semidwarf gene *sdw1/denso*: comparative mapping revealed that the *sdw1/denso* region in barley is syntenic to the *sd1* gene on chromosome 1 in rice and the gene (*Hv20ox2*) isolated from barley showed conserved gene structure and a high degree of sequence similarity with the rice *sd1* gene.

2.4 Linkage mapping and methods for genetic dissection of complex traits

The first concept of a genetic map was presented by Alfred H. Sturtevant (1913) who mapped sex-linked characters in a linear way on the Y chromosome of *Drosophila melanogaster*. Nowadays, whole genomes are being sequenced at increasing speed. In total, there are around 40 smaller and larger genome-sequencing projects in progress for plants, including species such as *Avena sativa*, *Medicago sativa* and *Medicago trunculata*, *Lotus corniculata*, different Brassica species, banana, barley, coffee, cotton, Eucalyptus, maize, Populus, soybean and tomato (Bernal *et al*., 2001) In addition, to the completely sequenced genomes of *Arabidopsis* and rice (Rounsley *et al.*, 2009).

A genetic map is constructed on the basis of recombination events between two non-sister chromatids of each pair of homologous chromosomes during meiosis. A genetic localization experiment determines the order of linked markers. The distance determination [in centiMorgans (cM) or percentage recombination] is relative. Recombination frequencies vary between different chromosome parts, physical conditions and sexes. As a result, the ratio between genetic and physical distance is not constant over the length of the chromosome. Also, genetic distance depends on the parental combination used, as closely related lines will exhibit an intrinsically higher recombination frequency

than distantly related lines do.

In the early days, progress in mapping was hindered by the lack of sufficient markers. However, with the advent of high-throughput genomics technologies the availability of markers is no longer a bottleneck. Moreover, sequencing projects enable us to assign markers a physical position on the map.

Linkage maps provide powerful tools for genetic dissection of quantitative traits into the genomic regions that subtend their variation: Quantitative Trait Loci (QTLs) are genetic loci where different functional alleles are segregating and causing significant effects on a quantitative trait (Salvi and Tuberosa, 2005).

2.4.1 Quantitative trait loci (QTL) analysis

Over the last 20 years, advances in molecular marker technology have made the mapping of quantitative traits much more feasible. QTL mapping creates the possibility for modeling quantitative traits at the individual gene level (Falconer and Mackay 1996).

The theory of QTL mapping was first described by Sax (1923), where he noted a relation between two traits in bean, seed size (a complex trait) was associated with seed coat color (a monogenic trait). This concept was further investigated by Thoday (1961), who suggested that it is possible to characterize all QTLs involved in complex traits, even though the segregation of simply inherited monogenes could be used to detect linked QTLs. In the early stages and before the advent of modern QTL mapping, Kearsey and Farquhar (1998) studied traits representing quantitative variation by statistical analysis of experimental populations based on the means, variances and co-variances of relatives, with some proposed knowledge of the number and location of the genes underlying such traits. These studies focused on phenotypic distributions of populations and correlations in phenotypes among related individuals or lines. With the advent of molecular DNA markers, QTL analyses have been greatly improved in efficiency and resolution allowing for the identification of genes determining quantitative variation (Tuberosa and Salvi, 2006).

Knowledge of such genes provides the ideal tool for marker assisted breeding (Figure 5).

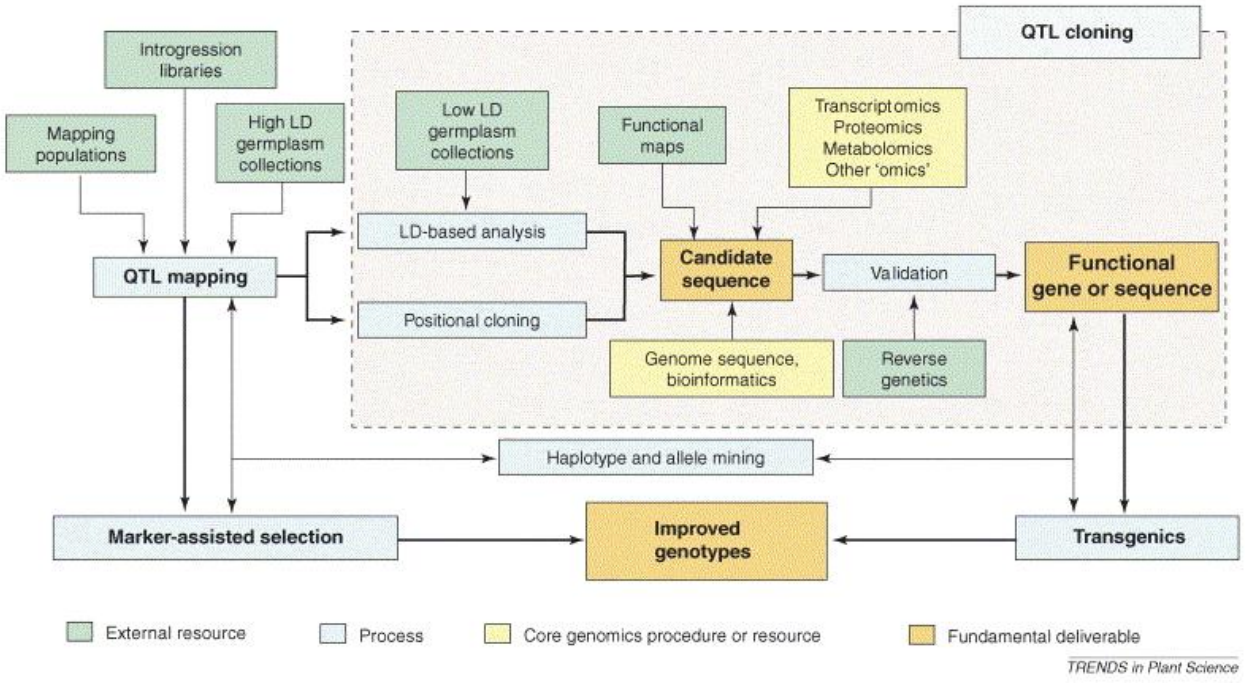


Figure 5: A schematic representation of the position and role of QTL cloning in the current framework of marker-assisted breeding activities aimed at crop improvement for quantitative traits such as drought tolerance.

Thornsberry *et al*., 2001 propose that the association mapping (AM) is powerful approach, since it tries to establish a statistical association between allelic variation at a locus and the phenotypic value of a trait across large unrelated accessions. In essence, AM exploits historical and evolutionary recombination at the population level. Following its wide application for the hereditary and dissection of human diseases. Association genetics has recently been applied to crops including barley (Cockram *et al*., 2008). In the context of plant breeding, AM has several advantages over classical linkage analysis using segregating populations (Kraakman *et al*., 2004). Firstly, broader genetic variation in a more representative genetic background can be included in the analyses. Secondly, already available multi-trial phenotypic data can be linked to marker characterizations of the involved cultivars. Thirdly, AM may attain a higher resolution. According to Rostoks *et al*., 2006; reviewed in Rafalski, 2002 alleles at a few selected candidate genes may be tested for association with a phenotype, or the whole genome may be scanned to identify regions that are associated with a particular phenotype (Figure 6). In barley, the first approach was successfully adopted to associate the vernalization requirement with the VRN-H1 and VRN-H2 genes (Cockram *et al*., 2008), and frost tolerance with a variant of a *CBF*

gene (Fricano *et al*, 2009). Sokheh *et al*., 2008 proposed integrating the two approaches, using linkage mapping for a preliminary genome wide scan for QTLs, and LD mapping to obtain more precise location of individual QTL.

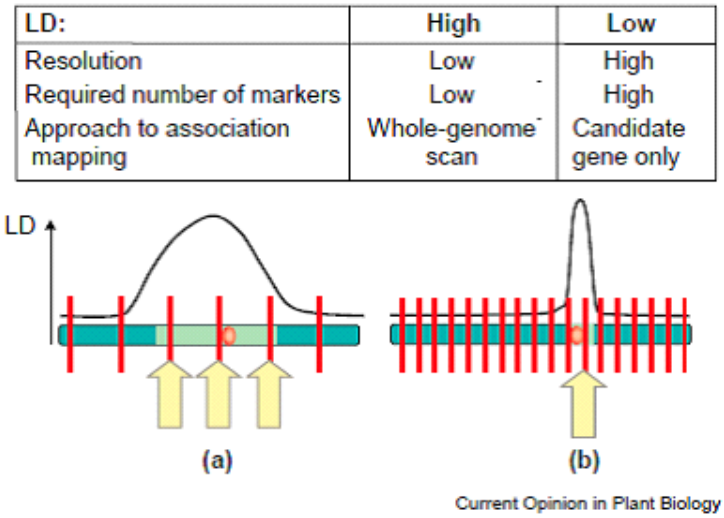


Figure 6: Relationship between the extent of LD and the resolution of association studies. In (a), LD declines slowly with increasing distance from the gene responsible for the phenotype (red oval) on a chromosome. In this case, even a low density of markers (shown as red vertical bars) is sufficient to identify associated markers (yellow arrows). In (b), LD declines very rapidly around the causative gene, and a much greater density of markers is required to identify an associated marker (yellow arrows) (Rafaliski 2002. *Current Opinion in Plant Biology*, 5:94–100)

2.4.2 Association mapping (AM)

In the 20th century, a theory for association was first given by Robbins. By the 21st century these and other indications of Association Mapping (AM) came to completion with the Human Genome Project when the physical map had been provided for the first time at nucleotide resolution.

In the past decade, crop geneticists have started applying association mapping (AM) as a complementary approach for QTL dissection. This method was initially developed for dissection of hereditary diseases in human genetics (Risch and Merikangas, 1996; Weiss & Clark, 2002). Thornsberry *et al.* (2001) proposed it as powerful approach also for plants, since it tries to establish a statistical association between allelic variation at a locus and the phenotypic value of a trait across large unrelated accessions. In essence, AM exploits historical and evolutionary recombination at the population level. Association genetics has recently been applied to various crops including barley (Cockram *et al.*, 2008). In the context of plant breeding, AM has several advantages over classical linkage analysis using segregating populations (Kraakman *et al.*, 2004). Firstly, broader genetic variation in a more representative genetic background can be included in the analyses. Secondly, already available multi-trial phenotypic data can be linked to marker characterizations of the involved cultivars. Thirdly, AM may attain a higher resolution. Alleles at a few selected candidate genes may be tested for association with a phenotype, or the whole genome may be scanned to identify regions that are associated with a particular phenotype (Rostoks *et al.*, 2006; Rafalski, 2002, Figure 7). In barley, the first approach was successfully adopted to associate the vernalization requirement with the VRN-H1 and VRN-H2 genes (Cockram et al., 2008), and frost tolerance with a variant of a CBF gene (Fricano et al, 2009). Sokheh et al., 2008 proposed integrating the two approaches, using linkage mapping for a preliminary genome wide scan for QTLs, and LD mapping to obtain more precise location of an individual QTL.

In association mapping studies, detection of significant association relies predominantly on genetic marker coverage, the number of individuals studied, and linkage disequilibrium (LD) between causative and linked polymorphisms (Mackay and Powell, 2007). Although genetic stratification in the majority of human studies is low (Rosenberg *et al*., 2010), inbreeding crops such as barley commonly display highly complex population structure because of their primarily inbreeding reproductive

strategy, population history, and close kinship (Rostoks *et al*., 2006). For example, association mapping was exploited using DArT markers to identify stem rust resistance genes in both wild and cultivated barley germplasm (Steffenson *et al.*, 2007). Kraakman *et al.* (2004) reported several marker-trait associations for yield and yield stability in elite lines of modern 2-row spring barley collection representing evaluated materials in official Danish variety trials. Comadran *et al.*, 2011 studied a panel of barley accessions representing cultivated germplasm in Mediterranean basin to localize QTLs controlling grain yield and related traits.

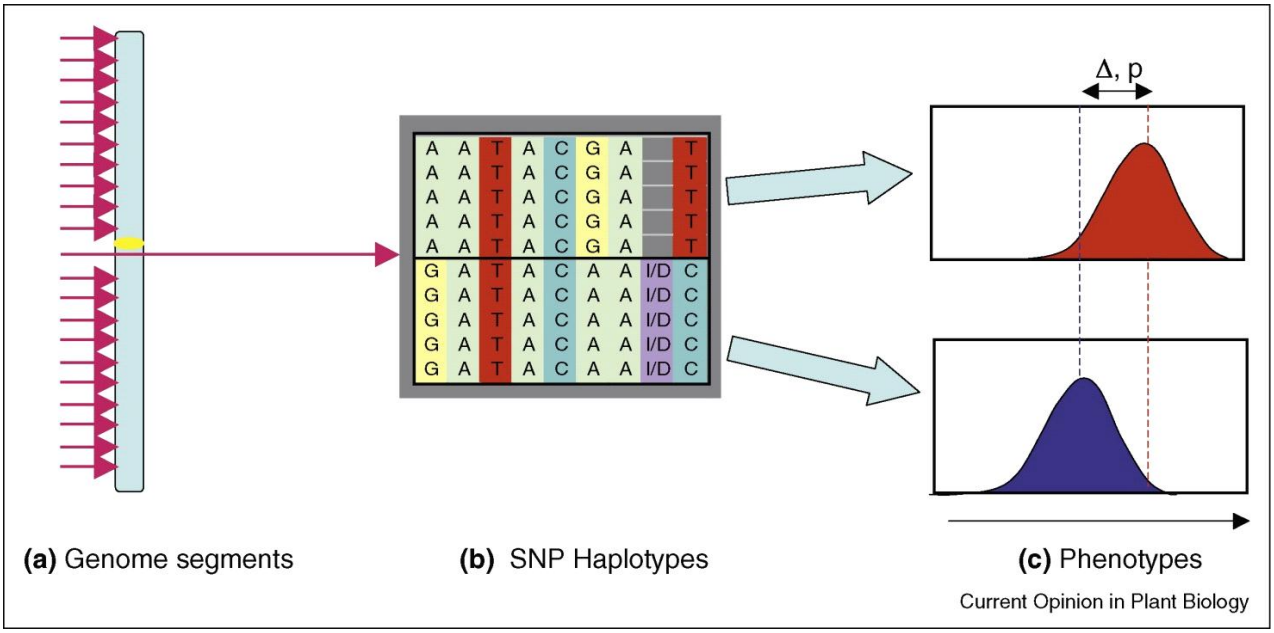


Figure 7: Principle of association analysis. (a) A collection of genetically diverse individuals is genotyped at densely spaced loci distributed throughout the genome; (b) the genotypes are divided into groups sharing SNP haplotypes (shown) or individual SNPs, at each locus in turn; (c) the distributions of phenotypic values for each of the haplotypes (or alleles) are compared and probability of null hypothesis (the distributions are equivalent) is evaluated statistically.

The accessibility of ultra-high density SNP maps opens the door for studying important complex traits by association genetic factors, taking advantage of the fact that genetic markers in close proximity to mutant genes may be in linkage disequilibrium (LD) to them.

2.4.3 Candidate genes-based association mapping

The candidate gene (CG) approach has emerged in plant genentics in the past two decades for the characterization and cloning of QTL as well as Mendelian traits (Pflieger *et al.*, 2001). CG analysis is

based on the hypothesis that known-function genes could correspond to loci controlling traits of interest. Such CGs refer either to cloned genes supposed to affect a given trait ('functional CGs') or to genes suggested by their close proximity on linkage maps to loci controlling the trait 'positional CGs'. These candidate genes can be used for identifying favorable alleles as well as following their inheritance in segregating populations (Tuyen and Prasad, 2008). Association studies can be done with a candidate gene approach (using some biological knowledge to prioritize the parts of the genome for the study) (Collins *et al*., 1997).

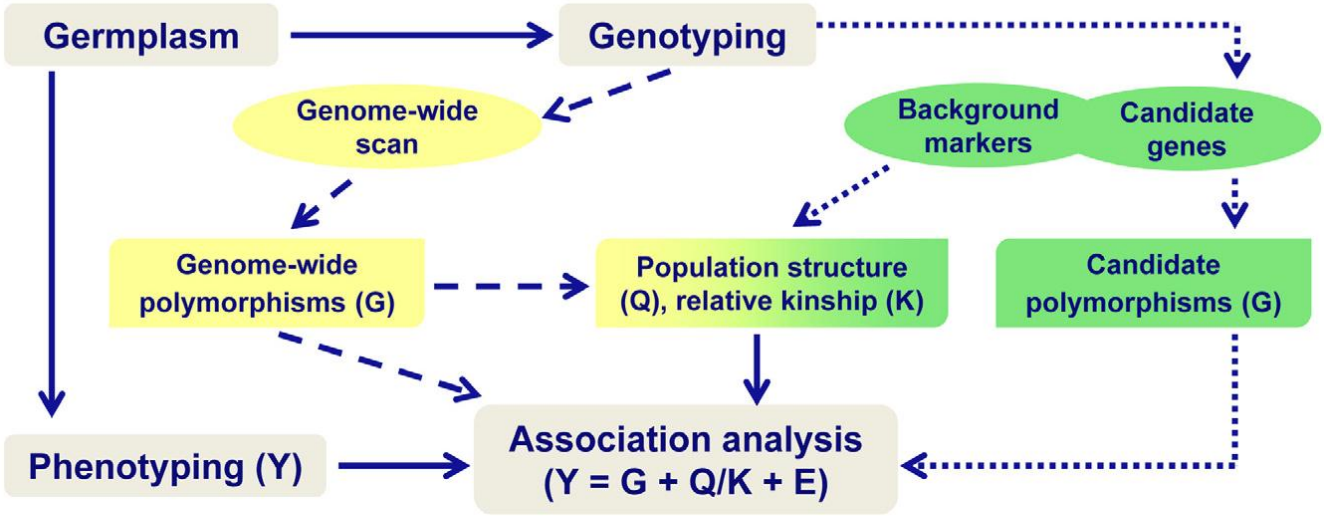
The current wealth of genomic information and tools have provided new methods for identifying candidate genes for the improvement of plants.

Statistical association analyses between molecular polymorphisms of the CG and variation in the trait of interest have been carried out in a few studies. AM grants the opportunity to dissect candidate genes underlay important agronomic QTLs using large germoplasm collections instead of family based crosses (Laird and Lange, 2006) as it can be performed with genome-wide molecular markers designed to target CGs. The final validation of a CG can be provided through genetic transformation and/or sexual complementation and, physiological analyses. In barley and maize, several studies were conducted to find statistical association between genetic variants and complex traits such as yield, flowering time (Kraakman, 2004; Thornsberry, 2001) and with kernel size in the case of maize (Breseghello and Sorrells, 2006). For instance, a set of 20 *CBF* genes in barley have been identified by Skinner *et al.* (2006).

2.4.4 Genome Wide Association Study

Genome-wide association studies (GWAS) are a powerful approach for the exploitation of the natural variation to detect the genomic causes for phenotypic variance by testing the relationship between genotypic and phenotypic variations. GWAS analysis achieved several successes in many organisms through identification of experimentally determined associations (Aranzana *et al*., 2005; Zhao *et al*., 2007) and associations that were subsequently experimentally confirmed (Klein et al., 2005; Sladek *et al*., 2007). They also highlight many feasible novel associations (Atwell *et al*., 2010; Todesco *et al*., 2010). Association studies can be done with a genome-wide approach (without assuming one region of

the genome is more likely to harbor the associated genetic factor) (Collins *et al*., 1997). The construction of association mapping in comparison with candidate genes approach is illustrated in Figure 8.



Genome-wide association mapping	Candidate-gene association mapping
It is a comprehensive approach to systematically search the genome for causal genetic variation. A large number of markers are tested for association with various complex traits, and prior information regarding candidate genes is not required. It works best for a research consortium with complementary expertise and adequate funding.	Candidate genes are selected based on prior knowledge from mutational analysis, biochemical pathway, or linkage analysis of the trait of interest. An independent set of random markers needs to be scored to infer genetic relationships. It is a low cost, hypothesis-driven, and trait-specific approach but will miss other unknown loci.

Figure 8: Schematic diagram and contrast of genome-wide association mapping and candidate-gene association mapping. The inclusion of population structure (Q), relative kinship (K), or both in final association analysis depends on the genetic relationship of the association mapping panel and the divergence of the trait examined. E stands for residual variance (Source: Zhu *et al*., 2008)

Genome-wide association studies (GWAS) are a powerful approach for the exploitation of the natural variation to detect the genomic causes for phenotypic variance by testing the relationship between genotypic and phenotypic variations. GWAS analysis achieved several successes in many organisms through identification of experimentally determined associations (Aranzana *et al*., 2005; Zhao *et al*., 2007) and associations that were subsequently experimentally confirmed (Klein et al., 2005; Sladek *et*

al., 2007). They also highlight many feasible novel associations (Atwell *et al*., 2010; Todesco *et al*., 2010). Association studies can be done with a genome-wide approach (without assuming one region of the genome is more likely to harbor the associated genetic factor) (Collins *et al*., 1997). The construction of association mapping in comparison with candidate genes approach is illustrated in Figure 8.

In a study using 1536 SNPs on 500 UK barley cultivars having a strong population structure, interesting marker-trait associations were assessed for some phenotypes by careful application of mixed model analysis (Cockram *et al*., 2010). Similarly, in a study using 615 cultivars and 32 morphological and 10 agronomic traits, Wang *et al*. (2011) concentrated on comparing various statistical approaches for association mapping in barley. They showed the superiority of mixed model methodology for GWA analysis to assess marker-trait association for complex traits in barley.

In a recent study (Varshney *et al*., 2012) exploring marker-trait associations for yield and agronomic traits under drought conditions, the allelic diversity present in the germplasm collection of barley held in the ICARDA genebank was analyzed through an LD-based GWA mapping approach. A set of 223 accessions sampled from ICARDA genebank was grown in two contrasting environments in Syria, namely in a favorable site (Tel Hadya) and a dry site (Breda). Genotyping of the germplasm collection with SSR and SNP markers (Varshney *et al*., 2010) was integrated with additional 710 DArT markers to perform statistical analysis of population structure, LD decay and marker-trait associations.

3 Agronomic traits and their genetic bases

3.1 Flowering date, importance, QTLs and genes

Flowering time reflects the adaptation of a plant to its environment and is a major factor determining the reproductive success in plants, impacting on grain yield in crop species. It is a complex trait that shows continuous variation and is affected by numerous QTLs both in outbreeding (Buckler *et al*., 2009) and in inbreeding species (Wang *et al*., 2010). Barley has evolved a late flowering phenotype in temperate environments to benefit from an extended vegetative period for resource storage. On the

contrary, early flowering has evolved as an adaptation to short growing seasons and as an escape mechanism from hot and dry summers. The variation in response to low temperature (vernalization) and day length (photoperiod) has been selected in barley to provide adaptation to different environments and farming practices (Wang *et al*., 2010).

Three major vernalisation genes act to control flowering time in response to temperature (Yoo *et al*., 2007). In the barley vernalization pathway, the protein products of the alleles of three genes interact to determine growth habit: *Vrn-H1*, *Vrn-H2*, and *Vr-H3* loci (Takahashi and Yasuda 1971, Yan *et al*., 2003; Oliver *et al*., 2009)). The vernalization response in cultivated barley is mainly due to an epistatic interaction between *Vrn-H1* and *Vrn-H2*, a model that has been validated by studies of the molecular variation at the two loci (Szucs *et al*. 2007).

Despite the analogous vernalization-response phenotypes of Arabidopsis and temperate grasses, genes governing vernalisation response in barley are not orthologous to the major vernalisation response genes in *Arabidopsis* suggesting that the two ancestries have evolved vernalization pathways independently (Yoo *et al*., 2007). In contrast to the unproven relevance of model species in the identification of cereal vernalisation genes, orthologous genes involved in the photoperiod pathway have been detected in *Arabidopsis* and grasses. Homologues of these genes, mainly GIGANTEA (GI), CONSTANS (CO), and FLOWERING LOCUS T (FT), have been associated with grass photoperiod pathway loci (Wang *et al*., 2010). Also, the major photoperiod response gene in barley, *Ppd-H1* turned out to be a homologue of a pseudo response regulator involved in flowering time control and phasing the clock in *Arabidopsis* (Turner *et al*. 2005).

Many significant SNP markers were reported to be associated with the trait FD. Some of these QTLs hit genomic regions that were previously reported to harbor major genes including *HvFT3*, *PpdH1*, *HvFT4*, *eps2*, *HvGI*, *HvCO3*, *HvFT1* and *HvCO1* (Wang *et al*., 2010; Laurie *et al*., 1995; and Griffiths *et al*., 2003). In a study by Pasam *et al*., 2012 using a collection of 224 spring barleys of worldwide origin, fragments from three flowering time candidate genes were re-sequenced and SNPs within the gene *PpdH1* revealed the largest effects on such trait (Stracke *et al*., 2009).

Early flowering, with or without fulfillment of the vernalization requirement in the winter and spring

crops, respectively, can be achieved through deploying alleles at various major gene loci affecting flowering time, such as *Ppd-H1* affecting photoperiodic response (Turner *et al*. 2005).

In GWA study by Wang *et al*. (2011) three heading date QTL, QHd2H.64, QHd3H.126-127 and QHd7H.37-41, were consistently detected. All these QTLs have been detected in previous mapping studies as QHd2H.64 was detected by SNP markers located on the long arm of chromosome 2H near the centromere. (Lorenz *et al*., 2010). In a recent study, SNP marker 11_20438 linked to heading date QTL was in the same chromosomal location as *eam6* (Comadran *et al*., 2011).

3.2 Plant height, importance in breeding, QTLs and genes

Plant height is among the most important biomass yield components. Reduction of plant height was a major target for cereal breeding programs during the Green Revolution to reduce lodging and improve harvest index. In a study to investigate the optimum plant height, Abeledo *et al*. (2002) found that for modern barley cultivars the optimum height ranges from 70 to 90 cm while greater height increases the risk of crop lodging and lower height complicates harvest and reduces the ability of the crop to compete with weeds. Height is highly influenced by the environment, particularly by drought (Baum *et al*., 2003).

Plant height in barley appears to be controlled by many genes, including dwarfing, semi-dwarfing, and other loci (Yu *et al*., 2009). Because dwarfing genes depress vigor and grain yield, they are not useful in breeding programs. Semi-dwarfing (*sdw*) genes are useful and more common than dwarfing genes. Three alleles at the *sdw1* locus [*sdw1.a* (Jotun), *sdw1.c* (*denso*), and *sdw1.d* (Diamant)] have been used to reduce plant height in many semi-dwarf cultivars and are known to also delay heading (Ren *et al*. 2010). Jia *et al.* (2009) proposed GA-20 oxidase as a candidate for the *sdw1/denso* gene. Barley malting varieties carrying the *sdw1* gene have not been approved in North America (Hellewell *et al*., 2000; Kuczyńska *et al*., 2012). However, the short culm *hcm* gene has been mapped to centromeric region of chromosome 2HL and has been proposed to reduce plant height in Upper Midwestern U.S. malting barley varieties (Franckowiak, 2000). In a doubled haploid population derived from a cross

between two winter barley cultivars Igri and Danilo, QTL for plant height were found on chromosomes 4HL, 6HL and 5H (Backes *et al*., 1995).

3.3 Yield and yield components, QTLs and genes

Yield is a complex, polygenic trait strongly influenced by environmental conditions, eg it can be severely reduced by drought (Andrade *et al*., 1996). Drought during seedling establishment or during the period of leaf area expansion causes a decrease in crop leaf area.

Thousand grains weight is one of the major yield components having direct effect on the final yield. Pasam *et al*. (2012) revealed 21 QTLs for this trait, some of which are in vicinity of row type genes. Some of the QTLs were consistent with previously mapped QTLs in the same genomic regions. Spike number per unit area is considered another important yield component for barley under dry land conditions regardless of the influence of water stress (Arnon, 1972).

A European germplasm collection of 146 two-rowed spring barley cultivars was used to carry out LD mapping of yield traits using 236 AFLP markers (Kraakman *et al*., 2004). Associated markers were identified in regions where QTLs for yield had already been found in barley (Romagosa *et al*., 1999 and Li *et al*., 2006). Li *et al*. (2006) reported several QTLs for yield and its components**,** such as number of grains per spike on chromosome 1H. In addition Cakir *et al*. (2003) reported three QTLs for grain yield in barley on chromosomes 2H, 3H and 5H, respectively. Additionally, studies by Franckowiak and Lundqvist (2002), Buck-Sorlin (2002) and Babb and Muehlbauer (2003) have resulted in identification of Mendalian loci for tiller number on chromosome 3HL and 6HL and Vinod *et al*. (2006) identified EXP15 as a candidate gene on chromosome 1 controlling number of tillers under well-watered conditions.

High throughput SNP genotyping and association mapping in barley identified several important QTL for yield and yield components (Comadran *et al*., 2011). Associations for grains per spike identified SNPs tightly linked to known major genes determining spike morphology in barley (*int-c*) (Ramsay *et al*., 2011). In addition, the largest QTL for heading date co-locates with *eam6*, a major locus for

heading date in barley for autumn sown conditions, which has been reported to have significant effects on yield (Cuesta-Marcos *et al*., 2009).

4 Genetic x Environment interactions

Selection for many traits is not only being complicated by their quantitative nature, but also by the interaction between genotype and environment (GE). As a result of this interaction, the ranking order of varieties may change as the growing conditions (environments) change. Yield is a complex, polygenic trait that is strongly influenced by environmental factors. For example, the changes of yield in relation to environmental changes are studied in the context of the concept of yield adaptability. Adaptability can be described as the reaction of the genotype to environmental factors, often defined in terms of linear or quadratic functions (Lin *et al*., 1986). A well-known measure for adaptability is the slope of the regression of yield for an individual cultivar on the mean yield (over all cultivars) across environments (Finlay and Wilkinson, 1963; Eberhart and Russell, 1966).

Several researchers have conducted multi-environment trials for various traits in different plant species, e.g. drought resistance in cotton (Saranga *et al*. 2001), growth and yield in rice (Hittalmani *et al*. 2003), and yield in barley (Teulat *et al*. 2001; Romagosa *et al*. 1996; Voltas *et al*. 2001; Malosetti *et al*. 2004). They all succeeded in identifying loci that interacted with the environment, so loci underlying GE. Some loci for GE co-localized with loci for the trait mean expression, while others appeared at positions where no QTLs for the mean expression were found.

4.1 Importance of drought as a limiting factor in barley production

4.1.1 Genetic studies of drought response in barley

Drought is defined as the absence of adequate moisture necessary for normal plant growth and completion of the life cycle (Zhu, 2002). Generally, when plants are exposed to environmental stress

such as drought, cold, or high salt, they undergo physiological and biochemical adaptations (Bray, 1993; Ingram and Bartels, 1996; Thomashow, 1999).

Drought stress is the most common adverse environmental condition that can seriously reduce crop productivity and is considered the main limiting factor of crop productivity. The Mediterranean region is sensitive to drought and potentially vulnerable to future climate changes (Rizza *et al.*, 2004). Drought might become more frequent with prolonged summer (Bolle, 2003). As water resources for agronomic uses become more limiting, the development of drought-tolerant lines becomes increasingly more important (Bruce *et al*., 2002) to improve agricultural productivity and to reduce agricultural use of fresh water resources. As a result, understanding the mechanisms of drought tolerance and breeding for drought-resistant crop plants has been the major goal of plant biologists and crop breeders. Improving the tolerance of crops to drought compared with other abiotic stresses, requires a broader interdisciplinary approach, involving an understanding of the factors (e.g. availability of water during the crop cycle) determining yield in a particular target population of environments (Collins *et al*., 2008). Plant water deficits may occur as a consequence of a seasonal decline in soil water availability, developing in the long term, or may result from drought spells. An increased evaporative demand of the atmosphere occurring mostly on a daily basis, affects total carbon gain by the crops, even irrigated ones. The timing, intensity and duration of stress episodes are pivotal to determine the effects produced by drought. Plant strategies to control water status and resist drought are numerous (Schulze, 1986).

Consequently, efforts are directed towards a better understanding of the genetic basis of the adaptive response of plants to drought and how best to exploit this knowledge for breeding purposes. The essence of good drought management is to use this range of responses to best advantage (Hafid *et al*., 1998).

However, drought tolerance is recalcitrant to molecular genetics study mainly due to the limited awareness of specific traits linked to drought tolerance. Furthermore, it is difficult to conduct drought stress treatments in a quantitative and reproducible way. These difficulties have significantly impeded research on plant drought tolerance. Consequently, the biological basis for drought tolerance is still largely unknown and few drought tolerance determinants have been identified (Ludlow and Muchow 1990; Bohnert *et al*., 1995; Araus *et al*., 2002; Bruce *et al*., 2002). The slow pace in revealing drought

tolerance mechanisms has hampered both traditional breeding efforts and use of modern genetics approaches in the improvement of drought tolerance of crop plants.

Adaptability is a natural reaction of genotype in order to survive and reproduce. Dimitrijević *et al*. (2002) proved that stability and adaptability represent genotype reaction to environmental variation. Stability means very small genotypic reaction to environmental changes, and in a broad sense, could not be considered as evolutionary favorable in natural conditions. However, in agriculture, stability represents a desirable reaction of cultivated genotypes, forced and supported by humans, ensuring the similar yield level in different environmental conditions through small genotype-environmental interaction. The border between adaptability and stability is quite hazy, reflecting in different and sometimes mixed up definitions of these two.

Drought responses can be classified into different categories including drought escape, exemplified in the early flowering and drought tolerance, represented in stem remobilization and resurrection of the plant (Levitt, 1972). Genetic, molecular and physiological approaches have provided insight into stress signal perception and responses, leading to the identification of signalling molecules, stress-inducible genes and transcription factors that regulate them. Main players in these molecular networks include DREB transcription factors (Agarwal *et al*., 2006) and the hormone ABA (Seki *et al*., 2007).

In a study by Shinozaki and Yamaguchi- Shinozaki (1997) plants were shown to have at least two major pathways, abscisic acid (ABA)-dependent and ABA-independent, for the induction of moisture deficit stress-inducible genes. ABA plays a significant role throughout dormancy and seed development (Seiler *et al*., 2011). ABA accumulation is triggered while the plant is exposed to drought which results in stomatal closure and induces expression of stress-related genes (Shinozaki and Yamaguchi-Shinozaki, 2007). ABA-independent gene activation often involves a *cis*-acting element called a dehydration response element (DRE; also known as a C repeat [CRT]) that responds to drought and low temperature (Baker *et al*., 1994; Yamaguchi-Shinozaki and Shinozaki, 1994) and has been found in many plants (Jiang *et al*., 1996; Dunn *et al*., 1998; Choi *et al*., 1999).

Stockinger *et al*., (1997) identified a transcription factor that binds the DRE/CRT element. This protein, designated CBF1 (C-repeat binding factor 1), has a potential nuclear localization sequence (NLS), an

AP2-DNA-binding domain, and an acidic activation domain. The *Arabidopsis CBF* (*DREB1*) genes are a small multigene family consisting of six paralogs that include three intensively studied genes (*CBF1/DREB1B*, *CBF2/DREB1C*, and *CBF3/DREB1A*) in an 8.7-kb region on chromosome 4 (Gilmour *et al*., 1998; Liu *et al*., 1998), and lesser studied genes on chromosome 5 (*CBF4/DREBID*; Nakamura *et al*., 1998; Thomashow *et al*., 2001) and chromosome 1 (DREB1E and DREB1F; Sakuma *et al*., 2002). The expression patterns of these genes have notable differences. For example, only the three *CBF/DREB1* genes on chromosome 4 have been shown to be chilling induced (*CBF/DREB1*) (Gilmour *et al*., 1998; Sakuma *et al*., 2002). In comparison, *HvDREB1* (Xu *et al*., 2009) encoding dehydration-responsive element binding protein 1.

In a study by Rizza *et al.* (2004) over three years, a collection of winter/spring, 2-rows/6-rows European barley cultivars was evaluated for yield performance under different water regimes in a Mediterranean environment. In general, most reproductive genotypes were 2-rowed types. However, further analyses in this research considering the average grain yield of the three years showed eight genotypes among all cultivars ranking the heightest in yield potential and the minimal G×E interaction in both treatments. Remarkably, most of genotypes with superior yield capacity were spring 2-row types. Notably, genotypes studied are favoured with genes for a wide range of adaptability under favorable and stress environments. These genotypes can be used for further investigation to understand which metabolic processes and morphophysiological traits are important to declare assure yield performance under different environments. Interestingly, this research has provided us an ideal starting point for the present project of our research.

Scope and objectives

Drought is recognized as the most common environmental factor influencing barley production in the Mediterranean area, where this phenomenon is expected to worsen with ongoing climate changes (Rizza *et al.*, 2004). To secure stable crop yields, breeding for drought resistance is an important objective. As most drought-related traits in crops are quantitative, molecular and genetic dissection of drought tolerance is expected to lead to the identification of key markers/loci and favorable alleles through exploration of biodiversity. Genetic, molecular and physiological approaches have provided insight into stress signal perception and responses, leading to the identification of signaling molecules, stress-inducible genes and transcription factors that regulate them. Main players in these molecular networks include DREB transcription factors (Agarwal *et al*., 2006) and the hormone ABA (Seki *et al*., 2007).

This project intends to identify genetic factors associated with key agronomic traits such as flowering date, plant height, and grain yield in barley by exploiting a cultivar collection previously phenotyped for yield performance under different water regimes (Rizza *et al*., 2004). An association genetics approach was adopted in order to correlate allelic variation of molecular markers with the trait variation.

An initial objective of the project was to evaluate the possible association between selected candidate genes for drought tolerance and yield performance under different water conditions. To this end, the following tasks were developed:

- Structure analysis of the Rizza germplasm collection based on AFLP fingerprinting
- Re-sequencing and identification of SNPs and haplotypes in barley CBF genes
- Association analyses were subsequently expanded to the whole genome taking advantage of a newly developed genome-wide Illumina panel comprising 7,864 SNPs. To this end, the following tasks were completed: Genome-wide genotyping of Rizza germplasm collection with Illumina SNP panel (in collaboration with EXBARDIV Consortium)
- Analysis of population structure based on a subset of 260 SNPs
- Genome-wide association scans for FD, PH and GY
- Comparison of results with previously published data for the same traits

Study workflow

Initial objective

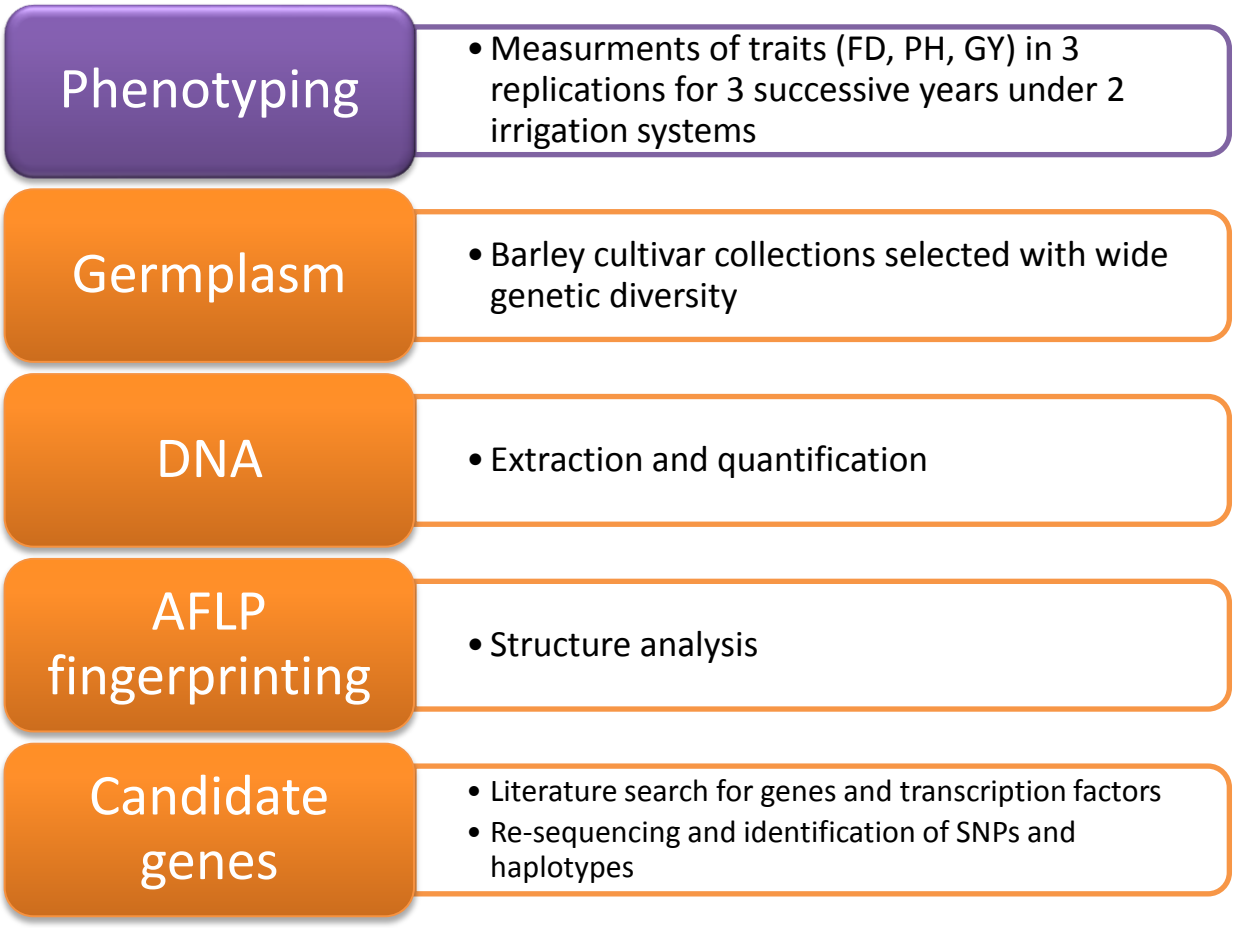


Chart 1: Orange colored boxes are the steps carried out in this study. Purple colored box only the data was utilized to carry out this research study.

Genome wide association study workflow

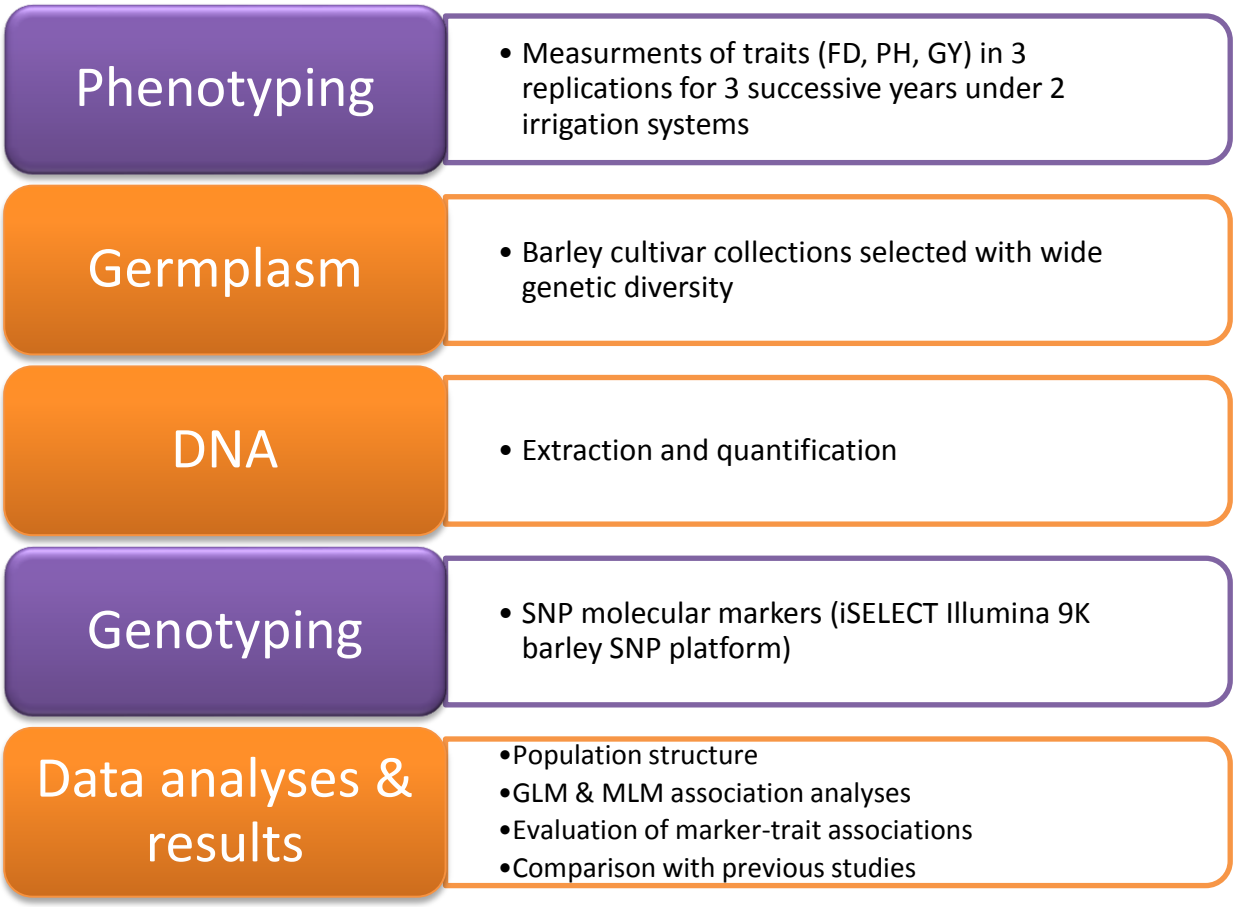


Chart 2: Orange colored boxes are the steps carried out in this study, the purple colored boxes only the data was utilized to carry out the association analysis.

CHAPTER 2

MATERIALS AND METHODS

Materials and methods

1 Genetic materials

A panel of 83 barley (*Hordeum vulgare* L.) cultivars (Table 1) was chosen to represent diversity of cultivated European germplasm for drought tolerance (Rizza *et. al.*, 2004). These 83 accessions are grouped according to their growth habit into 3 categories: 19 spring 2-rows, 24 winter 2-rows, 40 winter 6-rows genotypes. Together, these cultivars originated from 9 European countries (Table 1).

Table 1: Genotypes used for this study with their row types, growth habit and origins

ID	Accession name	Growth habit*	Row type*	Origin***	ID	Accession name	Growth habit*	Row type*	Origin	ID	Accession name	Growth habit*	Row type*	Origin
1	ARAMIR	**S	2	NLD	**29**	ARDA	W	2	ITA	**57**	SAMSON [s]	W	6	FRA
2	BARKE	S	2	ITA	**30**	FJORD [s]	W	2	-	**58**	SONORA	W	6	FRA
3	EXTRA	S	2	UK	**31**	ARCO [s]	W	2	UK	**59**	TREBBIA	W	6	ITA
4	APEX	S	2	DEU	**32**	ALPHA	W	2	FRA	**60**	ABONDANT [s]	W	6	FRA
5	GROSSO [s]	S	2	NLD	**33**	ISACCO [s]	W	6	ITA	**61**	GAIANO [s]	W	6	ITA
6	PRISMA [s]	S	2	NLD	**34**	TRASIMENO	W	2	YOG	**62**	GOTIC [s]	W	6	FRA
7	ALEXIS	S	2	DEU	**35**	TIPPER [s]	W	2	UK	**63**	PIRATE [s]	W	6	FRA
8	STEFFY	S	2	DEU	**36**	KELIBIA	W	2	FRA	**64**	SERENO [s]	W	6	ITA
9	MAGDA	S	2	DEU	**37**	ORCHIDEA	W	2	UK	**65**	PASSPORT	W	6	FRA
10	CHERI	S	2	DEU	**38**	ALFEO [s]	W	2	ITA	**66**	DAHLIA [s]	W	6	FRA
11	ATEM [s]	S	2	NLD	**39**	NURE	W	2	-	**67**	DJEBHEL [s]	W	6	FRA
12	FORMULA [s]	S	2	SWD	**40**	ULTRA	W	2	ITA	**68**	LETIZIA [s]	W	6	ITA
13	AURA	S	2	DEU	**41**	PASTORAL [s]	W	2	FRA	**69**	MAJESTIC [s]	W	6	FRA
14	DIGERSANO	S	2	ITA	**42**	ONICE	W	6	ITA	**70**	NIKEL [s]	W	6	FRA
15	APHRODITE	S	2	FRA	**43**	JAIDOR [s]	W	6	FRA	**71**	TAMARIS [s]	W	6	FRA

Table 1: Continued

ID	Accession name	Growth habit*	Row type*	Origin***	ID	Accession name	Growth habit*	Row type*	Origin	ID	Accession name	Growth habit*	Row type*	Origin
16	TREMOIS	S	2	ITA	**44**	AMILLIS	W	2	FRA	**72**	AGER	W	6	FRA
17	MARISOTTER	**W	2	UK	**45**	ETRUSCO s	W	6	ITA	**73**	BALKAN	W	6	FRA
18	LOMBARD	W	2	FRA	**46**	VERTIGE	W	2	FRA	**74**	CANORO s	W	6	ITA
19	TIDONE	S	2	ITA	**47**	MIRCO	W	6	ITA	**75**	CRIMONT s	W	6	BEL
20	ALISEO	W	6	ITA	**48**	ASSO	W	2	ITA	**76**	EXPRESS	W	6	FRA
21	MICUCCIO s	W	6	ITA	**49**	MATTINA s	W	6	FRA	**77**	GERBHEL s	W	6	FRA
22	RED s	W	2	ITA	**50**	PERGA s	W	6	DUE	**78**	GLENAN	W	6	FRA
23	GEORGIE s	S	2	UK	**51**	BARAKA	W	2	FRA	**79**	PLAISANT	W	6	ITA
24	KASKADE s	W	2	DEU	**52**	FEDERAL s	W	6	FRA	**80**	REBHEL s	W	6	FRA
25	KRONA	S	2	DEU	**53**	BARBEROUSSE	W	6	FRA	**81**	SELVAGGIO s	W	2	HUN
26	PILASTRO	W	6	FRA	**54**	CRITER s	W	6	FRA	**82**	VETULIO s	W	6	ITA
27	MAGIE	W	2	FRA	**55**	BALDA	W	6	ITA	**83**	TEA	W	2	ITA
28	PUFFIN s	W	2	FRA	**56**	ARMA s	W	6	FRA					

*European barley database (http://barley.ipk-gatersleben.de/ebdb.php3).
**S = spring cultivar, and W = winter cultivar.
***Country of origin abbreviations: NLD = Netherlands, ITA = Italy, UK = United Kingdom, DEU = Germany, SWD = Sweden, FRA = France, YOG = Yugoslavia, BEL = Belgium, and HUN = Hungary.
Colors legends: Blue = winter 6 rows, red = winter 2 rows, and green = spring 2 rows barleys.
s Genotypes used for sequencing (candidate genes work).

2 Preliminary phenotypic information

Phenotypic information used in this study was outsourced from Rizza *et al.* (2004). Experimental design and procedures are briefly summarized here. A field experiment was carried out in Foggia (southern Italy) through three successive growing seasons (1999-2000-2001) under 2 different water regimes: rainfed non-irrigated (R) and controlled irrigated (I) conditions (Rizza *et al.*, 2004). Experimental design was randomized complete block with three replications for each irrigation condition and each year. Each experimental unit consisted of a 4 m^2 plot. The sowing dates were 3, 13, and 21 December respectively in 1998, 1999, and 2000. Phenotypic data were recorded for different agronomic traits. Date of heading was recorded when spikes emerged from about half of the culms in a plot. Plant height (excluding spike) was measured in all plots for 3 plants per genotype, at Zadoks growth stage 7.5 (Figure 9). For characterizing the genotypes for grain yields under R and I conditions, this was calculated by standardizing the relative yield under each condition for each year, and the grain yield of each genotype to the average yield of the entire genotypes in the panel (Rizza *et al.*, 2004). These existing data were organized in spread sheets for statistical and association analyses (Supplementary material table S1).

3 Cultivation of barley plants

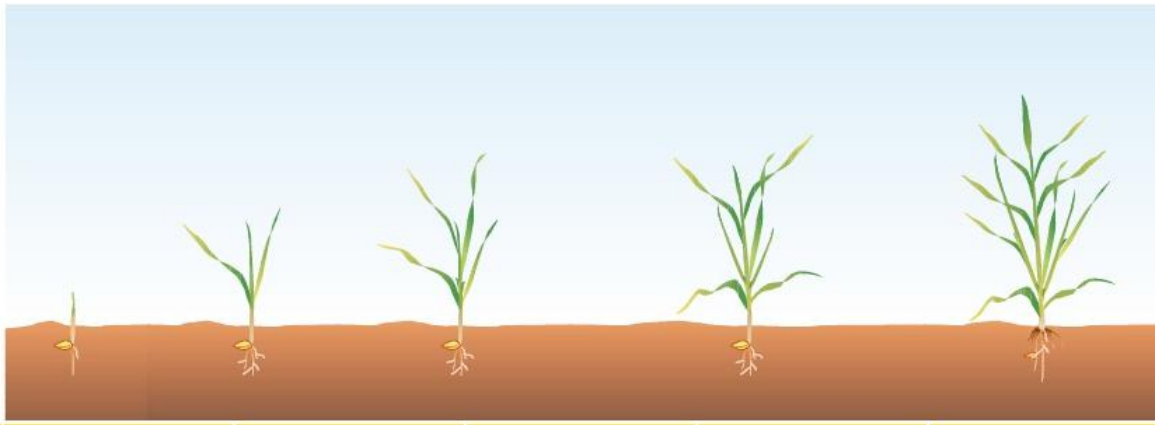
Seed stocks for the 83 barley genotypes of interest (Table 1) were obtained from the Centro di Ricerca per la Genomica e la Postgenomica Animale e Vegetale (CRA-GPG, Fiorenzuola d'Arda, PC), Italy (http://centrodigenomica.entecra.it/).

All these genotypes were re-grown in semi-controlled conditions in the CETAS University greenhouse at Tavazzano (Lodi, Italy) from mid December 2009 up to June 2010. Plastic pots (13.5 × 13.5 × 15 cm, ca. 2 L) were filled with soil in the following ratios: 4 peat-moss, 3 soil, 1 clay. Plastic cages were inserted in each pot to contain and support plants during their growth. Initially, 3-5 seeds from the same genotype were sowed in each pot; plants were then thinned to leave 1 or 2 plants maximum per genotype / pot. Irrigation was applied twice per week or 3 times in hot weather weeks. Fertilization

with 3-5 g urea/ pot was applied 3 times during plant growth at pre-germination, post-germination, and tillering stages ("Zadok's growth stage", Figure 9, Zadoks *et al*., 1974). During the maturation of grains, pots were fertilized with a balanced nutrient solution and treated with Folicur® (Tebuconazole 4,35%) fungicide. Fresh green leaves from each individual plant were collected at the 4-6 leaves stage and frozen on dry ice. Leaves were lyophilized under a vacuum for subsequent DNA extraction. Before lyophilization, leaf tissues were stored at -80°C.

In the next 2 years, seed stocks were propagated again by Single Seed Descent (i.e. harvesting progeny from one plant per genotype) as a backup for future research. Seeds were kept in a dedicated storage facility at -4°C at Parco Tecnologico Padano (PTP), Lodi, Italy. Seed stocks were inserted in the University of Milan (UMIMI) germplasm collection (GM) for future records.

Further information on the germplasm can be obtained from the European Barley Database (EBDB, http://barley.ipk-gatersleben.de/ebdb.php3).



Zadoks Growth Stage	GS 00 - 09	GS10 - 19	GS20 - 29	GS30 - 39	GS40 - 49
Development phase	Germination	Seedling growth	Tillering	Stem elongation	Booting

Zadoks Growth Stage	GS 50 - 59	GS60 - 69	GS70 - 79	GS80 - 89	GS90 - 99
Development phase	Ear emergence	Flowering	Milk Development (grain fill period)	Dough Development (grain fill period)	Ripening

Figure 9: Illustration for Zadoks decimal growth stages (Zadoks *et al*., 1974). Source: *Grains research & development corporation (GRDC)*, 2005.

4 Leaf lyophilization

Barley leaves were packed in specialized small bags 8 × 12cm (white envelops, BUSTE A SACCO KRAFT MONOLUCIDO code 303) dedicated for lyophilization purposes. Then envelopes containing the fresh leaves lyophilized under vacuum at -50^{o}C (Christ ALPHA 1-2 LD plus) for a minimum of 3 days, then stored in the dry fridge facility at 4^{o}C or -20^{o}C (PTP) Lodi, Italy.

5 DNA extraction and quantification

Genomic DNA was extracted from 15-20 mg of lyophilized leaf tissue and then ground at room temperature using a Retsch® MM300 Mixer Mill with metal tungsten carbide beads. Two DNA extraction protocols were initially compared for the yield and quality of DNA obtained. As a result, the MATAB protocol was finally selected.

Promega plant extraction Kit (Wizard® Magnetic 96 DNA Plant System). Extraction was carried out according to the manufacturer's instructions. The desired amount of ground lyophilized leavf material (20 mg) was supplemented with 300 µl of lysis/buffer A. After mixing, samples were centrifuged at 1700 *xg* for 10 min (temperature?) and the supernatants were transferred into 96-well plates. Next, 60 µl of magnetic beads (MagneSel®) with 40 µl of lysis/buffer B were added to the supernatant, then incubated for 5 min at room temperature. The 96-well plates were then transferred to a magnetic separation device (MagnaBot®) for 1 min, then liquid was discarded; 150 µl washing buffer were added and mixed for 10-15 sec. Plates were placed again on the MagnaBot® for 30 sec and liquid removed. The washing step and magnetic separation were repeated again. After removing as much liquid as possible, samples were left 5 min to dry at room temperature. Plates were removed from the MagnaBot® and 50 µl of nuclease-free pure water was added to each sample, to re-suspend the MagneSil®, then incubated at room temperature for 5 minutes. Plates were placed again on the MagnaBot®. The purified DNA was transferred to fresh nuclear-free vials of 1.5 ml.

This protocol yielded from 40-100ng of DNA /µl per sample (2-5-µg of total DNA after elution in 50µl of ultra-pure water) and quality of DNA was >1:1.65 absorption range at 260/280 absorbance.

MATAB extraction protocol. MATAB extraction buffer: 20 mM EDTA pH 8.0, 100 mM Tris-HCl pH 8.0, NaCl 1.5 M, 0.5% (W/V) Na_2SO_3, 2% (W/V) hexadecyltrimethylammonium bromide (MATAB), Polyethylene glycol (PEG) solution 1% (W/V).

Initially, 15-20 mg of leaf powder (previously lyophilized) were added to 300 µl of pre-warmed MATAB DNA extraction buffer at 62°C and mixed well. Samples were incubated at 62°C for 20 minutes, and then briefly centrifuged; 10 µl of RNasiA (20mg/ml Invitrogen) were added then samples were incubated 10 min at 65°C to eliminate RNA contamination. After adding 360 µl of Chloroform: Iso-amyl alcohol (ratio 24:1), tubes were gently inverted to homogenise the solution. After centrifugation for 20 minutes at 6200 *xg*, 250 µl of supernatant were transferred to a clean tube and the same procedure was repeated.

Samples were supplemented with 200 µl of isopropanol, mixed by inversion and centrifuged for 20 minutes at 6200 *xg*. The supernatant was discarded, 300 µl of 70% ethanol were added and tubes were centrifuged for 10 minutes at 6200 *xg*. The supernatant was discarded; the pellet was dried and re-suspended in 50 µl of water.

The MATAB protocol yielded 100-450ng of DNA /µl per sample (5-22.500µg of total DNA after elution in 50µl of ultra-pure water) and quality of at least 1.75:2 range of absorption at 260/280 absorbance.

DNA yield and integrity were confirmed by agarose gel electrophoresis and quantification using Quant-iT™ PicoGreen® (Invitrogen™) including a lambda DNA standard: 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) was prepared by diluting the concentrated buffer from the kit 20-fold with distilled DNase-free water. Aqueous working solution of Quant-iT PicoGreen reagent was prepared by making a 200-fold dilution of the concentrated dimethyl sulfoxide (DMSO) solution in the TE buffer (solution protected from the light by using dark colored tubes). DNA standard provided with the kit was diluted in the TE buffer to make the 2 µg/ml solution, then pipetted into the 96-well plate in 3 replications and then 100 µl of the Quant-iT™ PicoGreen® solution was added and mixed well. The reaction is then incubated in the dark for 5 minutes to permit the dye to bind to double stranded DNA.

The fluorescence of the samples is then measured on a plate reader at excitation/emission of 435/535 nm. Finally, the DNA concentration of each sample was estimated based on a standard concentration curve. The final concentrations of all DNA samples were adjusted to 50 ng/µl.

6 AFLP analysis

AFLP reaction was conducted in fluorescence following the published Vos *et al*., (1995) protocol with some modifications. Six AFLP primer combinations (E32M49, E32M55, E36M49, E36M55, E38M55, and E41M55) were used on 144 barley accessions (Table 3). This list is including 83 genotypes from Rizza *et al.,* 2004 (Table 1) panel and 61 wild accessions from Fricano *et al.,* (2009) panel. Digestion, ligation and pre-amplification were carried out according to Vos *et al*., (1995) protocol. Modifications following Fricano *et al.,* (2009) were applied to the original protocol for selective fluorescent labeling: customized forward primers were 5'-labelled with 6-FAM (Applied Biosystems®, Foster City, USA) (Table 2). Products of the selective amplification were diluted tenfold; 2 µl of each dilution were then mixed with 10 µl of deionized formamide and 0.15 µl of GeneScan 1200- LIZ internal size standard (Applied Biosystems®, Foster City, USA). Capillary electrophoresis was carried out using the ABI3730 DNA Analyzer and the GeneMapper 4.0 software (Applied Biosystems®) was used for scoring polymorphic peaks. Polymorphic peaks were then converted to a binary data matrix, as bands were scored as presence (1) or absence (0) to form the raw data matrix (Ghosh *et al.,* 2011). For diversity analysis, data were analyzed using DARwin (Diversity Analysis and Representation for windows) software (Perrier and Jacquemond-Collet, 2006). The similarity coefficient of Jaccard (Sneath and Sokal, 1973) was used for the calculation of genetic relatedness. The Neighbour-Joining (NJ) distance method was employed for the construction of the phylogenetic tree using DARwin v5.0.158 software (Perrier and Jacquemond-Collet, 2006), and statistical support was estimated with the bootstrap method (1000 replications) for the phylogenetic grouping.

Table 2: Primers sequences (from 5' end to 3' end) used in the selective amplification step for AFLP reactions.

Primer type	Sequence (5'->3')
*E32	GACTGCGTACCAATTC**AAC**
E36	GACTGCGTACCAATTC**ACC**
E38	GACTGCGTACCAATTC**ACT**
E41	GACTGCGTACCAATTC**AGG**
**M49	GATGAGTCCTGAGTAA**CAG**
M55	GATGAGTCCTGAGTAA**CGA**

* E = *EcoRI*-adapter
** M = *MseI*-adapter
E32, E36, E38, and E41 were the fluorescent labeled primers

Table 3: The barley 144 accessions used in AFLP experiment

#	*Genotype Name	Row type	**Origin	Growth habit
1	ISACCO	6	ITA	Winter
2	GROSSO	2	NLD	Spring
3	GOTIC	6	FRA	Winter
4	GERBHEL	6	FRA	Winter
5	GEORGIE	2	UK	Spring
6	FORMULA	2	SWD	Spring
7	ALFEO	2	ITA	Winter
8	LOMBARD	2	FRA	Winter
9	PASTORAL	2	FRA	Winter
10	TRASIMENO	2	YOG	Winter
11	DIOMEDE	6	ITA	Spring
12	ATEM	2	FRA	Spring
13	PRISMA	2	NLD	Spring
14	ULTRA	2	ITA	Winter
15	NIKEL	6	FRA	Winter
16	MATTINA	6	FRA	Winter
17	FJORD	2	FRA	Winter
18	PUFFIN	2	FRA	Winter
19	TIPPER	2	UK	Winter
20	BALDA	6	ITA	Winter
21	MAGIE	2	FRA	Winter
22	APHRODITE	2	FRA	Spring
23	AURA	2	DUE	Spring

Table 3: continued

#	*Genotype Name	Row type	**Origin	Growth habit
24	CANORO	6	ITA	Winter
25	CRITER	6	FRA	Winter
26	ARMA	6	FRA	Winter
27	KASKADE	2	DEU	Winter
28	ORCHIDEA	2	ITA	Winter
29	DAHLIA	6	FRA	Winter
30	MARIS OTTER	2	UK	Winter
31	SELVAGGIO	6	HUN	Winter
32	PERGA	6	DEU	Winter
33	LETIZIA	6	ITA	Winter
34	JAIDOR	6	FRA	Winter
35	GAIANO	6	ITA	Winter
36	FEDERAL	6	FRA	Winter
37	ETRUSCO	6	ITA	Winter
38	DJEBHEL	6	FRA	Winter
39	TREBBIA	6	ITA	Winter
40	VETULIO	6	ITA	Winter
41	ABONDANT	6	FRA	Winter
42	ARCO	2	ITA	Winter
43	CRIMONT	6	BEL	Winter
44	MAJESTIC	6	FRA	Winter
45	MICUCCIO	6	ITA	Winter
46	PIRATE	6	FRA	Winter
47	REBHEL	6	FRA	Winter
48	RED	2	ITA	Winter
49	SAMSON	6	FRA	Winter
50	SERENO	6	ITA	Winter
51	STEFFY	2	DEU	Spring
52	TAMARIS	6	FRA	Winter
53	AGER	6	FRA	Winter
54	ALEXIS	2	DEU	Spring
55	ALISEO	6	ITA	Winter
56	ALPHA	2	FRA	Winter
57	AMILLIS	2	FRA	Winter
58	ANGORA	2	DEU	Winter
59	APEX	2	DEU	Spring
60	ARAMIR	2	NLD	Spring
61	ARDA	2	ITA	Winter

Table 3: continued

#	*Genotype Name	Row type	**Origin	Growth habit
62	ASSO	2	ITA	Winter
63	BALKAN	6	FRA	Winter
64	BARAKA	2	FRA	Winter
65	BARBEROUSSE	6	FRA	Winter
66	BARKE	2	DEU	Spring
67	CHERI	2	DEU	Spring
68	DIGERSANO	2	ITA	Spring
69	ELAN	6	FRA	Winter
70	EXPRESS	6	FRA	Winter
71	EXTRA	2	AUT	Winter
72	GLENAN	6	FRA	Winter
73	KELIBIA	2	FRA	Winter
74	KRONA	2	FRA	Spring
75	MAGDA	2	DEU	Spring
76	MIRCO	6	NLD	Spring
77	NURE	2	ITA	Winter
78	ONICE	6	ITA	Winter
79	OTIS	2	ITA	Spring
80	PASSPORT	6	DEU	Winter
81	PILASTRO	6	FRA	Winter
82	PLAISANT	6	ITA	Winter
83	SOLEN	6	FRA	Winter
84	SONORA	6	FRA	Winter
85	TEA	2	FRA	Winter
86	TIDONE	2	ITA	Spring
87	TREMOIS	2	ITA	Spring
88	VERTIGE	2	FRA	Winter
89	AC_METCALFE	2	CANADA	Winter
90	ACI	2	ITA	Winter
91	AIACE	2	ITA	Winter
92	ALANNO	6	ITA	Alternative
93	ALBACETE	6	ESP	Winter
94	AQUILAI	6	ITA	Winter
95	ASSE	6	DEU	Spring
96	BARRAFRANCA	6	ITA	Winter
97	BAZANT	6	POL	Winter
98	BIDO	2	DEU	Spring

Table 3: continued

#	*Genotype Name	Row type	**Origin	Growth habit
99	BOMBAY	2	DEU	Winter
100	BRAEMAR	2	GB	Spring
101	BRUKER_STAMMII	6	AUT	Winter
102	BULBUL89	2	TUR	Spring
103	BURSZTYN	6	POL	Winter
104	CAROLA	6	AUT	Winter
105	CATANIA	6	ITA	Spring
106	CDC_ALAMO	2	CANADA	Spring
107	CDC_CANDLE	2	CANADA	Spring
108	CDC_FIBAR	2	CANADA	Spring
109	CHETIN	6	TUR	Winter
110	CLARA	2	DEU	Winter
111	COLONIA	6	DEU	Winter
112	DICKTOO	6	USA	Winter
113	DURA	6	DEU	Winter
114	EUIJEONGBUS	6	KOR	Winter
115	FROST	6	SVE	Winter
116	KESTREL	6	GB	Winter
117	KLAGES	2	-	Spring
118	LEGACY	6	-	Winter
119	LEONIE	2	DEU	Winter
120	LOMERIT	6	DEU	Alternative
121	MANOLIA	2	FRA	Winter
122	MANSHOLT_FLETUMER	6	NLD	Winter
123	MERLOT	6	DEU	Alternative
124	MOREX	6	-	Spring
125	NEMEX	2	ESP	Spring
126	NUDINKA	2	DEU	Spring
127	NUDO_LEONESSA	2	ITA	Spring
128	OKOS	2	ITA	Spring
129	PAMINA	6	DEU	Winter
130	PARIGLIA	2	-	Winter
131	PEARL	2	UK	Winter
132	PEWTER	2	UK	Spring

Table 3: continued

#	*Genotype Name	Row type	**Origin	Growth habit
133	RAGUSA	6	DEU	Winter
134	SAIGON	2	UK	Winter
135	SCARLETT	2	DEU	Spring
136	SINIS14	6	ITA	Spring
137	SOLETO	6	ITA	Spring
138	STANDER	6	-	Winter
139	TIFFANY	2	DNK	Winter
140	TIPPLE	2	-	Spring
141	TRADITION	6	-	Spring
142	VALLE_DA'OSTA	6	ITA	Winter
143	VOGELSANGHER_GOLD	6	DEU	Winter
144	ZACINTO	2	ITA	Winter

*European barley database (http://barley.ipk-gatersleben.de/ebdb.php3).
**Country of origin abbreviations: NLD = Netherlands, ITA = Italy, UK / GB = United Kingdom / Great Britain, DEU = Deutsch land (Germany), SWD / SVE= Sweden, FRA = France, YOG = Yugoslavia, BEL = Belgium, HUN = Hungarian, AUT = Austria, ESP = Spain, POL = Poland, TUR = Turkey, KOR = KOR = Korea, DNK = Denmark, and USA = United States.
Colors legends: Blue = winter 6 rows, red = winter 2 rows, and fluorescent green = spring 2 rows, dark green = spring 6 rows, and purple = alternative 6 rows barleys.

7 Candidate Genes analyses

A list of barley and rice candidate genes (CGs) previously known for their involvement in drought tolerance was assembled from the literature, with focus on transcription factors from DREB/CBF family and the ABA pathway (Table 4). The corresponding nucleotide sequences (Table 4) were obtained from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) or HarvEST barley (http://harvest.ucr.edu/, Version 1.83) databases.

In the case of sequences obtained from rice CGs: *OsDREB1A, OsDREB1G, OsDREB2A* and *OsDREB2B* (Table 4). We obtained the protein sequence of the CG in rice and used on IPK barley BLAST server (http://webblast.ipk-gatersleben.de/barley/viroblast.php) online search database [barley_HighConf_genes_MIPS] to identify similar nucleotide sequence in barley genome using (tBLASTn). Nucleotide sequences of the highest similarity were then recovered. Consequently, primers were designed on the basis of the homologous barley sequence.

The full length nucleotide sequence encoding amino acid sequences were selected for designing the primers and genomic DNA used for the amplification process. Gene-specific primers were designed (Table 5) with the default parameters using the Primer3 online tool (Rozen and Skaletsky, (1998); http://biotools.umassmed.edu/bioapps/primer3_www.cgi) initially for 6 CGs: *CBF2*, *CBF4A*, *CBF7*, *ABA3*, *DREB1*, *and DHN3* (Table 5). Three PCR programs were designed for the amplification process (Table 7) to amplify three CGs: *HvCBF2*, *HvCBF4A*, *HvCBF7* (Table 5). The extracted genomic DNA was used directly to amplify an initial core collection of 40 genotypes (Table 1) out of the panel as detailed below.

Table 4: candidate genes list from literature search with their NCBI reference numbers

Candidate gene	Species	NCBI Accession no.	Locus	*IPK rice accession	**IPK barley accession	Score	Identity (Query)	*E*-value	Reference
ABA3	*Hordeum vulgare*	X72748	-	-	-	-	-	-	Agarwal *et al.*, 2006
HvCBF1	*Hordeum vulgare*	AY785837	-	-	-	-	-	-	Skinner *et al.*, 2005
HvCBF2	*Hordeum vulgare*	AF442489	-	-	-	-	-	-	Xue 2002
HvCBF2A	*Hordeum vulgare*	AY785841	-	-	-	-	-	-	Skinner *et al.*, 2005
HvCBF2B	*Hordeum vulgare*	DQ097684	-	-	-	-	-	-	Skinner *et al.*, 2005
HvCBF3	*Hordeum vulgare*	AY785845	-	-	-	-	-	-	Skinner *et al.*, 2005
HvCBF4A	*Hordeum vulgare*	AY785849	-	-	-	-	-	-	Skinner *et al.*, 2005
HvCBF7	*Hordeum vulgare*	AY785864	-	-	-	-	-	-	Skinner *et al.*, 2005
HvSPY	*Hordeum vulgare*	AF035820	-	-	-	-	-	-	Robertson 2003
HvDREB1	*Hordeum vulgare*	DQ012941	-	-	-	-	-	-	Xu *et al.*, 2009
NCED2	*Hordeum vulgare*	AB239298.1	-	-	-	-	-	-	Chono *et al.*, 2006
HvDHN3	*Hordeum vulgare*	X15286	-	-	-	-	-	-	Close *et al.*, 1989
OsDREB1A	*Oryza sativa*	AF300970	LOC_Os09g35030	AK374663	61%	193	156/255 (238)	9e-50	Dubouzet *et al.*, 2003; Chen *et al.*, 2008
OsDREB1G	*Oryza sativa*	AY785896	LOC_Os11g13840.1	AK375953	82%	256	180/219 (219)	6e-69	Chen *et al.*, 2008
OsDREB2A	*Oryza sativa*	AF300971	LOC_Os01g07120.2	MLOC_12223.1	67%	111	59/88 (274)	4e-25	Dubouzet *et al.*, 2003
OsDREB2B	*Oryza sativa*	AK099221	LOC_Os05g27930.1	AK249060.1	42%	189	140/328 (298)	2e-48	Chen *et al.*, 2008

E-value shows the significant relation to *Hordeum vulgare* species with the aminoacid identity score for the sequence hit (max. identity to barley).

*IPK accession corresponding to rice and barley from protein search

**IPK accession corresponding to barley from protein search

7.1 PCR amplification

PCR was processed for a core collection of 40 genotypes from the research panel representing diversity in their growth habits and row types (highlighted in Table 1).

Table 5: Primers designed on coding sequences for some CGs, except the reverse primer designed for *HvCBF2* that was including a part of 3'UTR (un-translated region) via NCBI bioinformatics programming online tool (Rozen and Skaletsky, 2000); using primer 3 software (Rozen and Skaletsky, 1998).

CGs	**Primer Sequence 5′->3′**	**Primer orientation**	**Melting temp. (Tm)**	**Expected amplicon size in bp**
HvCBF2	AGGGGCAAGACTACATGACG	Fwd*	60.13	817
	CATGGTTTGAGATTGCTTGC	Rev**	59.28	
HvCBF4A	TACTCAACCACGCACTCCAG	Fwd	59.9	920
	AGGGAAGGAAATAACTGTTTTAAGT	Rev	57.23	
HvCBF7	CCCCAACTACTAACTCCACCAC	Fwd	59.79	898
	GCTAACCCCAATTTGTACATGG	Rev	60.47	

*Fwd = forward primer
**Rev = reverse primer
Re-sequencing obtained from the forward primers for the listed CGs

PCR amplification was carried out in a total volume of 25µl for each genotype (Table 6).

Table 6: The standard PCR reaction master mix preparation used for all CGs work in a total volume of 25 µl reaction

Reagent with stock concentration	**Final concentration in 25µl reaction mixture**	**Volume added (µl)**
Template DNA (Stock conc.)	50 ng	1 µl
Forward primer (10 µM)	400 nM	1 µl
Reverse primer (10 µM)	400 nM	1 µl
dNTPs mix* (2.5 mM each)	200 µM	2 µl
PCR buffer* (5×)	1 ×	5 µl
*$MgCl_2$ (50 mM)	1.5 mM	0.5 µl
Taq polymerase* (5 U/µl)	1 unit	0.25 µl
Ultra-pure nuclease free water*		Reaction mix completed with water up to 25 µl

*Promega® products were used

Thermal-cycling was carried out using a BIORAD PCR thermal cycler^{TM} 96 well machine (Table 7).

Table 7: Thermal cycling programs used for each CG

PCR program for *HvCBF2* amplification

95°C × 4 min × 1

94°C × 45 sec
64°C × 30 sec × 34
72°C × 1 min

75°C × 5 min × 1

PCR program for *HvCBF4A* amplification

96°C × 2 min × 1

95°C × 45 sec
54°C × 30 sec × 34
72°C × 1 min

75°C × 5 min × 1

PCR program for *HvCBF7* amplification

96°C × 2 min × 1

95°C × 45 sec
56°C × 30 sec × 34
72°C × 1 min

75°C × 5 min × 1

7.2 Sequencing process

Re-sequencing was initially focused on a subset of 3 candidate genes. PCR product of one replicate for each accession was first purified to eliminate unincorporated primers and dNTPs so

they do not interfere with downstream sequencing. To this end, we used the ExoSAP-IT® (Exo-nucleases) PCR clean-up protocol (Applied Biosystems®): 2 µl from ExoSAP-IT® solution were added to 5µl of DNA amplicon and incubated 15 min at 37°C; then the enzymatic reaction blocked were then inactivated by incubation at 85°C for 15 min.

Sequencing was performed using automated fluorescent sequencing by BigDye™ Terminator v3.1 Matrix Standard Sequencing Kit (Applied Biosystems®, ABI PRISM® 3700 DNA Analyzer, Foster City, USA). The final reaction was set up in 10 µl volume (Table 8): DNA sequencing reactions were run in 96 multi-well plates on a PTC-200 thermal cycler (MJ Research, USA) using a thermal cycling program of 95°C for 15 sec, 45°C for 5 sec, and 60°C for 2 min for 35 cycles. A precipitation process was carried out afterwards by adding 2.5 µl of125 mM) EDTA and 30 µl of 100% ethanol, mixing by inverting 4 times, then incubating for 15 min at room temperature in the dark. The reaction was then centrifuged at 2000 *xg* at 4 °C for 45 min. Plate then removed and inverted upside down on a tissue paper to get rid of the supernatant, this step was aided by a quick spin-off step at 185 *xg* for a minute. Next, supernatant was discarded and the pellet was washed with 30 µl of 70% ethanol. Samples were centrifuged at 1650 *xg* at 4°C for 15 min. Supernatant was discarded and the plate was left at room temperature until pellets were completely dried. Next, 10 µl of formamide added to each sample and the plate was covered with aluminum foil and stored at 4°C. Subsequently, the plate was submitted to PTP genomics platform (PGP) (Parco Tecnologico Padano, Lodi, Italy) for a final automated sequencing process by capillary electrophoresis.

Table 8: Reagents used for sequencing CGs with forward primers previously designed in this work (see Table 5).

Chemical reagent	**Initial conc.**	**Volume (µl)**	**Final conc.**
*BIG DYE v. 3.1 (Ready Reaction Mix)	10 ×	1	1 ×
Sequence buffer	5×	2	-
Primer	10 µM	0.8	800 nM
ultra-pure molecular grade water	-	5.2	-
DNA	from purified DNA stock (2500ng)	1	50 ng
Total reaction mix	-	10	-

* BigDye™ (Applied Biosystems®)

Raw electropherograms were analyzed using Sequencing Analysis® software (Applied Biosystems Foster City, USA) to obtain FASTA sequences. Subsequently, for the tested gene a consensus sequence for each accession was created, assembling the resulted sequences using *Bioedit* v7.0.9 bioinformatics software (Hall, 1999, http://www.mbio.ncsu.edu/bioedit/bioedit.html) to trim off low-quality regions. The consensus sequences from each gene and from different genotypes were aligned and compared (Figure 13) using the MultAlin web-based alignment (Corpet, 1988) bioinformatics tool: (http://multalin.toulouse.inra.fr/multalin).

8 Genome-wide SNP genotyping

Genotyping with 9K barley SNP array (iSELECT Infinium® Illumina technology) was carried out at TraitGenetics (http://www.traitgenetics.com/en/, Gatersleben, Germany) in collaboration with the EXBARDIV consortium (http://bioinf.hutton.ac.uk/iselect/app/, 2012) and resulted in identification of a total of 7,864 SNPs distributed over the whole barley genome (7 chromosomes). Filtering of monomorphic, failed markers and markers with Minor Allele Frequency (MAF) <10% resulted in the identification of a total of 4,661 informative SNPs.

8.1 Population structure analysis

Population structure analysis was conducted with the software *Structure* v2.3.3 (Pritchard *et al.*, 2000), based on Bayesian statistics using 260 SNP markers selected with less than 10 cM intervals as highly informative from the barley HarvEST Map (Close *et al.*, 2009; and Muñoz-Amatriaín *et al.*, 2011; Comadran *et al*., 2012). "Admixture model" of ancestry and correlated allele frequencies were adopted to analyze the data set. No preliminary information on the number of subpopulations was considered. The proportion of the ancestry of each individual was tested considering a number of K from 1 to 20, with 5 iterations for each value of K, to verify the consistency of the results (Pasam *et al.*, 2012). The settings for burning and MCMC (Markov Chain Monte Carlo) were 250,000 and 100,000, respectively based on the suggestion of Pritchard and Wen (2007). To determine the number K, the model by Evanno *et al.* (2005) was adopted using the Structure Harvest (Earl and vonHoldt, 2012). 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 (Pasam *et al.*, 2012).

8.2 Principle coordinate analysis (PCoA) and diversity tree

PCoA analysis with the same subset of genotypic 260 SNPs markers used to run *Structure* was carried out in *Past* software (Hammer *et. al.*, 2001) using a simple matching similarity matrix.

9 Genome wide association analyses

The selected SNPs were used for a genome wide association study (GWAS) of agronomical traits such as yield, flowering date, and plant height. Using *Tassel* v3.0.114 bioinformatics program (Bradbury *et al.*, 2007; www.maizegenetics.net), two models -General Linear Model (GLM) and Mixed Linear Model (MLM)- were used and compared for calculating the best *P*-values for testing possible associations between markers and the traits under the study (Pasam *et al*., 2012; and Mezaka *et al.*, 2011). A kinship matrix was estimated using *Tassel*. Accounting for population structure was considered for correction to eliminate spurious associations. A threshold for the significance of the association was calculated based on Bonferroni correction (Balding 2006) and used for all traits and all models (*P*-value threshold = 1.9×10^{-4}). Manhattan plots (scatter plot) were produced by TASSEL software; by plotting all SNP markers (each SNP considered as a point) in which they show their location on each of the 7 barley chromosomes on the X-axis against the negative logarithmic *P*-value of each SNP association displayed on the Y-axis. Thus, the result of the strongest association(s) will have the smallest *P*-value.

CHAPTER 3

RESULTS

Results

1 AFLP analysis

1.1 Molecular fingerprinting

For this study, we initially decided to analyze population structure based on AFLP-fingerprinting following the approach successfully used by Fricano *et al.* (2009). To facilitate comparison of results with those from this previous study, we assembled a panel of 144 genotypes including 32 genotypes from Rizza *et al.* (2004), 56 from Fricano *et al.* (2009) and 56 shared between the two studies. Six primer combinations were used (E38M55, E36M55, E32M49, E32M55, E36M49, and E41M55). Data analyses resulted in a total number of 71 polymorphic loci (Table 9).

Table 9: Primer combinations used in the AFLP experiment and peaks identified in this study

Primer combination	Overall detected peaks n.	Polymorphic peaks n.
*E38*M55	360	24
E36M49	290	8
E36M55	325	16
E32M49	100	7
E32M55	218	13
E41M55	171	3
Total n.	**1464**	**71**

* E = *EcoRI*-adapter
* M = *MseI*-adapter

In the previous work from Fricano *et al.* (2009) a total of 215 polymorphic peaks were identified in comparison to 71 peaks identified in this study (Table 9). Despite adoption of the same protocols, AFLP profiles differed in the two studies. Both the overall number of detected peaks

and the number of polymorphic peaks obtained from AFLP analysis varied significantly among primer combinations (Table 9). However, both studies considered all the polymorphic peaks that have major or minor allele frequency (MAF) of > 0.1 to perform the phylogenetic analyses.

1.2 Clustering analysis

All pairwise genetic distances among accessions were computed according to the Jaccard algorithm (Jaccard, 1908), and an un-rooted tree was constructed using the Neighbour-Joining method (Saitou and Nei, 1987) using DARwin v5.0.158 software showing 2 clades (Figure 10). Bootstrap values were calculated to represent how reproducible relationships are within the tree (Felsenstein, 1985). These would reflect how likely inferred relationships are to actually occur in nature (Hall and Salipante, 2007). Low bootstrap values indicate the tree structure is not well supported (Figure 10). In addition, the recovered population structure is atypical (Figure 10), in that it cannot be clearly reconciled with results from Fricano *et al.* (2009) and other classical subdivisions reported in the literature e.g. winter/spring and 6/2-rows groups (Cockram *et al.*, 2008). For example winter 2/6 rows barleys are inter-dispersed among the other cultivar types (spring 2/6 rows, alternative). In summary, AFLP has resolved 2 clades that show no correspondence with growth habit, or spike morphology (2/6 rows). However, nodes are weakly supported from the bootstrapping test (Figure 10). Although we selected our primer combinations based on the previous study carried by Fricano *et al.* (2009), we were not able to achieve similar AFLP profiles: the lower resolution of our phylogenetic analysis may be attributed to the modest number of polymorphic loci.

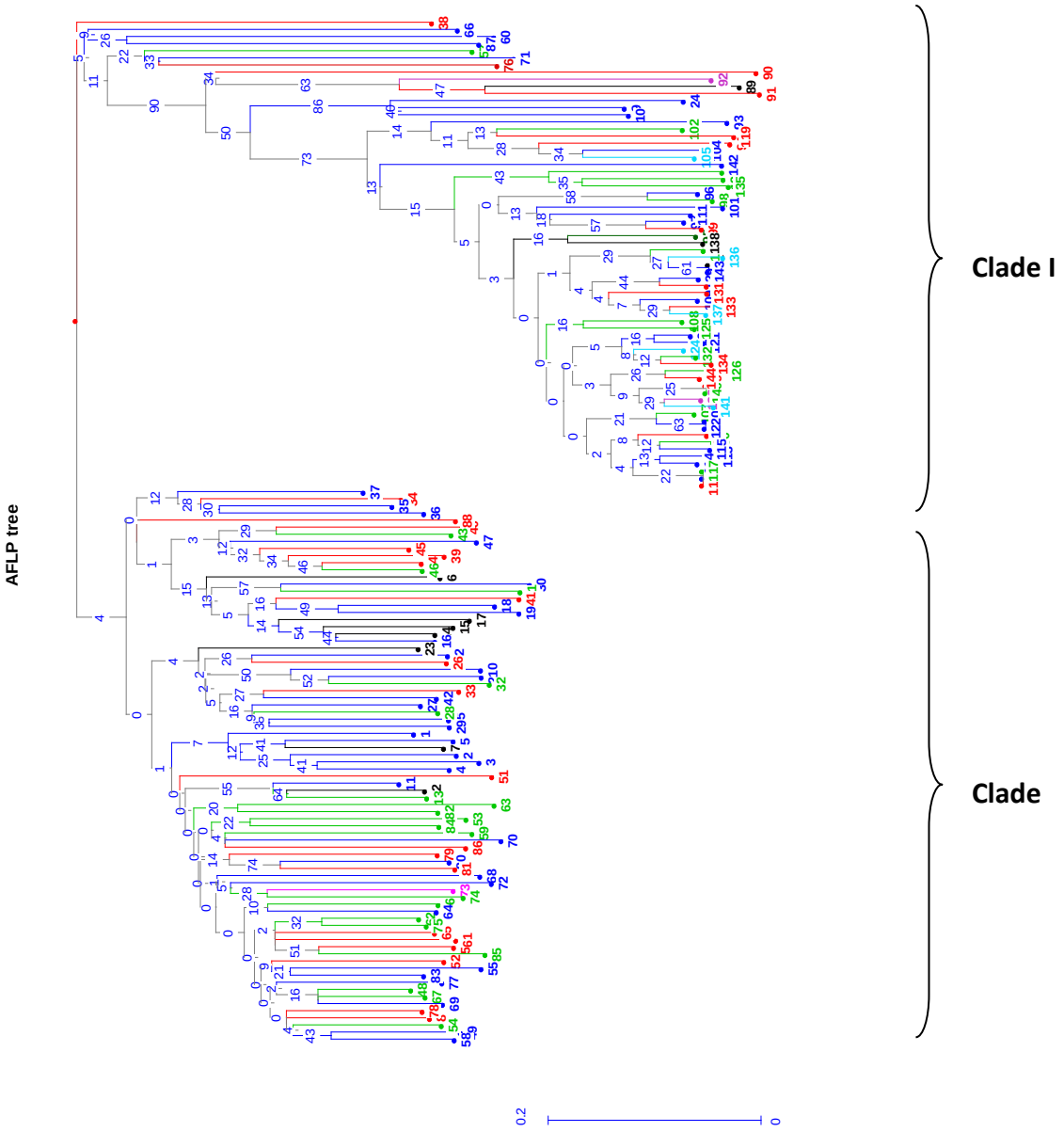


Figure 10: Un-rooted Neighbour joining phylogeny tree of 144 barley accessions, calculated from the analysis of 71 AFLP molecular markers using six primer combinations; the tree was constructed by means of Jaccard's genetic distance with boot strapping 1000 (% indicated in each node). Representing the clustering of the accessions: Blue = winter 6 rows, red = winter 2 rows, green = Spring 2 rows, light blue = Spring 6 rows, Purple = alternative cultivar 2/6 rows, and black = no information available

2 Candidate genes analyses

An initial objective of this project was to evaluate the potential association between allelic variants of candidate genes (CGs) selected for their known roles in drought responses and phenotypic variation for yield under different watering regimes, previously analyzed by Rizza *et al.* (2004). To this end, following an extensive literature search, we decided to focus on a list of 16 candidate genes (Table 4), belonging to the DREB family of transcription factors (Agarwal *et al.*, 2006) or involved in the ABA pathway (Shinozaki and Yamaguchi-Shinozaki, 2007). The reference sequences of CGs from barley were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov) according to the published accession numbers (Table 4). Candidate gene sequences from rice were used in similarity searches (blast-p and megablast) using the IPK barley BLAST server (http://webblast.ipk-gatersleben.de/barley/viroblast.php) to identify highly similar barley sequences [barley_HighConf_genes_MIPS]. Consequently, primers were designed on the basis of the recovered barley sequences, paying attention to avoid highly conserved domains that may lead to primer annealing to paralogous sequences (Table 4).

In order to gain a preliminary overview of nucleotide diversity, analyses were initially focused on resequencing of 3 genes – *HvCBF2*, *HvCBF4A* and *HvCBF7*- in a diverse subset of 40 genotypes from the germplasm panel. To ensure maximum diversity of this core-set, we selected contrasting genotypes according to their growth habit (winter/spring), row type (2/6 rows) and distribution in the AFLP phylogenetic tree produced by Fricano *et al.* (2009).

Sequences from genomic PCR products of the selected candidate genes were compared to identify polymorphisms and haplotypes in the subset of 40 accessions. As an example, a 478 bp high quality sequence was obtained (Figure 13) and four polymorphic loci and 4 haplotypes were identified for *HvCBF4A* (Figure 11 and Figure 13). In contrast, no polymorphisms were recovered for *HvCBF2* and *HvCBF7* (Figure 11). Polymorphisms in *HvCBF4A* are positioned within the translated region resulting in aminoacid substitutions in 2 cases: the first (SNP1) and third (SNP3) polymorphic loci (Figure 13) lead to changes in amino acid from Alanine to Valine, and from Alanine to Serine, respectively. In comparison, silent nucleotide changes leading to no

amino acid substitution were detected in the second (SNP2) and the forth (SNP4) polymorphic loci.

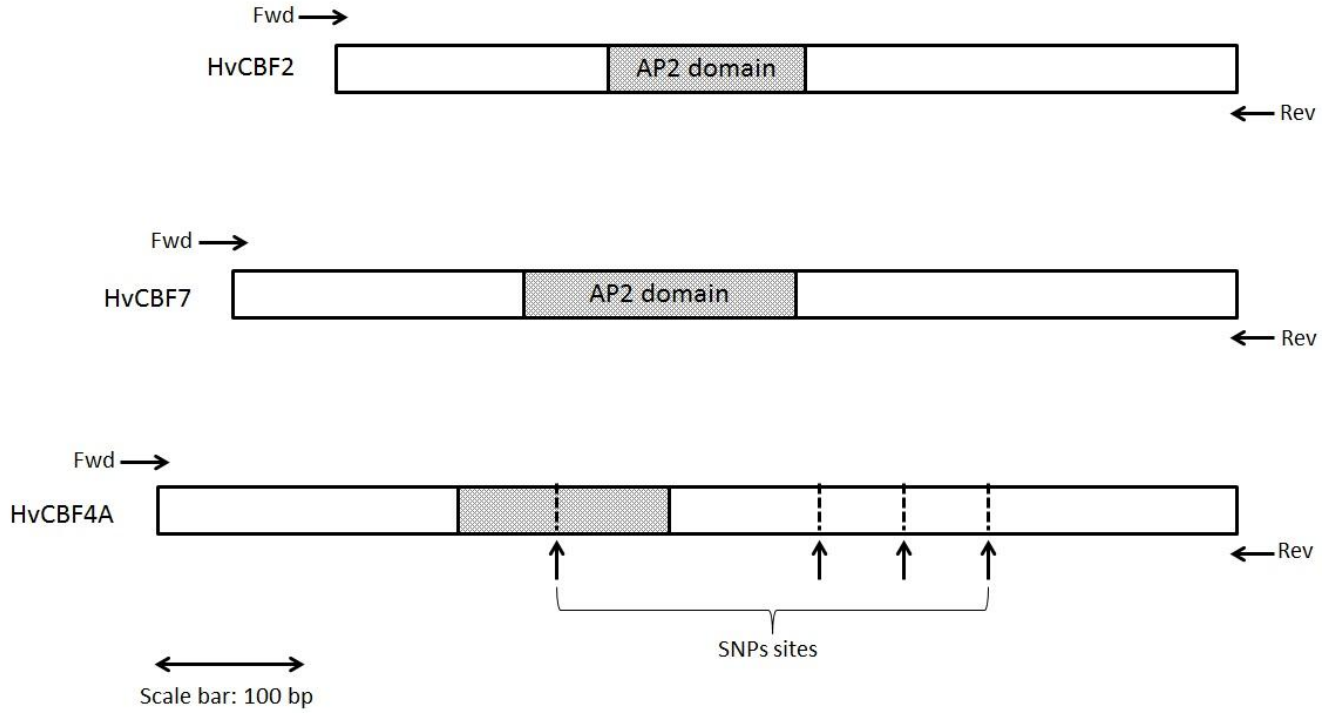


Figure 11: Illustration for the regions amplified in 3 candidate genes and SNPs detected: Three CGs (coding region, illustrated in horizontal bars) amplified against the core collection of the panel (defined for initial diversity analysis). Four SNPs detected (shown in vertical dashed bars) in *HvCBF4A* CG. No SNPs detected in *HvCBF2* nor *HvCBF7*. Arrows are representing primer positions: Fwd, Rev = forward and reverse primers, respectively. Hatched bars representing APETALA2 (AP2) domain.

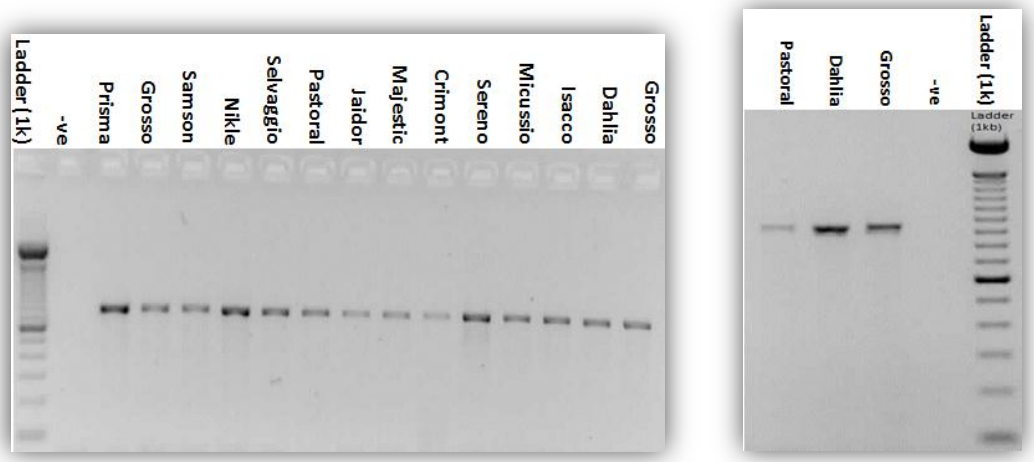


Figure 12: Examples of HvCBF2 amplification (image on left) with (product size of 700bp) and CBF4A amplificatio size of 920bp). This result was on some accessions (Table 1) with negative control showed in blank. Amplifications DNA on 2% agarose gel.
Legends: Invitrogen® ladder (1kb), -ve = negative control (PCR reaction with no DNA template)

2.1 DNA sequencing and alignment

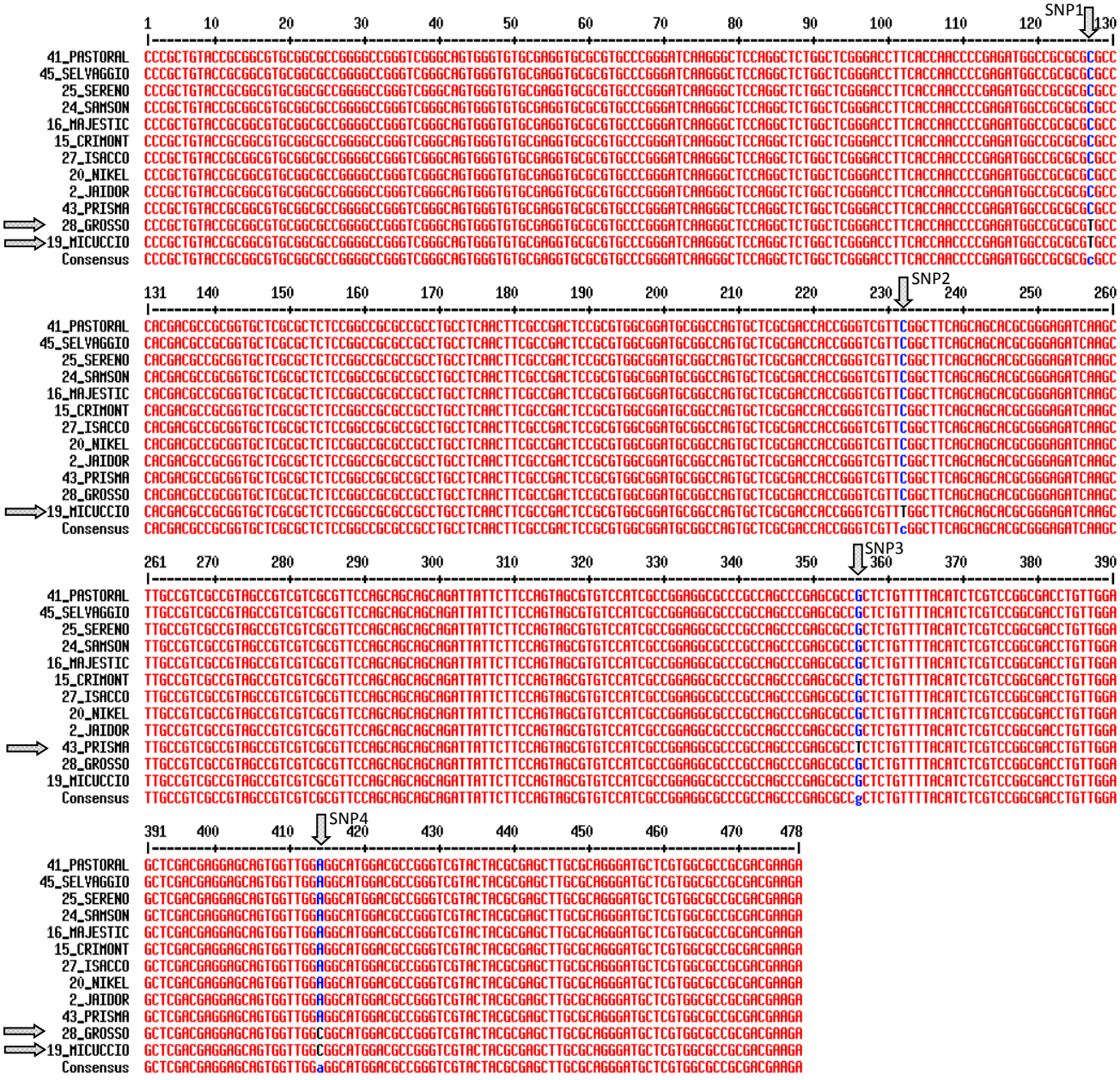


Figure 13: Sequence alignment of the amplified region on *HvCBF4A* for some genotypes, showing the identified polymorphic loci (highlighted in the figure) for genotypes: Grosso, micuccio, and prisma

3 Genome wide association analyses

The rational for expanding to utilize GWAS approach was the necessity to overcome: 1) inconsistency of results obtained from AFLP analysis compared to well-established partitions of cultivated barley germplasm; 2) limitations linked to CG-based association approaches.

3.1 SNP Genotyping

In collaboration with the EXBARDIV consortium, we used the "iSELECT 9K Infinium" SNP Chip that contains a total of 7,864 SNPs to genotype 83 accessions in our panel (Comadran *et al.*, 2012; http://bioinf.hutton.ac.uk/iselect/app/). We excluded markers with missing data ≥ 10% and those with minor allele frequency (MAF) < 10%.

This resulted in a total number of 4,661 markers that were used in analyses of population structure and association between markers and phenotypic data analysis.

The design and mapping information of the markers in the iSELECT 9K Infinium SNP panel have been recently published (Comadran *et al.*, 2012; http://scisoc.confex.com/scisoc/2012am/webprogram/Paper73334.html). Most SNPs were mapped using a barley segregating population from the cross Barke x Morex. Other SNPs were mapped using linkage disequilibrium (LD) with respect to the mapped SNPs (Kilian and Graner 2012). Moreover, some markers were mapped by Close *et al* (2009) and the Genomics-Assisted Analysis and Exploitation of Barley Diversity consortium (EXBARDIV: http://www.erapg.org). About 20% of SNPs are not mapped.

The distribution of the used 4,661 SNPs by chromosome is shown in Table 10.

Table 10: SNPs mean coverage and their distribution across all the 7 chromosomes in our barley accessions.

Chr. no	**No. Markers**	**Dist**(cM)**	**No. Markers by cM**
1H	468	133.14	3.51
2H	632	149.5	4.22
3H	751	155.03	4.84
4H	501	115.23	4.34
5H	927	169.65	5.46
6H	716	126.63	5.65
7H	666	141.36	4.71
Total	**4661**	**990.54**	**4.7**

*Chr.: Chromosome number
**Dist.: maximal distance between two markers in centiMorgans

3.2 Population structure analysis

Population structure considers the presence of subpopulations in the sample in which individuals are more closely related to each other than the average pair of individuals taken at random in the population (Breseghello and Sorrells, 2006).

To determine whether our barley accessions could be grouped into genetic clusters or groups and to infer the number of clusters that best fit the data, we used the K value (Figure 15 and Table 11) as it best describes the population structure based on the criteria of maximizing the logarithmic probability of data [called 'LnP(D)'] (Falush *et al.*, 2003). K is obtained by first computing the log likelihood of the data at each step of the MCMC; in the following step, the average of K values is then computed and half their variance is subtracted from the mean. This gives 'LnP(D)', the model choice criterion to which the program refers as (K).

Structure software was used implementing a Bayesian clustering approach in which it assumes a number of K populations, each of which is characterized by a set of allele frequencies at each locus and attempts to assign individuals to populations on the basis of their genotypes, while instantaneously approximating population allele frequencies (Pritchard *et al.*, 2000).

The model assumes that some fraction (q_k) of each individual's genome originates from each of the K subpopulations. Individuals are probabilistically assigned to the subpopulations in such a way as to achieve Hardy-Weinberg equilibrium for the loci within subpopulations (Smiko and Hu, 2008). Although the individual accessions can belong to multiple subpopulations, the sum of *q* values (Table 12) across all subpopulations for an individual is equal to 1 ($\sum_k q_k = 1$).

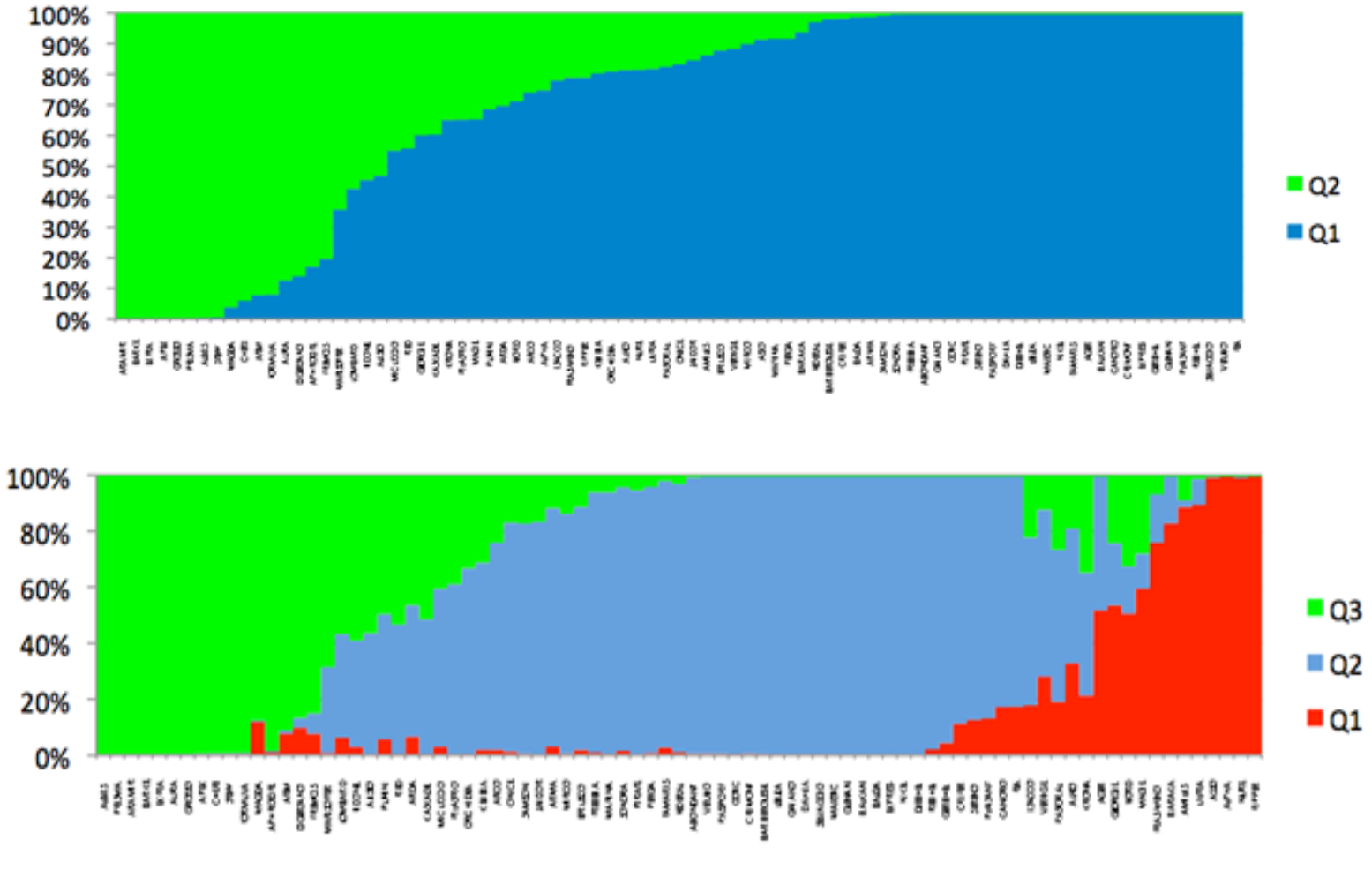


Figure 14: Population structure results for the 83 barley accessions, assessed with a subset of 260 SNP markers selected based on criteria that they are at least 10cM apart and with MAF > 0.3 (Mezaka *et al.*, 2011). Each color represents a population, and the color of individual haplotypes represents their proportional membership in the different populations. A. Population structure with 2 groups (2 populations Q1 in blue mainly represents winter barleys and Q2 in green mainly spring barleys). B. Population structure with 3 groups (3 populations Q1 in red mainly represents winter 2 rows, Q2 in blue mainly winter 6 rows, and Q3 in green mainly spring barleys).

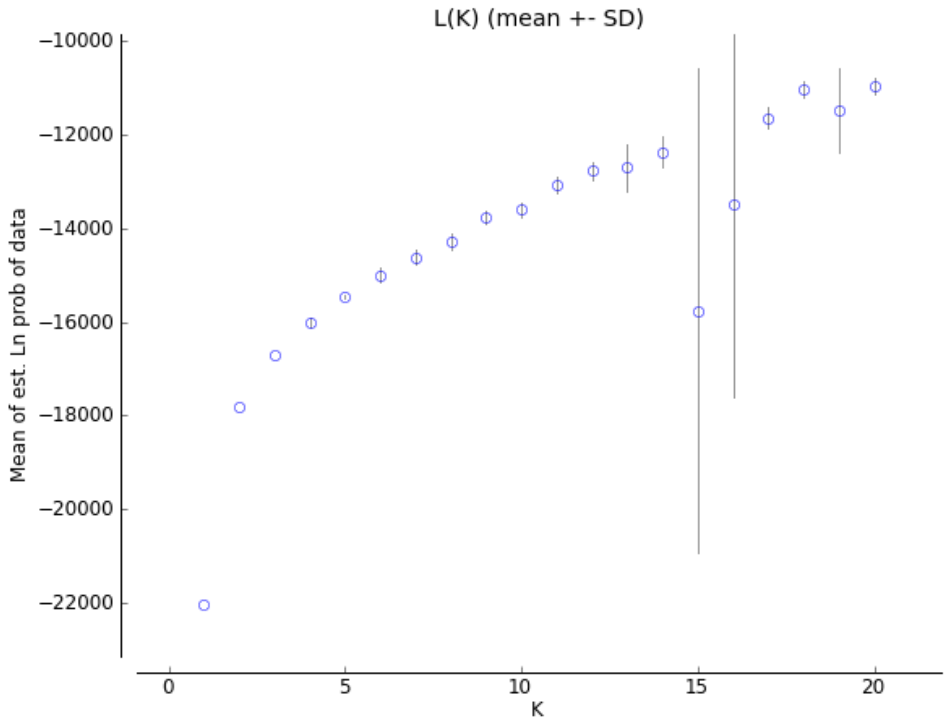


Figure 15: *Structure* results using a subset of 260 SNPs markers. Mean L(K) (±SD) over 5 runs for each assumed K value from 1 : 20. Log probability data (LnP(D)) as function of K (number of clusters) from the *Structure* run. The estimation of the true value of K continues to increase slightly when true K has been reached (Butts *et al.*, 2008). K = 2 or 3 indicates the minimum number of groups possible in the panel.

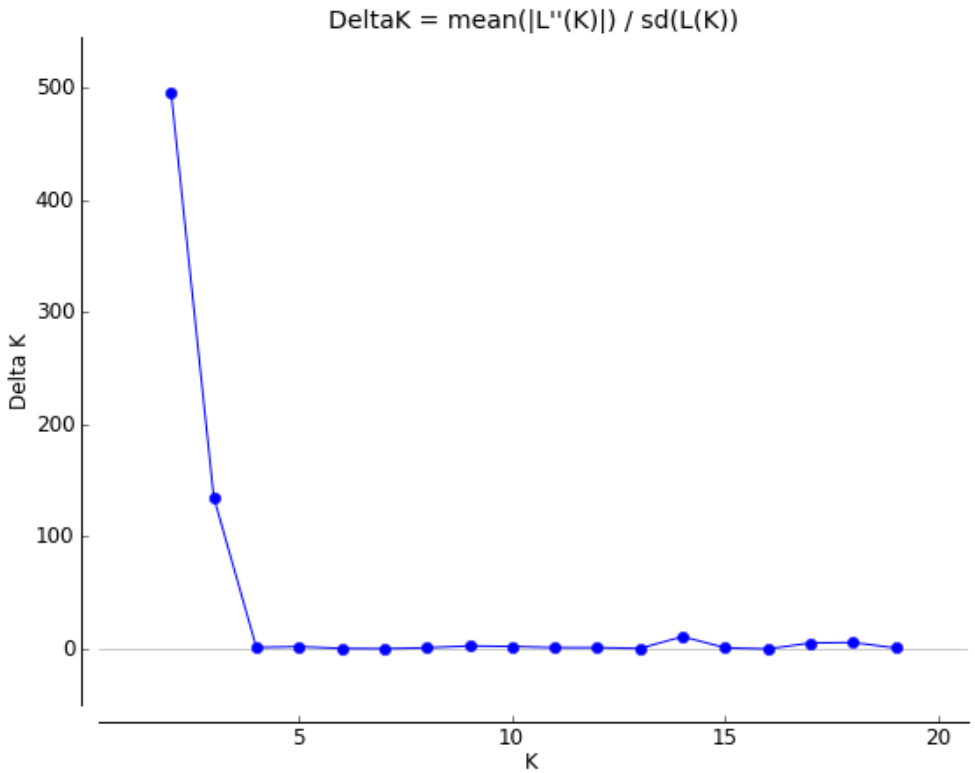


Figure 16: Detection of clusters number (K) by estimation of Delta K over 5 runs of each K value with software STRUCTURE. Identifying the appropriate sub-population number (K) obtained by the method of Evanno *et al.*, (2005). Sub-population number (K) against Ln(P|D) ± 1 sd (250,000 burn-in and 100,000 MCMC iterations).

*Delta K is expressed as a mean of the absolute values of ratio of the change in the likelihood function with respect to K (Mezaka *et al.*, 2011).
*Evanno *et al.*, 2005 provides a correct estimation of the number of clusters using an ad hoc statistic DeltaK based on the rate of change in the log probability of data between successive K values.

The number of subgroups (K) in our 83 accessions was detected by the estimation of ΔK over 5 replications of runs for each K value using the software *Structure* and the method by Evanno *et al*. (2005). The rationale for this ΔK is to make prominent the break in slope of the distribution of L(K) at the true K. It is calculated as ΔK = m(|L(K + 1)-2L(K)+L(K-1)|)/s[L(K)], where K symbolizes the assumed number of subgroups, L denotes the average of LnP(D) for the 5 iterations of the K, and s refers to the average standard deviation of the 5 replications of the K (Haseneyer *et al.*, 2010). The height of ΔK is the indicator of the strength of the signal detected by structure (Mezaka *et al.*, 2011). Therefore, K = 2 or 3 indicates the minimum number of subgroups possible in this study's barley panel.

In previous studies, the major phenotypic divisions in cultivated barley suggested K = 4; corresponding to the four possible growth habit and row-number groupings with their combinations (Cockram *et al.*, 2008). In this study, the best K values obtained according to the Evanno *et al.* (2005) method (Table 11) were K = 2 (2 groups based on growth habit winter/spring) and K = 3 (3 groups corresponding respectively to winter 2-rows, winter 6-rows and spring barleys) (Figure 14).

Table 11: Values obtained from structure results using methodology of Evanno *et al.*, (2005).

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	5	-22034.660000	0.709225	—	—	—
2	5	-17816.300000	6.279729	4218.360000	3109.780000	495.209245
3	5	-16707.720000	3.100323	1108.580000	418.860000	135.102071
4	5	-16018.000000	95.050302	689.720000	128.740000	1.354441
5	5	-15457.020000	43.540579	560.980000	93.160000	2.139613
6	5	-14989.200000	158.699370	467.820000	89.900000	0.566480
7	5	-14611.280000	143.417039	377.920000	43.340000	0.302196
8	5	-14276.700000	166.635425	334.580000	191.000000	1.146215
9	5	-13751.120000	132.368697	525.580000	368.080000	2.780718
10	5	-13593.620000	159.218457	157.500000	365.780000	2.297347
11	5	-13070.340000	167.372841	523.280000	221.160000	1.321361
12	5	-12768.220000	181.375665	302.120000	235.920000	1.300726
13	5	-12702.020000	497.787020	66.200000	277.760000	0.557990
14	5	-12358.060000	338.625823	343.960000	3747.220000	11.065961
15	5	-15761.320000	5174.639363	-3403.260000	5670.320000	1.095790
16	5	-13494.260000	4102.268287	2267.060000	407.960000	0.099447
17	5	-11635.160000	234.598493	1859.100000	1255.780000	5.352890
18	5	-11031.840000	180.065774	603.320000	1045.660000	5.807100
19	5	-11474.180000	906.718135	-442.340000	962.886667	1.061947
20	5	-10953.633333	182.234364	520.546667	—	—

**Stdev = standard deviation values*

Table 12: List of the barley accessions used in this study and their classification based on growth habit (winter or spring) and row types (2 or 6 rows) and their estimated fractions (q_2, q_2, q_3) of the accession's genome that originates from two or three inferred sub-populations (subpopulations 1, 2, and 3).

Accession	Subgroup (q_1 value)	Subgroup (q_2 value)	Accessions	Subgroup (q_1 value)	Subgroup (q_2 value)	Subgroup (q_3 value)
ARAMIR	0.001	0.999	DJEBHEL	0.001	0.998	0.001
BARKE	0.001	0.999	NIKEL	0.0016	0.9974	0.001
EXTRA	0.001	0.999	EXPRESS	0.0018	0.9972	0.001
APEX	0.002	0.998	GLENAN	0.002	0.997	0.001
GROSSO	0.002	0.998	BALKAN	0.002	0.997	0.001
PRISMA	0.002	0.998	BALDA	0.002	0.997	0.001
ALEXIS	0.0038	0.9962	MAJESTIC	0.0026	0.9964	0.001
STEFFY	0.0056	0.9944	DAHLIA	0.0028	0.9962	0.001
MAGDA	0.0382	0.9618	SELVAGGIO	0.0028	0.9962	0.001
CHERI	0.0602	0.9398	LETIZIA	0.0038	0.9952	0.001
ATEM	0.0776	0.9224	BARBEROUSSE	0.0054	0.9936	0.001
FORMULA	0.0796	0.9204	CRIMONT	0.0066	0.9924	0.001
AURA	0.1256	0.8744	GERBHEL	0.0436	0.9554	0.001
DIGERSANO	0.1396	0.8604	PLAISANT	0.1308	0.8682	0.001
APHRODITE	0.1708	0.8292	TEA	0.1728	0.8262	0.001
TREMOIS	0.1966	0.8034	AGER	0.5172	0.4818	0.001
MARISOTTER	0.3576	0.6424	BARAKA	0.828	0.171	0.001
LOMBARD	0.4256	0.5744	ASSO	0.9916	0.0074	0.001
TIDONE	0.4546	0.5454	ALPHA	0.997	0.002	0.001
ALISEO	0.4682	0.5318	TIPPER	0.9974	0.0016	0.001
MICUCCIO	0.5506	0.4494	CANORO	0.1721	0.8267	0.0012
RED	0.5572	0.4428	GAIANO	0.002	0.9962	0.0018
GEORGIE	0.6006	0.3994	REBHEL	0.0206	0.9776	0.0018
KASKADE	0.6026	0.3974	PASSPORT	0.0062	0.9918	0.002
KRONA	0.6506	0.3494	VETULIO	0.0068	0.9912	0.002
PILASTRO	0.6518	0.3482	CRITER	0.1132	0.8848	0.002
MAGIE	0.6526	0.3474	GOTIC	0.0048	0.9922	0.003

Table 12: continued

Accession	Subgroup (q_1 value)	Subgroup (q_2 value)	Accessions	Subgroup (q_1 value)	Subgroup (q_2 value)	Subgroup (q_3 value)
PUFFIN	0.686	0.314	SERENO	0.1262	0.8708	0.003
ARDA	0.6966	0.3034	NURE	0.993	0.002	0.005
FJORD	0.7116	0.2884	ABONDANT	0.0066	0.986	0.0074
ARCO	0.7406	0.2594	ULTRA	0.8946	0.0928	0.0126
ALPHA	0.7476	0.2524	TAMARIS	0.0272	0.9528	0.02
ISACCO	0.7796	0.2204	FEDERAL	0.0108	0.9584	0.0308
TRASIMENO	0.7886	0.2114	PERGA	0.0084	0.9498	0.0418
TIPPER	0.7892	0.2108	SONORA	0.0152	0.9418	0.043
KELIBIA	0.8026	0.1974	PIRATE	0.0024	0.9436	0.054
ORCHIDEA	0.8086	0.1914	MATTINA	0.0052	0.9336	0.0612
ALFEO	0.8136	0.1864	TREBBIA	0.0098	0.929	0.0612
NURE	0.8146	0.1854	TRASIMENO	0.7592	0.173	0.0678
ULTRA	0.8176	0.1824	AMILLIS	0.8862	0.024	0.0898
PASTORAL	0.824	0.176	ETRUSCO	0.0186	0.868	0.1134
ONICE	0.8328	0.1672	ARMA	0.0308	0.851	0.1182
JAIDOR	0.8456	0.1544	VERTIGE	0.2814	0.5956	0.123
AMILLIS	0.8616	0.1384	MIRCO	0.0064	0.8564	0.1372
ETRUSCO	0.8776	0.1224	JAIDOR	0.0026	0.8306	0.1668
VERTIGE	0.883	0.117	ONICE	0.013	0.8178	0.1692
MIRCO	0.8986	0.1014	SAMSON	0.0062	0.8214	0.1724
ASSO	0.9132	0.0868	ALFEO	0.3286	0.4818	0.1896
MATTINA	0.916	0.084	ISACCO	0.1779	0.5998	0.2223
PERGA	0.916	0.084	ARCO	0.019	0.7405	0.2404
BARAKA	0.9378	0.0622	GEORGIE	0.5351	0.2226	0.2424
FEDERAL	0.9706	0.0294	PASTORAL	0.1898	0.5441	0.2661
BARBEROUSSE	0.9796	0.0204	MAGIE	0.5954	0.1244	0.2802
CRITER	0.9816	0.0184	KELIBIA	0.0192	0.6687	0.3121
BALDA	0.9866	0.0134	FJORD	0.5058	0.1683	0.3259
ARMA	0.9874	0.0126	ORCHIDEA	0.0044	0.6631	0.3325
SAMSON	0.9918	0.0082	KRONA	0.2126	0.4393	0.3481

Table 12: continued

Accession	Subgroup (q_1 value)	Subgroup (q_2 value)	Accessions	Subgroup (q_1 value)	Subgroup (q_2 value)	Subgroup (q_3 value)
SONORA	0.9952	0.0048	PILASTRO	0.003	0.6077	0.3893
TREBBIA	0.996	0.004	MICUCCIO	0.0304	0.5642	0.4054
ABONDANT	0.997	0.003	ARDA	0.0662	0.47	0.4638
GAIANO	0.997	0.003	PUFFIN	0.0568	0.4471	0.4961
GOTIC	0.997	0.003	KASKADE	0.0046	0.4807	0.5147
PIRATE	0.998	0.002	RED	0.005	0.4619	0.5331
SERENO	0.998	0.002	ALISEO	0.003	0.4332	0.5638
PASSPORT	0.998	0.002	LOMBARD	0.063	0.3705	0.5665
DAHLIA	0.9986	0.0014	TIDONE	0.0298	0.3795	0.5907
DJEBHEL	0.9986	0.0014	MARISOTTER	0.0086	0.3062	0.6852
LETIZIA	0.9986	0.0014	TREMOIS	0.0764	0.0732	0.8503
MAJESTIC	0.9986	0.0014	DIGERSANO	0.0981	0.0352	0.8667
NIKEL	0.9986	0.0014	MAGDA	0.1202	0.0038	0.876
TAMARIS	0.9986	0.0014	ATEM	0.077	0.0124	0.9106
AGER	0.999	0.001	APHRODITE	0.0114	0.004	0.9846
BALKAN	0.999	0.001	FORMULA	0.0056	0.003	0.9914
CANORO	0.999	0.001	STEFFY	0.0048	0.003	0.9922
CRIMONT	0.999	0.001	CHERI	0.0046	0.003	0.9924
EXPRESS	0.999	0.001	APEX	0.0046	0.002	0.9934
GERBHEL	0.999	0.001	GROSSO	0.0034	0.002	0.9946
GLENAN	0.999	0.001	AURA	0.0024	0.002	0.9956
PLAISANT	0.999	0.001	BARKE	0.003	0.001	0.996
REBHEL	0.999	0.001	ARAMIR	0.0026	0.001	0.9964
SELVAGGIO	0.999	0.001	EXTRA	0.002	0.0012	0.9968
VETULIO	0.999	0.001	PRISMA	0.0016	0.001	0.9974
TEA	0.999	0.001	ALEXIS	0.0012	0.001	0.9978

When an accession has $q > 0.3$ for two subpopulations, the accession is assigned into mixed sub-population (q_1/q_2, q_1/q_3, q_2/q_3) (Simko and Hu, 2008).

*Colors legends: Blue = winter 6 rows, red = winter 2 rows, and green = spring 2 rows barleys as derived from information in the European barley database (http://barley.ipk-gatersleben.de/ebdb.php3).

3.3 Principle coordinate analysis (PCoA)

To further understand the distribution of accessions in our panel, we performed a principal coordinates analysis (Figure 17): the first (Coordinate 1) and second (Coordinate 2) principal components accounted for 24.7 % and 6.3 % respectively of the observed genetic variation (Table 13). Superimposing row type and growth habit information (see color codes, Figure 17 and table 13), three groups can be distinguished in the graph, corresponding to winter 6-rows, winter 2-rows and spring 2-rows genotypes, respectively. This result is consistent with the Bayesian model-based clustering presented above.

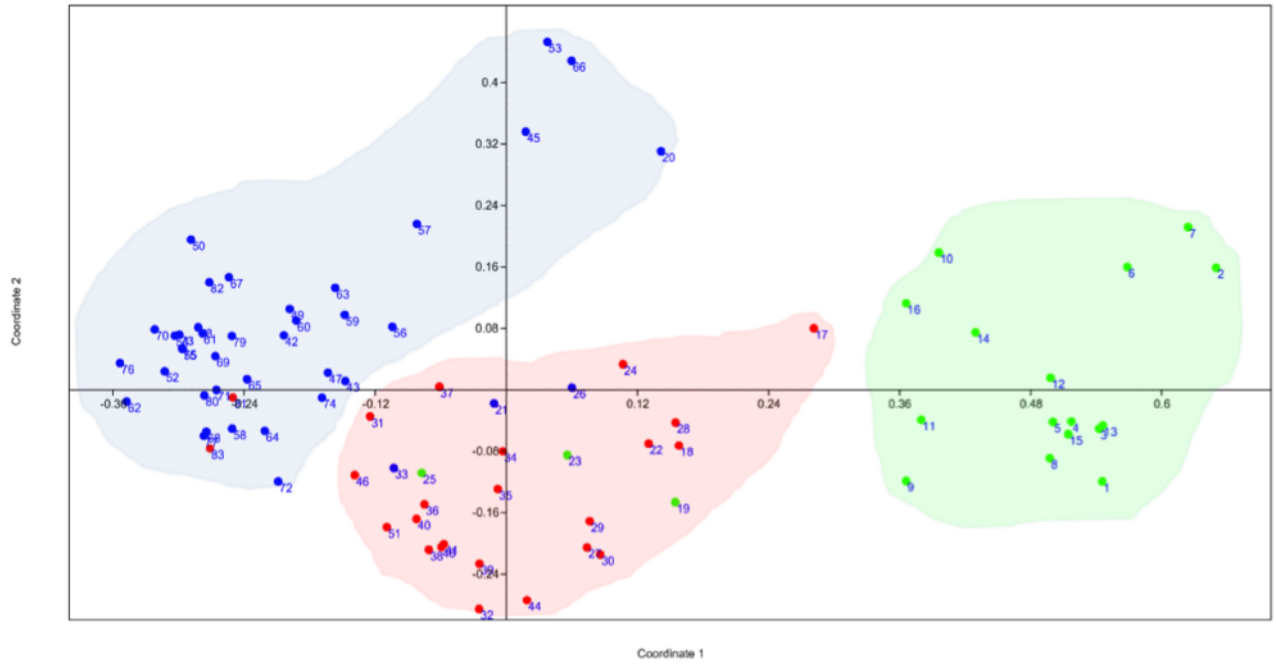


Figure 17: Principle coordinate analysis (2 coordinates) for 83 barley cultivars shows 3 clusters separating different barley accessions based on their growth habit and row type.

* *Past* software (Hammer *et. al.*, 2001).

Legend

winter 6 rows
winter 2 rows
spring 2 rows

Few genotypes falling outside the expected clusters can be explained due to particularities in their pedigrees or alternative growth habit (European barley database; http://barley.ipk-

gatersleben.de/ebdb.php3). For example, the position of the spring 2-rowed cultivar "Tidone" (n. 19 in Figure 17) amongst winter 2-rows may be due to the presence of winter type "Igri" in its pedigree. In the case of "Georgie" (n. 23), the presence of some heterozygosity in genotyping data may be responsible for its atypical position. Multiple crossbreeds characterize the pedigree of "Krona" (n. 25). These results are in contradiction with *Structure* findings since "Tidone" was clustered with q_3 group when K = 3 which corresponds to spring barley. In conclusion, *Structure* did not place "Tidone" in winter 2-rows as PCoA. Thus, in the case of this genotype, *Structure* results seem more reliable. In the case of "Krona" and "Georgie", *Structure* indicates a tendency to cluster with group q_2 (winter 6-rows) and q_1 (winter 2-rows), respectively.

Two winter 2-rows cultivars: "Selvaggio and Tea" (n. 81, and 83 in Figure 17), clustered within winter 6-rows in PCoA. This position may reflect the presence of winter 6 rows accessions in the pedigree of "Selvaggio", while "Tea" is considered an alternative cultivar. In comparison, *Structure* assigned both these cultivars to q_2 when K = 3 (Table 12), indicating a tendency to be grouped within winter 2 rows.

In conclusion, PCoA separates the accessions along the primary axis according to growth habit and further grouping is related to the row types.

Table 13: PCoA results - 3 components shown: The first 2 PCs represent 24.7 % and 6.3 % respectively of the observed genetic variation.

Axis	Eigenvalue	Percent %
*PC1	6.6002	24.759
PC2	1.6947	6.357
PC3	1.3203	4.9528

*PC: Principle component. This result was produced using *Past* software (Hammer *et. al.*, 2001).

4 Genome wide association study of barley flowering date, plant height, and grain yield traits.

We used the barley phenotypic data published by Rizza *et al* (2004, supplementary material Table S1).

4.1 Genome wide association scans

One of the principal goals of our study was to test for possible association analysis between grain yield and the genome-wide SNPs represented in the 9k iSELECT panel. Flowering date and plant height were also included in our study as benchmarks to check the ability to detect known loci as these two traits were studied in many previous works (Haseneyer *et al.*, 2010; Pasam *et al.*, 2012, Zakhrabekova *et al.*, 2012).

Agronomic traits such as flowering time and yield are influenced by QTLs, environmental effects, as well as the interactions between them (Zhu *et al.*, 2008). Accordingly, in this work, we considered both the genetic and environmental effects (Salvi and Tuberosa 2005). Furthermore, we dissected the environmental effect into separate years (1999, 2000, and 2001) and different irrigation systems (rainfed and controlled irrigation). This strategy might be useful to the understanding of the genetic/environmental factors underlying the considered traits.

GLM was used to scrutinize yield, plant height and flowering date traits; the latter trait was analyzed also with MLM model for a comparison between the two models.

For all the analyses, we looked at a confidence interval as it was considered in previous works (Pasam *et al.*, 2012). Therefore, we considered the potential loci falling in an interval ± 5-10 cM harboring the significant SNPs coming out from the analysis with a threshold of $-\log_{10}$ (*P*-value) > 4. When many SNPs fell in the same interval of ± 5-10 cM, we considered this genomic region or SNP cluster as a potential QTL and we referred to it with the most significant SNP.

Comparison of our results with previously identified QTLs/genes (Comadran *et al.*, 2011, Pasam *et al.*, 2012) was based on the map developed by Close *et al.* (2009), using as anchor markers a subset of BOPA1 and BOPA2 SNPs that are shared with our iSELECT map (Comadran *et al.*, 2012). Anchor markers (i.e. markers shared between maps) were used to align different maps and compare positions of QTLs from different studies.

4.1.1 Associations between SNPs and flowering date (FD) trait in barley

4.1.1.1 Analysis with the general linear model (GLM)

GLM is among the most used statistical approaches to deal with associations of a high number of SNPs with agronomic traits (Pasam *et al.*, 2012). We used GLM to perform the genome wide association analysis of FD with 4,661 SNPs. We corrected for population structure by including the kinship matrix into the model. K3 parameter for the kinship matrix was sufficient to cluster the different barley populations. In addition, we considered a Bonferroni threshold $-\log_{10}$ value at (*P*-value <= 0.0001) (Mezaka *et al.*, 2011). The genome wide association scan for FD revealed a total number of 49 and 393 statistically significant SNPs, excluding the shared SNPs among years, when we considered controlled irrigation and rainfed, respectively.

For the controlled irrigation, 35, 12 and 10 SNPs were statistically significant for years 1999, 2000 and 2001, respectively (Figure 18). For rainfed condition, 385, 31, and 131 SNPs were statistically significant for years 1999, 2000 and 2001, respectively (Figure 18). The intersections of the significant markers among the years and the irrigation systems are shown in Figure 18.

Loci involved in agronomic traits and mapped in the intervals harboring our significant SNPs were considered for a comparative approach as detailed below.

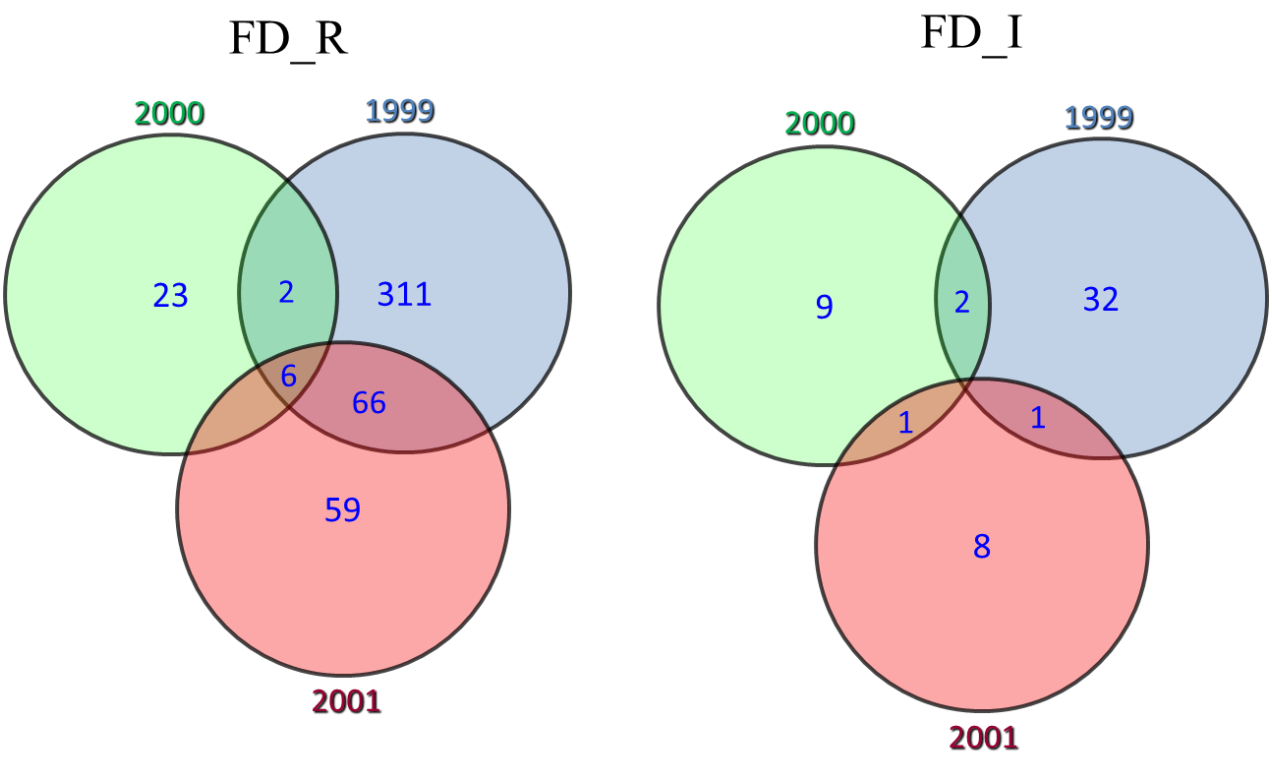


Figure 18: Venn diagram showing the most significant markers and their overlaps among years. All these markers are exceeding the Bonferroni threshold line (considered significant) and associated with flowering date (FD) trait in each year and each irrigation condition with their intersections among years (FD_R = flowering date in rainfed condition, FD_I = flowering date in controlled irrigation condition). The analysis was run by *Tassel* program following GLM.

4.1.1.1.1 SNP markers linked to FD under controlled irrigation condition and significant across years

A total number of 4 SNP markers exceeding the Bonferroni threshold were common between two years (1999, 2000, and 2001) as shown in Table 14.

Table 14: Markers intersected between years and associated with FD trait in controlled irrigation conditions.

Marker	Chr. no.	Locus position in iSELECT map (cM)	*P*-value	R^2	Years
BOPA1_3263-2865	1H	133	3,3535E-4	0,14936	1999, 2001
SCRI_RS_120529	2H	23.16	2,7689E-4	0,15319	1999, 2000
SCRI_RS_132388	3H	7.01	8.26E-04	0,1312	2000, 2001
BOPA1_10386-329	5H	162.5	3,7255E-4	0,14556	1999, 2000

R^2 = square of the correlation coefficient between two loci (variance explained by marker).
P-values = the probability of having a false association, as –log10 (P)

4.1.1.1.2 Most significant markers associated with flowering date trait in the controlled irrigation condition: comparative analyses with previously mapped genes and QTLs

In this analysis we summarized and grouped clusters harboring the most significant SNPs (exceeding the Bonferroni threshold) into QTLs. These QTLs were indicated according to the most significant SNP marker, with the lowest *P*-value in the cluster as presented in (Table 15). This analysis resulted into 13 potential QTLs (Table 15).

Table 15: The most significant SNP markers associated with FD trait amongst all years in controlled irrigation conditions. Indications of previously mapped genes and nearby markers potentially linked with the different genes or QTLs are shown in the table. Markers indicated herein showed the lowest *P*-values in the different years for each chromosome.

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
*FD_I_99	SCRI_RS_199945	1H	132.86	4.70E-04	0.174344	*Mat-a*, QTL1_HD	Zakhrabekova *et al.*, 2012; Pasam *et al.*, 2012	QTL.FD.2
FD_I_99	SCRI_RS_110647	2H	29.39	7.75E-05	0.176338	*Ppd-H1*, QTL2_HD	Turner *et al.*, 2005; Wang *et al.*, 2010; Pasam *et al.*, 2012	QTL.FD.4
FD_I_99	SCRI_RS_146429	3H	83.07	7.02E-04	0.132947			QTL.FD.7
FD_I_99	SCRI_RS_205975	4H	97.31	8.92E-04	0.161005			QTL.FD.10
FD_I_99	SCRI_RS_166296	5H	65.97	9.15E-05	0.173116			QTL.FD.12
FD_I_99	BOPA2_12_30329	7H	23.02	6.81E-05	0.180874			QTL.FD.18
FD_I_00	SCRI_RS_12516	2H	23.8	1.67E-04	0.165081	*Ppd-H1*, QTL2_HD	Turner *et al.*, 2005; Wang *et al.*, 2010; Pasam *et al.*, 2012	QTL.FD.4
FD_I_00	SCRI_RS_132388	3H	7.01	8.26E-04	0.131199	QTL7_HD	Pasam *et al.*, 2012	QTL.FD.5
FD_I_00	SCRI_RS_122057	4H	97.31	1.14E-04	0.168776			QTL.FD.10
FD_I_00	SCRI_RS_202800	5H	162.5	7.46E-04	0.133252			QTL.FD.15
FD_I_00	BOPA1_ABC14397-1-2-208	7H	85.98	6.12E-04	0.135682	*HvCO1*, QTL18_HD	Griffiths *et al.*, 2003; Wang *et al.*, 2010; Pasam *et al.*, 2012	QTL.FD.20
FD_I_01	BOPA2_12_30934	1H	133.14	3.35E-04	0.149361	*Mat-a*	Zakhrabekova *et al.*, 2012	QTL.FD.2
FD_I_01	SCRI_RS_192440	2H	5.38	6.14E-05	0.182893			QTL.FD.3
FD_I_01	SCRI_RS_132388	3H	7.01	3.01E-04	0.151503	QTL7_HD	Pasam *et al.*, 2012	QTL.FD.5
FD_I_01	SCRI_RS_145228	4H	1.06	8.91E-04	0.128148			QTL.FD.9
FD_I_01	SCRI_RS_143367	6H	67.92	8.91E-04	0.128156	QTL14_HD	Pasam *et al.*, 2012	QTL.FD.16
FD_I_01	SCRI_RS_154193	7H	66.36	2.49E-05	0.198084	QTL17_HD		QTL.FD.19

Significant markers associated with flowering date with GLM model corresponding, *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 genes or QTLs from literature are shown.

*FD_I_99 = Flowering date under controlled irrigation condition in the year 1999 and 00, 01 for the years 2000, and 2001.

R^2 = square of the correlation coefficient between two loci (variance explained by marker).

P-values = the probability of having a false association, as –log10 (P).

QTL.FD = Quantitative trait loci for most significant markers associated with FD trait were grouped according to this study.

Our analyses recovered markers congruent with previously mapped QTLs and known genes controlling the trait (Table 15), as well as QTLs that can be considered potentially novel. For instance, one QTL was identified at chromosome 1H and named QTL.FD.2 (represented by the SNP marker SCRI_RS_199945 at 132.86 cM, Table 15). This QTL region encompasses 3 significantly associated SNPs (Supplementary table S2) and corresponds to the recently reported mapping of the "Bowman" introgression line (BW289), carrying the *Mat-a* allele *eam8.k* (Zakhrabekova *et al.*, 2012). Significantly, *Mat-a* is known to play a role in early flowering and it has been used in breeding programs for conferring photoperiod insensitivity (Zakhrabekova *et al.*, 2012).

We identified another important QTL in chromosome 2H, which we named QTL.FD.4. This QTL harbors 13 statistically significant SNPs in our study (Table 15) and is congruent with QTL2_HD identified in chromosome 2H by Pasam *et al.*, 2012 in the region 27.29-33.73 cM, which corresponds with our marker BOPA1_7144-973 anchored within our nominated interval region at 28.68 cM. This region harbours the *Ppd-H1* gene known to be a major determinant of long day response in barley (Turner *et al.*, 2005). In a recent study, *Ppd-H1* was associated with a QTL affecting heading, height, lodging at harvest, and thousand-grain weight (Wang *et al.*, 2010).

Another statistically significant SNP marker (BOPA1_2895-1064) on chromosome 2H (57.86 cM) is in concordance with previously mapped SNP 11_20438 which is associated with *eam6* gene, known to confer earlier heading (Comadran *et al*., 2011).

A potentially novel SNP that was significant in two years (2000 and 2001) was identified on chromosome 3H (SCRI_RS_132388, QTL.FD.5, Table 15). A locus on chromosome 7H was identified at 66.36 cM, and grouped with QTL.FD.19. Such locus may correspond to the previously identified QTL17_HD as shown by Pasam *et al* (2012).

4.1.1.1.3 SNP markers linked to FD under rainfed conditions and significant across years

A total number of 74 SNP markers exceeding the Bonferroni threshold were shared at least by two years (1999, 2000, and 2001), 6 of which were common to all three years as shown in Table 16.

Table 16: Markers intersected among years and associated with FD trait in rainfed conditions.

Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Years
BOPA1_8613-278	1H	47.8	5.65E-05	0.182443	1999, 2001
SCRI_RS_160466	1H	48.51	5.65E-05	0.182443	1999, 2001
SCRI_RS_229932	1H	48.51	5.65E-05	0.182443	1999, 2001
SCRI_RS_125407	1H	48.58	7.69E-04	0.131101	1999, 2001
BOPA1_4716-1205	1H	48.94	7.69E-04	0.131101	1999, 2001
BOPA2_12_30110	1H	48.94	7.69E-04	0.131101	1999, 2001
BOPA2_12_30694	1H	48.94	7.69E-04	0.131101	1999, 2001
BOPA1_3689-1101	1H	49.08	7.69E-04	0.131101	1999, 2001
BOPA2_12_30406	1H	49.08	7.69E-04	0.131101	1999, 2001
SCRI_RS_17256	1H	49.08	7.69E-04	0.131101	1999, 2001
SCRI_RS_109060	1H	49.43	3.22E-04	0.150156	1999, 2001
BOPA1_ABC13652-1-2-156	1H	49.58	5.53E-04	0.139325	1999, 2001
BOPA1_7284-710	1H	49.58	3.22E-04	0.150156	1999, 2001
SCRI_RS_14834	1H	50.57	9.74E-04	0.159165	1999, 2001
BOPA1_1865-396	2H	19.05	1.47E-05	0.207994	1999, 2000
BOPA1_2029-1143	2H	24.5	2.81E-05	0.19579	1999, 2001
BOPA1_7144-973	2H	28.68	1.53E-08	0.331403	1999, 2000, 2001
SCRI_RS_110647	2H	29.39	1.77E-09	0.362219	1999, 2000, 2001
BOPA1_4037-916	2H	66.29	4.15E-06	0.233997	1999, 2001
BOPA1_4659-1261	2H	57.01	5.12E-07	0.268906	1999, 2001
SCRI_RS_127347	2H	57.01	5.12E-07	0.268906	1999, 2001
SCRI_RS_177375	2H	57.01	5.12E-07	0.268906	1999, 2001
SCRI_RS_4969	2H	57.01	5.12E-07	0.268906	1999, 2001
BOPA1_6510-1430	2H	60.84	5.32E-06	0.2269	1999, 2001
SCRI_RS_4802	2H	60.84	5.30E-07	0.274082	1999, 2001
BOPA1_6911-866	2H	60.69	4.69E-07	0.270444	1999, 2001
SCRI_RS_1502	2H	57.15	5.12E-07	0.268906	1999, 2001
SCRI_RS_222769	2H	57.15	5.12E-07	0.268906	1999, 2001
SCRI_RS_231725	2H	57.72	2.99E-08	0.317135	1999, 2001
BOPA1_9191-263	2H	57.72	2.99E-08	0.317135	1999, 2001

Table 16 continued

Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Years
BOPA2_12_30297	3H	1.91	5.58E-05	0.18474	1999, 2001
BOPA2_12_30622	3H	51.2	3.77E-04	0.147009	1999, 2001
BOPA2_12_30130	3H	51.2	3.20E-04	0.14859	1999, 2001
BOPA1_10248-954	3H	51.35	8.70E-05	0.176072	1999, 2001
BOPA1_2861-1941	3H	51.35	8.70E-05	0.176072	1999, 2001
BOPA1_2897-208	3H	51.35	2.12E-04	0.16032	1999, 2001
BOPA1_4707-421	3H	51.35	8.70E-05	0.176072	1999, 2001
BOPA1_6171-956	3H	51.35	8.70E-05	0.176072	1999, 2001
BOPA2_12_30039	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_111312	3H	51.35	7.98E-05	0.193149	1999, 2001
SCRI_RS_120315	3H	51.35	7.86E-05	0.176062	1999, 2001
SCRI_RS_137787	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_150370	3H	51.35	7.86E-05	0.176062	1999, 2001
SCRI_RS_150800	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_151545	3H	51.35	6.43E-04	0.167836	1999, 2001
SCRI_RS_158967	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_167084	3H	51.35	6.43E-04	0.167836	1999, 2001
SCRI_RS_170765	3H	51.35	6.43E-04	0.167836	1999, 2001
SCRI_RS_171453	3H	51.35	6.43E-04	0.167836	1999, 2001
SCRI_RS_176264	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_188912	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_194233	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_214796	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_224702	3H	51.35	6.43E-04	0.167836	1999, 2001
SCRI_RS_229167	3H	51.35	8.70E-05	0.176072	1999, 2001
SCRI_RS_84470	3H	51.35	8.70E-05	0.176072	1999, 2001
BOPA1_7140-595	5H	48.19	7.81E-04	0.1308	1999, 2001
BOPA1_ABC09365-1-3-378	5H	48.19	7.81E-04	0.1308	1999, 2001
SCRI_RS_186111	5H	48.19	7.16E-05	0.179892	1999, 2001
SCRI_RS_236759	5H	48.19	6.47E-04	0.134574	1999, 2001
SCRI_RS_166296	5H	65.97	2.27E-06	0.242395	1999, 2000, 2001
SCRI_RS_218201	5H	94.72	3.75E-05	0.190286	1999, 2000

Table 16 continued

Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Years
SCRI_RS_154574	6H	24.01	3.24E-04	0.148356	1999, 2001
SCRI_RS_231372	6H	24.65	7.96E-04	0.130409	1999, 2001
SCRI_RS_8034	6H	100.4	7.51E-04	0.131599	1999, 2001
SCRI_RS_47197	7H	22.73	5.31E-07	0.271149	1999, 2000, 2001
SCRI_RS_160641	7H	22.73	4.84E-06	0.228636	1999, 2001
SCRI_RS_182092	7H	22.73	4.84E-06	0.228636	1999, 2001
BOPA2_12_30329	7H	23.02	3.89E-07	0.276594	1999, 2000, 2001
SCRI_RS_229727	7H	23.02	3.69E-06	0.233605	1999, 2001
SCRI_RS_169269	7H	23.02	3.00E-05	0.231771	1999, 2001
BOPA1_8365-454	7H	23.3	2.07E-07	0.287648	1999, 2000, 2001
SCRI_RS_142007	7H	23.8	1.47E-06	0.285251	1999, 2001
SCRI_RS_187590	7H	50.99	8.02E-04	0.130258	1999, 2001

R^2 = square of the correlation coefficient between two loci (variance explained by marker).
P-values = the probability of having a false association, as –log10 (P)

The SNPs significant through all the years are more likely to have a fundamental role in the genetic control of the trait.

4.1.1.1.4 Most significant markers associated with flowering date trait in the rainfed condition: comparative analyses with previously mapped genes and QTL*s*

During our analysis we summarized and grouped clusters harboring the most significant SNPs (exceeding the Bonferroni threshold) into QTLs. These QTLs were named according to the most significant SNP marker, with the lowest *P*-value in the cluster as presented in (Table 17). This analysis resulted into 15 potential QTLs (Table 17).

Table 17: The most significant SNP markers associated with FD trait amongst all years and with rainfed condition. Indications of previously mapped genes and nearby markers potentially linked with the different genes or QTLs are shown in the table. Markers indicated herein showed the lowest *P*-values in the different years for each chromosome.

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
**FD_R_99	SCRI_RS_229932	1H	48.51	5.65E-05	0.182443			QTL.FD.1
FD_R_99	SCRI_RS_192657	2H	29.39	3.04E-12	0.457664	*Ppd-H1*, QTL2_HD	Turner *et al.*, 2005; Wang *et al.*, 2010; Pasam *et al.*, 2012	QTL.FD.4
FD_R_99	SCRI_RS_146429	3H	83.07	7.32E-05	0.177433			QTL.FD.7
FD_R_99	SCRI_RS_14498	4H	60.69	4.64E-05	0.186211			QTL.FD.9
FD_R_99	SCRI_RS_166296	5H	65.97	2.27E-06	0.242395			QTL.FD.12
FD_R_99	SCRI_RS_218201	5H	94.72	8.42E-06	0.2184	*Fr-H2/CBF*, QTL12_HD	Knox *et al.*, 2010; Francia *et al.*, 2007; Pasam *et al.*, 2012	QTL.FD.13
FD_R_99	SCRI_RS_6399	6H	116.01	3.66E-06	0.233729			QTL.FD.17
FD_R_99	SCRI_RS_47197	7H	22.73	5.31E-07	0.271149			QTL.FD.18
FD_R_00	SCRI_RS_120529	2H	23.16	5.92E-05	0.183595			QTL.FD.4
FD_R_00	SCRI_RS_1793	3H	133.29	7.79E-04	0.130856			QTL.FD.8
FD_R_00	SCRI_RS_122057	4H	97.31	3.10E-06	0.236782			QTL.FD.10
FD_R_00	SCRI_RS_232930	5H	15.56	8.07E-05	0.17555			QTL.FD.11
FD_R_00	BOPA1_ABC14397-1-2-208	7H	85.98	4.15E-05	0.188332	*HvCO1*,QTL18_HD	Griffiths *et al.*, 2003; Wang *et al.*, 2010; Pasam *et al.*, 2012	QTL.FD.20
FD_R_01	BOPA2_12_30934	1H	133.14	6.18E-06	0.226601	*Mat-a*	Zakhrabekova *et al.*, 2012	QTL.FD.2
FD_R_01	SCRI_RS_192657	2H	29.39	4.44E-06	0.232723	*Ppd-H1*, QTL2_HD	Turner *et al.*, 2005; Wang *et al.*, 2010; Pasam *et al.*, 2012	QTL.FD.4
FD_R_01	SCRI_RS_84470	3H	51.35	8.70E-05	0.176072			QTL.FD.6
FD_R_01	SCRI_RS_151735	4H	52.2	9.26E-04	0.12738			QTL.FD.9
FD_R_01	SCRI_RS_230675	5H	118.75	9.13E-04	0.130697			QTL.FD.14
FD_R_01	SCRI_RS_169672	6H	60.06	7.23E-05	0.17767	QTL14_HD	Pasam *et al.*, 2012	QTL.FD.16
FD_R_01	BOPA1_8365-454	7H	23.3	4.00E-06	0.234658			QTL.FD.18

Significant markers associated for flowering date with GLM model corresponding, *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 genes or QTLs from literature are indicated.

**FD_R_99 = Flowering date under rainfed condition in the year 1999 and 00, 01 for the years 2000, and 2001.

R^2 = square of the correlation coefficient between two loci (variance explained by marker).

P-values = the probability of having a false association, as –log10 (P).

QTL.FD = Quantitative trait loci for most significant markers associated with FD trait were grouped according to this study.

In this analysis we recovered markers consistent with previously mapped QTLs as well as known genes controlling the trait (Table 17), also QTLs that can be considered potentially novel. For instance, two QTLs were identified on chromosome 1H. One of these loci named QTL.FD.2 (represented by SNP marker BOPA2_12_30934 at 133.14 cM, Tables 16 and 18) is in accordance with the position of the *Mat-a* gene which was recovered in the nearby genomic region of 2_0915 and 3_0231 SNP markers (Zakhrabekova *et al.*, 2012, Close *et al.*, 2009). Significantly, QTL.FD.2 was recovered also in our association analysis for the controlled irrigation condition under the same year (2001) (Table 15). Another important QTL in chromosome 2H (1999 and 2001) was named QTL.FD.4 (SNP marker BOPA1_7144-973 at 28.68 cM, Tables 16 and 17) and is congruent with QTL2_HD by Pasam *et al.* (2012) and the *Ppd-H1* gene (Turner *et al.*, 2005). An interesting locus on chromosome 5H (QTL.FD.13, SCRI_RS_218201 and BOPA1_6315-914SNP markers at 94.72 cM) co-localizes with a cluster of genes encoding C-repeat binding factors (CBFs) affecting freezing tolerance and winter hardiness of the temperate-climate cereals (Knox *et al.*, 2010). Our marker is corresponding with the recovered QTL12_HD by Pasam *et al.* (2012). Our study also recovered QTL.FD.16 on chromosome 6H at 60.06 cM, consistent with a QTL region in Pasam *et al.* (2012) (Table 17). An interesting locus (QTL.FD.20) identified by four SNP markers on chromosome 7H (BOPA1_ABC14397-1-2-208, BOPA2_12_31395, SCRI_RS_146157, SCRI_RS_148722, 85.98 cM) may correspond to QTL18_HD as shown by Pasam *et al.* (2012) This regions harbors previously mapped flowering time QTLs and genes such as *HvCO1* (Griffiths *et al.*, 2003; Wang *et al.*, 2010). Although, this QTL appeared in one year (2000) under one condition (rainfed), further experiments may confirm its relevance in our germplasm panel. In parallel, a potentially novel QTL region was identified on chromosome 7H from 22.73 to 23.3 cM (QTL.FD.18). Noticeably, this QTL was recovered under rainfed condition in years 1999 and 2001, as well as for the controlled irrigation in 1999 (Table 15).

Significant markers and known genes and QTLs are indicated in the GLM Manhattan plots in Figure 19 (Bradbury *et al.*, 2007).

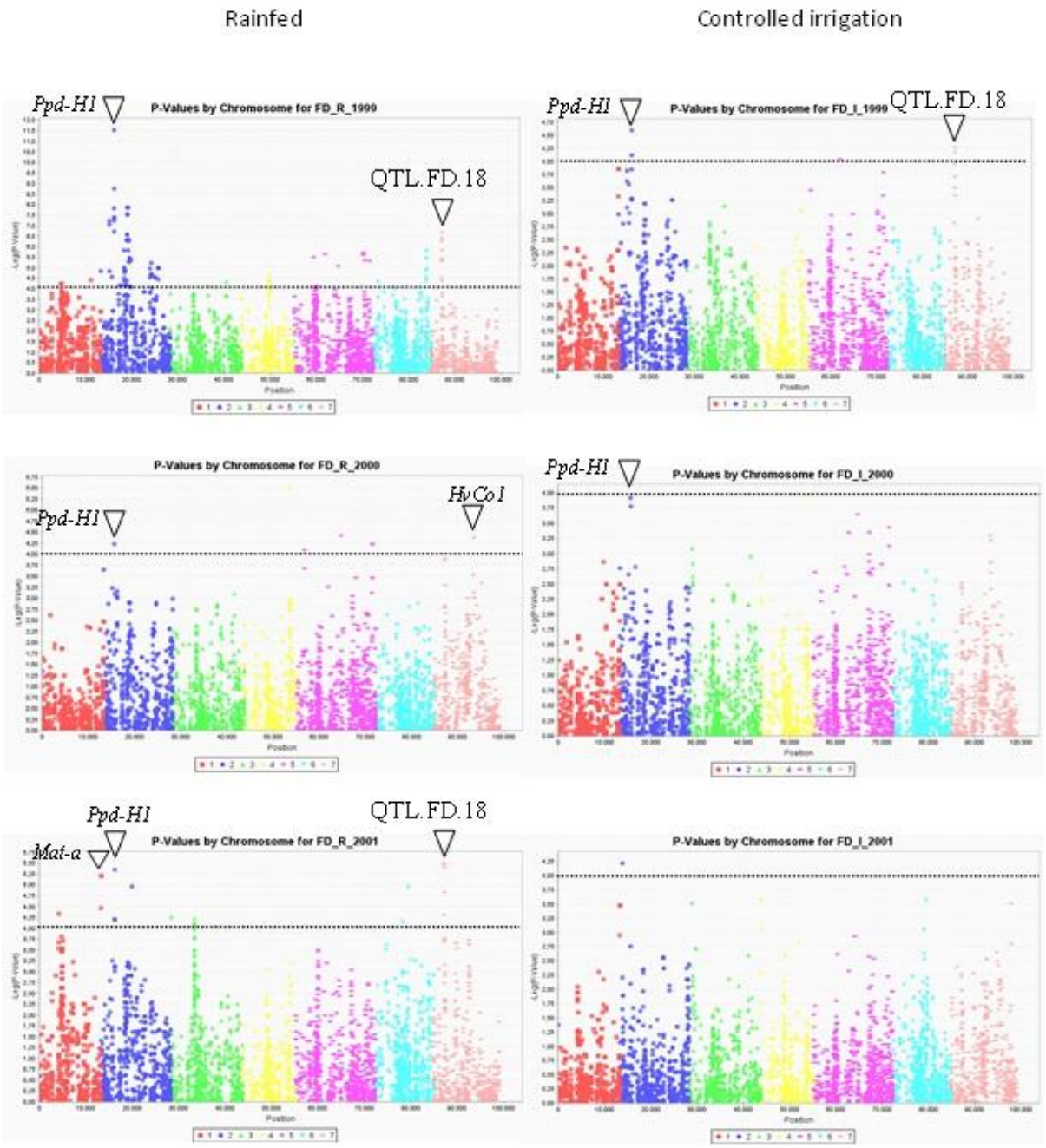


Figure 19: Manhattan plot: Whole genome association scan results showing (-log10 (p)) values for marker associations with flowering date trait in separate environments. Regions where significant SNPs are potentially linked with known genes are indicated.

Years and irrigation condition indicated above the graphs. Dashed line intersecting Y axis represents significance threshold for Bonferroni correction (The peaks above minimum threshold of 10^{-4} (*P*-value = 0.0001). SNPs from all seven barley chromosomes are in linear order on X axis (different chromosomes in different colors representing each chromosome respectively from left to right: 1H, 2H, 3H, 4H, 5H, 6H, and 7H). GLM model was used.

In conclusion, we identified 20 potential QTLs associated to the flowering date trait. Of these, four QTLs (QTL.FD.5, QTL.FD.15, QTL.FD.19, and QTL.FD.20) were recovered under controlled irrigation conditions only (Table 17). Conversely, seven QTLs (QTL.FD.1, QTL.FD.6, QTL.FD.8, QTL.FD.11, QTL.FD.13, QTL.FD.14, and QTL.FD.17) were identified under rainfed conditions but not under controlled irrigation (Table 17).

4.1.1.2 Analysis with the mixed linear model (MLM) and comparison with GLM results

MLM as a statistical model approach assumes the errors to be normally distributed and also includes both fixed and random effects. This approach was demonstrated as an improved method to simultaneously account for population structure and unequal relatedness among individuals (Pressoir *et al.*, 2006).

Applying this model for FD trait association analysis yielded few significant SNPs (N=20) under both controlled irrigation and rainfed conditions. All these SNPs were exceeding the Bonferroni threshold value (*P*-value <= -log 0.0001). These markers were considered to be significantly associated with the FD trait. Of these, Three SNPs were also found with GLM analysis.

The differences of the output between the two models (GLM and MLM) might be due to the strict Bonferroni correction in the MLM model as reported by Mezaka *et al* (2011). On the other hand, the two approaches are based on different algorithms. Therefore, the difference of significant SNPs between GLM and MLM is inherent to the statistical approach essential for each model. In GLM case, stratification is addressed using structured association. However, In MLM case, population structure is fitted as a fixed effect, whereas kinship among individuals is incorporated as the variance-covariance structure of the random effect for the individuals.

The following figure (Figure 20) summarizes results from the MLM analysis.

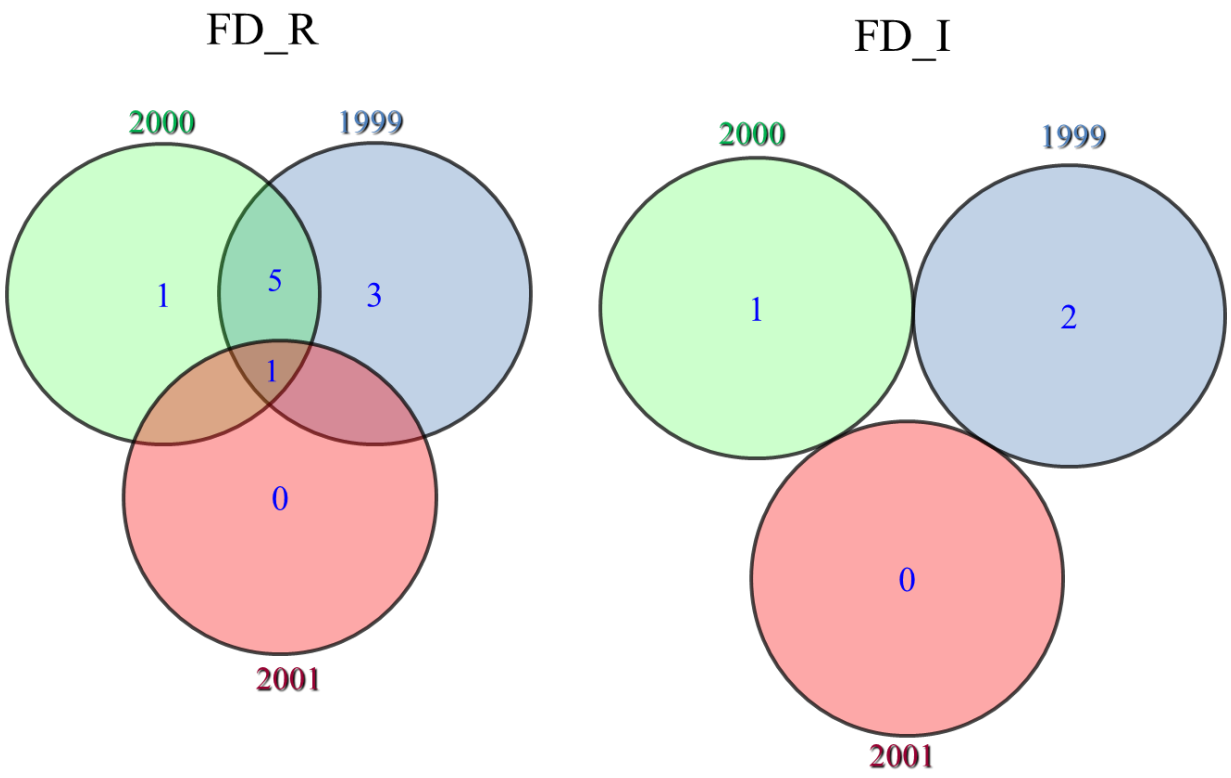


Figure 20: Venn diagram showing the most significant markers overlap. The number of markers in all 7 chromosomes exceeding the Bonferroni threshold line (considered significant) and associated with flowering date (FD) trait in each year and each irrigation condition with their intersections among years (FD_R = flowering date in rainfed condition, FD_I = flowering date in controlled irrigation condition). This result was recovered from the analysis by Tassel program following MLM approach.

A total number of 3 significant SNP markers were identified in the controlled irrigation condition (Figure 20). In comparison, a total number of 17 SNP markers were recovered in the rainfed condition, with 6 intersected across different years (Figure **20**).

Rainfed

Controlled irrigation

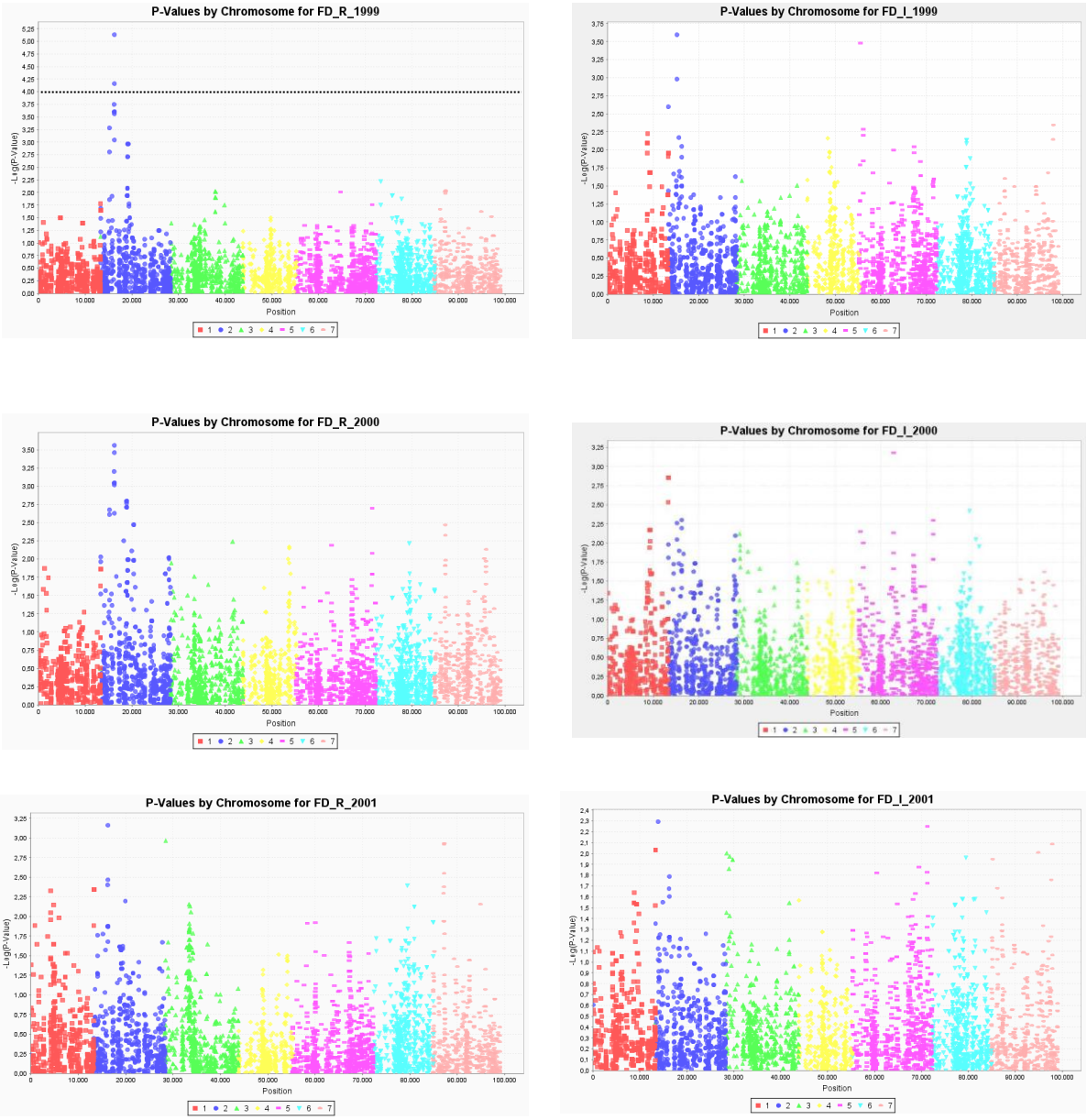


Figure 21: Manhattan plots for whole genome association scans showing (-log10(p)) values for marker associations with flowering date trait in separate environments.

Years and irrigation condition indicated above the graphs. Dashed line intersecting Y axis represents significance threshold for Bonferroni correction (The peaks above minimum threshold of 10^{-4} (*P*-value = 0.0001). SNPs from all seven barley chromosomes are in linear order on X axis (different chromosomes in different colors representing each chromosome respectively from left to right: 1H, 2H, 3H, 4H, 5H, 6H, and 7H). MLM model was used.

4.1.2 Associations between SNPs and plant height (PH) trait in barley

4.1.2.1 Analysis with the general linear model (GLM)

We applied GLM statistical analysis to carry out the genome wide association analysis of the PH with 4,661 SNPs. The entire procedure used for analyzing this trait was similar to that used for FD trait. As a result, 385 SNPs and 49 SNPs were statistically significant, excluding the shared SNPs among years, when we considered controlled irrigation and rainfed conditions, respectively.

Considering the three years of analysis, 11, 72 and 334 SNPs were statistically significant under the controlled irrigation for 1999, 2000 and 2001, respectively (Figure 22). In the case of rainfed condition, the 49 SNPs were statistically significant only for 1999 (Figure 22). The intersections of the significant markers among the years and the irrigation systems are presented in Figure 22.

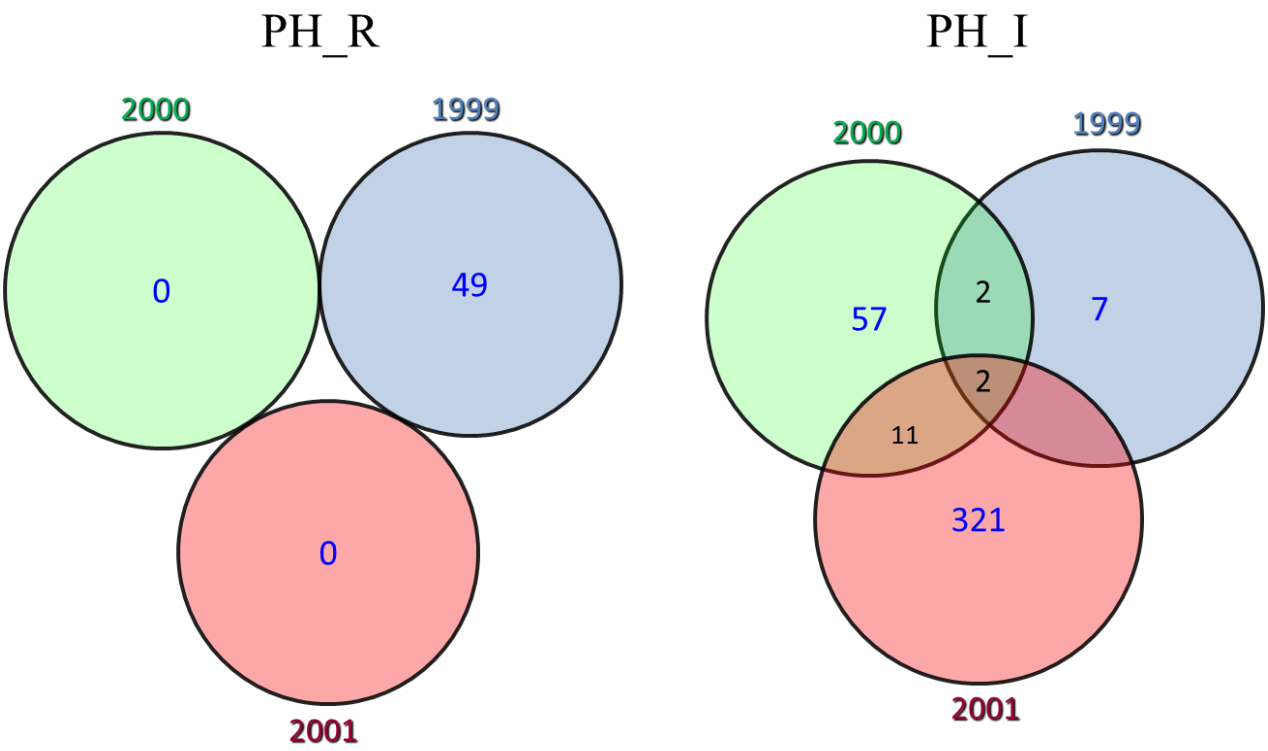


Figure 22: Venn diagram showing the most significant markers and their overlaps among years. All these markers are exceeding the Bonferroni threshold line (considered significant) and associated with plant height (PH) trait in each year and each irrigation condition with their intersections among years (PH_R = plant height in rainfed condition, PH_I = plant height in controlled irrigation condition). The entire analysis was run by *Tassel* program following GLM.

4.1.2.1.1 SNP markers linked to PH in controlled irrigation condition and significant across years

A total number of 15 SNP markers exceeding the Bonferroni threshold were in common between at least two or three years (1999, 2000, and 2001) as shown in Table 18.

Table 18: SNP markers intersected among years and associated with (PH) trait in the controlled irrigation condition

Marker	Chr. n.	Locus position (cM)	*P*-value	R^2	Years
BOPA1_2877-867	1H	59.14	9,9817E-7	0,25713	1999, 2000, 2001
BOPA2_12_10166	1H	65.23	3,5881E-5	0,19113	1999, 2000, 2001
SCRI_RS_224335	3H	59.63	9,9012E-4	0,16065	1999, 2000
SCRI_RS_237939	3H	62.96	5,3056E-4	0,13853	2000, 2001
BOPA2_12_30399	3H	62.96	7,9145E-4	0,13206	2000, 2001
BOPA1_42-512	3H	75.21	9,026E-4	0,16076	2000, 2001
BOPA1_1272-459	7H	74.43	6,3092E-4	0,17017	2000, 2001
SCRI_RS_103902	7H	76.56	3,9515E-4	0,14607	2000, 2001
SCRI_RS_149650	7H	77.83	3,9515E-4	0,14607	2000, 2001
BOPA1_3140-491	7H	77.97	7,5107E-4	0,13312	2000, 2001
SCRI_RS_109893	7H	77.97	7,5107E-4	0,13312	2000, 2001
BOPA1_12027-128	7H	91.93	9,0584E-4	0,12783	1999, 2000
SCRI_RS_144462	7H	108.07	1,0589E-5	0,21652	2000, 2001
SCRI_RS_162972	7H	125.28	7,2923E-5	0,17953	2000, 2001
SCRI_RS_180889	7H	125.28	7,2923E-5	0,17953	2000, 2001

R^2 = square of the correlation coefficient between two loci (variance explained by marker).
P-values = the probability of having a false association, as –log10 (P)

4.1.2.1.2 Most significant markers associated with plant height trait in the controlled irrigation condition: comparative analyses with previously mapped genes and QTLs

We grouped clusters harboring the most significant SNPs (exceeding the Bonferroni threshold) into QTLs. We proposed a name for each QTL in our study according to the most significant SNP marker (with the lowest *P*-value) in the cluster, as presented in (Table 19). This analysis resulted into 47 potential QTLs (Table 19).

Table 19: GWAS result: The most significant SNP markers associated with PH trait amongst all years and with controlled irrigation condition. Indications of previously mapped genes and nearby markers potentially linked with the different genes or QTLs are shown in the table. Markers indicated herein showed the lowest *P*-values in the different years for each chromosome.

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
*H_I_99	BOPA1_2877-867	1H	59.14	9,7225E-4	0,12641			QTL.PH.1
H_I_99	SCRI_RS_224335	3H	59.63	6,2184E-4	0,17048			QTL.PH.19
H_I_99	BOPA1_ConsensusGBS0654-4	5H	46.18	4,9111E-4	0,14007			QTL.PH.30
H_I_99	SCRI_RS_235055	5H	125.63	9,5504E-4	0,12676			QTL.PH.34
H_I_99	BOPA1_6523-1691	6H	119.33	6,4872E-4	0,13452			QTL.PH.42
H_I_99	BOPA1_12027-128	7H	91.93	9,0584E-4	0,12783			QTL.PH.47
H_I_00	BOPA1_2877-867	1H	59.14	9,9817E-7	0,25713			QTL.PH.1
H_I_00	SCRI_RS_224335	3H	59.63	9,9012E-4	0,16065			QTL.PH.19
H_I_00	BOPA1_42-512	3H	75.21	4,0018E-4	0,17765			QTL.PH.20
H_I_00	SCRI_RS_180027	3H	117.63	8,1477E-5	0,17537			QTL.PH.22
H_I_00	SCRI_RS_9736	7H	71.25	4,4658E-5	0,22398			QTL.PH.46
H_I_00	BOPA1_12027-128	7H	91.93	5,2134E-4	0,13888			QTL.PH.47
H_I_00	SCRI_RS_144462	7H	108.07	6,2923E-4	0,1367			QTL.PH.48
H_I_00	SCRI_RS_162972	7H	125.28	5,1274E-5	0,18639			QTL.PH.49
*H_I_01	BOPA1_9638-619	1H	59.42	9,1246E-4	0,12918			QTL.PH.1
H_I_01	BOPA1_3201-603	1H	86.97	5,9252E-5	0,18152			QTL.PH.3
H_I_01	BOPA1_2711-234	1H	95.89	2,4565E-4	0,1538			QTL.PH.4
H_I_01	SCRI_RS_201865	1H	116.78	8,837E-5	0,17379			QTL.PH.5

Table 19: continued

Trait	**SNP Marker**	**Chr. no.**	**Locus position (cM)**	***P*-value**	R^2	**Associated nearby gene/QTL previously mapped**	**Reference from literatures**	**QTL name**
H_I_01	SCRI_RS_192552	2H	8.57	4,8385E-4	0,14037	QTL1_PHT	Pasam *et al.*, 2012	QTL.PH.6
H_I_01	SCRI_RS_147371	2H	26.77	4,3799E-4	0,14235			QTL.PH.7
H_I_01	SCRI_RS_154981	2H	41.22	6,6696E-4	0,13396			QTL.PH.8
H_I_01	BOPA1_6804-1197	2H	62.46	9,5793E-4	0,1267	QTL3_PHT	Pasam *et al.*, 2012	QTL.PH.9
H_I_01	BOPA2_12_31293	2H	73.73	7,2825E-4	0,13375	QTL4_PHT, *sdw3*	Pasam *et al.*, 2012; Gottwald *et al.*, 2004	QTL.PH.10
H_I_01	SCRI_RS_203799	2H	86.76	5,4349E-4	0,13805			QTL.PH.11
H_I_01	SCRI_RS_223885	2H	106.44	9,4702E-4	0,12693			QTL.PH.12
H_I_01	SCRI_RS_211291	2H	117.99	6,6308E-4	0,13408			QTL.PH.13
H_I_01	SCRI_RS_215471	2H	129.32	4,5591E-4	0,1432	QTL5_PHT, QHt.StMo-2H.2	Pasam *et al.*, 2012; Hayes *et al.*,1993	QTL.PH.14
H_I_01	SCRI_RS_161281	2H	145.61	1,1388E-4	0,16887			QTL.PH.15
H_I_01	SCRI_RS_214280	3H	15.16	8,592E-5	0,17434			QTL.PH.16
H_I_01	SCRI_RS_199987	3H	40.65	7,4927E-4	0,16465	QTL6_PHT, QHt.HaMo-3H	Pasam *et al.*, 2012; Marquez-Cedillo *et al.*, 2001; Hayes *et al.*,1993	QTL.PH.17
H_I_01	BOPA1_5183-924	3H	51.63	4,4327E-5	0,22413			QTL.PH.18
H_I_01	BOPA1_6883-203	3H	68.06	8,6905E-5	0,17412			QTL.PH.19
H_I_01	SCRI_RS_185339	3H	86.19	5,3661E-5	0,18342			QTL.PH.20
H_I_01	SCRI_RS_159125	3H	103.75	9,5446E-4	0,12678			QTL.PH.21
H_I_01	SCRI_RS_167755	3H	118.2	2,9165E-5	0,19727			QTL.PH.22
H_I_01	BOPA2_12_31251	3H	123.37	7,8707E-5	0,17604			QTL.PH.23
H_I_01	SCRI_RS_194527	3H	148.37	7,9769E-4	0,13507			QTL.PH.24
H_I_01	BOPA2_12_31310	4H	50.99	7,8929E-4	0,13059			QTL.PH.26
H_I_01	SCRI_RS_10818	4H	102.12	4,0703E-4	0,14381			QTL.PH.27

Table 19: continued

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
H_I_01	BOPA1_10207-1024	5H	9.31	3,4288E-4	0,14721			QTL.PH.28
H_I_01	SCRI_RS_108416	5H	29.1	2,0766E-4	0,15711			QTL.PH.29
H_I_01	SCRI_RS_147462	5H	46.32	5,3351E-4	0,14003			QTL.PH.30
H_I_01	SCRI_RS_11206	5H	73.33	6,3081E-4	0,16824			QTL.PH.31
H_I_01	SCRI_RS_168467	5H	87.36	1,0367E-5	0,21454			QTL.PH.32
H_I_01	SCRI_RS_45011	5H	98.89	1,238E-5	0,21123			QTL.PH.33
H_I_01	BOPA1_1697-636	5H	122.36	9,0826E-5	0,17726			QTL.PH.34
H_I_01	SCRI_RS_216751	5H	134.72	9,6274E-5	0,17213			QTL.PH.35
H_I_01	BOPA1_6054-1050	5H	143.4	8,6745E-4	0,13495			QTL.PH.36
H_I_01	BOPA2_12_30673	6H	30.1	9,8341E-4	0,12618			QTL.PH.37
H_I_01	SCRI_RS_121633	6H	49.08	9,9723E-4	0,1259			QTL.PH.38
H_I_01	BOPA1_8048-952	6H	62.75	8,5322E-4	0,13054			QTL.PH.39
H_I_01	SCRI_RS_182275	6H	105.31	7,1962E-5	0,17777			QTL.PH.41
H_I_01	SCRI_RS_126069	6H	116.01	7,9269E-4	0,13203			QTL.PH.42
H_I_01	SCRI_RS_166511	7H	0.92	2,0766E-4	0,15711			QTL.PH.43
H_I_01	SCRI_RS_47197	7H	22.73	8,4344E-4	0,13077			QTL.PH.44
H_I_01	BOPA1_497-386	7H	57.93	8,0951E-4	0,16304			QTL.PH.45
H_I_01	BOPA1_3140-491	7H	77.97	7,1992E-4	0,13398			QTL.PH.46
H_I_01	SCRI_RS_144462	7H	108.07	1,0589E-5	0,21652			QTL.PH.48
H_I_01	SCRI_RS_180889	7H	125.28	7,2923E-5	0,17953			QTL.PH.49
H_I_01	BOPA1_1847-1745	7H	140.86	3,4793E-4	0,14692			QTL.PH.50

Significant markers associated with plant height (GLM model): the corresponding *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 genes or QTLs from literature are indicated.
*H_I_99: Plant height under controlled irrigation condition in the year 1999 and 00, 01 for the years 2000, and 2001.
R^2: square of the correlation coefficient between two loci (variance explained by marker).
P-values: the probability of having a false association, as –log10 (P).
QTL.PH: Quantitative trait loci for most significant markers associated with PH trait were grouped according to this study.

Forty two potentially novel QTLs have been recovered from our analysis. They are mainly grouped in chromosomes 1H, 4H, 5H, 6H, and 7H (Table 19). Remarkably, we were able to identify 5 QTLs corresponding with previously mapped genes or QTLs (Table 19). For instance, QTL.PH.10 on chromosome 2H coincides with the map position of the *sdw3* gene, which is known to play a major role in gibberellin-insensitive dwarfing in barley (Gottwald *et al.*, 2004). This result also confirms QTL4_PHT identified by Pasam *et al.* (2012) at 73.75 cM in chromosome 2H. In addition, QTL.PH.6, QTL.PH.9, QTL.PH.14, QTL.PH.17 are also consistent with previous findings (Pasam *et al.*, 2012; Marquez-Cedillo *et al.*, 2001; Hayes *et al.*,1993).

4.1.2.1.3 SNP markers linked to PH in the rainfed condition

No SNP markers associated with PH were significant for more than one year when analyzing the association for this trait under the rainfed condition.

4.1.2.1.4 Most significant markers associated with plant height trait in the rainfed condition: comparative analyses with previously mapped genes and QTLs

We summarized each clustering group harboring the most significant SNPs (exceeding the Bonferroni threshold) into QTLs. These QTLs were identified according to the most significant SNP marker (with the lowest *P*-value in the cluster) as presented in (Table 20). This analysis resulted into 35 potential QTLs (Table 20).

Table 20: The most significant SNP markers associated with PH trait amongst all years under rainfed condition. Indications of previously mapped genes and nearby markers potentially linked with the different genes or QTLs are shown in the table. Markers indicated herein showed the lowest *P*-values in the different years for each chromosome.

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
*H_R_99	SCRI_RS_238263	2H	5.38	8,6832E-4	0,12868	QTL1_PHT	Pasam *et al.*, 2012	QTL.PH.6
H_R_99	BOPA1_7032-201	2H	26.77	9,5489E-5	0,17229			QTL.PH.7
H_R_99	SCRI_RS_78277	2H	39.38	7,3683E-4	0,13351	*Ph2*, QTL2_PHT	Qi *et al.*, 1998; Yan *et al.*, 1998; Pasam *et al.*, 2012	QTL.PH.8
H_R_99	BOPA1_4241-445	2H	123.65	7,4434E-4	0,13177			QTL.PH.13
H_R_99	SCRI_RS_151056	2H	149.15	8,3409E-4	0,1427			QTL.PH.15
H_R_99	SCRI_RS_224377	3H	64.87	7,2648E-4	0,13538			QTL.PH.19
H_R_99	SCRI_RS_167755	3H	118.2	5,1072E-5	0,18646			QTL.PH.22
H_R_99	BOPA1_2146-2256	5H	47.22	6,3866E-4	0,16798			QTL.PH.30
H_R_99	SCRI_RS_219608	5H	73.33	4,1981E-4	0,14486			QTL.PH.31
H_R_99	SCRI_RS_238417	5H	87.64	6,1234E-4	0,13567			QTL.PH.32
H_R_99	SCRI_RS_188785	5H	121.25	9,2564E-4	0,12889			QTL.PH.34
H_R_99	SCRI_RS_226875	6H	116.01	5,7211E-4	0,13863			QTL.PH.42
H_R_99	BOPA2_12_10652	7H	102.34	7,089E-4	0,13274			QTL.PH.48
H_R_99	SCRI_RS_181575	7H	124.58	4,3288E-4	0,14258			QTL.PH.49
H_R_00	BOPA1_7299-183	1H	61.47	5,8357E-5	0,18181			QTL.PH.1
H_R_00	SCRI_RS_161281	2H	145.61	7,3659E-5	0,17732			QTL.PH.15
H_R_00	SCRI_RS_108543	3H	64.87	8,6559E-4	0,12874			QTL.PH.19
H_R_00	SCRI_RS_106728	3H	128.61	7,3659E-5	0,17732	QTL8_PHT, *sdw1/denso*	Pasam *et al.*, 2012; Jia *et al.*, 2001; Yin *et al.*, 1999	QTL.PH.23
H_R_00	SCRI_RS_156016	5H	30.56	4,5992E-4	0,14138			QTL.PH.29
H_R_00	SCRI_RS_7368	5H	51.46	2,2418E-4	0,1556			QTL.PH.30
H_R_00	SCRI_RS_159611	5H	71.67	8,6694E-4	0,1616			QTL.PH.31

Table 20: continued

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
H_R_00	BOPA1_2188-425	6H	24.36	8,8219E-4	0,12986			QTL.PH.37
H_R_00	SCRI_RS_219810	6H	53.61	8,4767E-4	0,12916			QTL.PH.38
H_R_00	SCRI_RS_16773	6H	69.26	8,5844E-4	0,12891			QTL.PH.39
H_R_00	BOPA2_12_31126	6H	115.93	8,9607E-4	0,12804			QTL.PH.42
H_R_00	BOPA2_12_10652	7H	102.34	7,3659E-5	0,17732			QTL.PH.48
H_R_01	SCRI_RS_181239	1H	71.18	6,6064E-4	0,13415			QTL.PH.2
H_R_01	BOPA1_3201-603	1H	86.97	8,4412E-4	0,12924			QTL.PH.3
H_R_01	BOPA2_12_30532	1H	100.92	2,6085E-4	0,18852			QTL.PH.4
H_R_01	SCRI_RS_158687	2H	39.66	3,5525E-4	0,14994	*Ph2*, QTL2_PHT	Qi *et al.*, 1998; Yan *et al.*, 1998; Pasam *et al.*, 2012	QTL.PH.8
H_R_01	SCRI_RS_221992	2H	67.92	7,1195E-4	0,13421	QTL3_PHT	Pasam *et al.*, 2012	QTL.PH.9
H_R_01	SCRI_RS_235860	2H	76.91	8,3546E-4	0,12945	QTL4_PHT, *sdw3*	Pasam *et al.*, 2012; Gottwald *et al.*, 2004	QTL.PH.10
H_R_01	BOPA1_3906-558	3H	15.3	2,8522E-4	0,15085			QTL.PH.16
H_R_01	BOPA1_2838-663	3H	49.29	8,9707E-4	0,12952			QTL.PH.17
H_R_01	SCRI_RS_189322	3H	145.33	4,8612E-5	0,18532			QTL.PH.24
H_R_01	BOPA1_12128-313	4H	26.77	5,0193E-4	0,13964			QTL.PH.25
H_R_01	SCRI_RS_219608	5H	73.33	4,4771E-4	0,14356			QTL.PH.31
H_R_01	SCRI_RS_157897	5H	98.12	4,0674E-4	0,14382			QTL.PH.33
H_R_01	BOPA1_1697-636	5H	122.36	9,5328E-4	0,12981			QTL.PH.34
H_R_01	BOPA1_6260-183	5H	135.35	8,4412E-4	0,12924			QTL.PH.35
H_R_01	BOPA1_ABC08769-1-1-205	6H	50.85	9,039E-4	0,12937			QTL.PH.38
H_R_01	SCRI_RS_159133	6H	95.04	8,5172E-4	0,16197			QTL.PH.40
H_R_01	SCRI_RS_179580	6H	116.01	7,9303E-4	0,13202			QTL.PH.42

Table 20: continued

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
H_R_01	SCRI_RS_152122	7H	47.03	1,3201E-4	0,16599			QTL.PH.45
H_R_01	BOPA1_1212-890	7H	70.68	9,6894E-4	0,16111			QTL.PH.46
H_R_01	SCRI_RS_181575	7H	124.58	9,2633E-4	0,12738			QTL.PH.49
H_R_01	BOPA1_1847-1745	7H	140.86	6,4953E-4	0,13449			QTL.PH.50

Significant markers associated with plant height (GLM model): the corresponding, *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 genes or QTLs from literature are indicated.
*H_R_99 = Plant height under rainfed condition in the year 1999 and 00, 01 for the years 2000, and 2001.
R^2 = square of the correlation coefficient between two loci (variance explained by marker).
P-values = the probability of having a false association, as –log10 (P).
QTL.PH = Quantitative trait loci for most significant markers associated with PH trait were grouped according to this study.

We recovered important QTLs for PH coincident with previously mapped QTLs and genes. For example, QTL.PH.6 and QTL.PH.8 in chromosome 2H are collinear with recently identified QTL loci QTL1_PHT, and QTL2_PHT respectively recovered by Pasam *et al.* (2012). The dwarfing gene *sdw1* has been deployed in America and Australia, while its allelic form *denso* is regularly seen in European two-rowed germplasm (Pasam *et al.*, 2012). Interestingly, the region spanning QTL.PH.23 on chromosome 3H harbors the *sdw1* gene (Jia *et al.*, 2001).

QTL.PH.10 on chromosome 2H coincides with the region hosting the *sdw3* gene, which plays a main role in gibberellin-insensitive dwarfing barley (Gottwald *et al.*, 2004), as well as QTL4_PHT in Pasam *et al.* (2012). The remaining QTLs presented in Table 19 and 21 could be potentially novel. Further experiments are needed to confirm and refine analysis of these QTLs.

The following Manhattan plots (Figure 23) illustrate QTLs/genes recovered from our analyses, under controlled irrigation and rainfed conditions, as recovered from GLM analysis using *Tassel* program (Bradbury *et al.*, 2007).

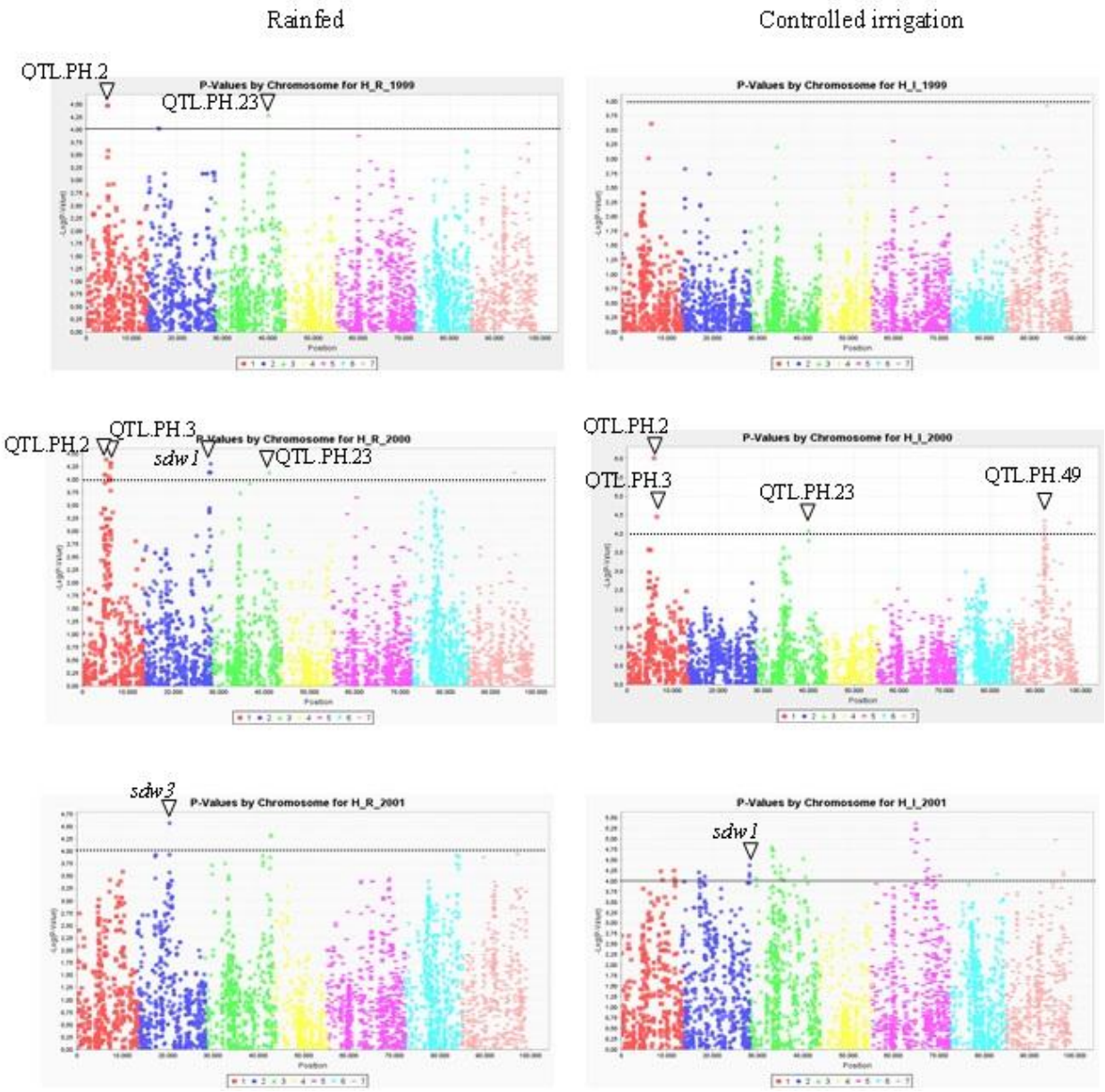


Figure 23: Manhattan plots for whole genome association scans showing (-log10 (p)) values for marker associations with plant height trait in separate environments. Regions where significant SNPs are potentially linked with known genes are indicated.

Years and irrigation condition indicated above the graphs. Dashed line intersecting Y axis represents significance threshold for Bonferroni correction (the peaks above minimum threshold of 10^{-4} (*P*-value = 0.0001). SNPs from all seven barley chromosomes are in linear order on X axis (different chromosomes in different colors representing each chromosome respectively from left to right: 1H, 2H, 3H, 4H, 5H, 6H, and 7H). GLM model was used.

In conclusion, the recovery of PH-QTLs in our study varied based on the year of investigation or with the different irrigation condition. QTL.PH.1 was recovered in three years under the controlled irrigation system and under rainfed condition in the year 2000 only. Notably, the

climate condition in the year 2000 was generally favorable with relatively high precipitation levels of rainfalls (Rizza *et al.*, 2004). QTL.PH.11 and QTL.PH.2 were identified in year 2001 under controlled irrigation and rainfed conditions, respectively.

4.1.3 Associations between SNPs and grain yield (GY) trait in barley

4.1.3.1 Analysis with the general linear model (GLM)

We applied GLM statistical analysis to carry out genome wide association scans for GY. All the procedures considered for analyzing this trait were similar to those previously used for FD and PH traits: 129 SNPs and 262 SNPs markers displayed significant associations with GY, excluding the shared markers amongst different years, when we considered controlled irrigation and rainfed treatment, respectively.

The controlled irrigation showed total numbers of 13, 96 and 38 SNPs statistically significant for the years 1999, 2000 and 2001, respectively (Figure 24). With respect to rainfed condition, 16, 212, and 47 SNPs were statistically significant for the years 1999, 2000 and 2001, respectively (Figure 24). The intersections of the significant markers among years are presented in Figure 24.

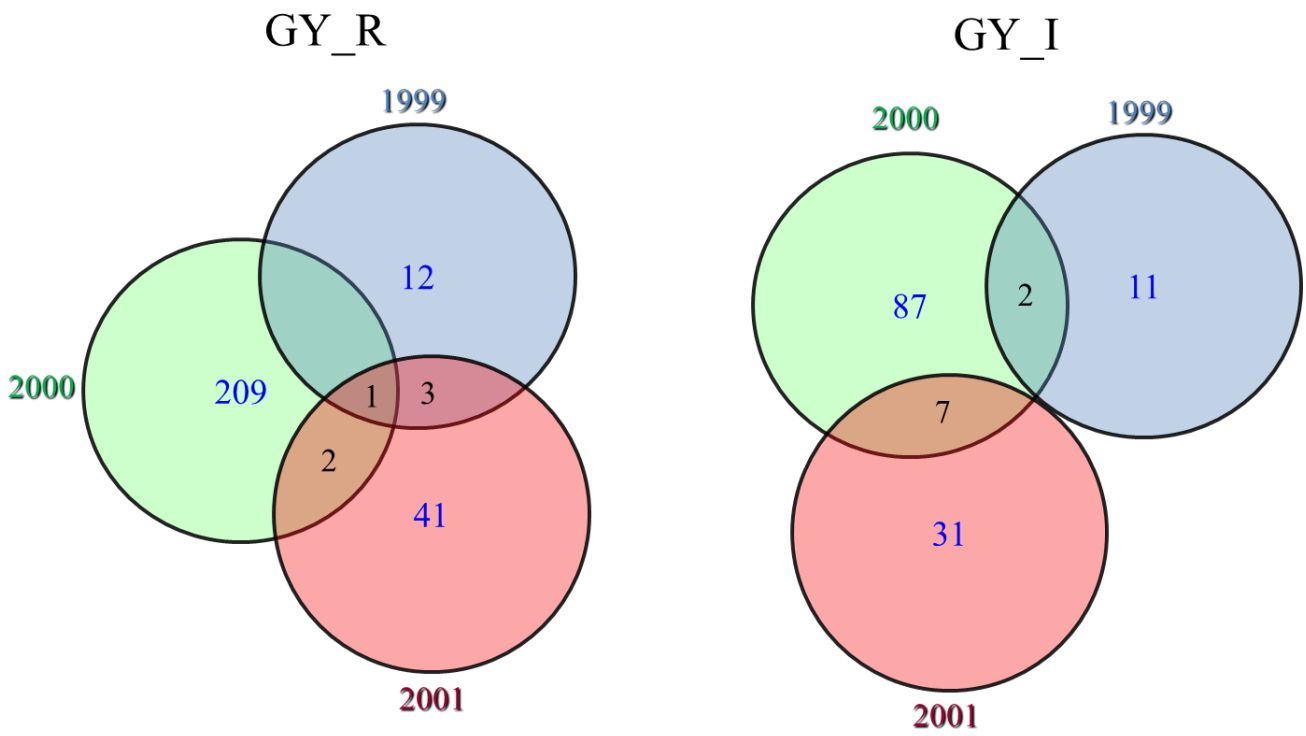


Figure 24: Venn diagram showing the most significant markers and their overlaps among years. All these markers are exceeding the Bonferroni threshold line (considered significant) and associated with grain yield (GY) trait in each year and each irrigation condition with their intersections among years (GY_R = grain yield in rainfed condition, GY_I = grain yield in controlled irrigation condition). The entire analysis was run by *Tassel* program following GLM.

4.1.3.1.1 SNP markers linked to GY under controlled irrigation conditions

A total number of 9 SNP markers exceeding the Bonferroni threshold were in common across years (1999, 2000, and 2001) as shown in Table 21.

Table 21: Markers intersected among years and associated with (YLD) trait in the controlled irrigation condition

Marker	**Chr. n.**	**Locus pos. (cM)**	***P*-value**	R^2	**Years**
BOPA1_4787-1746	3H	143.48	4,9928E-6	0,23575	1999, 2000
BOPA1_6450-755	5H	129.44	3,5516E-6	0,23685	2000, 2001
BOPA1_6970-462	5H	122.64	8,97E-4	0,12802	1999, 2000
BOPA1_ConsensusGBS0704-2	5H	129.44	3,5516E-6	0,23685	2000, 2001
SCRI_RS_128407	5H	129.44	3,5516E-6	0,23685	2000, 2001
SCRI_RS_133453	5H	129.44	3,5516E-6	0,23685	2000, 2001
SCRI_RS_147315	5H	122.43	9,358E-4	0,16	2000, 2001
SCRI_RS_148120	5H	129.44	2,4221E-4	0,15408	2000, 2001
SCRI_RS_174710	5H	121.74	4,1386E-4	0,17696	2000, 2001

R^2 = square of the correlation coefficient between two loci (variance explained by marker).
P-values = the probability of having a false association, as –log10 (P)

4.1.3.1.2 Most significant markers associated with grain yield trait in the controlled irrigation condition: comparative analyses with previously mapped genes and QTLs

We grouped clusters harboring significant SNPs into QTLs. We proposed a name for each QTL in our study according to the most significant SNP marker (with the lowest *P*-value in the cluster) as presented in (Table 22). This analysis resulted into 28 potential QTLs (Table 22).

Table 22: Most significant SNP markers associated with GY trait amongst all years and with controlled irrigation condition. Indications of previously mapped genes and nearby markers potentially linked with the different genes or QTLs are shown in the table. Markers indicated herein showed the lowest *P*-values in the different years for each chromosome.

Trait	SNP Marker	Chr. n.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
YLD_I_99	SCRI_RS_154153	2H	149.36	4,2355E-4	0,17648			QTL.GY.7
YLD_I_99	BOPA2_12_30482	3H	128.75	7,9037E-4	0,13526			QTL.GY.10
YLD_I_99	BOPA1_4787-1746	3H	143.48	4,9928E-6	0,23575			QTL.GY.11
YLD_I_99	BOPA1_ABC03900-1-2-406	5H	125.76	6,5792E-5	0,18154	*vrn1*	Sutka *et al.*, 1999	QTL.GY.18
YLD_I_00	BOPA1_2036-1027	1H	66.29	5,6644E-4	0,13723	QTL2_TGW, *vrs3*	Pasam et al., 2012; Pourkheirandish and Komatsuda 2007	QTL.GY.1
YLD_I_00	SCRI_RS_235724	1H	95.82	4,771E-5	0,18568			QTL.GY.2
YLD_I_00	BOPA2_12_10905	1H	106.16	1,4859E-4	0,2			QTL.GY.3
YLD_I_00	BOPA2_12_30631	2H	12.11	4,8552E-4	0,1403			QTL.GY.4
YLD_I_00	BOPA1_1447-464	2H	38.1	9,1918E-4	0,12753	QTL4_TGW	Pasam *et al.*, 2012	QTL.GY.5
YLD_I_00	SCRI_RS_91810	2H	79.89	6,2767E-4	0,13518			QTL.GY.6
YLD_I_00	SCRI_RS_138848	2H	149.36	6,8256E-4	0,1335			QTL.GY.7
YLD_I_00	BOPA2_12_30616	3H	59.63	8,821E-4	0,16124			QTL.GY.8
YLD_I_00	SCRI_RS_237846	3H	117	8,2819E-4	0,12963			QTL.GY.9
YLD_I_00	BOPA1_4787-1746	3H	143.48	1,9146E-4	0,16431			QTL.GY.11
YLD_I_00	BOPA1_1513-514	4H	14.73	3,9022E-4	0,14464			QTL.GY.14
YLD_I_00	BOPA1_4098-758	4H	26.35	9,6412E-5	0,1721	QTL12_TGW, *int-c*, SNP11_20680	Pasam *et al.*, 2012; Ramsay *et al.*,2011; Comadran *et al.*, 2011	QTL.GY.15
YLD_I_00	BOPA2_12_30993	4H	48.65	2,2972E-4	0,15512			QTL.GY.16
YLD_I_00	BOPA1_299-163	4H	100.64	6,8395E-4	0,13346			QTL.GY.17

Table 22: continued

Trait	SNP Marker	Chr. n.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
YLD_I_00	BOPA1_ABC04322-1-3-208	5H	122.43	7,9456E-5	0,17786	*vrn1*	Sutka *et al.*, 1999	QTL.GY.18
YLD_I_00	BOPA1_7337-388	5H	143.26	5,6332E-4	0,13734			QTL.GY.19
YLD_I_00	SCRI_RS_146663	6H	13.67	9,8571E-4	0,12613			QTL.GY.20
YLD_I_00	BOPA1_4642-1124	6H	68.91	8,3409E-4	0,12948			QTL.GY.21
YLD_I_00	SCRI_RS_8252	6H	86.26	9,5658E-5	0,17226			QTL.GY.22
YLD_I_00	BOPA2_12_31357	7H	54.82	6,6961E-4	0,13545			QTL.GY.24
YLD_I_00	BOPA1_1674-468	7H	76.98	4,8431E-5	0,18539			QTL.GY.25
YLD_I_00	BOPA1_ABC14397-1-2-208	7H	85.98	4,5665E-5	0,18652	QTL19_TGW	Pasam *et al.*, 2012	QTL.GY.26
YLD_I_00	SCRI_RS_214028	7H	110.27	2,4567E-5	0,20056			QTL.GY.27
YLD_I_01	BOPA2_12_30444	1H	103.82	9,1149E-4	0,16431			QTL.GY.3
YLD_I_01	BOPA1_7728-341	3H	51.2	9,1416E-4	0,16049			QTL.GY.8
YLD_I_01	SCRI_RS_236603	3H	154.82	4,2503E-6	0,23101			QTL.GY.12
YLD_I_01	SCRI_RS_231066	4H	3.61	1,4071E-4	0,16663			QTL.GY.13
YLD_I_01	BOPA1_ConsensusGBS0704-2	5H	129.44	3,5516E-6	0,23685	*vrn1*	Sutka *et al.*, 1999	QTL.GY.18
YLD_I_01	BOPA1_ConsensusGBS0369-1	6H	62.75	6,7964E-4	0,13359			QTL.GY.21
YLD_I_01	SCRI_RS_101976	7H	1.49	4,6433E-4	0,14623			QTL.GY.23
YLD_I_01	SCRI_RS_229041	7H	54.96	7,753E-4	0,16777			QTL.GY.24
YLD_I_01	SCRI_RS_16316	7H	128.54	2,4457E-4	0,18985	QTL20_TGW	Pasam *et al.*, 2012	QTL.GY.28

Significant markers associated for grain yield with GLM model: *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 genes or QTLs from literature.

**YLD_I_99: Grain yield under the controlled irrigation condition in the year 1999 and 00, 01 for the years 2000, and 2001.

R^2: square of the correlation coefficient between two loci (variance explained by marker).

P-values: the probability of having a false association, as –log10 (P).

QTL.GY: Quantitative trait loci for most significant markers associated with GY trait were grouped according to this study.

A total of 147 markers yielding 28 QTLs were significantly associated with grain yield trait (Table 22).

Remarkably, QTL.GY.1, QTL.GY.15 and QTL.GY.18 are co-localized to genomic regions where genes *vrs3*, *int-c*, and *vrn1* (VRN-H1) (Table 22) were previously identified, respectively. A comparative study of the genomic region spanning the wheat *vrn1* gene revealed this to be co-linear with *Sh2* on chromosome 5H of barley (Stuka *et al.* 1999; Laurie et al., 1995). Hence, our identified region encompassing QTL.GY.18 could be harboring the barley *vrn1* gene. *VRN-H1* acts as a promoter of heading induced by vernalization (Trevaskis 2010), regulating the transition of the shoot apex to the reproductive stage (Hemming *et al.*, 2009). QTL.GY.5, QTL.GY.26 and QTL.GY.28 are matching previously identified QTLs (Table 22) from Pasam *et al* (2012). In addition, 22 potential novel QTLs were identified under this condition of analysis (Table 22).

4.1.3.1.3 SNP markers linked to GY under rainfed condition

In total 6 SNP markers exceeding the Bonferroni threshold were shared by at least two years (1999, 2000, and 2001), as shown in Table 23.

Table 23: Markers intersected among years and associated with FD trait in the rainfed condition

Marker	**Chr. no.**	**Locus position (cM)**	***P*-value**	***R*^{2}**	**Year**
BOPA1_4787-1746	3H	143.48	3,2201E-5	0,19987	1999, 2001
SCRI_RS_148120	5H	129.44	4,3327E-5	0,18753	2000, 2001
SCRI_RS_150232	5H	120.21	3,7943E-7	0,27413	1999, 2000, 20001
SCRI_RS_158259	5H	120.35	1,5412E-5	0,20713	1999, 2001
SCRI_RS_159474	5H	120.35	2,1636E-5	0,20074	1999, 2001
SCRI_RS_194291	7H	77.41	6,8725E-4	0,13336	2000, 2001

R^2 = square of the correlation coefficient between two loci (variance explained by marker).
P-values = the probability of having a false association, as –log10 (P)

4.1.3.1.4 Most significant markers associated with grain yield trait in the rainfed condition: comparative analyses with previously mapped genes and QTLs

We summarized each clustering group harboring the most significant SNPs into QTLs. These QTLs were named according to the most significant SNP marker (with the lowest *P*-value in the cluster) as presented in (Table 24). This analysis resulted in 37 potential QTLs (Table 24).

Table 24: GWAS result: The most significant SNP markers associated with GY trait amongst all years and with rainfed condition. Indications of previously mapped genes and nearby markers potentially linked with the different genes or QTLs are shown in the table. Markers indicated herein showed the lowest *P*-values in the different years for each chromosome.

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
YLD_R_99	BOPA1_4787-1746	3H	143.48	3,2201E-5	0,19987			QTL.GY.11
YLD_R_99	SCRI_RS_192689	4H	103.97	7,3889E-4	0,16495			QTL.GY.17
YLD_R_99	SCRI_RS_122458	5H	125.49	7,6618E-4	0,13272	*vrn1*	Sutka *et al.*, 1999	QTL.GY.18
YLD_R_99	SCRI_RS_149556	6H	65.72	8,0205E-5	0,21004			QTL.GY.21
YLD_R_00	BOPA1_4020-643	1H	48.23	9,6777E-4	0,1265			QTL.GY.1A
YLD_R_00	BOPA1_12492-541	1H	72.52	8,2791E-5	0,17506			QTL.GY.1B
YLD_R_00	BOPA1_2711-234	1H	95.89	6,3788E-7	0,26505			QTL.GY.2
YLD_R_00	BOPA2_12_10905	1H	106.16	9,9253E-4	0,1606			QTL.GY.3
YLD_R_00	SCRI_RS_182947	2H	8.29	7,8705E-4	0,13374			QTL.GY.4
YLD_R_00	BOPA2_12_30631	2H	12.11	7,0927E-4	0,13273			QTL.GY.4
YLD_R_00	BOPA2_12_30042	2H	56.37	6,8711E-4	0,13337			QTL.GY.5A
YLD_R_00	SCRI_RS_4930	2H	77.27	5,9904E-7	0,26615			QTL.GY.6
YLD_R_00	BOPA2_12_10937	2H	142.63	1,0563E-5	0,21657			QTL.GY.7
YLD_R_00	BOPA1_15141-257	3H	39.38	6,8028E-4	0,13513			QTL.GY.8A
YLD_R_00	BOPA1_3674-1352	3H	88.81	4,1844E-4	0,14326			QTL.GY.9A
YLD_R_00	SCRI_RS_237846	3H	117	9,8702E-4	0,1261			QTL.GY.9
YLD_R_00	SCRI_RS_194148	3H	128.05	8,5901E-4	0,12889			QTL.GY.10
YLD_R_00	SCRI_RS_236603	3H	154.82	3,2273E-4	0,14841			QTL.GY.12
YLD_R_00	SCRI_RS_100278	4H	1.13	2,4903E-5	0,19807			QTL.GY.13
YLD_R_00	BOPA1_1513-514	4H	14.73	6,1997E-6	0,22407			QTL.GY.14
YLD_R_00	BOPA1_12128-313	4H	26.77	4,2618E-7	0,27211	QTL12_TGW, *int-c*, SNP11_20680	Pasam *et al.*, 2012; Ramsay *et al.*,2011; Comadran *et al.*, 2011	QTL.GY.15
YLD_R_00	SCRI_RS_74014	4H	51.27	3,6962E-7	0,27459			QTL.GY.16
YLD_R_00	SCRI_RS_163033	4H	59.99	3,797E-6	0,23823			QTL.GY.16
YLD_R_00	BOPA1_ConsensusGBS0589-1	4H	67	8,2732E-4	0,16258			QTL.GY.16A
YLD_R_00	BOPA1_299-163	4H	100.64	9,6412E-5	0,1721			QTL.GY.17

Table 24: continued

Trait	SNP Marker	Chr. no.	Locus position (cM)	*P*-value	R^2	Associated nearby gene/QTL previously mapped	Reference from literatures	QTL name
YLD_R_00	SCRI_RS_192396	5H	14.24	4,2465E-4	0,14296			QTL.GY.18A
YLD_R_00	SCRI_RS_108416	5H	29.1	1,3442E-4	0,16563			QTL.GY.18B
YLD_R_00	SCRI_RS_148120	5H	129.44	4,3327E-5	0,18753	*vrn1*	Sutka *et al.*, 1999	QTL.GY.18
YLD_R_00	SCRI_RS_235652	5H	159.51	8,0778E-4	0,13013			QTL.GY.19A
YLD_R_00	SCRI_RS_176	6H	49.08	8,0697E-4	0,13015	QTL17_TGW	Pasam *et al.*, 2012	QTL.GY.21A
YLD_R_00	BOPA1_4642-1124	6H	68.91	4,0445E-5	0,18884			QTL.GY.21
YLD_R_00	SCRI_RS_159133	6H	95.04	9,105E-4	0,16057			QTL.GY.22
YLD_R_00	SCRI_RS_17542	6H	119.33	2,1768E-4	0,15798			QTL.GY.22A
YLD_R_00	BOPA2_12_31357	7H	54.82	4,7867E-4	0,14222			QTL.GY.24
YLD_R_00	BOPA1_1674-468	7H	76.98	2,1907E-6	0,24306			QTL.GY.25
YLD_R_00	SCRI_RS_134640	7H	84.56	7,9684E-5	0,1758	QTL19_TGW	Pasam *et al.*, 2012	QTL.GY.26
YLD_R_00	SCRI_RS_214028	7H	110.27	1,8717E-6	0,24856			QTL.GY.27
YLD_R_00	BOPA1_1847-1745	7H	140.86	8,9521E-4	0,12806	QTL21_TGW, QTw.HaTR-7H.1	Pasam *et al.*, 2012; Pillen *et al.*, 2003	QTL.GY.29
YLD_R_01	SCRI_RS_135248	2H	94.9	6,5124E-4	0,13444			QTL.GY.6A
YLD_R_01	BOPA1_4787-1746	3H	143.48	6,1433E-4	0,14047			QTL.GY.11
YLD_R_01	SCRI_RS_143514	5H	92.99	2,699E-5	0,19876			QTL.GY.18C
YLD_R_01	SCRI_RS_150232	5H	120.21	3,7943E-7	0,27413	*vrn1*	Sutka *et al.*, 1999	QTL.GY.18
YLD_R_01	SCRI_RS_202438	6H	100.85	9,442E-4	0,12699			QTL.GY.22
YLD_R_01	SCRI_RS_159529	6H	117.49	9,4243E-5	0,20917			QTL.GY.22B
YLD_R_01	SCRI_RS_160723	7H	23.8	4,2175E-4	0,1431			QTL.GY.23A
YLD_R_01	SCRI_RS_124478	7H	77.27	8,3837E-4	0,12938			QTL.GY.25

Significant markers associated with grain yield (GLM model): corresponding *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 genes or QTLs from literature are shown.

**YLD_R_99: Grain yield under the rainfed condition in the year 1999 and 00, 01 for the years 2000, and 2001.

R^2: square of the correlation coefficient between two loci (variance explained by marker).

P-values: the probability of having a false association, as –log10 (P).

QTL.GY: Quantitative trait loci for most significant markers associated with GY trait were grouped according to this study.

Under the rainfed condition we identified a total of 278 markers (Figure 24) significantly associated with the GY trait, corresponding to 37 QTLs (Table 24 and Figure 25). The majority of these QTLs were identified in the year 2000. Several QTL locations are consistent with previously identified QTLs in various mapping populations (Table 24). For instance, QTL.GY.15 co-localizes with QTL12_TGW (thousand grains weight) and SNP11_20680 (thousand kernel weight) in Pasam *et al.* (2012) and Comadran *et al.* (2011), respectively in a region spanning the *INTERMEDIUM-C* (*int-c)* gene (Table 24): *int-c* encodes an ortholog of the maize domestication gene TEOSINTE BRANCHED 1 and acts as a modifier of lateral spikelet fertility in barley (Ramsay *et al.*, 2011). The position of QTL.GY.18 corresponds to the *vrn1* (VRN-H1) gene (Table 24), encoding an APETALA1 and FRUITFULL-like MADS-box transcription factor essential for the initiation of reproductive development at the shoot apex in cereals (Wang et al 2010; Sasani et al 2009; Trevaskis *et al.*, 2003, Ferrandiz *et al.*, 2000). In addition, our analysis for this condition uncovered a total of 32 potential novel QTLs.

In summary, we identified a total of 28 and 37 QTLs when we analyzed the GY data considering controlled irrigation and rainfed conditions, respectively (Tables 22 and 24). Many of these QTLs (N = 21) are shared between conditions except 7 and 16 that are specific to controlled irrigation and rainfed conditions, respectively.

The following Manhattan plots (Figure 25) illustrate QTLs/genes regions recovered from our analyses, under controlled irrigation and rainfed conditions.

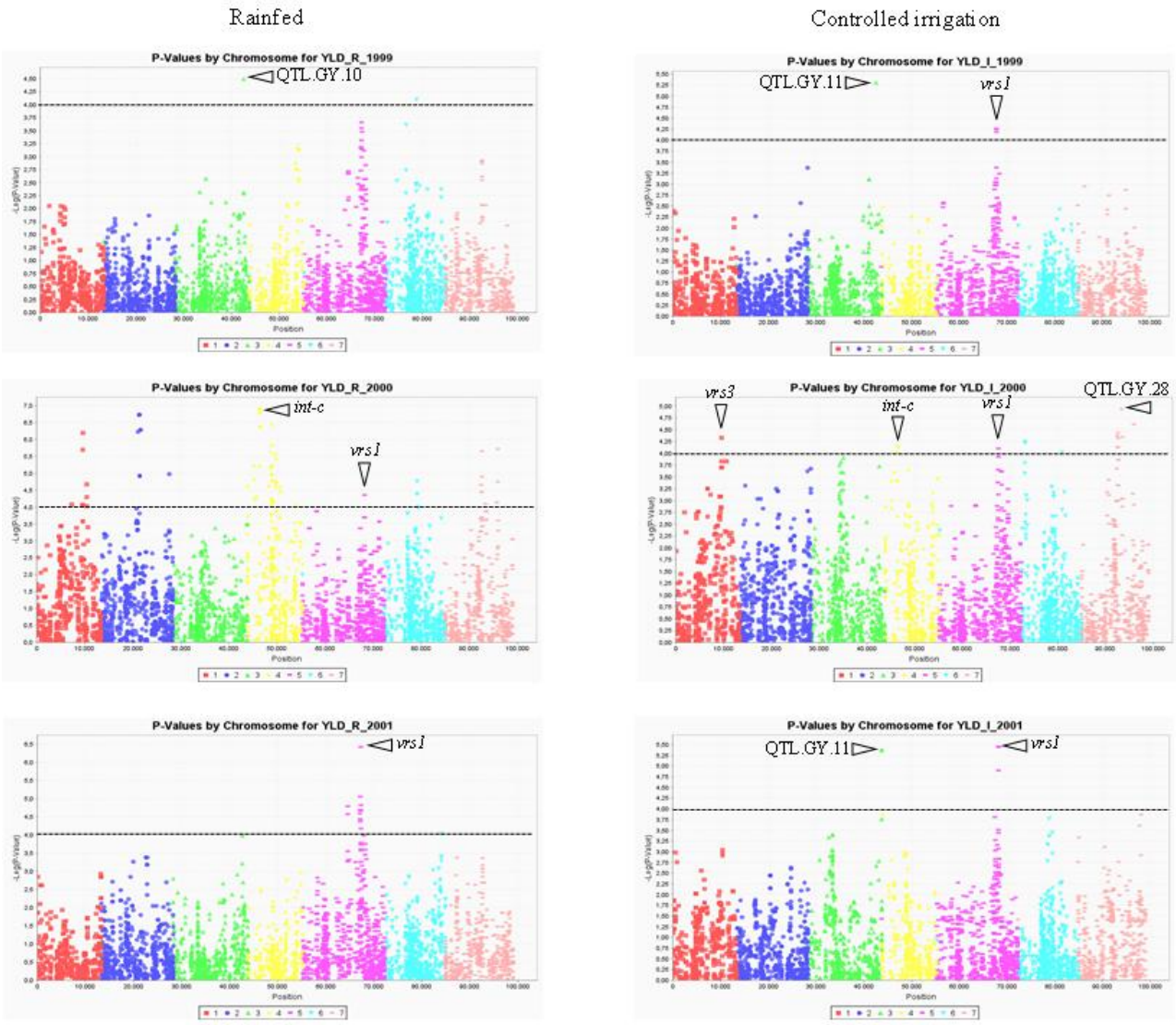


Figure 25: Manhattan plots for whole genome association scan results showing (-log10 (p)) values for marker associations with grain yield trait in separate environments. Regions where nearby significant SNPs potentially linked with known genes are indicated.

Years and irrigation condition indicated above the graphs. Dashed line intersecting Y axis represents significance threshold for Bonferroni correction (The peaks above minimum threshold of 10^{-4} (*P*-value = 0.0001). SNPs from all seven barley chromosomes are in linear order on X axis (different chromosomes in different colors representing each chromosome respectively from left to right: 1H, 2H, 3H, 4H, 5H, 6H, and 7H). GLM model was used.

In conclusion, genomic regions recovered in our analyses revealed numerous QTLs (Tables 22 and 24). Several were novel and some were in agreement with previously identified genes and QTLs. Accordingly, we have 20 potential novel QTLs that are stable across irrigation conditions. Importantly, *vrn1* (VRN-H1) gene is located within the region spanning QTL.GY.18, which seems to be very stable across all years under both irrigation conditions (Figure 25). QTL.GY.10 appeared in both irrigation conditions in two years 1999 and 2000, respectively. These and other QTLs can be further verified and dissected to evaluate their potential importance for breeding for grain yield traits in various environmental conditions.

CHAPTER 4

DISCUSSION

Discussion

Cultivated barley (*Hordeum vulgare* ssp. *vulgare* L.) is an established model species for genetic studies (Koorneef *et al.*, 1997): it is an annual crop with a short life cycle, a diploid genome with only seven pairs of chromosomes, allowing multiple testing as a result of being true breeding, exhibiting wide diversity in terms of morphology and genetics. Barley is also an important cereal crop species ranking fourth in the worldwide production among cereals after rice, wheat and maize (according to FAO in 2012; http://www.fao.org).

1 Amplified Fragment Length Polymorphism (AFLP)

DNA-based markers have been extensively used in many areas of genetic research such as genetic diversity (Erschadi *et al.*, 2001; Godt and Hamrick, 1999) or genetic kinship (Mace *et al.*, 1999; Roa *et al.*, 1997). The AFLP technique does not require any prior sequence knowledge and produces dominant markers, preferably used for fingerprinting because of the high amount of polymorphism it can detect (Mueller and Wolfenbarger, 1999; Hongtrakul *et al.*, 1997; Weising *et al.*, 1995). In addition, such powerful markers are not influenced by environmental factors (Cheong, 2012) and are convenient in the early stages of selection within large progenies (Paun and Schönswetter, 2012).

For these reasons, we initially chose the AFLP technique to assess the genetic diversity of our germplasm panel (Rizza *et al.*, 2004), as a starting point to finally investigate associations of interest (Ersoz *et al.*, 2009). A set of 144 barley accessions, previously characterized by AFLP profiling by Fricano *et al.*, (2009) included 61 genotypes from the germplasm panel considered in this PhD project (Rizza et al., 2004). Thus, we used the same protocols and primer combinations used by Fricano *et al.* (2009) to carry out AFLP fingerprinting of 83 genotypes which are representing diversity for drought tolerance of cultivated European germplasm and previously characterized under different water regimes, (Rizza *et al.*, 2004) (Tables 1 and 3).

Despite adherence to the same protocols of Fricano *et al.* (2009), different AFLP profiles were obtained and the total number of polymorphic peaks (71 peaks in this study) was inferior compared with 215 peaks in the previous study (Fricano *et al.*, 2009). In our study, different primer

combinations displayed heterogeneous numbers of polymorphic peaks. For example, primer combinations E38M55 and E36M55 resulted in 24, 16 polymorphic peaks respectively. In comparison, primer combinations of E36M49 and E41M55 resulted in a final number of 8 and 3 polymorphic peaks, respectively. This may be a result of the choice of primer combinations which yield low numbers of unambiguous polymorphisms (Schut *et al.*, 1997). Inconsistencies and low-reproducibility of AFLP profiles is not uncommon. However, discrepancies may be induced to personal or laboratory-related criteria (Herrmann *et al.*, 2010). Scoring of polymorphic peaks depending on the human eye can result in errors of inaccuracy in scoring AFLP-bands due to missing observations (Schut *et al.*, 1997). Therefore, the number of polymorphic bands obtained can vary. Automated scoring methods have been proposed as a solution to ameliorate this problem in which these methods mainly concerns bin width and peak height to finally minimize the error in scoring (Whitlock *et al.*, 2008; Herrmann *et al.*, 2010; Arthofer *et al.*, 2010; Kuck *et al.*, 2012). Another explanation would be the low quality runs that can create noisy signals while scoring process (i.e. "false- positive" or "false- negative" peaks) within the dataset (Arrigo *et al.*, 2009; Price and Casler, 2012).

The use of the different tissue types might result in different AFLP profiles (Boiteux *et al.*, 1999, Aranzana *et al.*, 2001, Arnau *et al.*, 2002). This can be due to differences in DNA purity obtained from different tissues (Benjak *et al.*, 2006). In our study we used leaf tissue following Fricano et al. (2009) so this may not be a key issue in our case.

The use of AFLP marker requires improved statistical methods to determine the genetic structure of the population (Zhu *et al.*, 2008). Furthermore, AFLP as a dominantly inherited marker require special statistical methods if used to assess population genetic parameters (Falush *et al.*, 2007; and Ritland, 2005). In our study, phylogenetic analyses based on AFLP data resulted in an atypical and population structure which could not be reconciled with classical partitions of barley germplasm (see results section 1.2. Clustering analysis, Chapter 3). In conclusion, in our study, the AFLP technique did not show the levels of consistency and reliability required for analysis of population structure and subsequent association analyses.

Finally, SNP markers offered us a possible solution to overcome the problems in reconstruction of population structure based on AFLP data.

2 Candidate genes approach

Drought tolerance is a key trait for increasing and stabilizing barley productivity in dry areas worldwide. Identification of the genes associated to drought tolerance will facilitate understanding of the molecular mechanisms of stress responses, and support the genetic improvement of barley through marker-assisted selection or gene transformation. Association genetics analyses have been successfully applied to correlate allelic variation at selected candidate genes with phenotypic variation for traits of interest (eg Thornsberry *et al.*, 2001, Fricano *et al.*, 2009).

Candidate genes may be selected based on prior information from biochemical pathways, mutational analysis, physiology studies in model and non-model plant species (Mackay, 2001; Risch and Merikangas, 1996) or linkage analysis for the trait of interest.

CBF genes are considered major players in drought responses. Expression of *CBF* genes increases under cold or drought stress but it is low under the normal growth condition (Haake, 2002). *CBF* genes have been successfully used to engineer abiotic stress tolerance in a number of different species (Neffar *et al.*, 2011). Barley comprises at least 20 *HvCBF* genes forming three multi-gene groupings: *HvCBF1*-, *HvCBF3*-, and *HvCBF4*-subgroups (Skinner *et al.*, 2005). Barley CBFs share many functional characteristics with dicot CBFs, including a general primary domain (AP2 domain) structure and also C repeat CRT/dehydration responsive element (DRE) motif which is involved in dehydration responsive gene expression (Agarwal *et al.*, 2006).

In this project we set out to explore the possible association between nucletide variants in HvCBF genes and variation for agronomic traits under different watering regimes. The most straightforward method of identifying candidate gene alleles/polymorphisms relies on re-sequencing of amplicons from genetically distinct individuals of a larger population: selection of such individuals is crucial, as in general more individuals are needed to identify rare SNPs compared to those required to identify common SNPs.

In our case, we resequenced genomic amplicons from *HvCBF2*, *HvCBF4A* and *HvCBF7* in a subset of 40 barley accessions selected as genetically diverse based on growth habit, row-type and AFLP fingerprinting data (Fricano *et al.*, 2009). We identified 4 polymorphic loci in *HvCBF4A* coding region and a part from UTR. To this end our identified loci representing one SNP every 125 bp when considering our high quality sequenced part of 478 bp. Our result is comparable to Fricano *et*

al. (2009) with the total number of haplotypes versus SNPs revealed in both studies, when Fricano *et al.* (2009) worked on some of the *Hvcbf* gene families, as concluded there was a SNP location every 128 bp in his barley cultivars tested. It has been estimated previously an average frequency of 1 SNP per 70 bp in maize (Rafalski *et al.*, 2001).

Resequencing of the other two CGs, *HvCBF2* and *HvCBF7*, resulted in no SNPs for the tested genomic part and accessions used. Success of CG-based association analysis is dependent mainly on the recovered SNPs within the accessions considered. In our case, the time required and limited level of allelic diversity revealed by this approach became limiting factors towards our objective of performing an association analysis.

The relatively hight level of nucleotide diversity have been observed before in wild barley (Morrell *et al.*, 2006; Caldwell *et al.*, 2006). These have been interpreted as a consequence of genetic hitchhiking, since selection targeted at individual loci led to the reduction in genetic diversity in linked loci (Fricano *et al.*, 2009). In addition, the loss of genetic diversity increases significantly overtime as a result of domestication and intensive breeding, especially when taking modern cultivated varieties into account.

In general, the candidate gene approach has been proven in many studies to be powerful for characterization and cloning of Mendelian loci as well as quantitative trait loci (Pflieger *et al.*, 2001; Tabor *et al.*, 2002). Cost-effectiveness can be an advantage of such trait-specific and hypothesis-driven approaches. Nevertheless, the practicability of traditional candidate gene approach is largely limited by its confidence on existing knowledge about the acknowledged biology of the phenotype under investigation. In addition, the detailed molecular anatomy of most biological traits remains unknown.

For these reasons, we decided to expand to a genome-wide analysis taking advantage of a newly developed SNP panel as discussed in the next section.

3 Genome wide association analyses

In several cereal species, germplasm collections are being established for genome-wide association analyses (rice: Mather *et al.*, 2007, maize: Yu and Buckler 2006, sorghum: Casa *et al.*, 2006; Hamblin *et al.*, 2005). Comparing results from different studies of research communities will add a deeper understanding of genetic architecture and mechanism of adaptation, and consequently facilitates the mapping of functional variations (Buckler and Gore 2007).

Barley is an ideal candidate plant for association mapping. It has a long history of recombination events and conserved linkage disequilibrium, and is highly autogamous (Caldwell *et al.*, 2006). Thus fewer markers can scan the whole genome compared to the case in outbreeding species such as maize (Remington *et al.*, 2001).

Genome wide association mapping is a comprehensive approach scans the genome for genetic variation. Interestingly, this may not require previous information about candidate genes. On one hand, a large number of markers are being tested for association with various complex traits. On the other hand, the quality of the phenotypic data can affect the resolution of the genome wide association study result (Rafaliski, 2010). To fully exploit the potential of this approach, a research consortium and adequate funds are required (Zhu *et al.*, 2008).

In comparison with other genetic markers, SNP markers are can be efficiently and reliably genotyped with high-throughput detection systems and are thus rapidly becoming the markers of choice for complex trait dissection studies (Zhu *et al.*, 2008; Ersoz *et al.*, 2009). Such technologies have provided an extensive number of high quality SNPs for dissecting the genetic basis of complex quantitative traits in plants, constructing dense genetic maps, opening new perspectives towards gene or allele discovery for traits of agricultural importance (Mackay *et al.*, 2009; and Hall *et al.*, 2010). In this study, the Illumina iSELECT 9K SNP genotyping array (Comadran et al. 2012) was used in collaboration with the EXBARDIV consortium allowing the identification of 4,661 polymorphic markers in the 83 accessions studied, proving the efficiency of this technology and the variability of the accessions. Similar to previous works (Mezaka *et al.*, 2011; and Pasam *et al.*, 2012), we performed a filtering step to discard failed SNPs and those with MAF frequency <10 %. The availability of a large number of mapped high quality SNP markers (Kilian and Graner 2012; Muñoz-Amatriaín *et al.*, 2011; Close *et al.*, 2009) has allowed us to achieve a high marker coverage yielding 1 marker per approximately 0.2 cM as a mean of our markers coverage per chromosome.

That was much higher marker coverage comparing to the 1.18 cM of (Pasam *et al.*, 2012) achievement. Restricting the germplasm base may preclude important alleles from an association study. On the other hand, a highly diverse collection may comprise too many rare alleles (allele frequency <5 %) which can increase the risk of detecting spurious associations (Abecasis *et al.*, 2001). Importantly, increasing MAF > 10% may be disadvantageous when performing the association study: Pasam *et al* (2012) proposed that individual QTL may account only for a small portion of phenotypic variation, which can be due to the insufficient marker coverage or the elimination of rare alleles prior to analysis or both effects together.

3.1 Population structure

The presence of un-recognized population structure is a significant problem while carrying out association mapping. Also the different levels of relatedness in barley cultivars correlate with segregation of the genetic determinants of growth habit. As a consequence, this can give rise to false-positive associations which can lead to a failure to detect genuine associations (Simko and Hu, 2008), particularly in highly selfing species (Iwata *et al.*, 2007). To avoid such effects, we used different statistical approaches to obtain the population structure of our panel. The *Tassel* software was used to determine association between phenotypes and segregating sites, while accounting for relative kinship (Yu *et al.*, 2006) and population structure (Thornsberry *et al.*, 2001).

It was necessary to estimate the number of groups (K) in the analysis achieved by *Structure* software, in order to obtain the actual population structure. This is to determine whether our barley accessions could be grouped into genetic clusters and to infer the number of such clusters that best fit the data. This Evanno *et al.*, (2005) developed a method based on DeltaK, where the actual number of groups (K) occurs at the maximum value and at which LnP(D) reaches a true value of K (Falush *et al.*, 2003; Butts *et al.*, 2008) (Figure 15). Evanno *et al.*, 2005 provides a correct estimation of the number of clusters using an ad hoc statistic DeltaK to assign accessions based on the rate of change in the log probability of data between successive K values. This result was described our barley collection at K = 2 based on the growth habit (winter/spring) and at K = 3 when considering both, the growth habit and the row types (2/6 rows). The major phenotypic divisions in the cultivated barley suggest K = 4 (corresponding to the four possible growth habit and row-number groupings with their combinations) might be enough to confine most of the sub-population stratification exist (Cockram *et al.*, 2008), since accessions used in this study had no spring 6 rows barley therefore K = 4 cannot be considered. Comparable results obtained in this study by principle

coordinates analysis (PCoA), and it was obvious that the primary axis separates the accessions based on growth habit and further grouping is related to the row types (Figure 17). Remarkably, it was shown that the power to detect stratification increases with sample size (Smiko and Hu, 2008). The obtained results in this study are comparable with known classical partitioning of barley germplasm in previous studies (Cockram *et al.*, 2008; Cuesta-Marcos *et al.*, 2010; Wang *et al.*, 2012; Pasam *et al.*, 2012). In many previous studies (Rostoks *et al.* 2006; Zhang *et al.* 2009, Hamblin *et al.* 2010), the major factors reflecting population structure in barley were shown to be growth habit, spike morphology and geographical origin. In our present study, population substructure was explained mainly by barley growth habit and spike morphology.

Besides considerations regarding population structure, a sufficient number of individuals per subpopulation are necessary to provide adequate power for the statistical test (Haseneyer *et al.*, 2010).

3.2 Associations between SNPs and traits (FD, PH, and GY)

One of the main goals of our study was to test for possible association analysis between grain yield and the SNPs in the iSELECT panel. The intention for the inclusion of flowering date and plant height traits was to check the ability for recovery of known loci thus validating the panel for association purposes. These traits were investigated in many previous works (Haseneyer *et al.*, 2010; Pasam *et al.*, 2012, Zakhrabekova *et al.*, 2012).

GLM was used to run genome-wide scans for grain yield (GY), plant height (PH) and flowering date (FD); the latter trait was also analyzed with MLM model for a comparison between the two models.

The differences of the output between the two models might be due to the stringency of Bonferroni-corrections considered in MLM model as reported by Mezaka *et al* (2011). This stringent correction would result in the elimination of a big set of markers that cannot be considered significant under Bonferroni threshold (Balding, 2006). In addition, MLM considers multiple levels of relatedness integrated in the analysis through the matrix of population effects and kinship matrix (Sun *et al.*, 2010; Souza, 2011). As population structure is fitted as a fixed effect, whereas kinship among individuals is incorporated as the variance-covariance structure of the random effect for the individuals. However, in GLM case, stratification is addressed using structured association.

3.2.1 Association analysis for FD

The total numbers of the significantly associated SNP markers with flowering date trait in both irrigation conditions among years vary. This can be due to the different stress conditions that our barley cultivars were exposed to (Rizza *et al.*, 2004). For example, different QTLs may be detected under dry vs. well-watered conditions. On the other hand, stability of some QTLs across years provides support for their significance in controlling the trait under specific conditions. One such example is represented by QTL.FD.13 that was identified in two successive years (1999 and 2000) under rainfed conditions but not under controlled irrigation, suggesting a role for this locus in response to limited water availability. This genomic region reported to co-localize with a cluster of genes encoding C-repeat binding factors (CBFs) affecting freezing tolerance and winter hardiness of the temperate-climate cereals (Knox *et al.*, 2010). This can give us a glimpse for the importance of this QTL under stress conditions. The utilization of this information can be considered in breeding schemes for instance, for arid areas.

QTLs are recovered across different watering regimes. For example, the recovery of QTL.FD.2 under both water regimes suggests this locus plays a general role in flowering regardless of water availability. This is supported by co-localization with the *Mat-a* gene that has a significant role in the early flowering, which been used in breeding programs for conferring photoperiod insensitivity (Zakhrabekova *et al.*, 2012). Wang *et al* (2010) reported that the *VRN-H3* gene on chromosome 7H is associated with flowering time QTLs; this gene is known for its role in flowering in barley, as an integrator of the vernalization pathways in temperate cereals (Yan *et al.*, 2006). This result is supported with our novel finding of QTL.FD.18 on the same chromosome. In a recent study by Ponce-Molina *et al* (2012) the same gene was localized in the nearby genomic region in the doubled haploid wild barley population (SBCC145 × Beatrix).

Findings of Haseneyer *et al* (2010) in the genetic variation for adaptive traits determines the ability of a barley species to conform to diverse environments. Early flowering, for instance, is an advantage in regions where the summers are hot and dry (e.g. in West Asia and North Africa) because the plants can complete their life cycle before they are exposed to severe drought. In Central Europe, where summer seasons are comparatively cool and humid, late flowering is an advantage because the longer growing period allows the crops to produce higher yields (Hershey 2005). This broad range of eco-climatic variation may explain the large variation in flowering time.

In summary, recovery of QTLs and genes previously associated with flowering date is supporting our idea of utilizing this trait as verification for our panel and indicates that despite the small size of the panel under study, valid marker-traits associations can be identified.

3.2.2 Association analysis for PH

Under the controlled irrigation condition, we recovered 5 QTLs corresponding with previously mapped genes or QTLs for PH. For instance, our recovered QTL.PH.10 on chromosome 2H coincides with the mapping position of *sdw3* gene which is known to play a major role in gibberellins-insensitive dwarfing in barley (Gottwald *et al.,* 2004). This result also is confirming the identified QTL by Pasam *et al* (2012): QTL4_PHT on the same chromosome. In addition, four statistically significant QTLs harboring several loci were identified associated with this trait (PH): on chromosome 2H, QTL.PH.6 involves 2 SNP markers, QTL.PH.9 harbors 7 SNPs, and QTL.PH.14 includes 8 SNPs, while QTL.PH.17 on chromosome 3H incorporates 5 SNPs. They are confirming previous findings of similar work carried out by many groups (Pasam *et al.,* 2012; Marquez-Cedillo *et al.,* 2001; Hayes *et al.*,1993).

Association analysis pf PH under rainfed condition recovered interesting QTLs findings. Such as QTL.PH.6 and QTL.PH.8 (harboring 3 SNPs) on chromosome 2H which are in homogeny of recently identified QTL loci QTL1_PHT and QTL2_PHT, respectively recovered by Pasam *et al.*, 2012. Interestingly, we recovered QTL.PH.23 (7 SNPs within ± 5 cM intervals on chromosome 3H) potentially harboring *sdw1* gene (Jia *et al.,* 2001). Importantly, *sdw1* has been deployed in America and Australia, while it has an allelic form (*denso*), which is regularly seen in European two-rowed germplasm (Pasam *et al.,* 2012). Notably, the breeding for dwarf and semi-dwarf cultivars have been developed worldwide to reduce lodging further more to improve the harvest index (Pasam *et al.,* 2012). Other QTLs are consistent with loci identified by Pasam *et al* (2012), while the remaining 45 QTLs we identified in both irrigation conditions separetly could be potentially novel. These QTLs are distributed over 5 chromosomes: 1H, 4H, 5H, 6H, and 7H. Significantly, QTL.PH.1 on chromosome 1H was detected in the three years successively, under the controlled irrigation system, and appeared once under rainfed condition in the year 2000. Notably, the climate condition in 2000 was generally favorable with relatively high precipitation levels of rainfalls (Rizza *et al.,* 2004). QTL.PH.11 and QTL.PH.2 displayed once at the same year 2001 under the controlled

irrigation, and rainfed, respectively. These findings can be useful in breeding programs for plant height related traits under in a wide range of environments.

3.2.3 Association analysis for GY

In the association analysis for this trait we suggested a total of 28, 37 QTLs considering both controlled irrigation and rainfed conditions, respectively. Twenty one of these QTLs were shared between both conditions. However, 7 and 16 QTLs are specific to the controlled irrigation and rainfed conditions, respectively.

We revealed 20 potential novel QTLs which are stable across all irrigation conditions. QTL.GY.18 (5H) spans the chromosomal location of the *vrn1* gene, a promoter of heading induced by vernalization, and regulating the transition of the apex to the reproductive stage (Hemming *et al.*, 2009). Previous study by Ramsay *et al.* (2011) revealed that *int-c* is an orthologous of the maize domestication gene TEOSINTE BRANCHED 1 and acts as a modifier of lateral spikelet fertility in barley. However, further studies are required to achieve higher resolution of our QTL and validate if it really corresponds to the *VRN-H1* gene.

Besides confirming some QTLs previously identified by Pasam et al. (2012), potential novel QTLs were identified under the controlled irrigation condition of analysis. Remarkably, QTL.GY.15 (harboring 8 significant associated SNP markers under this condition within ± 5 cM intervals on chromosome 4H), is co-localized to genomic regions where a previously identified gene is located *intermedium spike-c* (*int-c*) (Ramsay *et al.*, 2011). While fertility of lateral spikelet is controlled principally by the alleles at the *vrs1* locus on chromosome 2HL, it is modified by alleles at the *int-c* locus on chromosome 4HS. Loss of function of *vrs1* has occurred independently during barley domestication and has resulted in the full alteration of the sterile laterals into fully developed fertile spikelets (Komatsuda *et al.*, 2007). Furthermore, natural quantitative variation in fertility of the lateral spikelets has been noticed in progenies of two- by six-rowed crosses (Lundqvist and Lundqvist, 1989) and indicated this is due to the effect of *INT-C* alleles (Lundqvist *et al.*, 1997).

Another gene know to affect spikelet fertility is s*ix–rowed spike 3* (*vrs3*), a recessive gene located on the long arm of chromosome 1H (Lundqvist *et al.*, 1997), possibly colocalizing with QTL.GY.1. Considering QTL.GY.1 displayed once in year 2000 under the controlled irrigation condition, the association of this QTL with GY trait needs to be confirmed by further experiments.

Under the rainfed condition, several QTL locations are consistent with previously identified QTLs in various mapping populations (Table 24). For instance, QTL.GY.15 co-localizes with the recently identified QTL12_TGW (thousand grains weight) by Pasam *et al* (2012). In addition Comadran *et al.*, 2011 reported this locus and revealed SNP 11_20680 which is located in chromosome 4H and associated with small grains. Such a QTL appearance is in a region spanning of (*INTERMEDIUM-C*) *int-c* gene.

In addition, our analysis for this condition uncovered a total of 32 potential novel QTLs. The growing genomics resources available for barley and the Triticeae (Feuillet *et al.*, 2012, The International Barley Genome Sequencing Consortium 2012) will facilitate further confirmation and dissection of our novel QTLs. An advantage of the SNP platforms developed for barley is that markers were largely developed from transcribed gene SNPs (Close *et al.*, 2009, Comadran et al. 2012) allowing for efficient interspecies comparisons by sequence homology and synteny analyses, between barley and related grass genomes including rice (Dubcovsky *et al.*, 2001; Bennetzen and Ma 2003).

In conclusion, genomic regions recovered in our analyses revealed numerous QTLs (Tables, 22, 24). Several were novel and some were in agreement with previously identified QTLs. The significant markers identified in our work can be further verified for their potential importance for breeding for grain yield traits in various environmental conditions.

In early stages, plant breeders dealt with drought stress in crops through standard breeding practices such as field observations. The evolution to molecular breeding has allowed a deeper understanding of the interacting quantitative trait loci of the drought tolerance related complex traits and has exposed the underlying genetic variation (Holloway and Li 2010).

Accumulation of knowledge on quantitative trait loci has led to remarkable advances in breeding to develop traits of agronomic interest with the opportunity to dissect complex traits into component loci (Marza *et al.* 2006). Some QTLs affecting yields were identified under particular environmental conditions (Rodriguez et al 2007; Cattivelli *et al.*, 2008; Chenu *et al.*, 2011). However, the same yield QTLs are not usually found, even in the same population when examined in different environments. This may be as an effect of large genotype-by-environment interactions can be exhibited and consequently revealed in the QTL analysis. On the other hand, to achieve yield

stability, yield QTLs need to be stable across environmental conditions. Such loci have been identified for example in Durum wheat (Maccaferri *et al.*, 2008) two major QTLs affected grain yield showed significant effects in 16 field trials over 2 years. Such major QTLs on chrs. 2BL and 3BS were consistent across all environments. These QTLs on the mentioned chromosomes: QTL on chr. 2BL has not been previously described in wheat, while the one on chr. 3BS confirms the importance of this genomic region. Indeed these QTLs are being further investigated by the group of Roberto Tuberosa at University of Bologna. Comparatively, our identified QTL.GY.18 was stable across all environments where it spans *vrn1* gene.

Last month, an integrated and ordered physical, genetic and functional sequence resource that describes the barley gene-space in a structured whole-genome context has been made available to the community providing a powerful platform for trait dissection and breeding (http://barleygenome.org, The International Barley Genome Sequencing Consortium 2012). Although we were not able to take advantage of this resource in our project, but a promising perspectives are going to be in the way for barley breeding.

CHAPTER 6

CONCLUSIONS

CONCLUSIONS AND OUTLOOK

We explored the genetic regions that are associated with some agronomic traits such as flowering date, plant height, and grain yield in barley by exploiting a cultivar collection previously phenotyped for yield performance under controlled irrigation and rainfed conditions along three successive years (Rizza *et al*., 2004).

We initially used AFLP markers for assessing the population structure of our barley panel. This method has revealed atypical structure that cannot be clearly reconciled with the classical subdivisions reported in the literature e.g. winter/spring and 6/2-rows groups (Cockram *et al*., 2008). In comparison the use of SNP markers technology has facilitated the identification of our population structure to overcome the problem: we utilized a subset of 260 SNP markers covering the whole genome with an inter-marker distance of less than 10 cM. We performed the analysis using Structure and PCoA. Both analyses were able to recover the classical known partitioning of our barley groups (spring/winter and 2/6 rows). Genotyping with an iSELECT Infinium® Illumina 9K SNP panel and elimination of monomorphic and failed markers resulted in the identification of a total of 4,661 SNPs distributed over the whole genome. Several markers have been significantly associated with the different agronomic traits under the different water regimes and many of them were localized near some genes known to be highly involved in the genetic control of the studied trait.

Careful consideration of significant SNP clusters allowed us to identify several QTLs associated with the different agronomic traits of interest. All significantly associated SNP markers were considered above the Bonferroni correction ($-\log(P\text{-value}) = 1.9 \times 10^{-4}$). In addition, performance of General linear model was compared with the mixed linear model for the flowering date trait. Although few significantly associated markers were recovered from the mixed model analysis, all of them were included in the general model. The general linear model was utilized for the rest of association analyses.

We identified a total of 20, 50, and 44 potential QTLs associated to flowering date, plant height, and grain yield traits, respectively under either irrigation conditions. Several remarkable QTLs showed consistency with previously mapped loci for the respective trait under the study. For example, flowering date loci QTL.FD.4, QTL.FD.13 and QTL.FD.20 harbored significantly

associated markers potentially co-localizing with *PPD-H1*, *FR-H2* and *HvCO1* genes. QTL.PH.10 and QTL.PH.23 were potentially associated with *sdw3* and *sdw1* genes, respectively, that are reported in previous studies to play a role in gibberellin-insensitive dwarfing in barley. In addition, we identified QTLs underpinning grain yield trait: QTL.GY.1, QTL.GY.15 and QTL.GY.18 co-localized with genomic regions where genes *vrs3*, *int-c*, and *vrn1* were previously identified, respectively.

Despite the small panel and scale of work carried out in identifying QTLs in our study. We recovered relatively small power in seizing up and confirming our findings. Therefore, supporting our results with more investigations would solidify these findings and confirm them. Indeed with the emergence of barley new maps this will result in high quality QTLs linked to genes associated to different agronomic traits.

As genotyping and sequencing costs continue to decrease, GWA studies will become a standard tool for dissecting natural variation. Ultimately, genes responsible for the major QTLs need to be identified to gain insight into the molecular mechanisms underlying agronomic traits. Recent and future progress with Triticeae genomic resources will open unprecedented opportunities for discovery and functional analysis of novel genes (Brenchley *et al.*, 2012; Feuillet *et al.*, 2012; Mayer *et al.*, 2012).

Major challenges still are facing accurate phenotyping and high-throughput approach (Tuberosa 2010). The real issue is how to translate the flood of phenotypic and molecular data into improved cultivars for the livelihood of farmers (Collins *et al.*, 2008; Reynolds and Tuberosa 2008). Satisfying these challenges will only be possible through a multidisciplinary effort.

References

Abdel-Ghani H., Parzies K., Ceccarelli S., Grando S., Geiger H. (2005) Estimation of quantitative genetic parameters for outcrossing-related traits in barley. *Crop Science* 45: 98-105.

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

Abeledo G., Calderini D., Slafer G. (2002) Physiological changes associated with genetic improvement of grain yield in barley. In: Slafer G, Molina Cano J, Savin R, Aruas J, Ramagosa I (eds) *Barley science: recent advances from molecular biology to agronomy of yield and quality*. Food Products Press, New York, pp 361–385.

Adam A. (1961) Linkage between G-6-P-D deficiency and colour-blindness. *Proc. 2nd Inter. Cong. Hum. Genet*. Rome, Italy, pp. 565–567.

Agarwal K., Agarwal P., Reddy K., Sopory K. (2006) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Reports.* 25:1263-74.

Aghnoum R., Marcel C., Johrde A., Pecchioni N., Schweizer P., Niks E. (2010) Basal resistance of barley to barley powdery mildew: connecting QTLs and candidate genes. *Mol. Plant Microbe Interact*.23: 91–102.

Ahmadian A., Gharizadeh B., Gustafsson C., Sterky F., Nyren P., Uhlen M., Lundeberg J. (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem*, 280:103-110.

Amos W., Driscoll E., Hoffman I. (2010). Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proc. R. Soc. B.* doi: 10.1098/rspb.2010.1920.

Anderson-Taylor, G., Marshall, C. (1983). Root-tiller interrelationships in spring barley (*Horedeum distichum* (L.) Lam.). *Annals of Botany.* 51: 47-58.

Aranzana J., Abbassi K., Howad W., and Arús P. (2010). Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genet*., 11:69.

Aranzana J., De Vicente, M. C., Arús P. (2001) Comparison of fruit and leaf DNA extracts for AFLP analysis in Peach (*Prunus persica* (L.) Batsch). *Acta Hortic*. 546: 297-300.

Arnau G., Lallemand J., Bourgoin M. (2002) Fast and reliable strawberry cultivar identification using inter simple sequence repeat (ISSR) amplification. *Euphytica*. 129: 69-79.

Arnon I. (1972) Crop Production in Dry Regions. Leonard Hill, London.

Arrigo *et al.* (2009). Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinformatics*. 10:33 doi:10.1186/1471-2105-10-33.

Arthofer W (2010) TINYFLP and TINYCAT: software for automatic peak selection and scoring of AFLP data tables. *Mol. Ecol. Resources.* 10: 385–388.

Babb S., Muehlbauer GJ (2003). Genetic and morphological characterization of the barley uniculm2 (*cul2*) mutant. *Theor. Appl. Genet*. 106: 846–857.

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

Balding J. D. (2006) A tutorial on statistical methods for population association studies. *Genetics*. 7:781-791.

Baum M, Grando S, Backes G, Jahoor A, Sabbagh A, Ceccarelli S (2003). QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' *H. spontaneum* 41-1. *Theor Appl Genet* 107:1215–1225.

Baum M, Korff MV, Guo P, Lakew B, Hamwieh A, Lababidi S, Udupa SA, Sayed H, Choumane W, Grando S, Ceccarelli S (2007) Molecular approaches and breeding strategies for drought tolerance in barley. *Genomics applications in crops*. 2: 51 – 79.

Benjak A., Konradi J., Blaich R. and Forneck A. (2006) Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.). *Vitis.* 45: 15–21.

Bennett M. D. and Smith J. B. (1976) Nuclear DNA amounts in angiosperms. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 274:227-274

Bennett, J. H. (1965). Estimation of the frequencies of linked gene pairs in random mating populations. *Amer. J. Hum. Genet*. 17, 51–53.

Bennett, M.D., Finch, R.A., Smith, J.B., Rao, M.K. (1973). The time and duration of female meiosis in wheat, rye and barley. *Proceedings of the royal society of London series B: Biological Sciences* 183: 301-319.

Bennetzen JL, Ma J (2003). The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. *Curr. Opin. Plant. Biol.* 6:128–133.

Bernal A. et al. (2001). Genomes online database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Res*., 29 pp. 126–127.

Bhattramakki D., Rafalski A. (2001). Discovery and application of single nucleotide polymorphism markers in plants. In Henry RJ (ed). *Plant Genotyping: The DNA Fingerprinting of Plants*. Wallingford: CABI Publishing.

Boiteux, L. S.; Fonseca M. E. N.; Simon P. W.; (1999) Effects of plant tissue and DNA purification method on randomly amplified polymorphic DNA-based genetic fingerprinting analysis in carrot. *J. Am. Soc. Hortic. Sci*. 124, 32-38.

Bolle J. (2003) Climate, climate variability, and impacts in the Mediterranean Area: An overview. p. 5–86. In H.-J. Bolle (ed.) Mediterranean climate: Variability and trends. Springer, Berlin.

Bonin, A.; Bellemain, E.; Bronken Eidesen, P.; Pompanon, F.; Brochmann, C.; Taberlet, P. (2004) How to track and assess genotyping errors in population genetics studies. *Mol. Ecol*. 13:3261-3273.

Bothmer von R. and Komatsuda T. (2011). *Barley Origin and Related Species*. World Agriculture Series. Wiley-Blackwell. A John Wiley & Sons, Inc.

Bothmer von R., N. Jacobsen, C. Baden, R. B. Jørgensen & I. Linde-Laursen (1995). *An ecogeographical study of the genus Hordeum*, 2nd (ed). International Plant Genetic Resources Institute, Rome.

Botstein D. *et al.* (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet,* 32 pp. 314–331.

Bradbury P. J., Zhang Z., Kroon D. E., Casstevens T. M., Ramdoss Y. and Edward S. Buckler (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 23:2633-2635.

Bray E.A. (1993) Molecular responses to water deficit. *Plant Physiol* 103: 1035–1040

Brenchley *et al.,* (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*. 491: 705–710.

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

Briggs, D.E. (1978). *Barley*. Chapman and Hall Ltd, London.

Buckler E.S., Holland J.B., Bradbury P.J., Acharya C.B., Brown P.J., Browne C., Ersoz E., Flint-Garcia S., Garcia A., Glaubitz *et al*. (2009). The genetic architecture of maize flowering time. *Science*, 325(5941):714-718.

Buck-Sorlin G.H. (2002). The search for QTL in barley (*Hordeum vulgare* L.) using a new mapping population. *Cell Mol. Biol. Lett*. 7: 523–535.

Butts T., Handcock S., and Hunter R. (2008) network: Classes for Relational Data. R package version 1.4-1. Irvine, California.

Cakir M, Poulsen D, Galwey NW, Ablett GA, Chalmers KJ, Platz GJ, Park RF, Lance RCM, Panozzo JF, Read BJ, Moody DB, Barr AR, Johnston P, Li CD, Boyd WJR, Grime CR, Appels R, Jones MGK, langridge P (2003). Mapping and QTL analysis of the barley population Tallon × Kaputar. *Australian J. Agric. Res*. 54: 1155-1162.

Caldwell KS, Russell J, Langridge P, Powell W (2006). Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172:557–567.

Campoli C., Shtaya M., Davis J., von Korff M. (2012) Expression conservation within the circadian clock of a monocot: natural variation at barley *Ppd-H1* affects circadian expression of flowering time genes, but not clock orthologs. *BMC Plant. Biol.* 21: 12:97.

Casao C., Igartua E., Karsai I., Lasa M., Gracia P., Casas M. (2011b) Expression analysis of vernalization and day-length response genes in barley (*Hordeum vulgare* L.) indicates that VRNH2 is a repressor of PPDH2 (*HvFT3*) under long days. *J. Exp. Bot.* 62: 1939-49.

Casao C., Karsai I., Igartua E., Gracia P., Veisz O., Casas M. (2011a) Adaptation of barley to mild winters: a role for PPDH2. *BMC Plant Biol.* 11:164.

Cattivelli L, Rizza F, Badeck F, Mazzucotelli E, Mastrangelo A, Francia E, Mare C, Tondelli A, Stanca A. (2008) Drought tolerance improvement in crop plants: An integrated view from breeding to genomics (Review). *Field Crops Res.* 105: 1-14.

Chakravarti, A., Buetow, K. H., Antonarakis, S. E.,Waber, P. G., Boehm, C. D., and Kazazian, H. H. (1984). Nonuniform recombination within the human beta-globin gene cluster. *Am. J. Hum. Genet*. 36, 1239–1259.

Chaudhary, H.R., Jana, S., Acharya, S.N. (1980). Outcrossing rates in barley populations in the Canadian prairies. *Canadian Journal of Genetics and Cytology* 22: 353-360.

Chen ,G. X ,T. Krugman, T. Fahima, A.B. Korol, and E. Nevo (2002). Comparative study on morphological and physiological traits related to drought resistance between xeric and mesic *Hordeum spontaneum* lines in Israel. *Barley Genet. Newsl*. 32:22–33.

Chen A, Baumann U, Fincher GB, Collins NC (2009). Flt-2L, a locus in barley controlling flowering time, spike density, and plant height. *Funct Integr Genom* 9:243–254.

Chen G, Krugman T, Fahima T, Chen K, Hu Y, Röder M, Nevo E, Korol A. (2010) Chromosomal regions controlling seedling drought resistance in Israeli wild barley, *Hordeum spontaneum* C. Koch.*Genet. Resour. Crop. Evol.* 57: 85–99.

Chenu K, Cooper M, Hammer KL, Dreccer MF, Chapman, SC. (2011) Environment characterization as an aid to wheat improvement: interpreting genotype – environment interactions by modeling water-deficit patterns in North-Eastern. *Austral. J. Experim. Bot.* 62: 1743-1755.

Close T., Kortt A. and Chandler M. (1989) A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. *Plant Mol. Biol*. 13:95-108.

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

Collins *et al.* (2008) QTL approaches for improving crop performance under abiotic stress conditions: Where do we stand? *Plant Physiology*. 147: 469-486.

Collins FS, Guyer MS, Charkravarti A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1580–81.

Collins, A. and Morton, N. E. (1998). Mapping a disease locus by allelic association. *Proc. Natl. Acad. Sci. USA* 95, 1741–1745.

Collins, M. et al. (2010) Variability in fine structures of noncellulosic cell wall polysaccharides from cereal grains: potential importance in human health and nutrition. *Cereal Chem.* 87, 272–282.

Comadran J, Russell JR, Booth A, Pswarayi A, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, van Eeuwijk FA, Thomas WT, Romagosa I. (2011) Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in Hordeum vulgare in Mediterranean environments. *Theor. Appl. Genet.* 122:1363-1373.

Comadran J., Kilian B., Russell J., Ramsay L., Stein N., Ganal M., Shaw P., Bayer M., Thomas W., Marshall D., Hedley P., Tondelli A., Pecchioni N., Francia E., Korzun V., Walther A., Waugh R. (2012) A homolog of Antirrhinum CENTRODIALIS contributed to spring growth habit and environmental adaptation in barley. *Nature Genetics*, doi:10.1038/ng.2447.

Comai *et al.* (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J*. 37:778–786.

Corpet F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.*, 16:10881-10890.

Cuesta-Marcos A, Casas A, Hayes PM, Gracia MP, Lasa JM et al (2009). Yield QTL affected by heading date in Mediterranean grown barley. *Plant Breed* 128:46–53.

Cuesta-Marcos A., Szűcs P., Close T. J., Filichkin T, Muehlbauer G. J., Smith K. P., Hayes P. M. (2010). Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure. *BMC Genomics*, 11:707

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., and Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nat. Genet*. 29, 229–232.

Devlin, B. and Reisch, N. (2000). A comparison of linkage disequilibrium measures for fine scale mapping. *Genomics* 29, 311–322.

Dimitrijević M, Knežević D, Petrović S and Zečević V (2002). Variability and stability of harvest index in wheat (*Triticum aestivum* L.) *Kragujevac J. Sci.* 24 91-96.

Druka, A. Muehlbauer, I. Druka, R. Caldo, U. Baumann, N. Rostoks , A. Schreiber, R. Wise, T. Close, A. Kleinhofs, A. Graner, A. Schulman ,P. Langridge, K. Sato , P. Hayes, J. McNicol, D. Marshall, and R. Waugh. (2006) An atlas of gene expression from seed to seed through barley development. *Funct. Int. Genom.* 6: 202-211.

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001). Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiol* 125:1342–1353.

Earl, Dent A. and vonHoldt, Bridgett M. (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* DOI: 10.1007/s12686-011-9548-7.

Eastham, K. and Sweet, J. (2002). Genetically modified organisms (GMOs): The significance of gene flow through pollen transfer. Environmental issue No. 28, European Environment Agency (EEA), Copenhagen, Denmark.

Elberse, I.A.M., van Damme J.M.M., and van Tienderen P.H. (2003) Plasticity of growth characteristics in wild barley (*Hordeum spontaneum*) in response to nutrient limitation. *J. Ecol*. 91: 371-382.

Ellestrand, N.C. (2003). Dangerous Liaisons? When cultivated plants mate with their wild relatives. *The John Hopkins University Press*, Baltimore. pp 1-244.

Ersoz ES, Yu J, Buckler ES (2009) Applications of linkage disequilibrium and association mapping in maize Molecular Genetic Approaches to Maize Improvement, Springer Berlin Heidelberg, pp 173–195.

Falconer DS, Mackay TFC (1996*) Introduction to quantitative genetics*. Longman. Essex. 464 p.

Falush D, Stephens M, and Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567-1587.

Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.

Ferrandiz C., Q. Gu, R. Martienssen, M.F. Yanofsky (2000). Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development*. 127. 725–734.

Feuillet C. & N. Stein & L. Rossini & S. Praud & K. Mayer & A. Schulman & K. Eversole & R. Appels (2012). Integrating cereal genomics to support innovation in the Triticeae. *Funct Integr Genomics* 12(4): 573–583.

Feuillet C. and Keller B. (2002) Comparative Genomics in the Grass Family: Molecular Characterization of Grass Genome Structure and Evolution. *Ann. Bot.* 89: 3-10.

Feuillet C., Muehlbauer J. (2009) (eds.), Genetics and Genomics of the Triticeae, Plant Genetics and Genomics: *Crops and Models* 7, DOI 10.1007/978-0-387-77489-3_16.

Ford, E. B. (1940) Polymorphism and taxonomy, In: Huxley, J. (ed.), *the new systematics*. Clarendon Press, Oxford, UK, pp. 493-513.

Forster B., Franckowiak J., Lundqvist U., Lyon J., Pitkethly I. and Thomas W. (2007) The Barley Phytomer. *Ann. Bot.* 100: 725-733.

Forster BP, Ellis RP, Newton AC, Morris WL, Moir J, Lyon J, Keith R, Tuberosa R, Talame V, This D, Teulat B, El-Enein RA, Bahri H, Ben Salem M (2000) Stable yield in Mediterranean barley: application of molecular technologies in improving drought tolerance and mildew resistance. In: *Proceedings of the 8th international barley genetics symposium*, Adelaide, pp 273–274.

Francia E., Barabaschi D., Tondelli A., Laidò G., Rizza F., Stanca M., Busconi M., Fogher C., Stockinger J., and Pecchioni N. (2007) Fine mapping of a HvCBF gene cluster at the frost resistance locus *Fr-H2* in barley. *Theor. Appl. Genet.* 115:1083-91.

Franckowiak J.D., (2000) Notes on plant height in six-rowed barley and FHB resistance. In: Barley genetics VIII: *Proceedings of the Eighth International Barley Genetics Symposium*. pp 107–109.

Franckowiak JD, Lundqvist U (2002). Current list of new and revised barley genetic stock descriptions. *Barley Genet. Newslett*. pp.32- 47.

Garg K, Green P, Nickerson DA (1999). Identification of candidate coding region single nucleotide polymorphisms (cSNPs) using assembled expressed sequence tags (ESTs). *Genome Res*, 9:1087-1092.

Ghosh S., Majumder B. and Sen Mandi S. (2011) Species-specific AFLP markers for identification of *Zingiber officinale*, *Z. montanum* and *Z. zerumbet* (Zingiberaceae). *Genet. Mol. Res.* 10: 218-229.

Giles, R.J. (1989). The frequency of natural cross-fertilisation in sequential sowings of winter barley. *Euphytica* 43:125-134.

Gomez-Macpherson, H. (2001). *Hordeum vulgare*.

Gottwald S, Stein N, Borner A, Sasaki T, Graner A (2004). The gibberellic-acid insensitive dwarfing gene *sdw3* of barley is located on chromosome 2HS in a region that shows high colinearity with rice chromosome 7L. *Mol Genet Genomics*, 271(4):426-436.

Gottwald S., Bauer P., Komatsuda T., Lundqvist U., and Stein N. (2009) TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene *HvHox1*. *BMC Res. Notes.* 2:258.

Grando ,S.and H.G. Macpherson (2005). Food barley: importance, uses and local knowledge . *Proc. Int. Workshop on Food Barley Improvement*, Hammamet, Tunisia, January 14 – 17, 2002. ICARDA, Aleppo, Syria.

Graner A., Kilian A., and Kleinhofs A. (2011) Barley Origin and Related Species. 1st (ed) in Production, improvement, and uses. World Agriculture Series. Wiley-Blackwell. A John Wiley & Sons, Inc., Publication, ISBN 978-0-8138-0123-0.

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003). The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. *Plant Physiol*, 131(4):1855-1867.

Gujaria N., Kumar A., Dauthal P., Dubey A., Hiremath P., Prakash B., Farmer A., Bhide M., Shah T., Gaur P., Upadhyaya H., Bhatia S., Cook D., May G., and Varshney R. (2011) Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 122: 1577–1589.

Gull M., Hartings H., Raho G., Devos K., Marmiroli N., Maestri E. and Perrotta C. (1995) Isolation and characterization of abscisic acid inducible genes in barley seedlings and their responsiveness to environmental stress. *Life Science Advances-Plant Physiology*. 14: 89-96.

Gupta, S.K., Singh, D., Sharma, S.C. (2000). Genetic Variability for allogamic traits in barley (*Hordeum vulgare*). *Agricultural Science Digest* 20: 1-4.

Gut I.G. (2001). Automation in genotyping single nucleotide polymorphisms. *Hum Mutat*, 17:475-492.

Haake V., Cook D., Riechmann J.L., Pineda O., Thomashow M.F., Zhang J.Z., (2002) Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. *Plant. Physiol.* 130: 639-648.

Hall BG, Salipante SJ (2007) Measures of clade confidence do not correlate with accuracy of phylogenetic trees. *PLoS Comput Biol* 3(3).

Hall D., Tegstrom C., Ingvarsson K. (2010) Using association mapping to dissect the genetic basis of complex traits in plants. *Brief. Funct. Genomic*; 9:157–65.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser*. 41:95-98.

Hamblin M.T., Close TJ, Bhat PR, Chao SM, Kling JG, Abraham KJ, Blake T, Brooks WS, Cooper B, Griffey CA, et al (2010). Population structure and linkage disequilibrium in us barley germplasm: implications for association mapping. *Crop Sci*, 50(2):556-566.

Hammer O., Harper T., and Ryan R. (2001) PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. 4: 1-9, 2001.

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

Hayes P.M., Castro A., Marquez-Cedillo L., Corey A., Henson C., Jones B.L., Kling J., Mather D., Matus I., Rossi C., and Sato K. (2003). Genetic diversity for quantitatively inherited agronomic and malting quality traits. In: R. Von Bothmer, H. Knupffer, T. van Hintum, and K. Sato (ed.). *Diversity in Barley*. Elsevier Science Publishers, Amsterdam.

Hayes PM, Liu BH, Knapp SJ, Chen F, Jones B, Blake T, Franckowiak J, Rasmusson D, Sorrells M, Ullrich SE, *et al.* (1993). Quantitative Trait Locus Effects and Environmental Interaction in a Sample of North-American Barley Germplasm. *Theoretical and Applied Genetics*, 87(3):392-401.

Hellewell K.B., Rasmusson D.C., Meagher M.G. (2000) Enhancing yield of semidwarf barley. *Crop Sci.* 40: 352–358.

Hemming N., Peacock J., Dennis S., Trevaskis B. (2008) Low-temperature and daylength cues are integrated to regulate FLOWERING LOCUS T in barley. *Plant Physiol.* 147: 355-366.

Hena G. and Moose S. P. (2003). Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. *Plant Cell.* 15(5): 1143–1158.

Henry RJ (Ed) (2001). *Plant Genotyping. The DNA Fingerprinting of Plants*. Wallingford: CABI Publishing.

Herrmann D., Poncet N., Manel S., Rioux D., Gielly L., Taberlet P., and Gugerli F. (2010) Selection criteria for scoring amplified fragment length polymorphisms (AFLPs) positively affect the reliability of population genetic parameter estimates. *Genome: Ecological Genetics and Evolution.* 53: 302-310.

Hill,W. G. (1974). Estimation of linkage disequilibrium in randomly mating populations. *Heredity* 33, 229–239.

Hill,W. G. and Robertson A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet*. 38, 226–231.

Holly K. Tabor, Neil J. Risch and Richard M. Myers (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. 3(5):391-7.

Hongyun Wang, K P. Smith, E C T Blake, R D. Horsley, G J. Muehlbauer (2011). Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theor Appl Genet* 1691-8.

Huang, Q., A. Beharav, Y. Li, V. Kirzhner, and E. Nevo (2002). Mosaic microecological differential stress causes adaptive microsatellite divergence in wild barley, *Hordeum spontaneum,* at Neve Yaar, Israel. *Genome.* 45:1216–1229.

Ingram J, Bartels D (1996). The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Physiol* 47: 377–403

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.

Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster BP (2003) Association of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. *Plant breeding* 122: 300-304.

Iwata H, Uga Y, Yoshioka Y (2007) Bayesian association mapping of multiple quantitative traits loci and its application to the analysis of genetic variation among Oryza sativa L. germplasms. *Theor Appl Genet.* 114: 1437-1449.

Jaccard P. (1908): Nouvelle recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* 44: 223-270.

Jaccoud D., Peng K., Feinstein D., Kilian A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acid research*. 29(4):E25

Jacobsen T, Adams RM (1958) Distribution of wild wheats and barley. *Science*. 128: 1251–1258.

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

Jia Q.J, Zhang X.Q, Westcott S., Broughton S., Cakir M., Yang J.M., Lance R., Li C.D. (2011). Expression level of a gibberellin 20-oxidase gene is associated with multiple agronomic and quality traits in barley. *Theoretical and Applied Genetics*, 122(8):1451-1460.

Kerem, B., Rommens, J. M., Buchanan, J. A., et al. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science* 245, 1073–1080.

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

Kimura, M. and Ohta, T. (1973) Mutation and evolution at the molecular level. *Genetics Supplement* 73, 19–135.

Kirby, E.J.M. and M. Appleyard. (1987). *Cereal development guide*, 2nd Edition. Arable Unit, Stoneleigh, Warwickshire.

Kling, J. (2004). *An Introduction to Barley*. Notes from CSS 330 World Foods Class

Knox K., Dhillon T., Cheng H., Tondelli A., Pecchioni N., Stockinger J. (2010) CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theor. Appl. Genet.* 121: 21-35.

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

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

Kück P, Greve C, Misof B, Gimnich F (2012) Automated Masking of AFLP Markers Improves Reliability of Phylogenetic Analyses. *PLoS ONE* 7(11): e49119. doi:10.1371/journal.pone.0049119.

Kuczyńska A., Arkadiusz Kosmala, Maria Surma, and Tadeusz Adamski (2012) Identification of Tillering Node Proteins Differentially Accumulated in Barley Recombinant Inbred Lines with Different Juvenile Growth Habits. *Int. J. Mol. Sci.* 13: 10410–10423.

Kurowska M., Daszkowska-Golec A., Gruszka D., Marzec M., Szurman M., Szarejko I., Maluszynski M. (2011) TILLING - a shortcut in functional genomics. *J. Appl. Genet.* 52: 371–390.

Laird NM, Lange C. (2006) Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet.* 7: 385-94.

Laurie DA, Pratchett N, Snape JW, Bezant JH (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(3):575-585.

Lemieux B: Plant genotyping based on analysis of single nucleotide polymorphisms using microarrays (2001) In: Henry RJ (ed) *Plant Genotyping: The DNA Fingerprinting of Plants*. Edited by. Wallingford: CABI Publishing: 47-57.

Leonard , K.J. and L.J. Szabo (2005) Stem rust of small grains and grasses caused by *Puccinia graminis* . *Mol. Plant Pathol.* 6: 99 –111.

Levitt J. (1972) *Response of plants to environmental stress*. Academic Press, New York, NY.

Lewontin, R. (1988). On measures of gametic disequilibrium. *Genetics* 120, 849–852.

Li JZ, Huang XQ, Heinrichs F, Ganal MW, Röder MS (2006). Analysis of QTLs for yield components, agronomic traits, and disease resistance in an advanced backcross population of spring barley. *Genome*, vol. 49, No. 5: 454 - 466.

Lindlöf A (2003). Gene identification through large-scale EST sequence processing. *Appl Bioinformatics*, 2:123-129.

Lindsay A., Crawford, Daria k., and Nusha K. (2012) A call for more transparent reporting of error rates: the quality of AFLP data in ecological and evolutionary research. *Molecular Ecology*. 21: 5911–5917.

Lorenz AJ, Hamblin MT, Jannink JL, Baxter I (2010). Performance of single nucleotide polymorphisms versus haplotypes for genomewide association analysis in barley. *PloS* 5.

Lundqvist, U. and Lundqvist, A. (1989) The co-operation between intermedium genes and the six-row gene hex-v in a six-row variety of barley. *Hereditas.* 110: 227–233.

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

Mackay C., Stone A., and Ayroles F. (2009) The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10: 565–77.

Mackay I, and Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* 12:57–63.

Malecot, G. (1948) *Les mathématiques de l'hérédité*. Mason et Cie, Paris, France.

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

Maniatis, N., Collins, A., Xu, C. F., *et al*. (2002). The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc. Natl. Acad. Sci. USA* 99, 2228–2233.

Marquez-Cedillo LA, Hayes PM, Kleinhofs A, Legge WG, Rossnagel BG, Sato K, Ullrich SE, Wesenberg DM, Proj NABGM (2001). QTL analysis of agronomic traits in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. *Theoretical and Applied Genetics*, 103(4):625-637.

Matsumoto T., Tanaka T., Sakai H., Amano N., Kanamori H., Kurita K., Kikuta A, Kamiya K., Yamamoto M., Ikawa H., Fujii N., Hori K., Itoh T., Sato K. (2011) Comprehensive Sequence Analysis of 24,783 Barley Full-Length cDNAs Derived from 12 Clone Libraries. *Plant Physiol*. 156: 20–28.

McVean, G. A., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., and Donnelly, P. (2004) The fine scale structure of recombination rate variation in the human genome. *Science* 304, 581–584.

Mezaka I., Bleidere M., Legzdina L., and Rostoks N. (2011) Whole genome association mapping identifies naked grain locus *NUD* as determinant of β-glucan content in barley. *Žemdirbyste=Agriculture.* 98:283-292.

Mickelson S, See D, Fletcher D, Meyer FD, Garner JP, Foster CR, Blake TK, Fischer AM (2003). Mapping of QTL associated with nitrogen storage and remobilization in barley (*Hordeum vulgare* L.) leaves. *J. Exper. Botany*, Vol. 54, No. 383: 801- 812.

Miralles, D. J., and Slafer, G. A. (1995a). Yield, biomass and yield components in dwarf, semidwarf and tall isogenic liens of spring wheat under recommended and late sowing dates. *Plant Breeding* 114: 392–396.

Monica Rodriguez, Domenico Rau, Roberto Papa, Giovanna Attene (2007) Genotype by environment interactions in barley (*Hordeum vulgare* L.): different responses of l

andraces, recombinant inbred lines and varieties to Mediterranean environment. *Euphytica*. DOI 10.1007/s10681-007-9635-8.

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2006) *Genetics.* 173: 1705–1723.

Morton, N. E., Zhang,W., Taillon-Miller, P., Ennis, S., Kwok, P. Y., and Collins, A. (2001). The optimal measure of allelic association. *Proc. Natl. Acad. Sci. USA* 98, 5217–5221.

Muñoz-Amatriaín M., Moscou M., Bhat P., Svensson J., Bartoš J., Suchánková P., Šimková H., Endo T., Fenton R., Lonardi S., Castillo A., Chao S., Cistué L., Cuesta-Marcos A., Forrest K., Hayden M., Hayes P., Horsley R., Makoto K., Moody D., Sato K., Vallés M., Wulff B., Muehlbauer J., Doležel J., and Close T. (2011) An Improved Consensus Linkage Map of Barley Based on Flow-Sorted Chromosomes and Single Nucleotide Polymorphism Markers. *The Plant Genome*. 4:238–249.

Neffar F., Mare C., Hamenna B. (2011). Study of Barley (*Hordeum vulgare* L) transcript profiles under water stress conditions. *Australian Journal of Basic and Applied Sciences*, 5(12): 1481-1488.

Nelson, J.E., Kephart, K.D., Bauer, A., Connor, J.F. (2001). Growth Stages of Wheat, Barley, and wild Oat: A strategic step to Timing of Field Operations. http://weeds.montana.edu/crop

Nevo E (1992) Origin, evolution, population genetics and resources for breeding of wild barley, Hordeum spontaneum, in the Fertile Crescent. In: Shewry P, editor. Barley: genetics, molecular biology and biotechnology. Wallingford: *C.A.B. International*. pp. 19–43.

Newton, M., Peturson, B., and Meredith, W. O. S. (1945). The effect of leaf rust of barley on the yield and quality of barley varieties. *Can. J. Res., Sect. C*, 23, 212–218.

Oliver, S.N., Finnegan, E.J., Dennis, E.S., Peacock, W.J. & Trevaskis, B. (2009) Vernalization induced flowering in cereals is associated with changes in histone methylation at the VERNALIZATION1 gene. *Proc. Natl. Acad. Sci.* USA 106, 8386–8391.

Pakniyat, H., Kazemipour A., and Mohammadi A. (2003). Variation in salt tolerance of cultivated (Hordeum vulgare L.) and wild (*H. spontaneum* C. Koch) barley genotypes from Iran . *Iran Agric. Res*. 22:45 – 62.

Parkinson J, Blaxter M (2009). Expressed sequence tags: an overview. *Methods Mol Biol*, 533:1

Parzies, H.K., Schnaithmann, F., Geiger, H.H. (2005). Pollen viability of *Hordeum spp* genotypes with different flowering characteristics. *Euphytica* 145: 229-235.

Parzies, H.K., Spoor, W., Ennos, R.A. (2000). Outcrossing rates of barley landraces from Syria. *Plant Breeding* 119: 520-522.

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

Paun O. and Schönswetter P. (2012) Amplified fragment length polymorphism: an invaluable fingerprinting technique for genomic, transcriptomic, and epigenetic studies. *Methods Mol. Biol.* 862: 75-87.

Pennisi E. (2008) Plant genetics. The blue revolution, drop by drop, gene by gene. *Science*. 11; 320(5873):171-3.

Perrier X., Jacquemound-Collet P.: DARwin Software. http://darwin.cirad.fr/darwin 2006.

Peters J.L. et al. (2001). A physical amplified-fragment length polymorphism map of *Arabidopsis*. *Plant Physiol*., 127 pp. 1579–1589.

Peterson, P.D. (2001) Stem Rust of Wheat: from Ancient Enemy to Modern Foe . *American Phytopathological Society* , St. Paul, MN. *Plant Cell Tissue Organ Cult.* 84: 199–211.

Pflieger S., Lefebvre V. and Causse M. (2001). the candidate gene approach in plant genetics: a review. *Molecular Breeding* 7: 275–291.

Pillen K, Zacharias A, Leon J (2003). Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, 107(2):340-352.

Pillen K, Zacharias K, Léon J (2004) Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare* ssp. *spontaneum*. *Theor Appl Gene* 108(8):1591-601.

Ponce-Molina L. J., Casas A. M.,Gracia M B, Silvar C, Mansour E, William B.T, Schweizer T G, Herz M, and Igartua E (2012). Quantitative Trait Loci and candidate loci for heading date in a large population of a wide barley cross. *Crop Science*, vol. 52.

Pope, M.N. (1944). Some notes on technique in barley breeding. *The Journal of Heredity* 35: 99-111

Porter, I. H., Schulze, J., and McKusick, V. A. (1962.) Genetical linkage between the loci for glucose-6-phosphate dehydrogenase deficiency and colour-blindness in American Negroes. *Ann. Hum. Genet*. 26, 107–122.

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

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

Price and Casler (2012) Simple regression models as a threshold for selecting AFLP loci with reduced error rates. 13:268 doi:10.1186/1471-2105-13-268.

Pritchard J. K., Stephens M., and Donnelly P. (2000) Inference of Population *Structure* Using Multilocus Genotype Data. *Genetics*. *155: 945-959*.

Pritchard JK, Wen W (2007) Documentation for structure Software, version 2.2. Department of Human Genetics, *University of Chicago*, Chicago.

Qi X, Niks RE, Stam P, Lindhout P (1998). Identification of QTLs for partial resistance to leaf rust (*Puccinia hordei*) in barley. *Theoretical and Applied Genetics*, 96:1205-1215.

Rafalski A. (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*. 5:94-100.

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

Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WTB, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N, *et al.* (2011). INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. *Nat Genet*, 43(2):169-172.

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

Ren X., Sun D., Guan1 W., Sun G., and Li C. (2010) Inheritance and identification of molecular markers associated with a novel dwarfing gene in barley. *BMC Genetics*. 11: 89.

Reynolds M, Tuberosa R. (2008) Translational research impacting on crop productivity in drought-prone environments. *Current Opinion in Plant Biology*. 11: 171-179.

Richards, R. A. (1992). The effect of dwarfing genes in spring wheat in dry environments. I. Agronomic characteristics. *Aust. J. Agr. Res.* 43: 517–522.

Riddle, O.C., Suneson, C.A. (1944). Crossing studies with male-sterile barley. *Journal of the American society of* agronomy 36: 62-65.

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

Ritala, A., Nuutila, A.M., Aikasalo, R., Kauppinen, V., Tammisola, J. (2002).Measuring gene flow in the cultivation of transgenic barley. *Crop Science* 42: 278-285.

Robbins, R. B. (1918). Some applications of mathematics to breeding problems. III. *Genetics* 3, 375–389.

Roelfs, A.P., R. Singh , and E.E. Saari . (1992b) Rust Diseases of Wheat: Concepts and Methods of Disease Management. CIMMYT, Mexico, DF, Mexico.

Romagosa I, Han F, Ullrich SE, Hayes PM, and Wesenberg DM (1999). Verification of yield QTL through realized molecular marker-assisted selection responses in barley cross. *Molecular Breeding*, vol. 5, No. 2: 143 -152.

Rosenberg NA, et al. (2010) Genome-wide association studies in diverse populations. *Nat. Rev. Genet*. 11:356–366.

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

Rounsley *et al.* (2009) De Novo Next Generation Sequencing of Plant Genomes. *Rice.* 2: 35–43.

Royo, C., A´ lvaro, F., Martos, V., Ramdani, A., Isidro, J., Villegas, D., and Garc´ıa del Moral, L. F. (2007). Genetic changes in durum wheat yield components and associated traits in Italy and Spain during the 20th century. *Euphytica* 155: 259–270.

Rozen S. and Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL et al. (2001): A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 409:928-933.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL *et al.* (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409:928-933.

Saisho D., Ishii M., Hori K., and Sato K. (2011) Natural variation of barley vernalization requirements: implication of quantitative variation of winter growth habit as an adaptive trait in East Asia. *Plant Cell Physiol*. 52(5): 775–784.

Saitou N. and Nei M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 4: 406-425.

Salamini F., Ozkan H., Brandolini A., Schafer-Pregl R., and Martin W., (2002) Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*. vol. 3, no. 6, pp. 429–441.

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

Sato, K. and K. Takeda (2009) An application of high throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines. *Theor. Appl. Genet.* 119: 613–619.

Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, Sato K, Schulman AH, Waugh R, Wise RP, Stein N. (2009). The international barley sequencing consortium-at the threshold of efficient access to the barley genome. *Plant Physiol.* 149(1):142-7.

Schuster W. H. (1997) Welchen Beitrag leistet die Pflanzenzuchtung zur Leistungssteigerung von Kulturpflanzenarten. *Pflanzenbauwissenschaften*, vol. 1, pp. 9–18. sequence information independent genotyping. *Nucleic Acids Res.* 29:e25.

Schut J. W. á X. Qi á P. Stam (1997). Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. *Theor. Appl. Genet.* 95: 1161-1168.

Seki M, Umezawa T, Urano K, Shinozaki K (2007). Regulatory metabolic networks in drought stress responses. *Current Opinion in Plant Biology*, 10:296–302.

Seiler C., Harshavardhan V., Rajesh Kalladan, Reddy S., Strickert M., Rolletschek H., Scholz U., Wobus U. and Sreenivasulu N. (2011) ABA biosynthesis and degradation contributing to ABA homeostasis during barley seed development under control and terminal drought-stress conditions. *J. Exp. Bot.* doi:10.1093/jxb/erq446.

Shinozaki K. and Yamaguchi-Shinozaki K. (2007) Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany*, 58: 221–227.

Simko I. and Hu J. (2008) Population Structure in Cultivated Lettuce and Its Impact on Association Mapping. *J. Amer. Soc. Hort. Sci.* 133:61-68.

Sinascalco, M., Bernini, L., Latte, B., and Motulsky, A. G. (1961). Favism and thalassemia in Sardinia and their relationship to Malaria. *Nature* 190, 1179–1180.

Song R., Llaca V., Messing J. (2002) Mosaic Organization of Orthologous Sequences in Grass Genomes. *Genome Res.* 12: 1549–1555.

Sreenivasulu N., Graner A., and Wobus U. (2008) Barley Genomics: An overview. *International Journal of Plant Genomics.* doi:10.1155/2008/486258.

Steffenson BJ, Olivera P, Roy JK, Jin Y, Smith KP, Muehlbauer GJ. (2007) A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. *Australian Journal of Agricultural Research*. 58(6): 532–544.

Stephanie Pflieger, Veronique Lefebvre, and Mathilde Causse (2001). The candidate gene approach in plant genetics: a review. *Molecular Breeding*. 7: 275–291.

Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A, Piepho HP (2009). Association mapping reveals gene action and interactions in the determination of flowering time in barley. *Theor Appl Genet*, 118(2):259-273.

Stubbs, R.W. (1985) Stripe rust , pp. 61 – 101 . In A.P. Roelfs and W.R. Bushnell (eds.). The Cereal Rusts: Diseases, Distribution, Epidemiology, and Control , Vol. 2 . *Academic Press* , New York.

Sturtevant A.H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association *J. Exp. Zool.,* 14 pp. 43–59

Suprunova, T., T. Krugman,T. Fahima, G. Chen,I. Shams, A. Korol, and E. Nevo (2004). Differential expression of dehydrin genes in wild barley, *Hordeum spontaneum* associated with resistance to water deficit. *Plant Cell Env*. 27:1297–1308.

Sutka J, Galiba G, Vaguifalvi A, Gill BS, Snape JW (1999). Physical mapping of the Vrn-A1 and Fr1 genes on chromosome 5A of wheat using deletion lines. *Theor Appl Genet* 99:199–202.

Sved, J. A. (1970). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Pop. Biol*. 2, 125–141.

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

Szucs , P., Blake C., Bhat R., Close T., Cuesta A., Marcos , Muehlbauer J. , Ramsay L., Waugh R., and Hayes P. (2009) An integrated resource for barley linkage map and malting quality QTL alignment. *Plant Genome*. 2: 134–140.

Takahashi R, Yasuda S (1956). Genetic studies of spring and winter habit of growth in barley. *Bericht des Ohara Instituts für Landwirtschaftliche Biologie*, 10:245–308.

Takahashi, R. and S. Yasuda . (1971). Genetics of earliness and growth habit in barley. In: R.A. Nilan (ed.). *Barley Genetics II*. Proceedings of the Second International Barley Genetics Symposium . Washington State University Press , Pullman, WA.

Terwilliger, J. D. (1995). A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more marker loci. *Am. J. Hum. Genet*. 56, 777–787.

The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.

Thomas, Fukai, S. (1995). Growth and yield response of barley and chickpea to water stress under three environments in southeast Queensland. I. Light inception, crop growth and grain yield. *Australian Journal of Agricultural Research*. 46: 17-33.

Thomashow MF (1999). Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu Rev Plant Physiol* 50: 571–599.

Thornsberry M., Goodman M., Doebley J., Kresovich S., Nielsen D., Buckler S. (2001) *Dwarf*8 polymorphisms associate with variation in flowering time. *Nature Genetics*. 28:286-289.

Thornsberry, J.M. et al. (2001) Dwarf8 polymorphisms associated with variation in flowering time. *Nat. Genet.* 28, 286–289

Trevaskis B., D.J. Bagnall, M.H. Ellis, W.J. Peacock, E.S. Dennis (2003). MADS box genes control vernalization-induced flowering in cereals. *Proc Natl Acad Sci U S A*, 100 pp. 13099–13104.

Trevaskis B: The central role of the VERNALIZATION1 gene in the vernalization response of cereals. *Funct. Plant. Biol.* 2010, 37: 479-487.

Tuberosa R. (2010) Phenotyping drought-stressed crops: key concepts, issues and approaches. In: Monneveux, P., and Ribaut, J. M. (eds.) Drought Phenotyping in Crops: *From Theory to Practice. CIMMYT/Generation Challenge Programme*, Mexico, pp 3-35.

Tuberosa R. and Salvi S. (2006) *Trends in Plant Science*. 11: 405-412.

Turner A., Beales J., Faure S., Dunford P., and Laurie A. (2005) The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science*. 310 :1031-4.

Turner, A., J. Beales , S. Faure , R.P. Dunford , and D.A. Laurie . (2005). The pseudo - response regulator Ppd - H1 provides adaptation to photoperiod in barley. *Science* 310: 1031 – 1034.

Turuspekov, Y., Kawada, N., Honda, I., Watanabe, Y., Komatsuda, T. (2005). Identification and mapping of a QTL for rachis internode length associated with cleistogamy in barley. *Plant breeding* 124: 542-545.

Tuyen D. and Prasad D. (2008). Evaluating difference of yield trait among rice genotypes (*Oryza sativa* l.) under low moisture condition using candidate gene markers. *Omonrice* 16: 24-33.

Ullrich S. (2011). *Barley: Production, improvement, and uses*. World Agriculture Series. Wiley-Blackwell. A John Wiley & Sons, Inc., Publication, ISBN 978-0-8138-0123-0.

Varshney R., Tuberosa R. (2008) Genomics approaches and platforms. Published by Springer. ISBN 1402062966, 9781402062964. Pp. 61-62.

Varshney R.K., Paulo M.J., Grando S., van Eeuwijk F.A., Keizer L.C.P., Ceccarelli, Kilian A., aum M.B, Graner A., Guo P. (2012). Genome wide association analyses for drought tolerance related traits in barley (*Hordeum vulgare* L.). *Field Crops Research.* 126:171-180.

Vinod MS, N Sharma, K Manjunatha, A Kanbar, NB Prakash, HE Shashidhar (2006) Candidate genes for drought tolerance and improved productivity in rice (*Oryza sativa* L.). *J. Biosci*. 31(1): 69–74.

Vinod MS, Sharma N, Manjunatha K, Kanbar A, Prakash NB, Shashidhar HE (2006). Candidate genes for drought tolerance and improved productivity in rice (*Oryza sativa* L.). *J. Biosci.* 31(1): 69–74.

Vogel , J.P. , Gu Y.Q., Twigg P., Lazo G.R., Chingcuanco D.J., Hayden D.M., Donze T.J., Vivian L.A., Stamova B., and Coleman - Derr D. (2006b) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. *Theor. Appl. Genet*. 113: 186–195.

Von Bothmer R., Jacobsen N., Baden C., Jørgensen B., Linde-Laursen I. (1995) An Ecogeographical Study of the Genus Hordeum. 2nd Ed. Vol 7. Rome: *International Plant Genetic Resources Institute*. Systematic and Ecogeographic Studies on Crop Genepools.

Von Bothmer, R. (1992). The wild species of *Hordeum*: relationships and potential use for improvement of cultivated barley. Chapter 1. In: PR Shewry, ed. *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*. C.A.B International, Wallingford, Oxon.pp 3-18.

von Korff M, Grando S, Del Greco A, This D, Baum M, Ceccarelli S. (2008) Quantitative trait loci associated with adaptation to Mediterranean dryland conditions in barley. *Theor. Appl. Genet.*117:653–669.

Von Tschermak, E. (1901) Über Züchtung neuer Getreiderassen mittelst künstlicher Kreuzung: Kritisch-historische Betrachtungen. *Z. Landwirtsch. Versuchswes. Dtsch. Oesterr.* 4: 1029–1060.

Wagner, D.B., Allard, R.W. (1991). Pollen migration in predominantly self-fertilizing plants: barley. *J Hered* 82: 302-304.

Wang G, Schmalenbach I, von Korff M, Leon J, Kilian B, Rode J, Pillen K (2010). Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines. *Theor Appl Genet*, 120(8):1559-1574.

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

Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009). The emergence of whole genome association scans in barley. *Curr Opin Plant Biol*, 12(2):218-222.

Wenzl *et al.*, (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics*. 7:206.

Wenzl P., Carling J., Kudrna D., Jaccoud D., Huttner E., Kleinhofs A., and Kilian A. (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci.* USA, 101:9915-9920.

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol. Ecol. Res.* 8: 725–735.

Wilson, L.M. et al. (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell*. 16: 2719–2733.

Wright, S. (1931) Evolution in mendelian populations. *Genetics* 16, 97–159.

Xiong L., Wang R., Mao G., and Jessica Koczan M. (2006) Identification of Drought Tolerance Determinants by Genetic Analysis of Root Response to Drought Stress and Abscisic Acid. *Plant Physiology*. 142, pp. 1065–1074.

Xu S., Ni Y., Li Y., Li C., Chen M., Gao Y., Yu D., Liu P., Ma Z. (2009) Isolation and functional characterization of *HvDREB1*-a gene encoding a dehydration-responsive element binding protein in Hordeum vulgare. *Journal of Plant Research*. 122: 121-30.

Yan *et al*. (2003) Positional cloning of the wheat vernalization gene VRN1. *Proc. Natl. Acad. Sci.* USA 100, 6263–6268.

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, *et al* (2006). The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences of the USA*, 103:19581–19586.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, et al. (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell* 12: 2473–2484.

Yasuda S. (1969). Linkage and pleiotropic effects on agronomic characters of the genes for spring growth habit. Barley Newsletter.;12:57–58.

Yin X, Stam P, Dourleijn CJ, Kropff MJ (1999). AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley. *Theoretical and Applied Genetics*, 99(1-2):244-253.

Yoo SY, Kim Y, Kim SY, Lee JS, Ahn JH (2007) Control of Flowering Time and Cold Response by a NAC-Domain Protein in Arabidopsis. *PLoS ONE*. 2(7): e642. doi:10.1371/journal.pone.0000642.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 38(2):203-8.

Yu J. and Buckler E. (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*. 17:155-160.

Zadoks J. C., Chang T. T., and Konzak C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research.* 14:415-421.

Zakhrabekova S., Gough P., Braumann I., Müller H., Lundqvist J., Ahmanorpn K., Dockter C., Matyszczak I., Kurowska M., Druka A., Waugh R., Graner A., Stein N., Steuernagel B., Lundqvist U., Hansson M. (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 of the United States of America*. 109: 4326-31.

Zhang LY, Marchand S, Tinker NA, Belzile F (2009). Population structure and linkage disequilibrium in barley assessed by DArT markers. *Theor Appl Genet*, 119(1):43-52.

Zhang, W., Collins, A., Maniatis, N., Tapper, W. J., and Morton, N. E. (2002). Properties of linkage disequilibrium (LD) maps. *Proc. Natl. Acad. Sci. USA* 99, 17,004–17,007.

Zhu C., Gore M., Buckler M. and Yu J. (2008) Status and Prospects of Association Mapping in Plants. *The Plant Genome*. 1: p. 5-2.

Zhu K. (2002). Salt and drought stress signal transduction in plants. *Annual Rev. Plant Bio*. 53: 247-273.

Zhu M and Zhao S (2007). Candidate gene identification approach: progress and challenges. *Int J Biol Sci*; 3(7):420-427.

Zohary D. and Hopf M. (2000) Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe, and the Nile Valley (3rd ed.). *Oxford University Press*. pp. 59–69. ISBN 0-19-850357-1.

Zohary D. Proceedings of the First International Barley Genetics Symposium, Wageningen: Barley Genetics I; Wageningen, The Netherlands: Pudoc Centre for Agricultural Publications and Documentations; 1963. pp. 27–31.