M.Sc. – thesis February 2019

Earliness factors in Nordic spring barley (*Hordeum vulgare* L.)

Detection of QTL for growth and development traits in a Golf x Tampar barley mapping population

Naomi D. Bos

Faculty of Natural Resources and Environmental Sciences

M.Sc. – thesis February 2019

Earliness factors in Nordic spring barley (*Hordeum vulgare* L.)

Detection of QTL for growth and development traits in a Golf x Tampar barley mapping population

Naomi D. Bos

60 ECTS thesis submitted in partial fulfilment of a *Magister Scientiarum* degree in Agricultural Sciences

Academic advisors: Sæmundur Sveinsson and Morten Lillemo
Faculty supervisor: Bjarni Diðrik Sigurðsson

Agricultural University of Iceland
Faculty of Natural Resources and Environmental Sciences

Clarification of contributions

I hereby declare that the writing of the following thesis is my own work, done under the supervision of my advisor Sæmundur Sveinsson and co-advisor Morten Lillemo.

The barley mapping population examined in this study was created by Knut Aastveit and Åsmund Bjørnstad at NMBU, Norway. Jónatan Hermansson, Magnus Göransson and others have been involved in multiplying and maintaining this population and preparing the outdoor field trials at Korpa Experimental Station in Iceland. They were also responsible for phenotyping of the population in 2013 and 2014. Jónatan Hermansson also collected the weather data in 2014 and 2016 that were used in this study. My contribution to the phenotypic data collection is the phenotyping of the population in 2016. In addition, I prepared leaf samples for genotyping by SNP chip.

All further data analysis, including examination of field data, population structure and linkage disequilibrium decay analysis, GWAS and QTL analysis were performed by me.

Sæmundur Sveinsson and Morten Lillemo reviewed first versions of this manuscript. Åsmund Bjørnstad also contributed useful suggestions. Sæmundur Sveinsson translated the abstract into Icelandic. Bjarni Diðrik Sigurðsson acted as a formal faculty supervisor in the M.Sc. committee.

_____

Naomi D. Bos

Abstract

Barley *(Hordeum vulgare* L.) is the fourth most grown cereal crop in the world and can be successfully grown in a broad range of environments. Success of cultivation in subarctic regions such as Iceland depends largely on the ability of breeders to create high yielding cultivars that are adapted to the low temperatures, short growing season and strong winds that are often prevalent in these regions. An increased understanding of the genetic factors underlying traits such as earliness of flowering, maturation and height is therefore valuable. In this study, a Golf x Tampar barley mapping population consisting of 66 double haploid (DH) and 112 single seed descent (SSD) lines was phenotyped for three years in the field, and genotyped using the iSelect 9K barley SNP chip. The main aims were to characterize the population and to detect quantitative trait loci (QTL) for growth and development traits. Thereto, phenotypic data were analysed, population structure was assessed and linkage disequilibrium (LD) decay was calculated. Both genome wide association studies (GWAS) and QTL analysis were performed to link genotypic variation to the observed phenotypic variation.

Results showed large phenotypic variance in earliness of heading, maturity and height in the population. Year, line type, row type and line ID all significantly influenced phenotypic variation. Between-year correlation of traits was very strong for heading and moderately strong for maturity. Within-year correlation between traits was highest for heading-maturity. Analysis of population structure suggested the presence of two subgroups, containing DH and SSD lines respectively, as well as a slight partitioning based on row type, while LD decay analysis showed that LD decay was slower in DH lines than in SSD lines and varied considerably for each chromosome. Heading was predominantly affected by one major QTL in the centromeric region of chromosome 2H, for which *HvCEN* was the most likely candidate gene. Candidate genes for other, smaller, QTL for heading included among others *HvELF3*, *HvFT1* and members of the CONSTANS-like family of proteins. QTL associated with heading were sometimes also associated with maturity. Some candidate genes for QTL associated with maturity but not with heading include *Ppd-H1*, *HvCry1b* and *HvCO6*. Most candidate genes for QTL associated with height were part of the gibberellin pathway. Most variation in row type was caused by one QTL on 2H for which *Vrs-1*, a known major row type gene, was the best candidate gene. Sequencing of candidate genes detected in this study, *HvCEN* in particular, may give insight in the genetic variation responsible for the observed phenotypic variance of the studied traits.

Key words: Barley (*Hordeum vulgare* L.), flowering time, earliness factors, GWAS, QTL analysis.

Samantekt

[Flýtiþættir í norrænum bygglínum (*Hordeum vulgare* L.)]

Bygg (*Hordeum vulgare* L.) er fjórða mest ræktaða kornjurtin í heiminum og sú korntegund sem þrífst við fjölbreyttustu umhverfisaðstæður og ræktunarskilyrði. Ræktun byggs á norðlægum slóðum byggir á plöntuefniviði sem kynbættur hefur verið til að þroska korn við tiltölulega lágt hitastig og stutt vaxtartímabil ásamt því að þola hvassviðri. Aukin þekking á erfðafræðilegum þáttum byggs sem stjórna blómgun, þroska og stöngulhæð er því afskaplega verðmæt fyrir ræktun byggs á jaðarsvæðum. Efniviður þessarar rannsóknar eru 178 arfhreinar bygglínur. Þessar línur eru afkomendur víxlanna sem gerðar voru milli tveggja óskyldra byggyrkja: Golf og Tampar. Línunum var skipt í tvo hópa á grundvelli þeirrar aðferðafræði sem notuð var til að ná fram arfhreinum einstaklingum: DH (e. Double haploid) og SSD (e. Single seed descent). Allar línurnar voru svipgerðagreindar í akri og arfgerðagreindar á iSelect 9K barley genaflögu. Meginmarkmið rannsóknarinnar var að lýsa svipfari bygglínanna og beita tengslagreiningu til að finna erfðamörk sem hafa áhrif á vöxt og þroska. Svipgerðargögnin voru greind tölfræðilega og arfgerðargögnin nýtt til að kanna stofngerð. Ásamt því var hnignun tengslaójafnvægis mæld. Erfðamengis-tengslagreiningu og tengslagreiningu var beitt til að meta tengsl erfabreytileika við svipgerðarbreytileika byggs.

Mikill breytileiki reyndist vera í blómgunartíma, þroska og stráhæð á milli einstakara bygglína. Ræktunarár, uppruni línu (DH eða SSD) og axgerð voru allt þættir sem höfðu marktæk áhrif á breytileika í mældum svipgerðum. Stofngerðargreining leiddi í ljós að bygglínurnar skiptust í tvo hópa. Hópaskiptingin fylgdi uppruna línanna, þar sem bygglínur með DH uppruna aðskildust frá línum með SSD uppruna. Örlítil skipting reyndist vera á milli sexraða og tveggjaraða lína. Arfgerðagreiningar leiddu enn fremur í ljós að hnignun tengslaójafnvægis reyndist vera hraðari í línum með SSD uppruna heldur en í DH línum. Hnignun tengslaójafnvægis var jafnframt mjög breytileg milli litninga. Erfðabreytileiki innan litningasvæðis staðsettu nálægt þráðhafti á litningi 2H hafði marktæk áhrif á blómgun. Líklegt er að þessi áhrif stafi af breytileika innan *HvCEN*. Önnur möguleg gen sem gætu hafa haft áhrif á blómgun eru talin vera *HvELF3, HvFT1* og meðlimir CONSTANS próteinfjölskyldunnar. Líkleg gen sem höfðu áhrif á þroska en ekki blómgun voru *Ppd-H1*, *HvCry1b* og *HvCO6*. Raðgreining gena innan litningasvæða, sem tengslagreining sýndi fram á tengsl erfða- og svipgerðarbreytileika, t.d. *HvCEN*, er líkleg til að gefa frekari upplýsingar um erfðafræðilega stjórnun blómgunar í byggi.

Lykilorð: Bygg (*Hordeum vulgare* L.), blómgunartími, flýtiþættir, GWAS, QTL greining

Acknowledgements

I would like to thank my advisor Sæmundur Sveinsson for his help and advice over the course of this project and this study, and for suggesting this project to me in the first place. My sincere thanks also go to my co-advisor Morten Lillemo. Even though he joined the project only very late on, his help and advice have been indispensable.

Further, I would like to thank Åsmund Bjørnstad and Morten Lillemo for their advice on data analysis. Thank you for welcoming me to Norway, taking the time to help me and providing valuable feedback during the later stages of the writing of this manuscript.

My thanks also go to Magnus Göransson for his advice on GWAS and his support throughout the project, and to Þórdís Anna Kristjánsdóttir for advice on REML analysis.

Last but not least, I would like to thank my father Tjalling W. Bos and my sister Julia C. Bos for their never ending support.

This project was funded by Framleiðnisjóður Landbúnaðarins and the Barley Public Private Partnership pre-breeding project financed by the Nordic Council of Ministers (NMR).

List of tables

List of figures

List of abbreviations

AGDD	Accumulated growing degree days
AFLP	Amplified fragment length polymorphisms
AM	Association mapping
bp	Base pairs
DH	Double haploid
DNA	Deoxyribonucleic acid
ECMLM	Enriched compressed mixed linear model
GWAS	Genome wide association study
GxE	Genotype by Environment
K	Kinship
LD	Linkage disequilibrium
LOD	Logarithm of odds
MAF	Minor allele frequency
MAGIC	Multi-Parent Advanced Generation InterCross
MAS	Marker assisted selection
MLM	Mixed linear model
PC	Principle component
PCA	Principle component analysis
PGC	Parental genome contribution
QTL	Quantitative trait loci
RAPD	Random amplified polymorphism DNA
REML	Restricted maximum likelihood
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line
SNP	Single nucleotide polymorphism
SSD	Single seed descent
SSR	Simple sequence repeat (microsatellite marker)
QQ-plot	Quantile-quantile plot

Contents

1. Introduction

1.1 Origin and cultivation of Barley (*Hordeum vulgare* L.)

Cultivated barley (*Hordeum vulgare* ssp. *vulgare*, $2n = 2x = 14$) is a diploid annual cereal belonging to the *Triticeae* tribe of the *Poaceae* (*Gramineae*) family (von Bothmer, Jacobsen, Baden, Jørgensen & Linde-Laursen, 1995; Blattner, 2009). The closest relative, and most direct progenitor, of cultivated barley is wild barley (*H. vulgare* ssp. *spontaneum* (C. Kock) Thell.), which was formerly taxonomically treated as a separate species (Nevo, 1992). However, since wild barley and cultivated barley are fully interfertile, they have now been synonymized and are both treated as subspecies of *H. vulgare* (von Bothmer et al., 1995; Blattner, 2009). Cultivated barley is assumed to have been domesticated from multiple wild progenitor populations through a multicentric domestication process around 10 000-13 000 BCE (Badr et al., 2000; Poets, Fang, Clegg & Morrel, 2015; Allaby, 2015). Depending on vernalization requirement, two main types of barley can be differentiated: Spring- and winter barley (Verstegen, Köneke, Korzun & von Broock, 2014). Based on spike morphology, barley varieties can further be classified as either two-rowed or six-rowed (Briggs, 1978).

Barley is currently the fourth most grown cereal crop in the world (FAOSTAT, 2014). In 2014, barley was grown on around 49.5 million hectares worldwide, with an average yield of 2.9 tonnes/ha. The total world barley production was 144.5 million tonnes, the majority of which was spring barley (FAOSTAT, 2014). Approximately 70% of barley production is used for animal feeding, while the remainder is primarily used by the brewing and distilling industry, for sowing and human consumption (Verstegen et al., 2014).

As reviewed by Jónatan Hermannsson (1993), barley was first introduced to Iceland by the first settlers, around 874 CE. Barley appears to be have been successfully cultivated on several locations until 1300 CE, when a period of lower temperatures made it more difficult to grow barley. This, in combination with a decrease in the price of imported barley is likely to have led to the end of cultivation of barley in Iceland. It was only in the 20th century that cultivation of barley was resumed. In the 1920s, Klemenz Kr. Kristjánsson started experiments with barley cultivation, and since then, barley field trials have been conducted on and off at different sites throughout Iceland (Sigurbjörnsson, 2014; Hilmarsson et al., 2017). From 1987, barley field experiments have continuously been conducted by the Agricultural University of Iceland (AUI) (previously Agricultural Research Institute) (Hilmarsson et al., 2017). Between 1987 and 1989, a large number of spring barley accessions were tested as part of a Nordic barley project, and

around the same time a new Icelandic barley research and breeding project was initiated (Nurminiemi, Bjørnstad & Rognli, 1996; Sæmundur Sveinsson, personal communications, 2016). So far, this project has produced four relatively high yielding and early flowering cultivars that are adapted to the Icelandic climatic conditions, and has made them commercially available. Currently, about 32% of imported seed is of Icelandic origin (Hilmarsson et al., 2017). Since 2012, AUI is also involved in the Barley Public Private Partnership pre-breeding project financed by the Nordic Council of Ministers, which aims to support the development of Nordic barley breeding and cultivation (NordGen, 2017).

For the last decade, barley yields on Iceland have averaged at about 3.2 tonnes/ha (Intellecta ehf., 2009). Between 1992 and 2009, total annual production of barley has steadily increased from about 1,000 to 17,000 tonnes, before decreasing again in 2013 and 2014 to about 6,000 to 7,000 tonnes (Hilmarsson et al., 2017).

1.2 Challenges of cereal breeding in subarctic regions

Cultivation of cereal crops in subarctic regions such as Iceland faces many challenges including short growing season, low summer temperatures, occasionally very strong winds, risk of frost in late spring and early autumn, and high precipitation levels during the harvesting period (Reykdal et al., 2016; Hilmarsson et al., 2017). In regions with an oceanic climate, such as Iceland, summer temperatures are even lower than in regions with a continental climate. For example, on Iceland the average heat sum over the growing season is just 1300 accumulated growing day degrees (AGDD) (Jónatan Hermansson, personal communication). However, temperatures in Iceland, as in many other regions of the world, are expected to rise in the future. Halldór Björnsson et al. (2008) predict that the mean annual temperatures in the lowlands of Iceland are likely to increase by 1.5-2.5 °C in the 21st century. This may enable earlier sowing, and is likely to result in increased heat sum over the growing season. In Iceland, 'day of sowing' has been shown to only slightly affect heading and harvest dates (Hilmarsson et al., 2017). However, heat sum during the growing season has been reported as a major yield determining factor in Norway (Lillemo, Reitan & Bjørnstad, 2010). Similarly, a strong positive correlation between yield and temperature during the growing season has been observed in Icelandic barley trials (Hilmarsson et al., 2017). An increase in the heat sum during the growing season is therefore likely to positively affect barley growing conditions on Iceland. An advantage of the high latitude geographical position of Iceland and other subarctic areas, are the long

photoperiods during the growing season, which partially reduce heat sum requirements (Bergþórsson et al., 1987).

Barley, especially spring barley, is less demanding than many other cereal crops with regards to soil fertility and climatic conditions, and can be successfully grown in a broad geographical and climatological range. It is relatively well adapted to subarctic weather conditions, and is therefore often the preferred crop for cereal cultivation in subarctic regions (Fischbeck, 2003; Verstegen et al., 2014). For example, spring barley accounts for about 90% of total cereal cultivation on Iceland, while in Norway barley accounts for 49% of the total area used for cereals, making it the most important cereal crop in both countries (Reykdal et al., 2016; Statistics Norway, 2016a, b).

Successful cultivation of barley depends on the availability of high-yielding cultivars that are adapted to the local conditions. Some of the main target traits of barley breeders around the world therefore include yield, yield stability, disease resistance, straw strength, winter hardiness, drought tolerance, quality traits such as protein content as well as traits sought after by the malting industry (Verstegen et al., 2014). A number of pathogenic fungal species have been found in Iceland. Scald (*Rhynchosporium secalis)* is the fungal disease of most concern in Iceland since it is known to cause yield losses of economic importance (Helgi Hallgrímsson & Guðríður Gýða Eyjólfsdóttir, 2004; Jónatan Hermannsson, 2004). However, even though the importance of disease resistance is being recognised, it is not an important aim of Icelandic barley breeding programmes (e.g. Stefansson, Serenium & Hallsson, 2012). Instead, the focus of Icelandic barley breeding lies predominantly on adaptation to abiotic factors: The main sought-after traits have been earliness of flowering and maturity, maturation ability at low temperatures and lodging resistance.

The timing of reproductive processes such as flowering has a large impact on the ability of barley to grow and mature seeds. In order to increase reproductive success and grain yield in barley, flowering and maturation should take place within a timeframe during which resource availability and weather conditions are optimal for plant development (Nitcher, Distelfeld, Tan, Yan & Dubcovsky, 2013). Knowledge about the mechanisms regulating reproductive and developmental processes is therefore crucial to breeders who aim to target this optimal reproductive period to increase yield and reproductive success. Although multiple studies have previously addressed the genetic processes underlying developmental characteristics of barley, much still remains unknown (e.g. Kumlehn & Stein, 2014). An increased understanding of the

genes and pathways involved in regulatory processes has the potential to increase the efficiency of breeding programmes, and could therefore be of great agricultural benefit.

1.3 The barley genome and identification of trait loci

1.3.1 Sequencing of the barley genome

Barley is an inbreeding, diploid cereal crop with relatively simple genetics compared to some other, closely related, important cereal crops such as wheat (*Triticum* spp.) and oat (*Avena sativa*). For this reason, barley is an excellent experimental model to study the genetic mechanism underlying important traits and characteristics of cereal crops (Schulte et al., 2009). Furthermore, large genetic diversity is available in extensive collections of germplasm, including elite breeding lines, landraces as well as wild accessions, which undoubtedly contain alleles of high value for breeding programmes (IBSC, 2012). In order to facilitate the exploitation of these genetic resources, and to increase genetic and genomic knowledge about barley and its close relatives, The International Barley Sequencing Consortium (IBSC) (www.barleygenome.org) was established in 2006 with the aim to physically map and sequence the barley genome (Schulte et al., 2009). The barley genome is relatively large, approximately 5.1 Gbp (Dolezel et al., 1998), and has large amounts of highly repetitive DNA composition and low gene density, as well as severely reduced frequency of meiotic recombination in pericentromeric regions. This is also characteristic for other *Triticeae* cereal genomes, and complicates the generation of high quality reference genomes for these species (Künzel, Korzun & Meister, 2000; Schulte et al., 2009; IBSC, 2012). By integrating a variety of methods and including sequence data from large numbers of bacterial artificial chromosome (BAC) clones as well as transcriptome sequences, these problems were (partially) circumvented. As a result, a physical map of the cultivar Morex covering 4.98 Gbp of the barley genome was assembled. Furthermore, the positions of contiguous genome sequences (contigs) of a total size of approximately 3.9 Gbp were successfully anchored to a genetic map. The resulting genomic framework and reference genome for barley provides information on the physical position and distribution of genes and repetitive DNA sequences, and how they relate to a number of genetic characteristics, including gene expression and recombination frequency (IBSC, 2012). More recently, researchers made use of a technique called POPSEQ to increase the number of contigs anchored to a reference sequence (Mascher et al., 2013). In 2017, Mascher et al. presented an improved high-quality barley reference sequence covering 4.79 Gbp of the Morex genome. Apart from the Morex reference genome, a second reference genome for Tibetean hulless barley

(*H. vulgare* L. var. *nudum*) was also recently published (Dai et al., 2018), and draft genomes for three additional cultivars are also available (IBSC, 2012; Sato et al, 2016).

1.3.2 Molecular markers

In order to understand the genetic mechanisms underlying important traits in barley, and to be able to exploit these mechanisms to improve breeding efficiency, knowledge of the genomic location of causal genes and loci, as well as their effect is crucial. Using molecular markers, such genes and loci can be identified. Molecular markers are genetic tools that can be used to identify and track polymorphisms in the DNA. As reviewed by Mammadov, Aggarwal, Buyyarapu and Kumpatla (2012), different types of markers exist, each of which has its own advantages and disadvantages. Development and use of markers is largely dependent on the throughput-rate and cost of the method, as well as the level of reproducibility. The use of markers such as restriction fragment length polymorphisms (RFLPs), and PCR-based markers such as random amplified polymorphism DNA (RAPD), amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSR) have greatly increased knowledge about the genetics underlying agronomic traits, and made it possible to use marker assisted selection (MAS) in molecular breeding programmes (Somers, Isaac & Edwards, 2004). However, in recent years, single nucleotide polymorphisms (SNP) markers have become the marker-system of choice. SNPs are often bi-allelic, and have a lower mutation rate than several other types of markers, such as SSRs, and are therefore less polymorphic. The advantages of using SNP markers is their ubiquitous nature and that they are amenable to ultra-high throughput automation, which significantly reduces cost, labour and time (Mammadov et al, 2012). SNP markers can be used for studies on genetic variation and population structure, for linkage and association mapping. Genetic maps created using large numbers of SNPs enable the identification of quantitative trait loci (QTL) and genes of interest and their localization to a precise position, which can be valuable information for plant breeders (Ganal, Altmann & Roder, 2009; Ganal et al., 2012). A QTL is a locus (a section of DNA) which correlates with observed variation in a certain phenotype.

1.3.3 Methods for the identification and mapping of trait loci

Two main methods are currently used to map trait loci to the genome: Linkage mapping (often referred to as QTL analysis) and association mapping (AM), which is also called genome wide association study (GWAS). Both methods rely on fragmentation of the genome by recombination, and the genetic linkage of the used markers with the trait under investigation

(Myles et al., 2009). During meiosis, recombination of fragments of the genome generates genetic variation. Genome fragments in which no recombination occurs are called linkage blocks. Loci within the same linkage block are in high linkage disequilibrium (LD) with each other: They are more likely to be inherited together than would be expected based on random association of alleles at different loci in the genome. When a (SNP) marker and a trait locus are situated within the same linkage block, they are said to be linked. In such situations, the (SNP) marker can be used to detect this region (Slatkin, 2008).

Both linkage mapping and AM require geno- and phenotyped populations. However, the main difference between these methods, is that linkage mapping requires the generation of a mapping population in which relatedness is controlled by the experimenter, whereas AM can be applied on any type of population (Myles et al., 2009). In inbreeding species, such as barley, QTL analysis is often performed on bi-parental recombinant inbred line (RIL) populations. Such populations are usually created by crossing genetically divergent parent lines, followed by self-fertilisation to create homozygous lines (Pollard, 2012). The single seed descent (SSD) method, during which only one progeny of each plant is used in each successive generation until the preferred level of homozygosity has been reached, speeds up this process (Snape & Riggs, 1975). Another way of generating homozygous lines is to use the F_1 generation to create double haploid (DH) lines. As reviewed by Humphreys and Knox (2015), DH generation consists of two main steps: Haploid induction followed by chromosome doubling. An advantage of this method is that this creates plants that are homozygous at loci within a single generation, and therefore saves time. However, since only one recombination event takes place, F_1 DH lines have less recombination than for example SSD lines. Linkage blocks are therefore often large. In addition, LD generally decays slowly, especially in self-fertilising species. This may result in a low mapping resolution in such populations, especially compared to natural populations (Myles et al., 2009). One additional drawback of the use of bi-parental populations, is that this method can only detect loci that differ between the two parents. Multi-Parent Advanced Generation Inter-Cross (MAGIC) populations partially solve this problem: They are created using several founder parents that are crossed several times before homozygous lines are created, and consequently display a higher degree of polymorphisms, which allows for the detection of more different loci than in bi-parental populations (Huang et al., 2015). Once a population has been phenotyped and genotyped, QTL analysis aims to analyse the co-segregation between genetic markers and phenotypes in search for regions in the genome that are linked to a specific phenotype. Such regions are likely to contain causal genes for the trait

of interest (Myles et al., 2009). Yet another disadvantage of the use of controlled populations, such as RIL, is that the allele frequencies and combinations will differ from those in the natural population (Korte & Farlow, 2013).

As reviewed by Myles et al. (2009), during AM, correlations are sought between genetic markers and phenotypes within the genotyped and phenotyped population. There is no need for controlled populations, which have only gone through a small number of recombination events. The use of natural population, which can generally be expected to have gone through a higher number of recombination events, increases the potential for higher mapping resolution, and, due to higher degrees of polymorphisms in such populations, facilitates the detection of more different causative loci and alleles.

The risk of obtaining Type I or Type II errors in AM is higher than in bi-parental linkage mapping. Type I error is the reporting of "false positive" results, while Type II error (inversely related to statistical power) is the reporting of "false negative" results. False positives can for example occur as a result of failure to account for population structure. Population structure may for example be caused by admixture, mating system, genetic drift, artificial or natural selection (Crossa et al., 2007). Population stratification and relatedness between lines may result in false associations that are not caused by actual genetic linkage. This affects LD and therefore marker-trait associations (Gupta, Kulwal & Jaiswal, 2014). By incorporating relationship (Q, PCA) matrices and/or kinship (K) matrices in a mixed linear model (MLM), population structure can be accounted for, and the risk of Type I errors reduced (Myles et al, 2009). LD decay is often faster in natural populations, which may result in lower correlation between markers and traits, which may increase Type II error compared to bi-parental analysis (Carlson, Eberle, Kruglyak & Nickerson, 2004). Yet another disadvantage of AM is that rare alleles are often not detected, among others because alleles with a minor allele frequency (MAF) ≤ 5% are generally excluded at the beginning of the analysis (Gupta et al., 2014). Despite these short-comings AM is a very powerful method for the identification of trait loci, and is becoming increasingly more popular (Xiao, Liu, Wu, Warburton & Yan, 2017). One additional factor influencing both QTL analysis and AM is heritability of a trait. The lower the heritability (h^2), the more difficult it is to detect and identify QTL due to reduced reproducibility of phenotypic values (Cockram et al., 2010). Heritability is a measure of the proportion of observed variation in a phenotypic trait in a population that is due to genetic variation between the individuals in that population.

1.3.4 Analysis of population structure and linkage disequilibrium (LD)

As reviewed by Myles et al. (2009), population structure can have severe confounding effects on the results of AM. In case of genotype-phenotype covariance, genetic markers will appear to be associated with the phenotype, while in reality they capture genetic relatedness between lines. Although this is a problem for all traits, this can become especially problematic when a trait has been subject to adaptation. This is for example often the case with flowering time and may cause extremely high rates of false positive results. By estimating genetic relatedness between lines in a population, a baseline prediction of background QTL sharing between lines can be generated, and used to assess whether a marker explains more phenotypic variation than a random marker in this population would. This method is most useful for quantitative traits and less so for qualitative traits.

Two main methods to account for relatedness are structured association- and principal component analysis (PCA). Structured association analysis is most commonly performed using the programme STRUCTURE (Pritchard, Stephans & Donnely, 2000), which identifies sub-populations and then calculates the proportion of a line's variation that came from each subpopulation. The estimates in the resulting Q matrix are used as covariates in a mixed linear model (MLM) (Pritchard et al., 2000). PCA can be used to break down the genetic variation into a smaller number of components (dimensions). As with structured association analysis, the resulting matrix can then be added to an MLM (Price et al., 2006). Disadvantages of using STRUCTURE are that it is computationally intensive, and that it is designed for unrelated individuals from a population that is in Hardy-Weinberg equilibrium. PCA on the other hand does not make assumption about the population structure and is a lot faster. Furthermore, results are usually similar or better than from STRUCTURE analysis (Zhao et al., 2007).

As mentioned, LD affects both QTL analysis and AM. Faster LD decay allows for higher resolution mapping, but at the same time requires more molecular markers and increases the number of false negatives (Rafalski, 2002; Carlson et al., 2004; Myles et al., 2009). Although LD decay is generally slow in self-fertilizing species, large differences can also exist between cultivars or populations. Large differences in recombination frequencies have for example been observed in barley germplasm (Ruge-Wehling & Wehling, 2014; Bengtsson, The PPP Barley Consortium, Manninen, Jahoor & Orabi, 2017). Bengtsson et al. (2017) studied linkage disequilibrium within 180 Nordic spring barley breeding lines and found that LD decayed below a critical r^2 value of 0.2 within a range of 0-4 cM for the complete dataset. After separating the dataset in subgroups based on population structure, the LD decay for the different

groups ranged from 0-12 cM. Meanwhile, LD decay (in r^2) within subpopulations of US barley breeding germplasm ranged from 20-30 cM (Hamblin et al., 2010), and from 4-19.8 cM according to Zhou, Muehlbauer and Steffenson (2012). As reviewed by Vos et al. (2017), there are many ways to estimate LD decay, but no universal approach can be suggested. Depending on the material, different methods may be preferred.

Recombination frequency, and therefore LD decay, is not equal throughout the entire barley genome. Distribution of recombination is highly skewed to the distal ends of the chromosomes, and is considerably more skewed than the distribution of genes (Mayer et al., 2011). As a consequence, a large part of barley genes is suspected to be situated within genomic regions that rarely, or less frequently, recombine (Ramsay, 2014).

1.4 Regulation of developmental and structural traits in barley

Many different studies have addressed the genetic regulation of developmental and structural traits. Both by using QTL analysis and GWAS, large numbers of QTL and causal genes have been identified. In addition, homology-based studies have also proven useful, for example in the identification of genes involved in flowering time regulation in barley (e.g. Yan et al., 2003; Fauré, Higgins, Turner & Laurie, 2007; Maurer et al., 2015; Alqudah, Koppolu, Wolde, Graner & Schnurbusch, 2016). QTL associated with earliness can also be referred to as earliness factors.

1.4.1 Regulation of flowering time and maturity in barley

An understanding of the different phases of reproductive development, and the mechanisms underlying phase transitions is of great importance for barley breeders. The transition from the vegetative to the reproductive phase is a quantitative trait that is genetically controlled by genetic pathways integrating both environmental and endogenous factors (Alqudah & Schnurbusch, 2017). Based on morphological changes in the shoot apical meristem, pre-flowering development in temperate cereals such as barley is often divided into three main phases: The vegetative, early reproductive, and late reproductive phase. Each of these phases differs in their sensitivity to certain environmental signals. The sum of the duration of these stages determines earliness of flowering in the plant (Drosse, Campoli, Mulki & von Korff, 2014).

Photoperiod and vernalisation are two of the most important environmental factors involved in the regulation of flowering time during pre-flowering development (Verstegen et al., 2014).

Depending on their requirement for vernalisation, barley genotypes are either classified as winter or spring barley. Winter barley requires vernalisation, a period of cold temperature during winter, in order to be able to complete its life cycle and must therefore be sown in fall, whereas spring barley does not and is sown in spring (Nitcher et al., 2013). A third, less common, facultative genotype is the result of the presence of a certain allelic series at the *Vrn-H1* vernalisation locus. This type can be sown at any time of the year (Hemming, Fieg, Peacock, Dennis & Trevaskis, 2009). Barley is a facultative long-day plant. Some barley genotypes are photoperiod sensitive and have accelerated flowering under long days, while others are photoperiod insensitive and have only a limited response of flowering to long days (Verstegen et al., 2014).

Research has shown that the photoperiod, vernalisation and circadian clock pathways play a crucial role in the genetic regulation of flowering time (e.g. Yan et al., 2003; Fauré et al., 2007). Figure 1 shows a simplified model of how these pathways interact with each other to regulate earliness of flowering, and what major genes are involved. Much of the variation in photoperiod response that has been observed between genotypes is the result of sequence variation within the photoperiod regulator genes *Ppd-H1* and *Ppd-H2*, located on chromosome 2H and 1H respectively (Comadran et al., 2012; Drosse et al., 2014; Kikuchi, Kawahigashi, Oshima, Ando & Handa, 2012). *Ppd-H1* encodes a protein that is part of the circadian clock pathway, highlighting the interaction between these two pathways (Turner, Beales, Fauré, Dunford & Laurie, 2005). Under long day conditions, the dominant and recessive allele of this gene respectively up- and downregulate the expression of the flowering promoting *HvFT1* gene, which is situated on chromosome 7H (Turner et al., 2005; Fauré et al., 2007; Hemming, Peacock, Dennis & Trevaskis, 2008). *Ppd-H2* has a function similar to *HvFT1*. Under specific conditions, expression of this gene promotes flowering, although not as extreme as *HvFT1*. Conversely, *HvCO9*, which is located on chromosome 1HL, decelerates flowering, proposedly by downregulating *HvFT1* (Kikuchi et al., 2012; Nitcher et al., 2013; Drosse et al., 2014). Yet another paralog of *HvFT1* is *HvCEN* (not shown in Figure 1), which is proposed to have a function very similar to *HvFT1* or even interact with it (Loscos, Igartua, Contreras-Moreira, Gracia & Casas, 2014). *HvCEN* is the candidate gene for a locus in the centromeric region of chromosome 2H, which has also been associated with tiller biomass, tiller grain weight, ear grain number, and plant height (Obsa et al., 2016). Sequence variation at the circadian clock pathway gene *HvELF3*, located on chromosome 1H, affects expression of *Ppd-H1*, thereby indirectly influencing earliness of flowering of the plant (Fauré et al., 2012). Another circadian

clock controlled gene, *HvCO1*, acts in parallel to *Ppd-H1*, and influences expression of *HvFT1* (Campoli, Drosse, Searle, Coupland & von Korff, 2012a; Campoli, Shtaya, Davis & von Korff, 2012b). Not much is known about *HvGI*, but this gene is also suggested to play a role in flowering time regulation (Zakhrabekova et al., 2012).

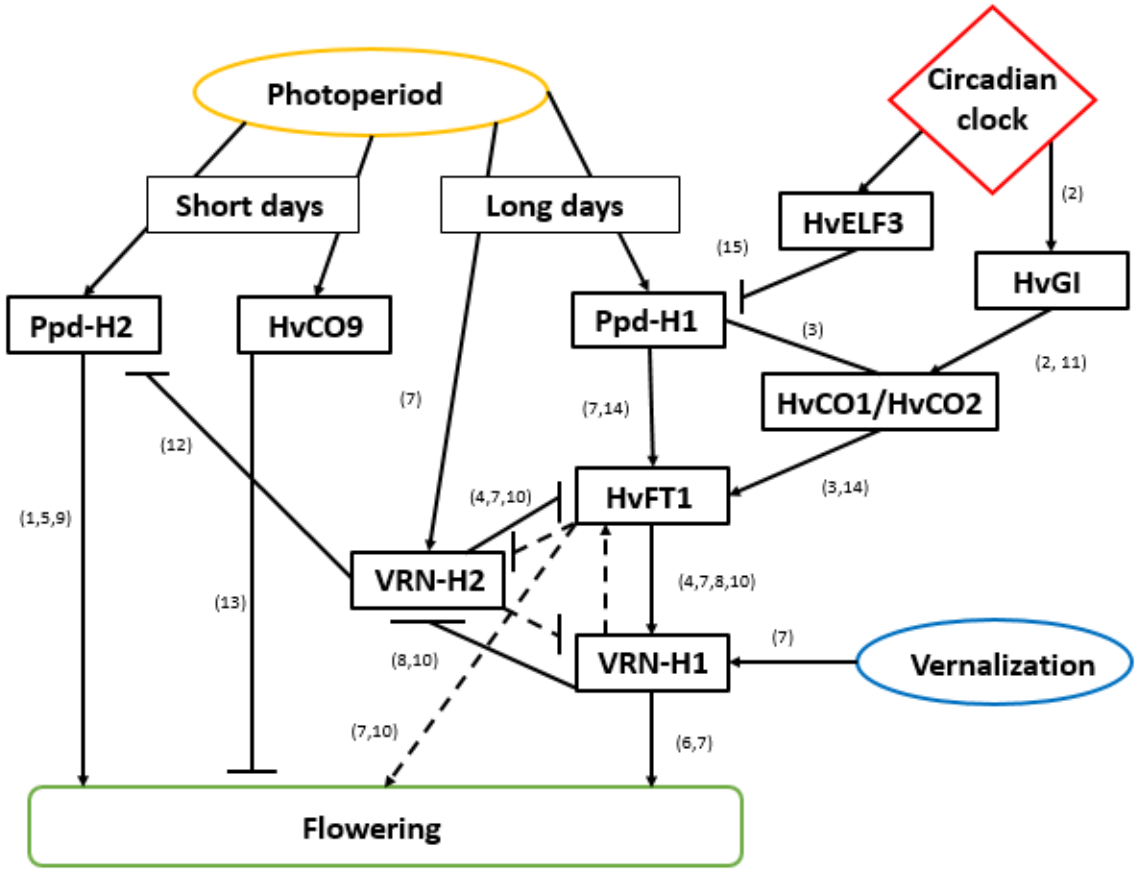


Figure 1. Model showing the interactions between genes of the photoperiod, vernalisation and circadian clock pathways and how they regulate flowering time in barley. Dashed lines indicate alternative models of interaction between the genes, and numbers in brackets indicate literature in which experimental evidences support this model; (1) Laurie et al. (1995), (2) Dunford et al. (2005), (3) Turner et al. (2005), (4) Yan et al. (2006), (5) Fauré et al. (2007), (6) Shitsukawa et al. (2007), (7) Hemming et al. (2008), (8) Li & Dubcovski (2008), (9) Kikuchi et al. (2009), (10) Shimada et al. (2009), (11) Shin-Young et al. (2010), (12) Casao et al. (2011), (13) Kikuchi et al. (2012), (14) Campoli et al. (2012a), (15) Fauré et al. (2012) (based on Drosse et al., 2014).

Most of the variation in vernalisation requirement that has been observed between genotypes is the result of sequence and copy number variation in the *Vrn-H1*, *Vrn-H2* and *HvFT1* (also known as *Vrn-H3*) genes of the vernalisation pathway (Drosse et al., 2014). *Vrn-H1* encodes a MADS box transcription factor required for the transition of the shoot apical meristem (SAM) from the vegetative to the reproductive phases and thus contributes to the initiation of flowering (Yan et al., 2003). Under long day conditions, *Vrn-H1* downregulates expression of *Vrn-H2*

and upregulates expression of *HvFT1* (Yan et al., 2004; Shimada et al., 2009). *Vrn-H2* is expressed under long day conditions and represses flowering, primarily by downregulating *HvFT1*. Downregulation of *HvFT1* causes downregulation of *Vrn-H1*, which leads to decelerated flowering. Vice versa, upregulation of *HvFT1* by other genes such as *Ppd-H1* and *HvCO1* results in upregulation of *Vrn-H1* and a subsequent acceleration of flowering (Nitcher et al., 2013). Other genes involved in flowering regulation are the red/far red light phytochrome barley genes *HvPhyA*, *HvPhyB* and *HvPhyC* (Szucs et al., 2006). Many more QTL for flowering time, including multiple *earliness per se* (*eps*) loci, have been detected, but only few of them have been characterized in barley (Alqudah & Schnurbusch, 2017).

Earliness of flowering is often assessed by recording 'heading date'. According to the Zadoks scale (Zadoks, Chang & Konzak, 1974), heading is the appearance/emergence of the spike out of the flag leaf sheath (Z50-Z59). However many studies use awn appearance (Z49) as proxy for heading date, while other use anther extrusion (Z60-Z69). Further, it is important to realise that heading date does not mean flowering time (flowering/fertilisation usually occurs at Z49). Inaccurate scoring of heading date has been shown to affect QTL detection, among others because these different stages are partially under different genetic control, and may therefore result in inaccurate interpretation of results. In order to prevent misunderstandings, it is therefore important to accurately describe how heading date was scored during phenotyping (Alqudah & Schnurbusch, 2017). Alqudah and Schnurbusch (2017) further suggest that it is better to quantify this heading date in accumulated growing degree-days (AGDD) than in days. AGDD is thought to be a more reliable and robust for particular developmental stages across a season or environment.

Breeding for early maturity has resulted in the generation of varieties and lines carrying mutations conferring increased earliness (Lundqvist, 2009). Although candidate genes have not yet been discovered for all these mutations, the *early maturity* 8 (*eam*8) mutation was for example shown to be located on the circadian clock gene *HvELF3* (Fauré et al., 2012). Maturity QTL have been found in other studies. For example, Obsa et al. (2016) identified 13 QTL for maturity, some of which co-located with the known phenology genes *HvAP2*, *HvTFL1*, *HvPhyC, HvPhyB, HvPhyC, HvCO1,* and *HvAP2*, respectively. Most of these QTL had low additive effect and showed QTL by Environment interaction. An overview of some of the main QTL and candidate genes known to affect flowering and maturity, and their position in the barley genome can for example be found in Drosse et al. (2014) and Alqudah et al. (2016).

Pleiotropic effects of genes and loci associated with flowering time and maturity have been observed in multiple studies (e.g. Drosse et al., 2014; Alqudah et al., 2016; Obsa et al., 2016; Alqudah & Schnurbusch, 2017). For example, variation at *Ppd-H1*, and overexpression of *HvCO1* have both been shown to affect not only inflorescence development but stem elongation (i.e. height) as well (Drosse et al., 2014). Furthermore, *HvCEN* was found to affect row type, and more recently *HvFT1* was found to affect root and shoot development (Comadran et al., 2012; Arifuzzaman et al., 2016). As reviewed in Alqudah and Schnurbusch (2017), the trait 'heading date' is highly associated with environmental adaptation and has been shown to be important for understanding other developmental traits such as leaf area, plant height, tillering and grain number. Furthermore, heading is considered a decisive stage for improvement of grain yield and yield components. It is therefore clear that an increased understanding of the genetic basis of pre-flowering development and contributions of flowering genes, is likely to contribute to increased yield in more than one way.

1.4.2 Genetic control of plant height and lodging

Plant height is influenced by many QTL as well as dwarfing, semi-dwarfing and other qualitative genes (Rossini, Okagaki, Druka & Muehlbauer, 2014). Dwarfing genes are usually not used in plant breeding because they are often associated with reduced vigour and grain yield. Semi-dwarfing genes can however be useful in breeding for reduced crop height (Mickelson & Rasmussen, 1994; Zhang, 2003). The main advantages of short-stature plants are increased grain yield due to repartitioning of assimilate from stem to grain production and reduced lodging (Khush, 2001). Lodging is the habit of cereal crops to bend over and lie on the ground. Lodging causes grain yield losses due to suboptimal grain-filling conditions and technical difficulties during harvesting leading to less efficient grain harvest. Furthermore, the chances of fungal infections, mycotoxin accumulation and pre-harvest germination of grains are higher (Caierão, 2006; Rajkumara, 2008). Certain genes and QTL influencing plant height are often correlated to barley lodging (Khush, 2001; Rossini et al., 2014).

Various genes influencing plant height in barley have been identified. Some, such as *sdw1/denso* are already being incorporated in barley breeding programmes while others have deleterious pleiotropic effects that result in low vigour and reduced yield, or have not yet been sufficiently genetically characterized and are for those reason not used in barley breeding (Rossini et al., 2014). Table 1 in Kuczynska et al. (2013) summarizes the most important (mostly semi-dwarf) genes that are used in barley breeding. Some genes cause only limited

reduction in plant height while others are capable of causing a height reduction as high as 50% of the original plant height (Yuo et al., 2012; Kuczynska et al., 2013). Although most studies have focused on analysis of height reducing factors, some mutations causing increased growth have been identified as well. For example, Franckowiak and Lundqvist (2002) reported that recessive mutations in the *Gigas2* gene increased plant height. And Chandler and Harding (2013) identified new mutations in the GA genes *Sln1* and *Spy1* that resulted in an increased growth rate. Apart from single genes that have been genetically mapped, many QTL associated with plant height and lodging, have been identified as well. For example, Gyenis et al. (2007) identified four QTL on chromosome 1H, 2H, 3H and 7H, Alqudah et al. (2016) detected 10 different QTL and more recently Mikolajczak et al. (2017) detected 9 different QTL on chromosomes 2H, 3H, 4H, 5H and 6H that affected the length of the main stem in a spring barley mapping population.

1.4.3 Row type regulation in barley

Barley varieties are usually classified as either two-rowed or six-rowed based on spike morphology. Two-rowed barley is a reduced version of the six-rowed barley where the side florets are sterile instead of fertile as in six-rowed (Briggs, 1978). The wild ancestor of cultivated barley was two-rowed, and research shows that the six-row phenotype is the result of breeding efforts very early in the domestication of barley (Rossini et al., 2014). Most spring barley that is currently grown in Central Europe is two-rowed, whereas in Northern Europe and North America, six-rowed spring barley is widely grown (Verstegen et al., 2014). For example, in Iceland two-rowed and six-rowed varieties both account for approximately 50% of total imported seed (Hilmarsson et al., 2017).

As summarized by Rossini et al. (2014), row type is regulated by a number of different genes. Sequence variation at the *six-rowed spike 1* (*Vrs-1*) gene alone is sufficient to control row type in barley. *Vrs-1* is located on chromosome 2H (Komatsuda et al., 2007) and encodes a HD-ZIP transcription factor that is required to suppress the development of the lateral florets. A loss of function mutation in *Vrs-1* prevents expression of the gene, and is consequently responsible for the six-row phenotype. However, phenotypes imposed by *Vrs-1* alleles can be modified by alleles of other genes such as *intermedium-c* (*Int-C*). This gene, which is situated on chromosome 4H (Komatsuda & Mano, 2002), affects the extent to which the lateral spikelets are developed. This often results in intermediate phenotypes that cannot clearly be defined as either two-rowed or six-rowed (Ramsay et al., 2011). Other loci known to modify the development of side spikelets include *Vrs-2* on chromosome 5H, *Vrs-3* on chromosome 1H,

Vrs-4 on chromosome 3H and *Vrs-5* on chromosome 4H (Pourkheirandish & Komatsuda, 2007; Li, Chen & Yan, 2015). Furthermore, additional QTL have been described by Pasam et al. (2012) and Muñoz-Amatriaín et al. (2014).

As mentioned previously, some of the genes and QTL regulating flowering time, are also involved in row type regulation. Alqudah et al. (2016) especially designed their study to reveal associations between the genetic control of different traits, and found the genetic mechanism regulating row type, heading time, tillering as well as plant height to be significantly associated. S*ix rowed spike 1* (*Vrs-1*) mutants and their two-rowed progenitors were examined in detail, which showed that wild type (two-rowed) plants were significantly taller, and had more tillers than mutants. This would suggest a negative pleiotropic effect of *Vrs-1* on these traits.

1.5 Study aims

The success of growing barley in subarctic regions such as Iceland depends on the availability of varieties adapted to the short growing season, low temperatures and strong winds. Developing early flowering and maturing varieties that are lodging resistant is therefore one of the main priorities of Icelandic barley breeding programmes. Although extensive research has taken place, regulation of developmental processes in barley is still not yet completely understood. The study of genetic factors contributing to different developmental characteristics of Nordic spring barley, such as earliness of flowering and maturation, as well as height could therefore potentially be of great use.

In this study, phenotypic and genotypic data were collected for a Golf x Tampar barley mapping population segregating for flowering time, and subsequently analysed with the main aim of increasing knowledge on the genetic regulation of selected growth and development traits in barley. The sub-aims of this study are as follows:

1.) To phenotypically characterize the mapping population.
2.) To study population structure and linkage disequilibrium in the population.
3.) To analyse the population for QTL for heading, maturity and height.
4.) And to identify the genetic characteristics of Tampar that infer the observed earliness in this landrace.

2. Materials and Methods

2.1 Plant material

In this study a Golf x Tampar barley mapping population was examined. This population was created by Knut Aastveit and Åsmund Bjørnstad (NMBU, Norway), and segregates for earliness of flowering. It consists of 66 double haploid (DH) lines, 112 single seed descent (SSD) lines and both parents. Golf is a high yielding, late flowering malting variety from Scotland that originates from the triple cross (Armella x Lud) x Luke (Nurminiemi et al., 1996). Tampar is an early flowering landrace from the Faroe Islands and is extensively represented in Icelandic breeding lines such as Skúmur and Lómur (Nurminiemi et al., 1996; Þórdís Anna Kristjánsdóttir, 2013). Along with the Faroe landrace Sigur, it may be the closest relative of the barley varieties that were grown in the first five hundred years after the first settlers first came to Iceland (Sigurbjörnsson, 2014). Tampar is known for its relatively low environmental stability, suggesting high Genotype x Environment interaction (Nurminiemi, 1995). Previous studies further show that the growth curve of Tamper deviates significantly from other Nordic spring barley accessions, including Golf (Figure 2). Tampar has a lower dry matter- and grain yield than Golf, but reaches heading and maturity earlier than any other accession, suggesting it contains a specific genetic characteristic that infers this earliness (Aastveit, unpublished data; Numiniemi, 1995, Åsmund Bjørnstad, personal communication, 2018).

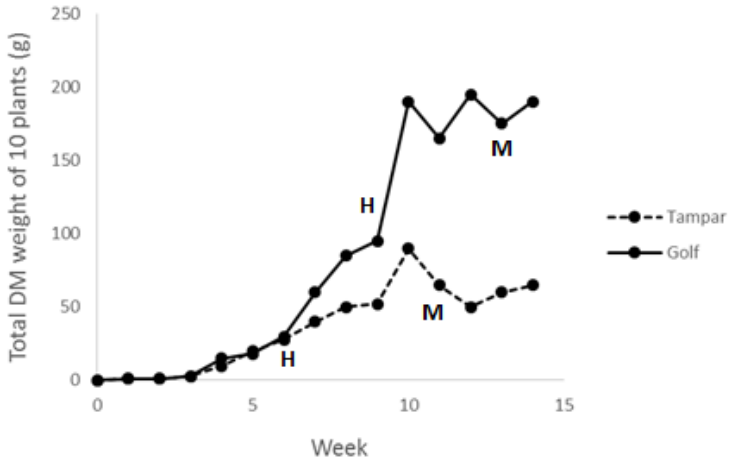


Figure 2. Growth curves of Golf and Tampar (H=heading time, M= maturity, DM= dry matter) (adapted from Aastveit, unpublished data).

Golf has never been experimentally grown on Iceland, but Tampar has been grown and several characteristics have been recorded. In 2014, average height of Tampar was recorded, and found to be 97 cm. This was well above the average height (77 cm) of the other eight lines, among

which three Icelandic cultivars, grown in the same trial. However, apart from this, none of the recorded traits coincide with the traits examined in the present study and no previous Icelandic data are therefore available for either parent line, other than those recorded in the present study (Þórdís Anna Kristjánsdóttir, 2015).

2.2 Phenotyping and genotyping of the study population

The study population was phenotyped for multiple traits during field experiments in 2013, 2014 and 2016 at Korpa Experimental Station (64°09'N, 21°45'W) in South-West Iceland. However, only data from 2014 and 2016 were used for further analysis. Total monthly precipitation and average temperature for Korpa for 2014 and 2016 are given in Figure S1.

In 2012, seeds from each developed line were sown in rows for amplification at Korpa (Þórdís Anna Kristjánsdóttir, 2013). DH seeds originated from the F_1 generation from the cross between Golf and Tampar, while SSD lines were all in the F_7 generation. In 2013, seeds from 2012 were used to sow experimental plots for each line and both parents. An augmented field design with two replications was used. One hundred and twenty lines were planted in both repetitions, but 60 lines only in one of the repetitions due to limited seed supply of these lines. Sixty kg N/ha was applied to all plots, and heading, maturity and height were recorded. During the growing season of 2013, temperatures were extremely low, and the summer especially wet, causing part of the experimental area to be flooded for part of the summer. As a result, recording of data was not always possible and seed harvest was very low. In the winter of 2013-2014, lines with low seed harvest were multiplied in the greenhouse at Korpa to increase the available seed volume (Þórdís Anna Kristjánsdóttir, 2014). In 2014, field experiments were designed as randomized designs with two replications. Plots were each 1.2 m^2 in size, and all 180 lines were tested in both repetitions. Figure S2 shows the experimental design. Sixty kg N/ha of 15-7-12 fertilizer was applied to all plots, and seeds were sown April 29th. The recorded traits include heading, maturity, height, lodging and straw breaking. Plots were harvested September 17th. Seed harvest that year was good, and was used for sowing of the experimental plots in 2016 (Þórdís Anna Kristjánsdóttir, 2015). In 2016, the same experimental field design was used as in 2014. All plots received 64 kg N/ha of fertilizer, and were sown April 27th. The traits heading, maturity, height, lodging, straw breaking, row type and colour were recorded. Plots were harvested September 26th and seeds stored.

The recorded traits analysed in this study are heading, maturity, height and row type. Both heading and maturity were described in two ways: As number of days from sowing, and as

accumulated growing degree days (AGDD). AGDD to heading and maturity were calculated as described by Bauer, Fanning, Enz and Eberlein (1984) and Bauer, Black, Frank and Vasey (1992), using heading and maturity dates in combination with weather data collected at Korpa (Jónatan Hermansson, unpublished data). Figure S3 details the definitions of the traits studied here. Due to bad weather conditions, a large part of the data for 2013 was missing and phenotypic data for 2013 were therefore excluded from further analysis.

In 2016, some lines showed a spike morphology where lateral floret size was intermediate between two- and six-rowed types, and which may therefore be described as an intermedium-spike (Ramsay et al., 2011). However, not in all cases it was clear whether to describe the row type phenotype as two-rowed, intermediate or six-rowed. Therefore, in cases where row type could not clearly be identified as being either six- or two-rowed, or where there were inconsistencies between repetitions, row type was reported as missing. Inconsistencies between repetitions were also found in the years 2013 and 2014. In 2013 and 2014, row type was described as either six-rowed or two-rowed, and in cases where lines were intermediate, this may have led to inaccurate defining of row type. For these years, too, row type data for lines in which inconsistencies between replications were described, were reported as missing. When inconsistencies across years occurred, row type data were reported as missing as well.

All lines were genotyped using the iSelect 9k Barley SNP Chip developed by Comadran et al. (2012). This genotyping platform is based on the Illumina Infinium genotyping assays, and consists of 7842 SNPP assays: 2832 barley OPA SNPs, and 5010 SNPs discovered using Next Generation Sequencing data (Comadran et al., 2012; The James Hutton Institute, 2012). A two cm long piece of a leaf of a young plant from each line grown in the greenhouse was cut off, frozen using liquid nitrogen and then transported on dry ice to TraitGenetics GmbH, Germany, where DNA isolation and genotyping were performed.

2.3 Data analysis

2.3.1 Analysis of field data

Restricted maximum likelihood (REML) variance components analysis was performed in order to find the predicted means for each trait per year. Different mixed models were tested for each trait, and the models with the lowest Akaike Information Criterion (AIC) score were chosen. In each case, ‘line ID’, ‘repetition’ and ‘column’ were treated as fixed factors, while ‘block’ was

treated as random factor. Output was generated using SAS software, version 6.1 Copyright © SAS Institute Inc., Cary, NC, USA.

Two-tailed paired t-tests and Pearson's correlation tests were performed on data recorded for all lines in 2014 and 2016 using JMP software version 13. SAS Institute Inc., Cary, NC, 1989-2017. For traits with a very strong ($r \geq |0.8|$) between-year correlation, the average of the observed measurements in 2014 and 2016 was used for further analysis in addition to year-specific data, while for traits with a weaker correlation each year was only examined individually. Using the same software, Multivariate Regression Analysis was used to analyse the correlation (Pearson's correlation coefficient) between different traits within years in the complete dataset. Furthermore, a Two-Way ANOVA was performed on the complete dataset (with exclusion from parent lines) for all investigated traits in order to assess the effect of line-type and row type on the selected trait. If a trait was not normally distributed, data were first transformed using the natural logarithm or square root. Boxplots showing the variation in phenotypic values per line type and row type were also generated. JMP was also used to calculate the following statistics for each trait: Mean, standard deviation, minimum, maximum, median, coefficient of variation (CV) as well as skewness and kurtosis of the distribution. R version 3.3.1 (www.R-project.org; R Core Team, 2016) and R Studio version 1.0.143 (www.rstudio.com; RStudio Team, 2015) were used to test for normality of distribution using the Shapiro-Wilk test, and for any other analyses performed using R.

2.3.2 Analysis of genotypic data

2.3.2.1 Preliminary analysis and analysis of population structure

Preliminary analysis of the received SNP marker data was performed using TASSEL v.5.2.30 (Bradbury et al., 2007). Quality control filters were applied to the dataset to remove low quality markers and lines: Markers with missing calls in > 22% of lines were removed, as were markers with a minimum allele frequency (MAF) of < 0.05, and lines with missing calls in >20% of the markers and/or a heterozygous proportion of >0.05 were also excluded. In addition, one line was excluded since the origin of the line was unclear. The remaining dataset was used for further population structure and linkage disequilibrium analysis. The same filters were applied to a dataset containing only DH lines and the parent line Tampar (henceforth referred to as the DH dataset), and a dataset containing only SSD lines and the parent line Golf (henceforth referred to as the SSD dataset). Tampar was added to the DH dataset because it was similar to DH lines, and Golf was added to the SSD dataset because it was similar to SSD lines. In addition, a second

version of each dataset was generated where the maximum number of markers per site was limited to four. These datasets were referred to as the 'thinned' datasets. Positions of markers were indicated in cM based on the consensus map provided by Muñoz-Amatriaín et al. (2014). Markers that were not mapped on this map were assigned to an artificial eighth chromosome every 0.2 cM.

Analysis of frequencies of the Golf and Tampar alleles was performed using TASSEL and Microsoft Excel. The SSD and DH Datasets were first sorted by chromosome using TASSEL. Markers for which data was missing for either Golf or Tampar were excluded. Subsequently, Microsoft Excel was used to generate plots of the allele frequency of the Golf respectively Tampar allele along each chromosome. The consensus map by Muñoz-Amatriaín et al. (2014) was used to position each marker. Markers without a position on this map were not included.

Population structure of each dataset was analysed using the software package STRUCTURE v.2.3.4 (Pritchard et al., 2000) under an admixture model. STRUCTURE was run 5 times for each hypothetical number of subpopulations (k) between 1 and 12, with a burn-in period of 10000 and 1000 Monte Carlo Markov Chain (MCMC) iterations. LnP(D) values were plotted and Δk values were calculated according to Evanno, Regnaut and Goudet (2005) to estimate the optimum number of subpopulations using the online Structure Harvester programme version 0.6.94 (Earl & vonHoldt, 2012). Additionally, Principal Component Analysis (PCA) using 5 components was conducted using TASSEL.

2.3.3.2 Analysis of LD and LD decay

Analysis of intra-chromosomal LD was performed by pairwise comparison of all mapped markers for each dataset using TASSEL. LD was calculated as the squared allele frequency correlation r^2 between marker pairs. The r^2 values were then plotted against the corresponding genetic distance between markers using R software, and LD decay was fitted according to the formula by Hill and Weir (1988). The average genome-wide LD decay was visualised using R software by plotting all significant intra-chromosomal r^2 values of all chromosomes against genetic distance and fitting LD decay according to Hill and Weir (1988). A critical r^2 value beyond which LD was assumed to be due to genetic linkage was set to 0.2 in accordance with Bengtsson et al. (2017). LD decay was calculated as the distance at which the fitted LD decay line intersected with LD=0.2, and as the half decay distance ($LD_{1/2}$), both of which using R software. Additionally, an LD plot was generated for chromosome 2H using the Full matrix setting in the Linkage Disequilibrium function of TASSEL. Heterozygous calls were set to

missing. The *Vrs-1* and *HvCEN* loci were identified using linked markers as found through GWAS, and indicated using an arrow.

2.3.2.3 Genome wide association analysis

GWAS was conducted using the enriched compressed mixed linear model (ECMLM) algorithm implemented in the GAPIT R package version 2.0 (Li et al., 2014; Tang et al., 2016). This model makes use of kinship and PCA (using 5 PCs) matrices generated by GAPIT. Two different statistical models were tested on the complete dataset: An ECMLM (K+Q) model and an ECMLM (K+PCA) model. The result of STRUCTURE for k=2 was used as a Q matrix for the complete dataset. The ECMLM (K+PCA) was the only model to be tested on the DH and SSD datasets. Both the unthinned and thinned DH, SSD and complete datasets were tested. Using GAPIT, Quantile-Quantile plots (QQ-plots) were generated for each of the tested models for all traits, and the model for which observed p-values were closest to expected p-values was then chosen. If the QQ-plots for both models were similar, the K+PC model and the unthinned dataset were used. The following traits were analysed: AGDD to heading (2014, 2016, average), AGDD to maturity (2014, 2016), height (2014, 2016) and row type. Normalization of traits that were not normally distributed by natural logarithm or square root transformation did not improve QQ-plots (or only slightly), and therefore no normalization was performed before subjecting the phenotypic data to GWAS. Results were visualised as Manhattan-plots using the 'qqman' package (Turner, 2014) in R software.

Several approaches were considered to determine the significance threshold for p-values. However, the Bonferroni, Bonferroni-Holm and False Discovery Rate (FDR) all proved conservative, creating a threshold only few (if any) markers managed to achieve, variable of trait, year and dataset. Therefore, after sorting markers by p-value, the 0.1 percentile of markers with the smallest p-value were considered significant, as suggested by Chan, Rowe and Kliebenstein (2010). In order to prevent markers with a very high p-value from being considered significant, an additional threshold was set at $-\log(\text{p-value})=2.5$. In addition, the Bonferroni-Holm ($\alpha=0.05$) threshold, calculated according to Holm (1979), was used to identify the most significant markers. Significant markers were initially merged into a single QTL if they were positioned within less than approximately 10 cM from each other according to the consensus map by Muñoz-Amatriaín et al (2014). The highest reported effect of a marker in a QTL was considered the effect of the QTL on a trait. QTL were named after the chromosome they were found on, followed by a number. The prefix AM_DH, AM_SSD or AM_ALL was

added to designate the method by which it was detected and the dataset the QTL was found in. Subsequently, each QTL was additionally positioned by recording the genetic position of each significant marker using the Comadran (Comadran et al, 2012), POPSEQ (Mascher et al., 2013), and IBSC_2012 (IBSC, 2012) genetic maps. The latter two were accessed through the BARLEYMAP pipeline (Cantalapiedra, Boudiar, Casas, Igartua & Contreras-Moreira, 2015). Markers without a position on the consensus map by Muñoz-Amatriaín et al. (2014) were also mapped using these additional maps, and added to the QTL corresponding to their respective location. In the table showing positions of all QTL, these markers were grouped together based on their position, but kept separated from the other markers in the QTL and tagged as '8H'. When no genetic position could be found for a marker, its physical position was looked up using the MorexGenome map available through the BARLEYMAP pipeline (Cantalapiedra et al., 2015).

QTL from all datasets were then combined into a single QTL if their genetic position overlapped according to any of the genetic maps or were positioned within 5 cM of each other. All combined QTL were then named after the chromosome they were found on, followed by a number. In addition, the prefix AM_ was added to designate the method by which the QTL was found.

Potential candidate genes for detected QTL were identified using the BARLEYMAP pipeline (Cantalapiedra et al., 2015). Markers were genetically positioned using the POPSEQ_2012 (Mascher et al., 2013) and IBSC_2012 (IBSC, 2012) maps. Subsequently, the 'enrich map with genes' function was used to identify genes on the interval created by the genetic position of all markers of a QTL. The gene search was then extended by ± approximately 5 cM. Genes known or suggested to be associated with the trait the QTL was associated with, or known to be involved in pathways associated with that trait, were then considered genes of potential interest. Of these genes, the genes for which previously published literature (e.g. Comadran et al., 2012; Maurer et al. 2015; Maurer, Draba & Pillen, 2016; Alqudah et al. 2016) suggested they play an important role in the regulation of the trait, were considered the most likely candidate genes for a QTL. Additionally, the genetic positions of QTL were compared with results from previously published literature such as Comadran et al. (2012), Maurer et al. (2015), Maurer et al. (2016) and Alqudah et al. (2016), in which the same/similar traits were studied as in the present study. When the genetic position of the QTL found in the present study and the genetic position of a QTL for the same trait in a previously published study overlapped, or the QTL were positioned

within a few cM from each other, candidate genes identified in the previously published study were also considered candidate genes for the QTL in the present study.

2.3.2.4 QTL analysis

Linkage maps were generated using Joinmap v. 4 (van Ooijen, 2006). Quality control filters were applied to the original genotypic dataset before the dataset was uploaded into the programme as a RIL (recombinant inbred line) population in the 7^{th} generation. Linkage groups and maps were generated for the SSD and DH datasets separately. The SSD dataset consisted of the SSD lines and both parent lines, and the DH dataset consisted of the DH lines and both parent lines.

QTL mapping was performed using MapQTL version 6 (van Ooijen, 2009). Analysis was performed both using the linkage maps created using DH lines and SSD lines on the following traits: AGDD to heading (2014, 2016, average), AGDD to maturity (2014, 2016), height (2014, 2016) and row type. The mapping of QTL was performed using the MQM (Multiple QTL Mapping) analysis function of the programme. A permutation test was run at 10 000 permutations, and the genome wide interval score at a relative cumulative count of approximately 0.95 was used as the threshold LOD score used to determine whether to regard a QTL as significant or not. QTL were named after the chromosome they were found on, followed by a number. The prefix QTL_DH_ or QTL_SSD_ was added to designate the method used to find the QTL and to indicate using which map the QTL was found. The highest percentage of explained phenotypic variance was recorded for all traits for each QTL. The genetic position of QTL was determined by looking up the position of the significant markers within a QTL on the consensus map by Muñoz-Amatriaín et al. (2014) as well as the Comadran (Comadran et al., 2012), POPSEQ (Mascher et al., 2013) and IBSC_2012 (IBSC, 2012) genetic maps, the latter two of which were accessed through the BARLEYMAP pipeline (Cantalapiedra et al., 2015). QTL overlapping according to their consensus map, POPSEQ and/or IBSC_2012 genetic position or located within 3 cM of each other were combined into a single QTL, which was named after the chromosome it was found on, followed by a number.

LOD score graphs were generated for selected parts of chromosomes using MapChart (Voorrips, 2001).

Potential candidate genes for detected QTL were identified as described for QTL found through GWAS.

In order to compare GWAS and QTL results, number of QTL found using each method were compared. Differences between datasets (DH vs SSD) and years (2014 vs 2016) were examined in the same way.

3. Results

3.1 Field data

Models used to calculate the REML adjusted predicted means for each trait per year showed that line ID significantly affected the REML model for each trait.

Differences between the means of 2014 and 2016 were significantly ($\alpha = 0.05$) different for all traits (Table 1). Apart from row type, which has a complete between-year correlation, the highest between-year correlations were found for heading (days) and heading (AGDD) (both r=0.86).

Table 1. REML adjusted mean values for 2014 and 2016, mean difference, standard error, p-value of a t-test to determine whether there is a significant difference between the mean values of 2014 and 2016, and between-year correlation for all traits between 2014 and 2016.

	Mean					
	2014	**2016**	**Mean diff.**	**Std. Error**	**p > |t|**	**r**
Heading (days)	79	81	2	0.21	< 0.0001	0.86
Heading (AGDD)	788	708	80	2.55	< 0.0001	0.86
Maturity (days)	121	127	6	0.58	< 0.0001	0.55
Maturity (AGDD)	1282	1258	24	6.20	< 0.0001	0.55
Height (cm)	80	69	11	0.66	< 0.0001	0.63
Row type	-	-	-	-	-	1.00

Within-year correlations between traits were highly similar for 2014 and 2016 for all traits (Table 2A and 2B). Almost perfect correlation existed between heading (days) and heading (AGDD) as well as between maturity (days) and maturity (AGDD). Heading and maturity were strongly correlated, but most other correlations were only very weak or moderately strong. Height was very weakly associated with all other traits. In 2016, most correlations with height were non-significant ($\alpha = 0.05$). Apart from a few exceptions, all other correlations were significant.

Table 2A. Correlation (Pearson's coefficient of correlation r) between investigate traits in 2014 and significance of the correlation.

	Heading (days)	Heading (AGDD)	Maturity (days)	Maturity (AGDD)	Height (cm)	Row type
Heading (days)	1					
Heading (AGDD)	0.997***	1				
Maturity (days)	0.620***	0.616***	1			
Maturity (AGDD)	0.623***	0.620***	0.999***	1		
Height (cm)	0.155 *	0.151 *	0.128 NS	0.127 NS	1	
Row type	-0.419***	-0.419***	-0.225 **	-0.233 **	-0.259***	1

NS = Non significant
* P < 0.05 (significant)
** P < 0.01 (highly significant)
*** P < 0.001 (extremely significant)

Table 2B. Correlation (Pearson's correlation coefficient r) between investigated traits in 2016 and significance of the correlation.

	Heading (days)	Heading (AGDD)	Maturity (days)	Maturity (AGDD)	Height (cm)	Row type
Heading (days)	1					
Heading (AGDD)	0.997***	1				
Maturity (days)	0.653***	0.659***	1			
Maturity (AGDD)	0.657***	0.663***	0.996***	1		
Height (cm)	-0.085 NS	-0.080 NS	-0.029 NS	-0.040***	1	
Row type	-0.402***	-0.403***	0.094 NS	0.094 NS	-0.049 NS	1

NS = Non significant
* P < 0.05 (significant)
** P < 0.01 (highly significant)
*** P < 0.001 (extremely significant)

The boxplots in Figure 3 display the variation and distribution of phenotypic values observed for each trait by line type and row type. All traits appeared to be approximately normally distributed. Normality tests run on the phenotype data showed however that heading (days) was not normally distributed in any of the datasets (Table 3. A-C, page 28-29). Heading (AGDD) was normally distributed in the complete dataset, but in the DH and SSD datasets this was only the case in 2016. Maturity (days) and maturity (AGDD) were normally distributed in all datasets. Height was normally distributed except in the complete dataset and in 2014 in the SSD dataset. Figure 3 further shows differences in phenotypic variation and distribution between DH and SSD lines, as well as between two-rowed and six-rowed lines. Six-rowed lines showed on average increased earliness of heading compared to two-rowed lines. Six-rowed lines were also earlier mature in 2014, but in 2016, six-rowed DH lines were on average slightly later mature than two-rowed lines.

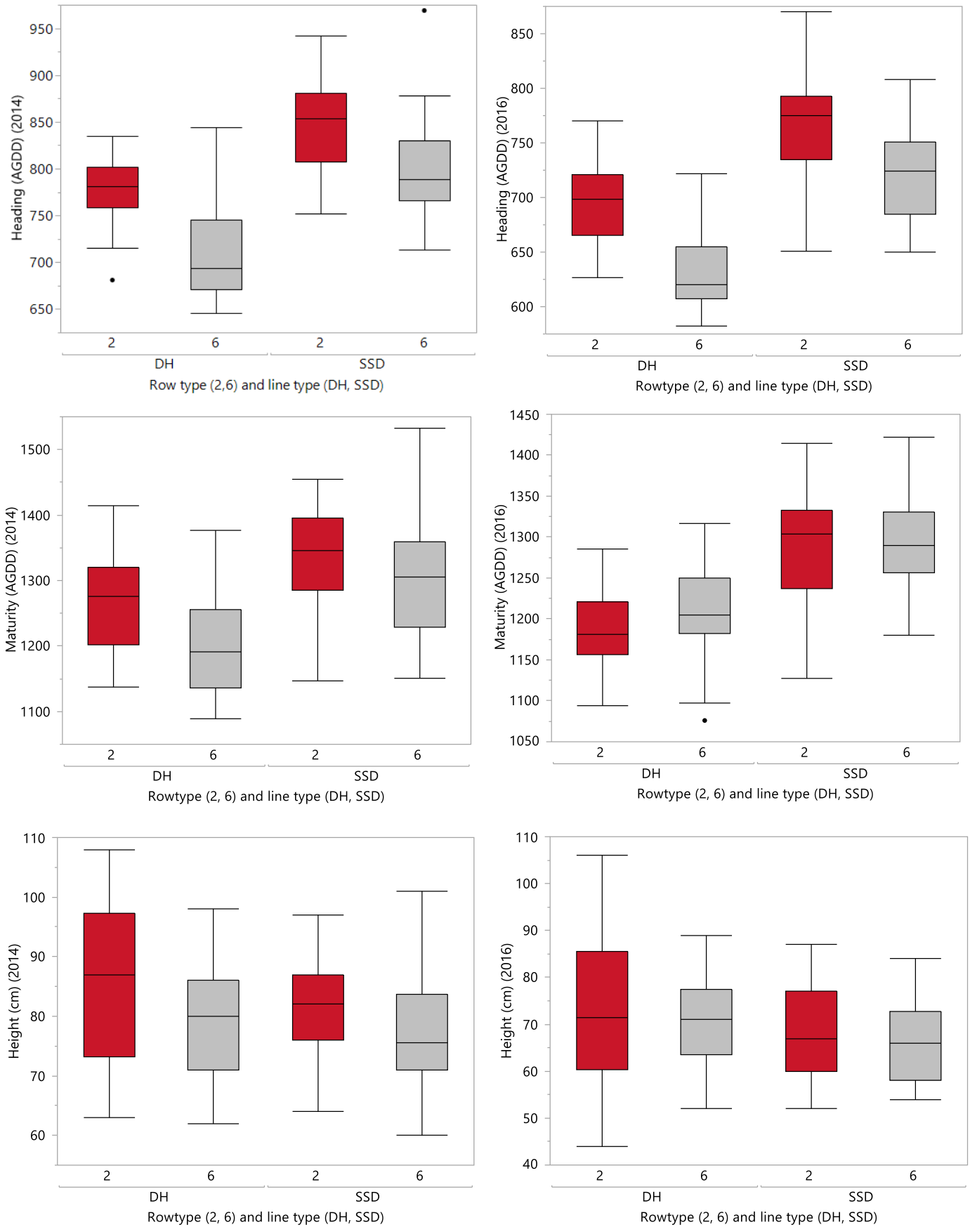


Figure 3. Boxplots showing distribution of observed phenotypic variance per trait, by line type (DH and SSD) and row type (two-rowed in red and six-rowed in grey) for 2014 and 2016.

More detailed descriptive statistics of all investigated traits are shown in Table 3 A, B and C. The average number of days to heading and maturity was slightly lower in 2014 than in 2016, but the average AGDD to heading and maturity was higher in 2014 than in 2016 in each dataset. Both average number of days and average AGDD to heading and maturity were lower in the DH dataset than the SSD dataset. Height ranged from 44 to 108 cm. Average height was greater in 2014 than in 2016, and was slightly greater in the DH dataset than the SSD dataset. Both the DH and SSD dataset contained more six-rowed lines than two-rowed lines.

Table 3. A-C. Summary statistics (minimum, median, mean, maximum standard deviation, variance, coefficient of variance (%)) as well as results of the Shapiro-Wilk Normality test for each of the phenotyped traits in 2014 and 2016, and for the average of these years for the heading, maturity and height traits in the complete, DH and SSD datasets respectively.

All lines + parents		**Summary statistics**											**Normality test (Shapiro-Wilk)**	
Trait	**Dataset**	**Min.**	**Median**	**Mean**	**Max.**	**Std. Dev.**	**Var.**	**CV(%)**	**Skewness**	**Kurtosis**	**NA***	**2/6 row****	**W**	**p-value**
Heading (days)	**2014**	67	79	79	93	5.3	28	6.72	-0.20	-0.15	-	-	0.982	0.018
	2016	70	82	81	94	4.8	23	5.89	-0.24	-0.52	-	-	0.980	0.012
	average	69	81	81	92	4.9	24	6.05	-0.27	-0.39	-	-	0.979	0.008
Heading (AGDD)	**2014**	646	786	788	969	65	4225	8.26	-0.04	-0.14	-	-	0.987	0.123
	2016	568	713	708	870	61	3721	8.72	-0.11	-0.63	-	-	0.985	0.059
	average	608	746	748	898	61	3721	8.16	-0.44	-0.10	-	-	0.987	0.102
Maturity (days)	**2014**	102	122	121	145	8.8	77	7.25	0.07	-0.42	2	-	0.992	0.389
	2016	111	128	127	144	7.1	50	5.58	0.15	-0.51	-	-	0.992	0.385
	average	109	124	125	144	7.0	49	5.65	0.11	-0.33	-	-	0.995	0.747
Maturity (AGDD)	**2014**	1068	1290	1282	1533	96	9216	7.50	0.04	-0.50	-	-	0.991	0.296
	2016	1076	1262	1259	1422	74	5476	5.89	-0.03	-0.55	-	-	0.987	0.093
	average	1094	1265	1271	1464	75	5625	5.92	-0.01	-0.43	-	-	0.993	0.568
Height (cm)	**2014**	60	79	80	108	9.8	96	12.2	0.46	-0.12	-	-	0.979	0.007
	2016	44	68	69	106	11	114	15.5	0.34	-0.22	-	-	0.983	0.026
	average	56	74	74	107	9.2	85	12.4	0.46	0.00	-	-	0.976	0.003
Row-type	**average**	-	-	-	-	-	-	-	-	-	16	60/98	-	-

* Number of missing values

** Number of two-rowed lines/number of six-rowed lines

DH lines + Tampar		Summary statistics											Normality test (Shapiro-Wilk)	
Trait	**Dataset**	**Min.**	**Median**	**Mean**	**Max.**	**Std. Dev.**	**Var.**	**CV(%)**	**Skewness**	**Kurtosis**	**NA***	**2/6 row****	**W**	**p-value**
Heading (days)	**2014**	67	76	75	85	4.9	24	6.47	-0.08	-1.12	-	-	0.952	0.011
	2016	70	77	78	86	4.1	17	5.28	0.15	-0.92	-	-	0.968	0.078
	average	69	77	77	85	4.3	18	5.55	-0.02	-1.18	-	-	0.949	0.008
Heading (AGDD)	**2014**	646	746	740	858	56.7	3215	7.67	-0.03	-1.12	-	-	0.956	0.018
	2016	568	653	659	770	50.8	2581	7.01	0.29	-0.83	-	-	0.968	0.086
	average	608	701	670	809	50.6	2560	7.22	0.07	-1.09	-	-	0.959	0.028
Maturity (days)	**2014**	102	116	116	133	7.5	56	6.46	0.13	-0.80	2	-	0.973	0.159
	2016	111	121	122	133	4.6	22	3.82	0.10	-0.18	-	-	0.986	0.645
	average	109	120	119	129	4.9	24	4.16	-0.07	-0.62	-	-	0.982	0.434
Maturity (AGDD)	**2014**	1068	1227	1234	1414	82	6801	6.74	0.13	-0.79	-	-	0.975	0.206
	2016	1076	1198	1200	1317	51	2634	4.28	-0.06	-0.15	-	-	0.992	0.951
	average	1094	1218	1212	1326	55	3015	4.53	-0.12	-0.53	-	-	0.987	0.734
Height (cm)	**2014**	62	83	82	108	11.3	127	13.7	0.32	-0.51	-	-	0.974	0.166
	2016	44	72	72	106	12	147	16.9	0.13	-0.06	-	-	0.992	0.936
	average	56	78	77	107	10.8	117	14.1	0.30	-0.26	-	-	0.975	0.193
Row-type	**average**	-	-	-	-	-		-	-	-	5	24/38	-	-

* Number of missing values

** Number of two-rowed lines/number of six-rowed lines

SSD lines + Golf		Summary statistics											Normality test (Shapiro-Wilk)	
Trait	**Dataset**	**Min.**	**Median**	**Mean**	**Max.**	**Std. Dev.**	**Var.**	**CV(%)**	**Skewness**	**Kurtosis**	**NA***	**2/6 row****	**W**	**p-value**
Heading (days)	**2014**	73	81	81	93	4.1	17	5.07	0.35	-0.44	-	-	0.968	0.009
	2016	76	84	84	94	3.6	13	4.31	-0.03	-0.45	-	-	0.976	0.036
	average	76	83	83	92	3.6	13	4.38	0.16	-0.80	-	-	0.970	0.011
Heading (AGDD)	**2014**	713	811	816	969	52.3	2740	6.42	0.46	-0.34	-	-	0.969	0.009
	2016	637	738	737	870	48.0	2300	6.51	0.02	-0.52	-	-	0.980	0.091
	average	690	771	776	898	47.4	2250	6.11	0.24	-0.76	-	-	0.970	0.011
Maturity (days)	**2014**	108	125	124	145	7.9	63	6.37	0.07	-0.31	-	-	0.990	0.610
	2016	116	130	131	144	6.2	38	4.72	-0.06	-0.19	-	-	0.985	0.296
	average	113	128	128	144	6.0	36	4.69	0.08	-0.12	-	-	0.993	0.837
Maturity (AGDD)	**2014**	1147	1324	1317	1533	86		6.56	0.03	-0.46	-	-	0.989	0.507
	2016	1127	1293	1294	1422	62	3884	4.82	-0.29	-0.04	-	-	0.987	0.339
	average	1144	1304	1306	1464	63	3928	4.80	-0.06	-0.29	-	-	0.995	0.983
Height (cm)	**2014**	60	78	79	101	8.6	74	10.9	0.34	-0.24	-	-	0.983	0.172
	2016	52	67	67	87	9.4	89	14.0	0.26	-0.98	-	-	0.957	0.001
	average	57	72	73	92	7.8	62	10.8	0.23	-0.81	-	-	0.975	0.035
Row-type	**average**	-	-	-	-	-		-	-	-	11	42/60	-	-

* Number of missing values

** Number of two-rowed lines/number of six-rowed lines

These relationships were investigated further using a Two-Way ANOVA on all traits (Table 4), which showed that line type and row type significantly affected heading (AGDD) in 2014, 2016 and the average of both years and maturity (AGDD) in 2016. Further, line type significantly affected height in 2016 while row type significantly affected height in 2014. Row-type and line type explained from 3% to as much as 56% percent of observed variation for a trait.

Table 4. Results of a Two-Way ANOVA showing the effect of line type and row type on each trait.

Trait	Variable	Least square mean	P-value	Rsquare
Heading 14* (AGDD)	**SSD**	822	<.0001	0.51
	DH	743		
	2	809	<.0001	
	6	756		
Heading 16* (AGDD)	**SSD**	743	<.0001	0.55
	DH	664		
	2	728	<.0001	
	6	679		
Heading AV* (AGDD)	**SSD**	783	<.0001	0.56
	DH	704		
	2	769	<.0001	
	6	717		
Maturity 14* (AGDD)	**SSD**	1321	<.0001	0.27
	DH	1231		
	2	1298	0.001	
	6	1254		
Maturity 16* (AGDD)	**SSD**	1293	<.0001	0.40
	DH	1197		
	2	1253	0.085	
	6	1237		
Height 14* (Ln cm)	**SSD**	4.37	0.052	0.09
	DH	4.40		
	2	4.42	0.001	
	6	4.35		
Height 16* (Ln cm)	**SSD**	4.20	0.025	0.03
	DH	4.25		
	2	4.23	0.518	
	6	4.22		

* 14 = 2014, 16 = 2016, AV = average

Table 5 shows all recorded phenotypic values for both parent lines during 2014 and 2016. Large differences were observed between parents: In 2016, heading in Tampar was on average 16 days (or approximately 200 AGDD) earlier than in Golf. In 2014, the difference was similar to

this in one of the repetitions but completely different in the other, this is most likely due to an error during data collection. In 2014, maturity was reached approximately 25 days (approximately 255 AGDD) earlier in Tampar than in Golf, while in 2016, Tampar was only about 16 days (approximately 135 AGDD) earlier. AGDD to heading was much longer in 2014 than in 2016. Tampar was up to 20 cm higher than Golf, but never smaller. Tampar is six-rowed while Golf is two-rowed.

Table 5. Phenotypic values of Golf and Tampar per replication observed in the field in 2014 and 2016, and REML adjusted averages per year.

Trait	Dataset	Replication	Golf	Tampar	Trait	Dataset	Replication	Golf	Tampar
Heading (days)	2014	1	85	66	Heading (AGDD)	2014	1	863	637
		2	85	94			2	863	978
		average	84	80			average	856	806
	2016	1	88	72		2016	1	794	593
		2	85	69			2	752	557
		average	86	70			average	768	571
Maturity (days)	2014	1	125	105	Maturity (AGDD)	2014	1	1329	1107
		2	125	101			2	1329	1042
		average	127	102			average	1346	1068
	2016	1	140	123		2016	1	1386	1221
		2	129	119			2	1277	1172
		average	134	118			average	1328	1163
Height (cm)	2014	1	60	80					
		2	70	75					
		average	67	78	Row-type			2	6
	2016	1	55	80					
		2	70	70					
		average	61	77					

3.2 Genotyping

Preliminary analysis of the received genotyping data showed that 6993 markers were successfully genotyped in one or more lines. After quality control filters were applied, the remaining complete dataset contained 175 lines, while the DH and SSD datasets contained 66 and 109 lines respectively. The proportion of heterozygosity was similar in all three datasets, ranging from 0.0036 in the complete and DH datasets to 0.0051 in the SSD dataset. The number of failed genotyping calls only varied slightly. The proportion of missing data was 0.0036 in the complete dataset, 0.0026 in the DH dataset and 0.0034 in the SSD dataset. Marker coverage and distribution across all chromosomes are shown in Table 6. In the complete and DH datasets, markers spanned 1100 and 1102.4 cM respectively across all 7 chromosomes, while they only

covered 1074 cM in the SSD dataset. In all three datasets, marker density (average number of markers/cM) was highest in chromosome 6H, and lowest in chromosome 1H. On average, marker coverage (total genetic distance covered by the markers) was lowest in the SSD set and highest in the complete dataset.

Table 6. Marker coverage and distribution across all chromosomes of the datasets used for population structure and linkage disequilibrium decay analysis.

Chromosome	Coverage (cM)			Number of markers			Marker coverage (cM/markers)		
	DH	SSD	ALL	DH	SSD	ALL	DH	SSD	ALL
1H	140.30	141.38	141.4	199	137	250	0.71	1.03	0.57
2H	177.43	173.42	174.4	351	377	441	0.51	0.46	0.40
3H	164.42	153.89	164.4	424	276	490	0.39	0.56	0.34
4H	128.18	128.18	128.2	231	189	251	0.55	0.68	0.51
5H	183.53	179.89	183.5	499	378	547	0.37	0.48	0.34
6H	139.39	139.39	139.4	397	302	437	0.35	0.46	0.32
7H	168.94	157.99	168.9	317	214	382	0.53	0.74	0.44
8H				676	673	888			
Total/average	1102.2	1074.14	1100	3094	2546	3686	0.49	0.63	0.41

Before application of quality control filters, SSD lines contained a higher number of markers for which genotyping failed than DH lines, resulting in a lower number of markers available for subsequent analysis.

Analysis of the frequency of the Golf- and Tampar alleles along each chromosome (Figure S4) showed that allele frequencies often vary from the expected 0.5, and that the pattern with which they fluctuate varies greatly per chromosome. In the SSD dataset, both the frequency of the Golf- and the Tampar allele fluctuated more or less around the 0.5 mark for chromosomes 1H, 3H, 4H and 7H. However on chromosome 2H, the frequency of the Tampar allele was higher than the frequency of the Golf allele for a long part of the chromosome. Vice versa, on chromosome 6H the frequency of the Golf allele was much higher than the frequency of the Tampar allele for a long distance. The allele frequency plots of the DH dataset looked very different from the plots for the SSD dataset. Golf- and Tampar frequencies did not fluctuate as much among each other, with one allele often maintaining a (much) higher frequency for a longer distance or even (almost) along the entire chromosome as was the case for chromosomes 1H, 6H and 7H, than the other. No specific pattern could be detected in the plots for the DH or

the SSD datasets at sites where significant QTL were detected. Furthermore, total frequency of the Golf and Tampar alleles combined did not always reach 1.0.

3.3 Population structure

Analysis of population structure using the complete dataset in STRUCTURE showed that the estimated natural log probability (LnP(D)) of the data for each k between 1 and 12 increased continuously without reaching a plateau. The maximum Δ*k* value was reached at *k*= 2 (Figure S5). Similarly, PCA performed using TASSEL showed that the first Principal Component (PC) separated the population into two subpopulations, containing DH and SSD barley lines respectively (Figure 4). When STRUCTURE was run on the DH and SSD datasets separately, no substructure was evident (Figures S6 and S7). PCA was also performed on the DH and SSD datasets separately, but no clear subgroups were found in the PC1xPC2 plots of either dataset. However, plots did indicate a slight partitioning of the germplasm based on row type in both the DH and SSD datasets (Figure S8).

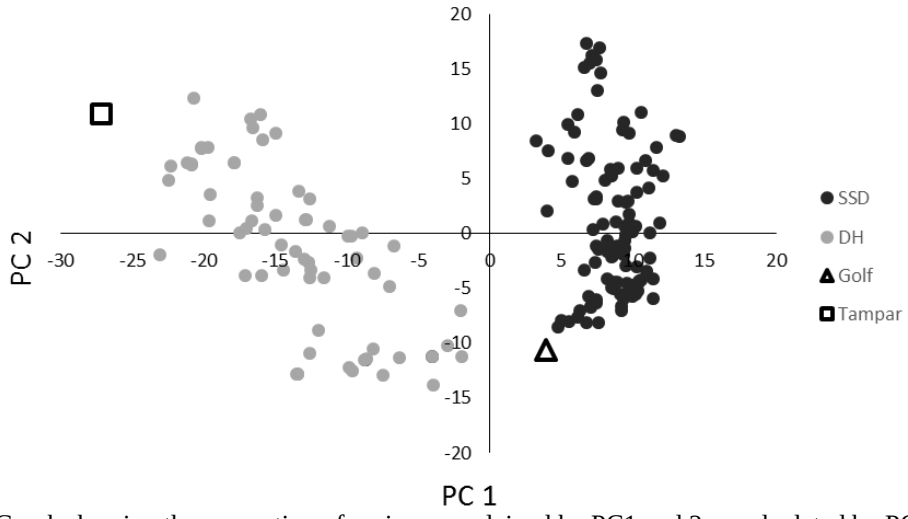


Figure 4. Graph showing the proportion of variance explained by PC1 and 2 as calculated by PCA analysis using TASSEL. Accessions are coloured by line type and parental lines indicated with symbols.

3.4 Linkage disequilibrium decay mapping

The LD decay calculated using LD decay mapping was more rapid in the complete and SSD datasets than in the DH dataset (Figure 5). The genome-wide half-decay value was 11.68 in the SSD dataset, 13.02 in the complete dataset and 27.02 in the DH dataset. The complete dataset contained a larger number of pair-wise marker linkages with an r^2 of approximately 0.2-0.5 compared to the SSD and DH lines, and in the DH dataset, fewer short-distance pair-wise marker linkages with LD< 0.2 were detected than in the other datasets.

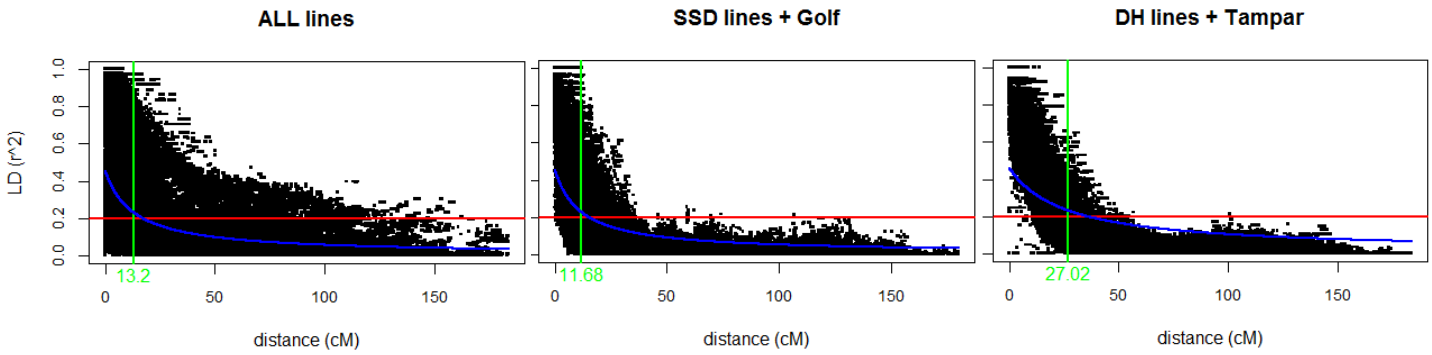


Figure 5. Intra-chromosomal LD decay (r^2) of marker pairs averaged over all chromosomes for each data set. A critical r^2 value beyond which LD was assumed to be due to genetic linkage was arbitrarily set to 0.2 (red line). The green line indicates the half-decay distance, and the blue curve shows the LD decay as a function of genetic distance (cM), as fitted by the formula by Hill and Weir (1988). From left to right: The complete dataset, the SSD dataset and the DH dataset.

Calculated intra-chromosomal LD decay varied per chromosome, and per dataset (Table 7). The half-decay distance was on average shorter than the distance to where the fitted LD decay line crossed LD=0.2. Further, intra-chromosomal LD decay was slower in the DH dataset than in the other datasets and large differences in LD decay were observed between chromosomes. In all three datasets, LD decay was fastest in chromosome 3H. The slowest LD decay was observed in chromosomes 6H and 2H of DH lines.

Table 7. Intra-chromosomal LD decay for each dataset given as half-decay distance (LD 1/2) and distance to LD=0.2 (both in cM).

	DH		SSD		ALL	
Chromosome	**LD 1/2**	**LD=0.2**	**LD 1/2**	**LD=0.2**	**LD 1/2**	**LD=0.2**
1H	19.15	25.21	16.16	21.01	12.15	15.91
2H	30.88	40.70	12.21	15.84	10.73	14.00
3H	26.22	34.56	12.98	16.83	13.39	17.00
4H	19.00	25.03	6.56	8.50	8.16	10.53
5H	19.57	25.78	12.88	16.69	10.38	13.37
6H	44.86	59.13	9.38	12.18	24.52	31.62
7H	25.63	33.77	12.85	16.65	12.36	15.91
Genome-wide	27.07	35.61	11.68	15.15	13.02	17.00

LD plots of chromosome 2H showed that The *HvCEN* and *Vrs-1* loci were linked. In the complete dataset, r^2 between these loci ranged from 0 – 0.2. However in the DH dataset, r^2 ranged from 0.4 – 0.7, and in the SSD dataset from 0 – 0.4 (Figure S9).

3.5 Genome wide association analysis (GWAS)

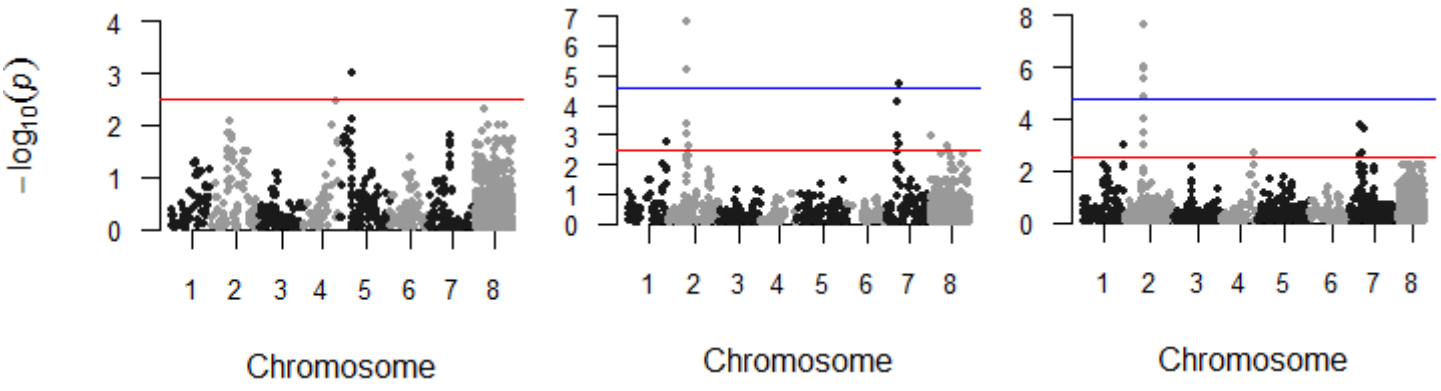
3.5.1 Quality and reliability of GWAS results

Visual inspection of QQ-plots showed that the QQ-plots for models containing a PC matrix as a measure of population structure gave better (i.e. showed smaller differences between observed- and expected p-values) or similar results compared to models where the PCA matrix was replaced by a Q matrix. Therefore only the ECMLM (PCA+K) model was used for GWAS analysis. Application of stricter quality filters (i.e. higher MAF and lower proportion of heterozygous or missing data) did on average not or only slightly improve (occasionally even worsen) QQ-plots. However, thinning of the dataset improved QQ-plots for the SSD and complete datasets: Although some increased inflation of the smallest p-values was observed, the remainder of observed p-values was closer to the expected p-values than in the unthinned datasets. This was not the case for the DH dataset, where using the thinned dataset did not improve QQ-plots. QQ plots are shown in Figure S10A-H.

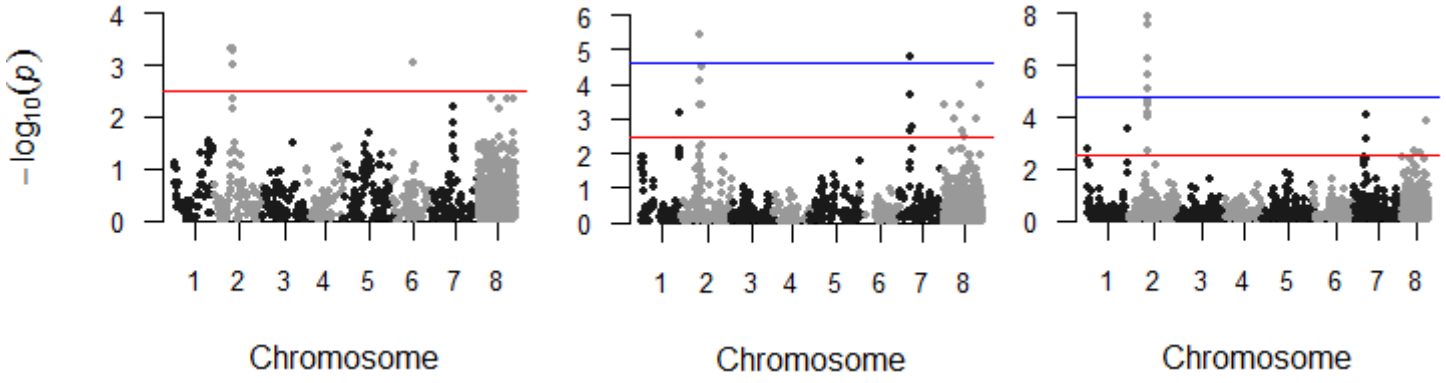
Using GWAS, numerous associated genomic regions were detected for all investigated traits, as shown in the Manhattan plots in Figure 6. Considerable differences were observed between results for the DH, SSD and complete dataset respectively. P-values observed in the DH dataset were on average higher, and peaks in the Manhattan plots could less clearly be defined than in the other datasets. Further, most markers failed to meet the –log(p-value) ≥ 2.5 significance threshold, while in the SSD and complete datasets more markers met this threshold.

Figure 6. Manhattan plots for each trait, in each tested subset of lines. The red line indicates the $-\log_{10}$(p-value)= 2.5 significance threshold, and the blue line indicates the Bonferroni-Holm significance threshold.

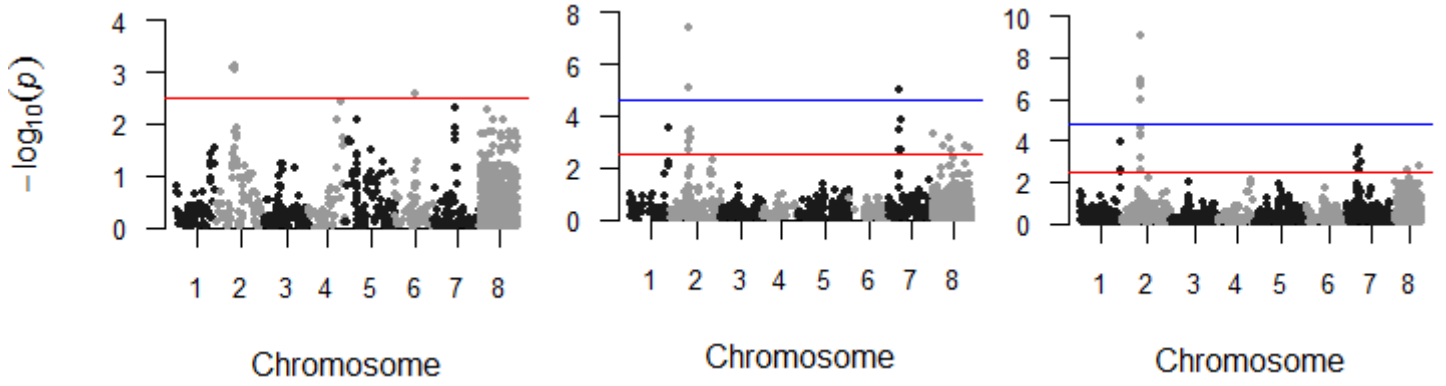
A**).** Heading (AGDD) (2014). From left to right: Plots for the DH, SSD and complete datasets.



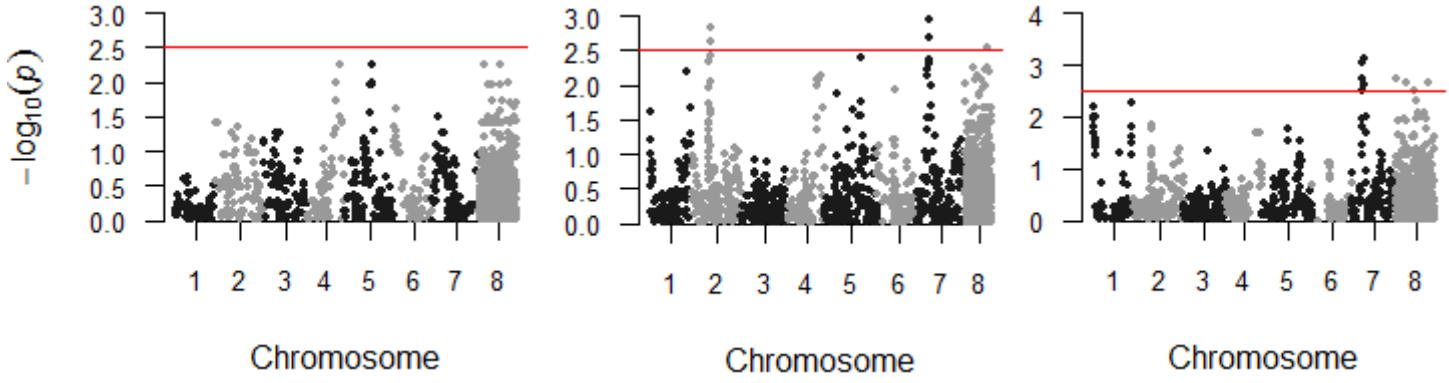
B). Heading (AGDD) (2016). From left to right: Plots for the DH, SSD and complete datasets



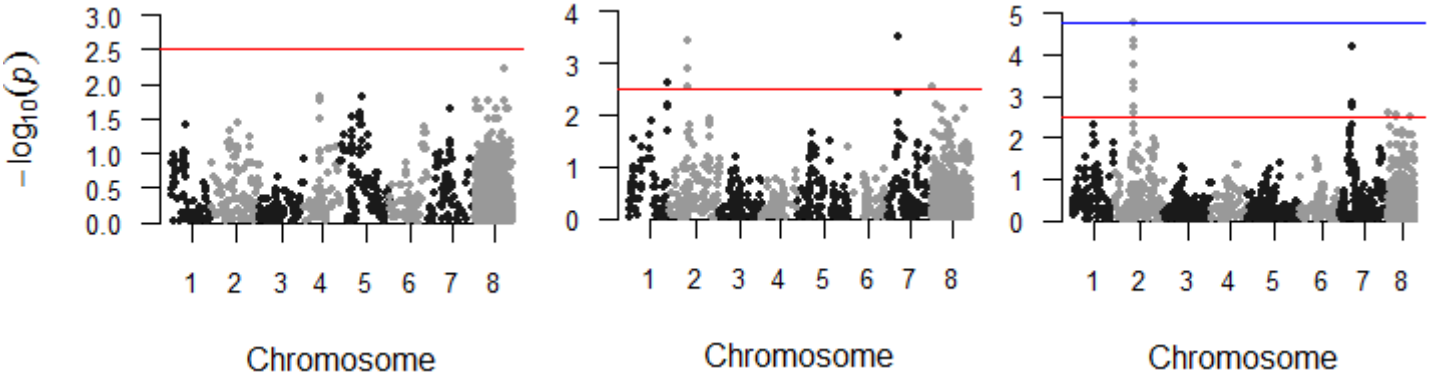
C). Heading (AGDD) (AV). From left to right: Plots for the DH, SSD and complete datasets



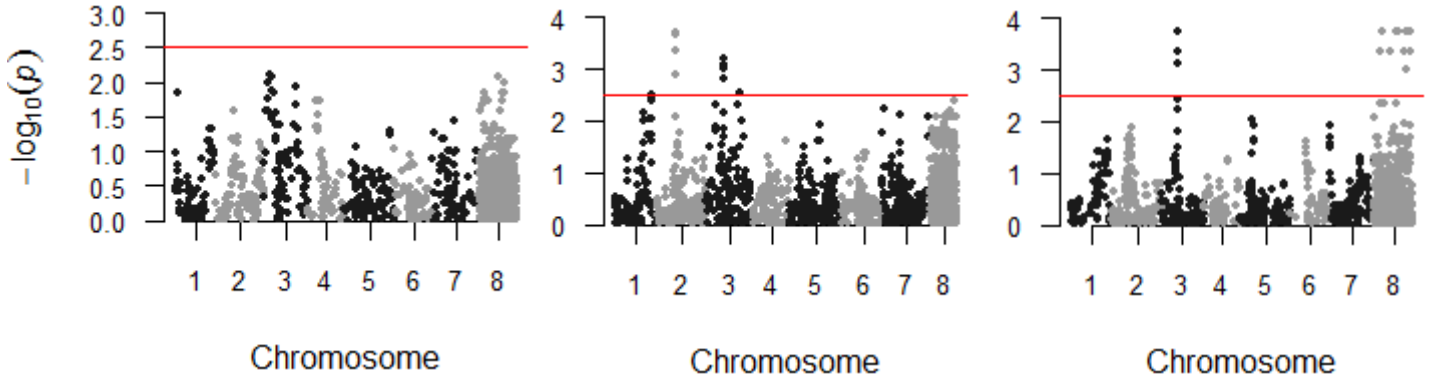
D). Maturity (AGDD) (2014). From left to right: Plots for the DH, SSD and complete datasets.



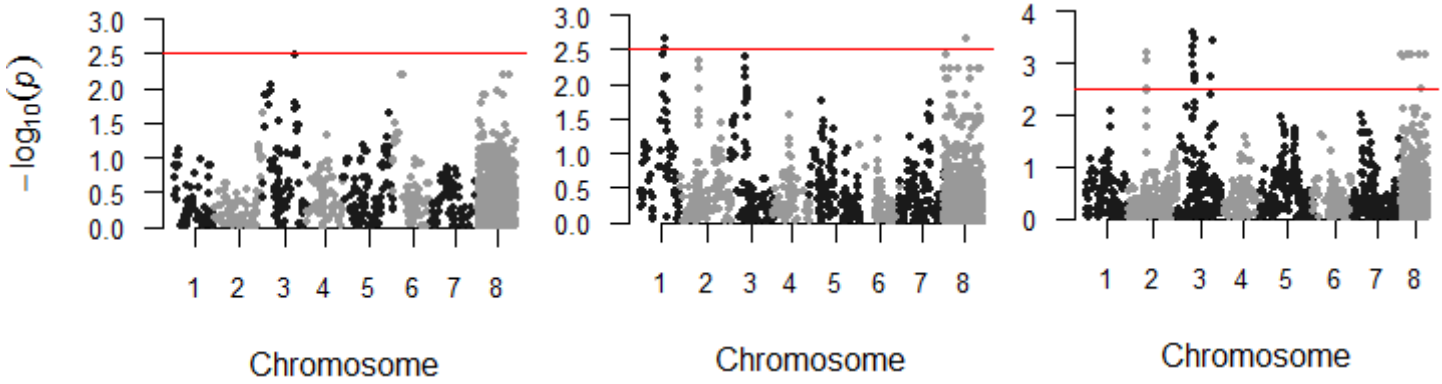
E). Maturity (AGDD) (2016). From left to right: Plots for the DH, SSD and complete datasets.



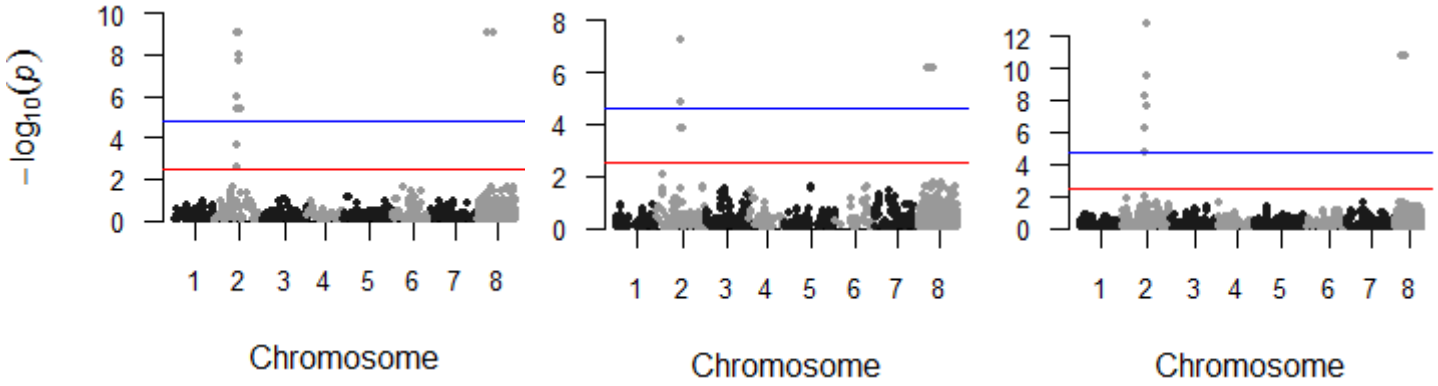
F). Height (2014). From left to right: Plots for the DH, SSD and complete datasets.



G). Height (2016). From left to right: Plots for the DH, SSD and complete datasets.



H). Row type. From left to right: Plots for the DH, SSD and complete datasets.



QQ-plots showed that both inflation and deflation of $-\log_{10}$(p-values) occurred (Figure S10). Deviation from expected p-value distribution was largest for heading in the SSD and complete datasets, where considerable inflation of $-\log_{10}$(p-values) occurred. Manhattan plots for heading looked similar in 2014, 2016 and in the average of these years, and peaks could reasonably well be defined in all three datasets. In 2014, QQ-plots for maturity were better (i.e. observed p-values were closer to expected p-values) than in 2016. In 2016, slight deflation of $-\log_{10}$(p-values) was observed in the DH dataset, but inflation in the other datasets. Peaks in the Manhattan plots could more easily be defined in the complete and SSD datasets than in the DH dataset. P-values were on average higher in 2014 than in 2016, and differences in the appearance of the Manhattan plots between years were larger than for heading. For height, the QQ-plots for the DH dataset showed deflation of $-\log_{10}$(p-values) while the SSD and complete datasets showed (slight) inflation of $-\log_{10}$(p-values). Although more peaks were observed in 2016 than in 2014, these peaks were less clearly defined than those in 2014. The Manhattan plots for row type showed one significant peak at chromosome 2H, while the QQ-plots showed a large inflation of the smallest $-\log_{10}$(p-values). Manhattan plots of the DH, SSD and complete datasets looked very similar.

3.5.2 Detection of QTL for growth and development

The results from GWAS performed on the DH dataset showed that 92 different markers were significantly associated with at least one of the investigate traits. As shown in Table 8, a total of five QTL were detected in the DH dataset, distributed over chromosomes 2H, 3H, 5H and 6H. Three QTL were detected for heading, one for height and one for row type. All QTL were trait specific, and detected only in one of the years the trait was phenotyped. Not one QTL met the Bonferroni-Holm (0.05) significance threshold. The percentage of variance explained by a

QTL varied from 5.8 to 10.1%. No significant markers were detected for maturity in both 2014 and 2016, or for height in 2014.

Table 8. QTL found in the DH dataset, their position according to the consensus map by Muñoz-Amatriaín et al. (2014) and maximum percentage of phenotypic variance explained by each QTL.

				Explained variance (%)				
		cM Interval		Heading			Height	Row
QTL	Chr.	From	Until	2014	2016	AV	2016	type
AM_DH_2H.1	2H	69.0	69.6	-	8.2	7.2	-	-
AM_DH_2H.2	2H	85.2	99.0	-	-	-	-	8.3
AM_DH_3H.1	3H	122.0	122.0	-	-	-	10.1	-
AM_DH_5H.1	5H	45.0	45.0	9.3	-	-	-	-
AM_DH_6H.1	6H	74.2	74.2	-	7.3	5.8	-	-

A total of 104 markers in the SSD were significantly associated with at least one trait. Six different QTL were found in the SSD dataset, distributed over chromosomes 1H, 2H, 3H and 7H (Table 9). Three QTL were found for heading, three for maturity, three for height and one for row type. Three QTL were trait specific, and three QTL were shared between traits. Further, QTL were not always found in both years that the trait was phenotyped in. All three heading QTL were however found in both 2014 and 2016 as well as the average of these years. AM_SSD_2H.1, AM_SSD_2H.2 and AM_SSD_7H.1 met the Bonferroni-Holm (0.05) significance threshold. The percentage of variance explained by each QTL varied from 6.1 to 19.3%.

Table 9. QTL found in the SSD dataset, their position according to the consensus map by Muñoz-Amatriaín et al. (2014) and maximum percentage of phenotypic variance explained by each QTL. Numbers in bold indicate where the Bonferroni-Holm (0.05) threshold was met.

				Explained variance (%)							
		cM Interval		Heading			Maturity		Height		Row
QTL	Chr.	From	Until	2014	2016	AV	2014	2016	2014	2016	type
AM_SSD_1H.1	1H	88.82	89.34	-	-	-	-	-	-	7.2	-
AM_SSD_1H.2	1H	142.7	142.7	6.1	6.3	6.9	-	7.0	-	-	-
AM_SSD_2H.1	2H	69.00	74.49	**19.3**	**12.3**	**17.3**	-	6.8	-	-	-
AM_SSD_2H.2	2H	86.44	95.24	-	-	-	-	-	-	-	**11.8**
AM_SSD_3H.1	3H	58.31	58.31	-	-	-	-	-	9.4	-	-
AM_SSD_7H.1	7H	34.74	45.14	**12.1**	10.6	10.6	8.6	10.1	7.8	-	-

In the complete dataset, GWAS analysis resulted in the detection of 125 markers that were significantly associated with at least one trait. A total of nine QTL were detected, distributed over chromosomes 1H, 2H 3H, 4H and 7H (Table 10). Five QTL were found for heading, two

for maturity, four for height and one for row type. Six QTL were trait specific, and only three QTL were shared between traits. Furthermore, three QTL were detected in one year only. AM_ALL_2H.1, and AM_ALL_2H.2 were the only QTL that did not fail to meet the Bonferroni-Holm (0.05) significance threshold. The percentage of variance explained by a QTL varied from 1.9 to 9.3%. In addition, one marker without a position on the consensus map was found to be significantly associated with maturity in 2014, explaining 3.5% of phenotypic variance. This marker is indicated as a QTL on artificial chromosome 8H, but was found to be located on chromosome 5H using other genetic maps.

Table 10. QTL found in the complete dataset, their position according to the consensus map by Muñoz-Amatriaín et al. (2014) and maximum percentage of phenotypic variance explained by each QTL. Numbers in bold indicate where the Bonferroni-Holm (0.05) threshold was met.

				Explained variance (%)							
		cM Interval		**Heading**			**Maturity**		**Height**		**Rowtype**
QTL	**Chrom.**	**From**	**Until**	**2014**	**2016**	**AV**	**2014**	**2016**	**2014**	**2016**	
AM_ALL_1H.1	1H	5.68	5.68	-	1.9	-	-	-	-	-	-
AM_ALL_1H.2	1H	133	133	-	-	-	-	-	3.3	-	-
AM_ALL_1H.3	1H	143	143	2.8	2.6	2.9	-	-	-	-	-
AM_ALL_2H.1	2H	69.0	74.5	**8.9**	**6.7**	**8.0**	3.6	6.1	5.4	5.6	-
AM_ALL_2H.2	2H	86.4	95.2	-	-	-	-	-	-	-	**9.3**
AM_ALL_3H.1	3H	55.4	61.9	-	-	-	-	-	4.5	6.4	-
AM_ALL_3H.2	3H	117.58	119.77	-	-	-	-	-	3.4	5.0	-
AM_ALL_4H.1	4H	117.1	117.1	2.6	-	-	-	-	-	-	-
AM_ALL_7H.1	7H	35.0	45.1	3.8	3.0	2.4	4.1	5.2	-	-	-
AM_ALL_8H.1	5H	-	-	-	-	-	3.5	-	-	-	-

3.5.3 Positioning of detected QTL

The genetic position of each QTL based on four different genetic maps is shown in Figure S11. QTL which overlapped or were positioned closer to each other than 5 cM according to either of the genetic maps are shown as combined QTL. Four different combined QTL were found on chromosome 1H, two QTL on 2H, two on chromosome 3H and one QTL on chromosome 4H, 5H, 6H and 7H each. In addition, QTL AM_8H.1 was positioned on chromosome 5H based on positioning according to the Morex_2016 physical map. In total, 13 combined QTL were detected during GWAS. Markers associated with each QTL are shown in Figure S12.

3.5.4 Candidate genes at detected QTL

As shown in Table 11, candidate genes known to be associated with the investigated traits were found for 11 of the 13 QTL detected using GWAS. In addition, for most QTL, a number of genes of possible interest were identified as well.

Table 11. Combined QTL, QTL and their associated traits, along with the most likely candidate genes for each combined QTL and other genes potentially of interest.

Combined QTL	Significantly associated traits*	Chr.	Most likely candidate gene(s)	Other genes potentially of interest
AM_1H.1	HD(16)	1H	-	-
AM_1H.2	HT(16)	1H	*HvHXK5*	-
AM_1H.3	HT(14)	1H	*Gibberellin 2-oxidase*	*Auxin responsive protein, MYB TFs*
AM_1H.4	HD(14, 16, AV), MT(16)	1H	*HvELF3, HvCMFa, HvCMF6b*	-
AM_2H.1	HD(14, 16. AV), MT(14, 16)	2H	*HvCEN, HvHOX2, HvGID2, HDG-2H, HvKNOX1, HvD11*	*HvSUSIBA2, FAR1-related sequence 5 and 11, GRAS family TF, Phototropic-responsive NPH3 family protein, Chlorophyll A/B binding protein, Cytochrome c6, Chlorophyll a-b binding protein 6, SAUR-like auxin-responsive protein family, Photosystem II 10 kDa polypeptide, Auxin signaling F-box 2, Gibberellin receptor GID1*
AM_2H.2	RT	2H	*Vrs-1*	-
AM_3H.1	HT(14, 16)	3H	*HvBRI1, HvGA3ox2, HvD2, HvHXK7, HvD18*	*Ethylene responsive transcription factor 1a, MYB related TF, MYB domain proteins*
AM_3H.2	HT(14,16)	3H	*Hv20ox2/denso/sdw1*	*GRAS family TF, MYB TF-like, MEI2-like protein 1, MYB domain protein, SAUR-like auxin-responsive protein family*
AM_4H.1	HD(14)	4H	*Vrn-H2*	*HvRIM1, FAR1-related sequence 6*
AM_5H.1	HD(14)	5H	*HvCO3, HvTFL1*	*HvTPS1, Auxin-responsive protein IAA, MYB domain protein 59, Ethylene-responsive TF, Abscicic acid insensitive 5-like protein, MYB family TF, Cytochrome b5-like, FAR1-related sequence 5, ELF4-like protein*
AM_6H.1	HD(16, AV)	6H	*HvCO2, HvCO11, HvCO14*	-
AM_7H.1	HD(14,16, AV), MT(14, 16), HT(14)	7H	*HvFT1, HvCO8*	*Photosystem II CP47 chlorophyll apoprotein, MYB family TF, HvSSI, MYB transcription factor, FAR1-related sequence 5*
AM_8H.1	MT(14)	5H**	-	*Auxin-responsive protein IAA13, Ethylene-responsive transcription factor ERF025, Photosystem II 11 kDa protein-related, FAR1-related sequence 11, Protein Far-red impaired response 1, Myb-like DNA-binding domain,*

* HD= Heading date, MT= maturity, HT= height, RT= row type

** According to Morex_2016 physical map, via T3/barley triticeae toolbox (https://triticeaetoolbox.org)

3.6 QTL analysis

3.6.1 Mapping of linkage groups

Linkage groups and maps were successfully generated for DH lines and SSD lines separately, resulting in eight respectively 14 linkage maps (Figures S13 and S14).

Combined, the linkage maps of the DH dataset contained 484 markers and spanned 614 cM across all seven chromosomes, while the linkage maps for the SSD dataset contained in total 366 markers and spanned 1153.5 cM across all seven chromosomes.

3.6.2 Detection of QTL for growth and development

3.6.2.1 Detection of QTL in the DH dataset of the Golf x Tampar population

In the DH dataset, a total of 15 significant (LOD ≥ 3.1) QTL were detected, the majority of which on chromosomes 2H and 3H (Table 12, page 43). No QTL were found on chromosome 7H. Only a few QTL were shared between traits, and many of the QTL were found in either 2014 or 2016, but not in both years. Some QTL were not significant for certain traits or in certain years, but explained none the less a large (>5%) percentage of observed phenotypic variance.

In total, five different maturity QTL, located on chromosomes 2H, 5H and 6H, were detected, each explaining between 11.9-22.2% of variance. Six different QTL were detected for heading, located on chromosomes 1H-5H and explaining between 4.4-58.8% of variance. In 2014, four significant QTL, distributed over chromosomes 2H and 3H, were found for height, each explaining between 10.8-17.3% of observed variance. Only one QTL, located on chromosome 2H, was discovered for row type, explaining up to 99.5% of observed phenotypic variance. LOD score graphs for the most important QTL can be found in Figure 7.

3.6.2.2 Detection of QTL in the SSD dataset of the Golf x Tampar population

In the SSD dataset, a total of 10 significant (LOD ≥ 3.1) QTL were detected, distributed over all chromosomes except chromosome 4H (Table 13, page 43). Three QTL were shared between traits, and seven QTL were trait-specific. QTL for a trait were not always found in both 2014 and 2016. Some QTL were not significant for certain traits or in certain years, but explained none the less a large percentage (>5%) of observed phenotypic variance.

In total, three different QTL, located on chromosomes 2H and 7H, were detected for maturity, each explaining between 14.3-20.3 % of variance. Six QTL, located on chromosomes 1H, 2H and 7H were detected for heading, each explaining between 5.0-44.5% of variance. Three QTL, located on chromosomes 2H, 3H and 5H, were detected for height. Each QTL explained between 11.8-17.3% of variance. Two QTL, located on chromosome 2H and 3H, were detected for row type, explaining 86.1 and 2.5% of variance respectively. LOD score graphs for the most important QTL can be found in Figure 7, displayed on pages 44-47.

Table 12. Results of QTL mapping in the DH dataset of the Golf x Tampar RIL mapping population. Chromosome, linkage group and range (cM) on which the QTL is situated and results from MQM mapping, showing the percentage of explained phenotypic variance and corresponding LOD values for each trait for each QTL. Significant LOD values are printed in bold, while QTL explaining more than 5% of phenotypic variance are printed in italics.

			DH linkage map range (cM)		Maturity (AGDD)				Heading (AGDD)						Height				Row type	
					2014		2016		2014		2016		AV		2014		2016			
QTL	Chr.	Linkage group			% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD
QTL_DH_1H.1	1H	1H.1	86.4	88.2	1.1	0.35	0.1	0.1	2.4	0.75	4.4	**3.6**	3.8	1.84	1.2	0.54	0.8	0.26	0	0.16
QTL_DH_2H.1	2H	2H	6.7	9.6	*14.5*	**4**	0.2	0.05	0.2	0.1	0.6	0.58	0.7	0.33	0.1	0.0	0.3	0.08	0	0.08
QTL_DH_2H.2	2H	2H	13.9	14.7	*17.9*	**4.75**	0.1	0.02	0.5	0.15	0.8	0.72	0	0.01	0.1	0.05	0	0.01	0	0.09
QTL_DH_2H.3	2H	2H	35.0	39.6	0.3	0.1	0.3	0.06	*43.8*	**10.03**	*58.8*	**22.81**	*54.2*	**17.27**	0.3	0.09	0.8	0.32	0	0.04
QTL_DH_2H.4	2H	2H	43.0	43.9	1.3	0.42	1.1	1.75	1.1	1.02	4.7	2.33	4.7	2.33	*14.7*	**5.32**	1.5	0.49	0	0.1
QTL_DH_2H.5	2H	2H	46.6	50.0	0.5	0.16	2.8	0.63	*10*	**3.47**	2.9	2.92	*7.6*	**3.95**	0.2	0.11	2.5	0.84	*99.5*	**337.4**
QTL_DH_2H.6	2H	2H	51.8	52.6	*22.2*	**5.69**	3.4	0.78	4.9	1.58	2.5	2.53	3.4	1.62	0.1	0.04	3.5	1.2	0	0.05
QTL_DH_3H.1	3H	3H	23.5	28.2	2.4	0.78	0.9	0.21	*6.3*	2.08	*13.2*	**9.08**	*10.9*	**4.44**	1.5	1.35	*17.3*	**4.79**	0	0.28
QTL_DH_3H.2	3H	3H	32.4	39.2	2.5	0.81	1	0.23	*5.4*	1.76	0.8	0.72	0.8	0.37	*11.3*	**4.25**	0.6	0.18	0	0.34
QTL_DH_3H.3	3H	3H	59.7	60.5	1.9	0.62	3.9	0.91	0.2	0.07	1	0.97	0.2	0.11	*13.9*	**5.41**	0	0.16	0	0.35
QTL_DH_3H.4	3H	3H	81.7	83.4	0.8	0.24	0.2	0.05	0.3	0.1	0.2	0.15	0	0.01	*14.2*	**5.19**	*10.8*	**3.17**	0	0.39
QTL_DH_4H.1	4H	4H	14.0	18.8	4.7	1.57	0.4	0.08	*14.8*	**4.14**	0.6	0.55	4.8	2.37	0.4	0.19	1.6	0.52	0	0.34
QTL_DH_5H-1.1	5H	5H-1	0.0	0.0	4.3	1.42	0.2	0.05	4.8	1.57	*6.1*	**4.78**	*6.8*	2.8	0.2	0.1	0.6	0.32	0	0.05
QTL_DH_5H-1.2	5H	5H-1	7.1	8.9	3.6	2.28	*18.3*	**3.6**	3.8	1.23	0.5	0.49	0.3	0.12	1	0.44	0.8	0.25	0	0.1
QTL_DH_6H.1	6H	6H	33.8	35.5	*11.9*	**3.33**	*15.6*	**3.12**	*5*	1.62	0.5	0.46	3.4	1.65	1.5	0.66	2.2	0.74	0	0.29

Table 13. Results of QTL mapping in the SSD dataset of the Golf x Tampar RIL mapping population. Chromosome, linkage group and range (cM) on which the QTL is situated and results from MQM mapping, showing the percentage of explained phenotypic variance and corresponding LOD values for each trait for each QTL. Significant LOD scores are printed in bold, while all QTL explaining more than 5% of phenotypic variance are printed in italics.

			SSD linkage map range (cM)		Maturity (AGDD)				Heading (AGDD)						Height				Row type	
					2014		2016		2014		2016		AV		2014		2016			
QTL	Chr.	Linkage group			% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD	% expl.	LOD
QTL_SSD_1H.1	1H	1H	0.0	2.9	3.4	1.06	0.9	0.32	*7.1*	**4.73**	*13.1*	**8.82**	*11.6*	**10.00**	3.7	1.33	3.3	1.11	0.1	0.11
QTL_SSD_1H-2.1	1H	1H-2	0.0	0.0	*5.7*	1.79	1.8	0.66	0.8	0.60	*5.0*	**4.40**	2.9	**3.25**	2.1	0.72	0.8	0.26	1.3	2.54
QTL_SSD_2H.1	2H	2H	81.5	90.4	*20.3*	**5.52**	*7.1*	2.68	*31.1*	**15.99**	*39.5*	**20.6**	*44.5*	**26.40**	*12.6*	**4.79**	*11.9*	**12.6**	0.0	0.01
QTL_SSD_2H.2	2H	2H	128.6	131.5	0.3	0.09	0.9	0.33	*5.3*	**3.54**	*5.5*	**4.13**	*6.9*	**6.47**	3.8	1.34	1.4	0.46	*86.1*	**47.11**
QTL_SSD_3H.1	3H	3H	14.8	19.2	0.3	0.11	2.5	0.91	0.1	0.06	0.3	0.33	0.2	0.18	3.5	1.23	0.4	0.12	2.5	**5.23**
QTL_SSD_3H.2	3H	3H	105.9	115.6	0.7	0.22	0.6	0.12	1.1	0.80	0.1	0.07	0.4	0.47	10.0	**3.23**	*17.2*	**5.14**	0.3	0.59
QTL_SSD_5H.1	5H	5H	228	229.4	0.6	0.17	0.3	0.12	0.6	0.45	1.6	1.36	0.2	0.19	*11.8*	**3.79**	0.4	0.13	0.0	0.04
QTL_SSD_7H.1	7H	7H	41.7	46.1	*6.0*	1.92	0.4	0.15	*11.3*	**7.55**	0.7	0.56	2.1	2.36	0.7	0.26	*6.0*	2.06	0.5	1.01
QTL_SSD_7H.2	7H	7H	49.2	52.6	*5.5*	1.75	*16.2*	**5.28**	2.1	1.54	*12.5*	**8.68**	*13.2*	**11.66**	0.3	0.10	4.4	1.49	0.3	0.53
QTL_SSD_7H-2.1	7H	7H.2	121.5	126.4	*7.1*	2.27	*14.3*	**4.68**	0.3	0.20	1.9	1.62	1.3	1.38	2.2	0.77	0.4	0.12	0.1	0.00

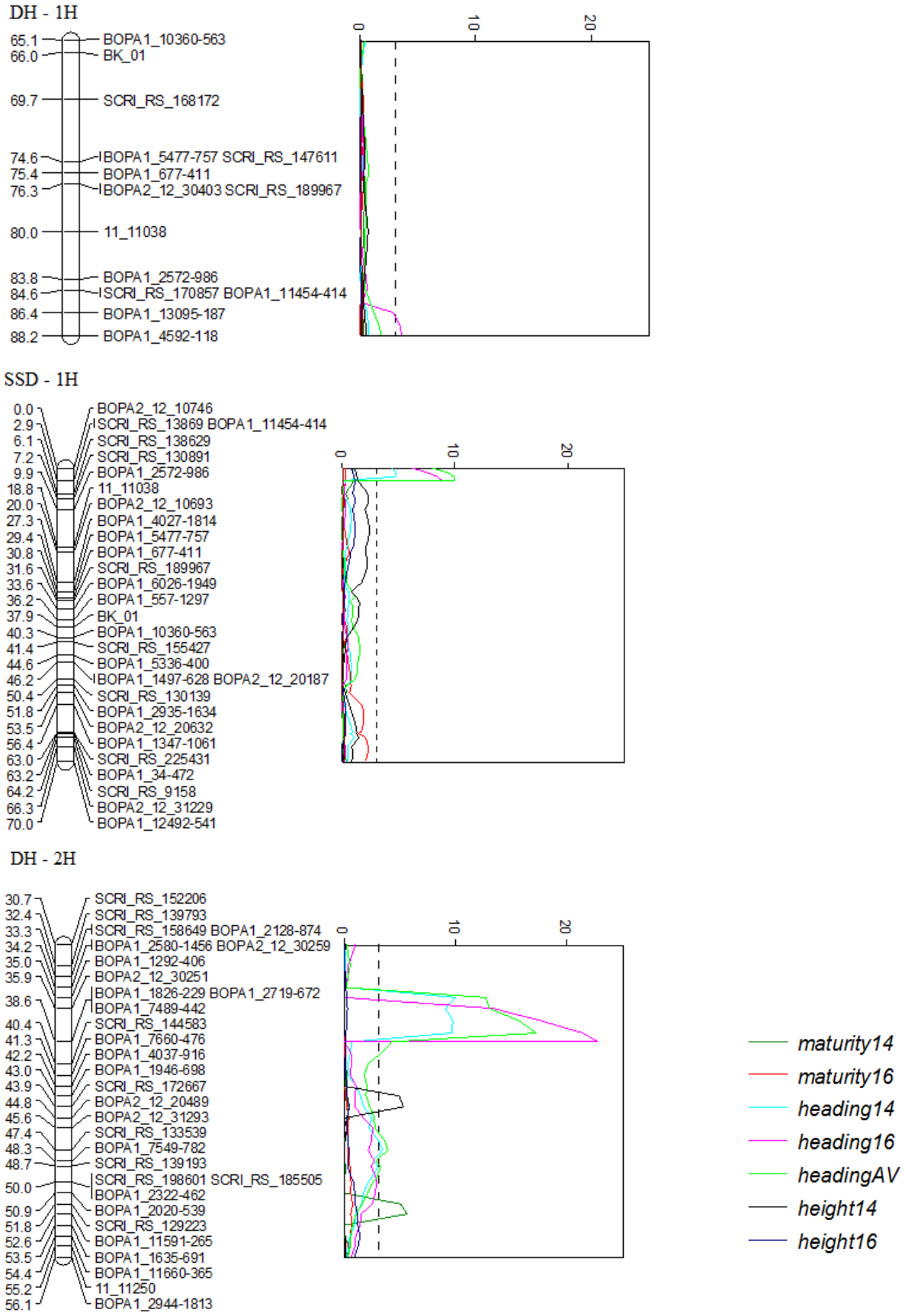


Figure 7. LOD score graphs for part of chromosome 2H in the DH and SSD datasets. Continuation on next page.

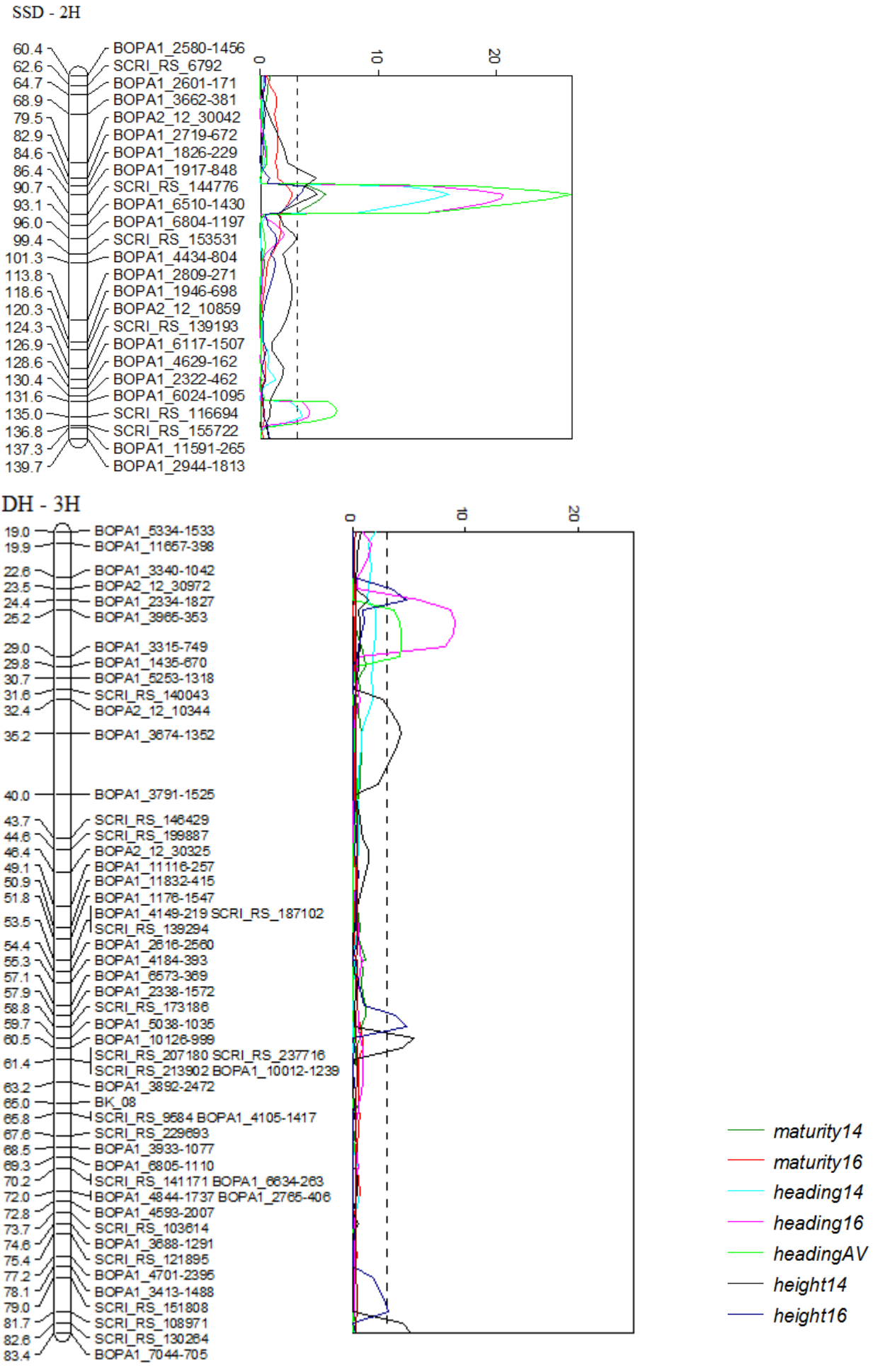


Figure 7 (cont.). LOD score graphs for part of chromosome 2H in the DH and SSD datasets. Continuation on next page.

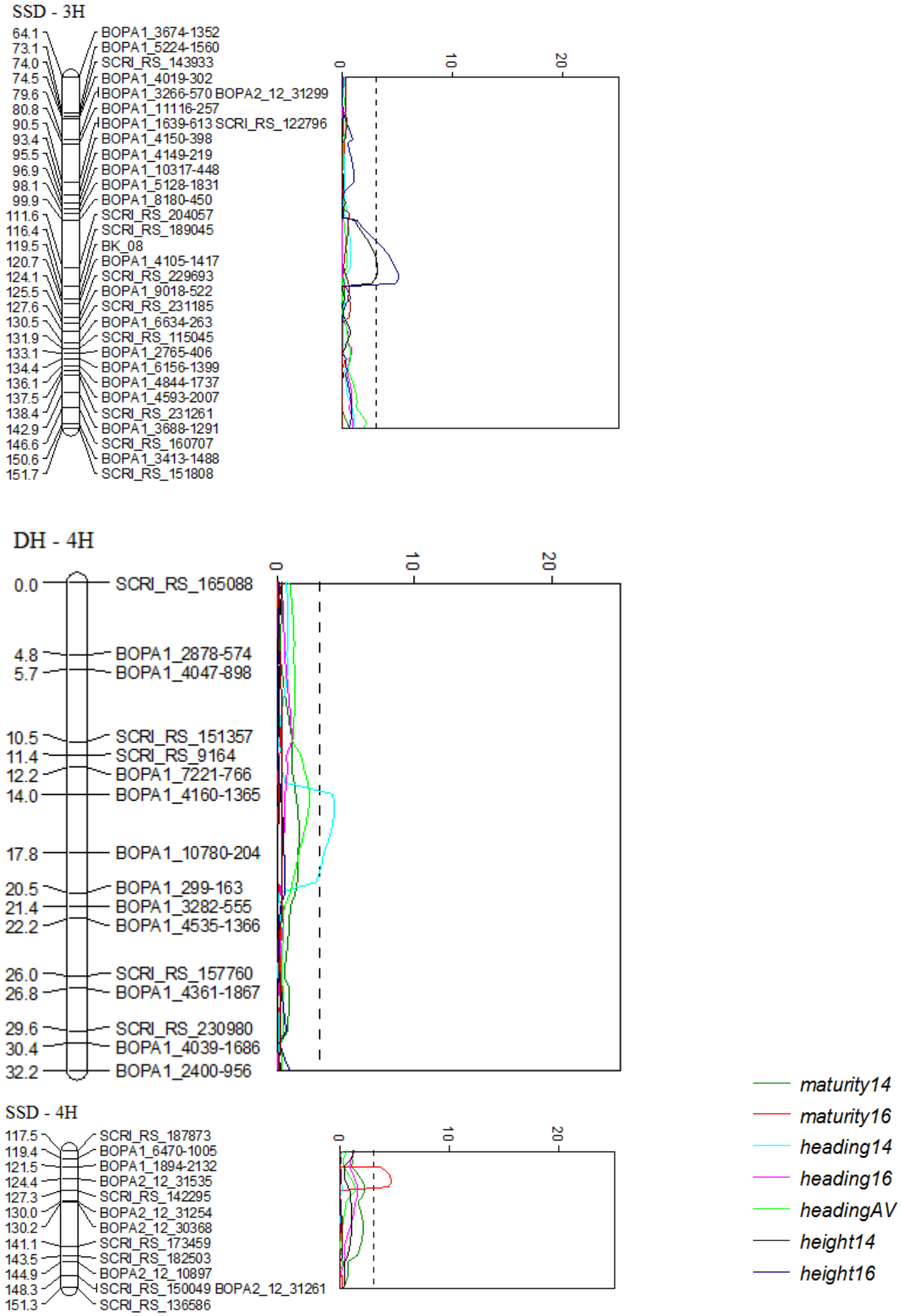


Figure 7 (cont.). LOD score graphs for part of chromosome 2H in the DH and SSD datasets. Continuation on next page.

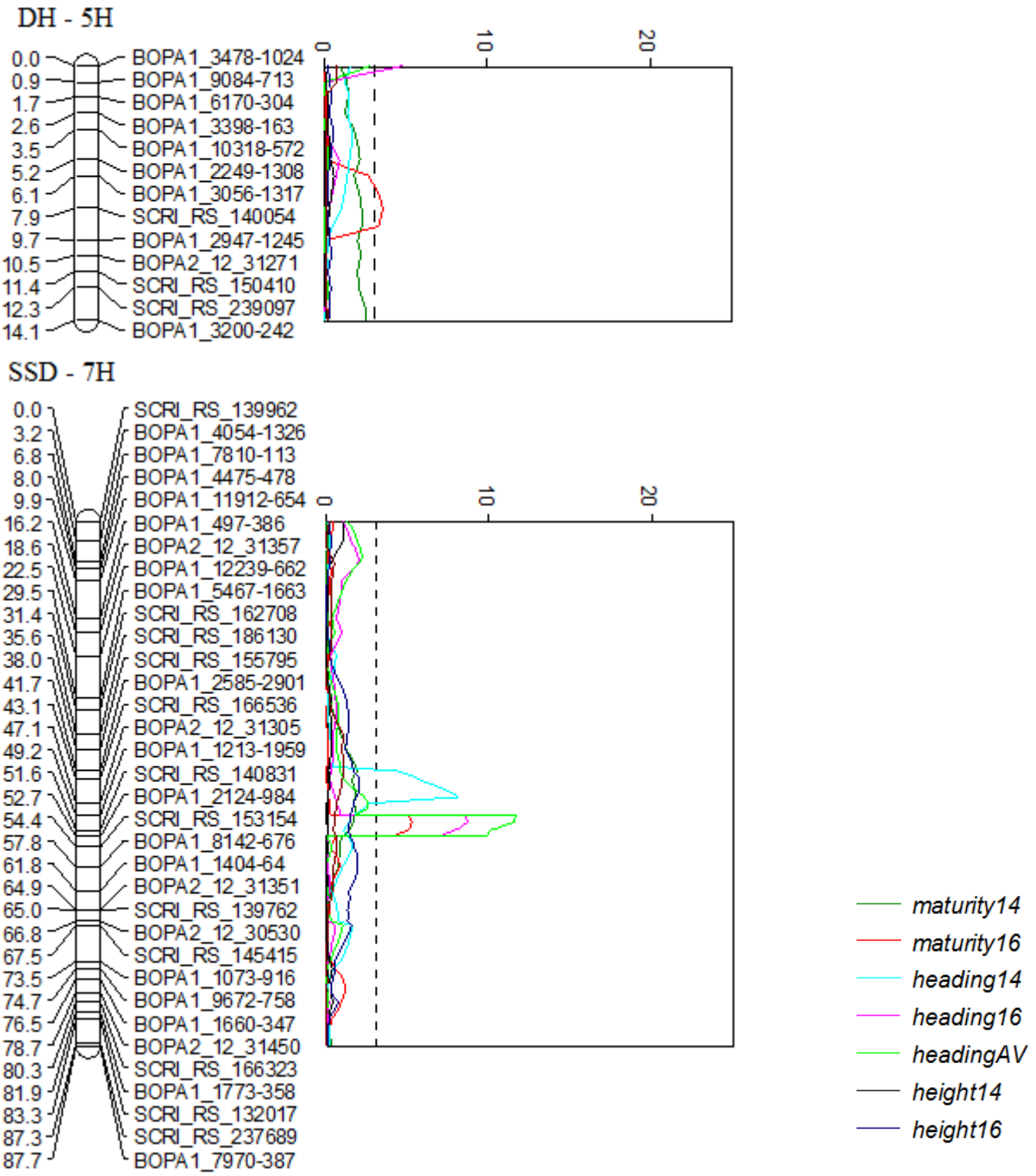


Figure 7 (cont.). LOD score graphs for part of chromosome 2H in the DH and SSD datasets.

3.6.3 Positioning of detected QTL

The genetic position of each QTL based on four different genetic maps in shown in Figure S15. QTL which overlapped or were positioned closer to each other than 3 cM according to either of the published genetic maps are shown as combined QTL. Only three such combined QTL were detected, each of which consisting of a QTL detected using the DH dataset and one or two QTL detected using the SSD dataset. In total, 21 QTL were detected. Markers associated with each QTL are shown in Figure S16.

3.6.4 Candidate genes at detected QTL

As shown in Table 14, candidate genes known to be associated with the investigated traits were found for 15 of the 21 QTL detected using QTL analysis. In addition, for most QTL, a number of genes of possible interest were identified as well.

Table 14. Combined QTL, QTL and their associated traits, along with the most likely candidate genes for each combined QTL and other genes potentially of interest.

Combined QTL	Significantly associated traits*	Most likely candidate genes	Other genes potentially of interest
QTL_1H.1	HD(16, AV)	*HvGA20ox2/HvSD1, HvCMF11/HvCO9*	*Ethylene-responsive TF*
QTL_1H.2	HD(14, 16, AV)	*HvELF3, HvCMF6a, HvCMF6b*	-
QTL_2H.1	MT(14)	*HvRAX1*	-
QTL_2H.2	MT(14)	*PPd-H1*	-
QTL_2H.3	HD(14, 16, AV), MT(14), HT(14,16)	*HvCEN, HvHOX2, HvGID2, HDG-2H, HvKNOX1, HvD11*	*HvSUSIBA2, FAR1-related sequence 5 and 11, GRAS family TF, Protein SPIRAL1-like1, Phototropic-responsive NPH3 family protein, Chlorophyll A/B binding protein, Cytochrome c6, Chlorophyll a-b binding protein 6, WRKY TF 29, SAUR-like auxin-responsive protein family, cellulose-synthase like D2, Photosystem II 10 kDa polypeptide, Auxin signaling F-box 2, Gibberellin receptor GID1*
QTL_2H.4	HT(14)	-	*AP2-like ethylene-responsive TF, IAA-amino acid hydrolase ILR1-like 3, cytokinin oxidase/dehydrogenase 1*
QTL_2H.5	RT, HD(14, 16, AV)	*Vrs1*	*MYB domain proteins, FAR1 related sequence 5, SAUR-like auxin responsible TF, ethylene responsive TF, GRAS family TF, 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein, IAA-amino acid hydrolase ILR1-like 3, Photosystem II 10 kDa polypeptide, WRKY TF*
QTL_2H.6	MT(14)	-	*HvMAX3, HvCCD7, HvHTD1, HvD17*
QTL_3H.1	HT(14, 16)	*HvSTE1/HvD7*	*Expansin, Ethylene-responsive TF*
QTL_3H.2	HT(14, 16)	*HvBRI1, HvHXK9*	*Ethylene responsive TF 1a, MYB related TF, MYB domain proteins*
QTL_3H.3	HT(14)	-	*MYB domain proteins, MYB TFs, HvLNT-1, HvCMF1*
QTL_3H.4	HD(16, AV), HT(16)	*HvGA20ox2/denso/ sdw1, HvGA20ox3*	*Cytochrome c and b6, 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein, GRAS family TF, FAR1-related sequence 5, MYB TF-like, MEI2-like protein 1, MYB domain protein, SAUR-like auxin-responsive protein family*
QTL_3H.5	RT	-	-
QTL_4H.1	HD(14)	*Vrn-H2*	*Auxin response factor, MYB domain proteins*
QTL_5H.1	HT(14)	*ari-e, HvTPS1, HvCPD, HvD53*	*HvND1/ TFL1, HvAS1, HvBC12/GGD1Auxin-responsive protein IAA, ABSCISIC ACID-INSENSITIVE 5-like protein, Ethylene-responsive TF, MYB domain proteins*
QTL_5H.2	MT(16)	-	*Auxin-induced like proteins, Auxin signalling F-box 2, Auxin efflux carrier, SAUR-like auxin repsonsive protein*
QTL_5H.3	HT(16)	-	*Ethylene-responsive TF*
QTL_6H.1	MT(14, 16)	*HvCry1b*	*Zinc finger protein CONSTANS-like protein, Ethylene-responsive TF, MYB domain proteins, Trehalose-6-phosphate phosphatase, Ethylene-responsive TF 1, MYB family TF, Auxin response factor 15*
QTL_7H.1	HD(16, AV), MT(16)	*HvFT1*	*HvSSI*
QTL_7H.2	HD(14)	*HvCO8*	*MYB TF*
QTL_7H.3	MT(16)	*HvCO6*	*HvESR1, MYB family TF like*

* HD= Heading date, MT= maturity, HT= height, RT= row type

3.7 Comparisons of QTL between methods, datasets and years

The QTL analysis resulted in the detection of 21 QTL, of which four were found both in the DH and SSD dataset. Fifteen QTL were found in the DH dataset, while only 10 QTL were found in the SSD dataset. Nine of the 21 detected QTL were also found using GWAS. Distribution of QTL across chromosomes was comparable using both QTL analysis and GWAS, with many of the QTL being found on chromosomes 2H and 3H. Performing of GWAS on the DH dataset resulted in the detection of five QTL, while GWAS on the SSD and complete dataset resulted in the detection of six and 10 QTL respectively.

QTL significantly associated with a trait in one year, were not always significantly associated with that trait in the other year it was phenotyped in. However, in some cases, the explained variance of the trait in the non-significant years was still comparatively high. This was the case both for results generated using GWAS and for those generated using QTL analysis.

4. Discussion

4.1 Field data

4.1.1 Distribution of phenotypic variance

Boxplots generated for each trait (Figure 3) showed that for most traits phenotypic values were approximately normally distributed. However, normality tests run on the data revealed that this was not always the case. Many statistical tests assume normality of data, and these results therefore indicated that transformation of data might be required at later stages of data analysis. Large variation was observed among the recorded phenotypic values, indicating that this population is suited for QTL and GWAS analysis.

On average, accumulated growing degree days (AGDD) to heading and maturity was somewhat lower in 2016 than in 2014, while days to heading and days to maturity were similar (Table 1). AGDD incorporates both time that has passed from sowing and accumulated heat sum, which is strongly related to development (Lillemo et al., 2010; Hilmarsson et al., 2017), and is regarded as a more reliable estimate of heading or maturity than days (Alqudah & Schnurbusch, 2017). It was therefore unexpected that days to heading and maturity varied less between years than AGDD to heading and maturity. Photoperiodic response may be suggested as a possible factor contributing to this.

Previous studies performed by Knut Aastveit (unpublished data) over the course of three years at four different locations in Scandinavia found both Golf and Tampar to be on average short-strawed and nearly equal in length. However, results of the present study indicate that Golf is on average shorter than Tampar (Table 5). The same study found Golf to mature on average 14.5 days later than Tampar. This differs slightly from the approximately 16 days observed in 2016 in this study (Table 5). However, the genotypes tested by Aastveit and those tested in the present study are not the same, furthermore Tampar is known to have low environmental stability (high Genotype by Environment interaction), both of which are likely to have contributed to this difference. In 2014, the recorded number of days to heading in the second replication appears to be an error made during data collection. Looking at the maturity data of this same replication, and comparing these data with the data recorded in 2016, it seems however likely that days to heading in 2014 was much shorter in Tampar than in Golf as well. Since these data were used in all subsequent analyses, this may have slightly affected results.

The models used to find the REML adjusted predicted means for each trait per year showed that for all traits, lines had a significant effect on phenotypes.

4.1.2 Correlations between traits

Between-year correlation was very strong for heading, but only moderately strong for maturity and height (Table 1). This is in accordance with multiple other studies that have also found that heading is more strongly related between years than other traits, i.e. has a higher heritability than most other investigated traits (Nice et al., 2017). Phenotypic data collection was performed by different people in 2014 and 2016, which may have influenced recording of phenotypic results, and may be responsible for some of the differences observed between 2014 and 2016. For example, height measurements were taken at different stages of development in 2014 and 2016, which may have contributed to a decrease in correlation between years for this trait. Genotype by Environment (GxE) interactions are also likely to affect these results, but they were not measured in this study. All investigated traits are affected by weather conditions. In the summer of 2016, precipitation levels were lower than in 2014, except in October, while temperatures were very similar, if somewhat colder in the first half of the growing season (Figure S1). Further, in the second half of the growing season, maximum wind velocity was somewhat higher in 2016 than in 2014. However, no clear associations between these weather data and recorded phenotypes could be identified. Additional years of phenotyping would allow for better estimation of GxE interactions in this population as well as calculation of heritability of traits.

Within-year correlation between traits showed near perfect correlation between heading date, measured as days from sowing, and heading, measured as AGDD (Table 2A, B). This was also the case with the correlation between maturity, measured as days from sowing, and maturity, measured by AGDD. Since AGDD values were calculated based on the number of days, this was expected. This strong correlation suggested that only analysing either metric is sufficient. Alqudah and Schnurbusch (2017) reviewed the best methods to measure heading and flowering time, and proposed that using AGDD should be preferred over days. Therefore, analysis was continued using AGDD to heading and AGDD to maturity. The strong correlation between heading and maturity was expected since these traits are two subsequent stages of development, both of which are affected by the duration of the phases that precede them (Drosse et al., 2014). Nevertheless, these traits had only few QTL in common. Heading and height were weakly associated, and this correlation was only significant in 2014. The perfect correlation for row type was due to the curation of data before analysis was performed. Although intermediate row

type phenotypes were observed in 2016, they were not recorded as such. Row type in 2014 had only been recorded as two or six-rowed, and since it was not possible to change these data post-hoc, this was also done this way in 2016. However, this is likely to have reduced the possibility of finding QTL or genes related to specific intermedium phenotypes.

4.2 Linkage Disequilibrium decay mapping

Linkage Disequilibrium decay mapping revealed that LD decay was faster in SSD lines than in DH lines. This is likely due to the fact that these lines have undergone more recombination events. Choice of minor allele frequency (MAF) cut-off value, and other types of marker selection, are known to affect LD decay calculations. Although quality control filters were the same for each dataset, there were large differences in number and identity of markers between datasets, and this is therefore likely an additional factor explaining the large differences in observed LD decay between datasets. Restriction on the MAF usually reduces the average r^2, which in turn influences LD-decay estimation when using the intersection of a fitted trend-line and a threshold as LD-decay distance (Vos et al., 2017). In the present study, it was decided to filter MAF at <0.05, since it is a common threshold in the literature. However, since allele frequencies close to 0.5 are expected for biparental crosses, this is not likely to have had a large effect on the calculated LD decay.

Short-range distance LD (r^2) in the DH dataset was almost entirely above 0.2 (Figure 5). This means that markers at short distance from each other are almost all significantly genetically linked. Also, the distribution of r^2 values may have impacted the fitting of the LD decay curve, thereby possibly confounding the correct estimation of LD in this dataset. The large differences observed between the DH and the SSD lines suggest that it may be better to use only one line type in a study, i.e. either DH or SSD. Since the amount for LD between loci determines the number of markers necessary to reach a certain resolution during association mapping (AM), a reasonably fast LD decay is often preferred to reach a high mapping resolution without a need for large numbers of markers (e.g. Vos et al., 2017). Since LD decay in barley is already slow, it may therefore be advisable to use SSD rather than DH lines in studies using bi-parental populations, especially when population size is limited, since that reduces statistical power.

Kraakman, Niks, van den Berg, Stam and van Eeuwijk (2004) found that LD stretched over a distance of at least 10 cM in a collection of 146 modern two-row spring barley cultivars, representing the current commercial germplasm in Europe, and Bengtsson et al. (2017) found that LD decayed below a critical level of r^2 =0.2 between 0-12 cM in subgroups of Nordic

spring barley lines. Wonneberger, Ficke and Lillemo (2017) found that genome-wide LD decay, calculated as distance to intersection of the fitted decay curve and r^2 =0.1, in a population of 209 Nordic barley accessions was 13 cM. Wonneberger et al. (2017) further found LD decay to be most rapid in chromosomes 4H and 6H (7.0 and 9.3 cM respectively), and slowest in chromosome 2H (55.9 cM). Since the methods to calculate LD in these studies differ from each other, and from the present study, this should be taken into account when comparing these values. However, it is clear that both these previous studies and the study presented here find that LD decay in barley is slow, and reaches at least as far as 10 cM. Since the present study solely examined a RIL biparental population while the previously mentioned studies examined a variety of different barley accessions, a much slower LD decay was expected. It is therefore not surprising that LD decay in the Golf x Tampar population appears to be slower (or similar) than previously reported values.

Theoretically, the SSD method of generating homozygous lines gives a higher probability to break gene linkage than the DH method, since these lines are the result of more recombination events than DH lines, and the required level of homozygosity is only reached after a six or seven of generations of selfing. This may be reflected in increased phenotypic variance of quantitative traits such as heading or maturity. However, this is not always the case: For example, when phenotypic variation is caused by small effects of segregating genes or by linkage of only some of them, the increased breaking of linkage and changes in the frequency of recombinants may not be enough to significantly alter statistical characteristics of the population (Surma, Adamski, Kaczmarek & Czajka, 2006). The coefficient of variation (CV%) was slightly larger in DH lines than in SSD lines for heading (days and AGDD) and height in both 2014 and 2016 and for maturity (days and AGDD) in 2014 (Table 3B, C). This suggests that the increased linkage decay of SSD lines compared to DH lines was not sufficient to cause a significant increase in phenotypic variance for these traits. However, the phenotypic data of the SSD lines appears to indicate that some selection towards reduced earliness has taken place, since the number of days and AGDD to heading and maturity is on average much greater than that of DH lines.

4.3 Population structure and linkage disequilibrium

Since the Golf x Tampar mapping population consisted solely of recombinant inbred lines (RIL), no population structure was expected in the mapping population. However, using STRUCTURE™, PCA and the kinship matrix function of GAPIT, a clear separation of lines into subpopulations containing DH respectively SSD lines was found (Figures S5-S9, Figure

4). Furthermore, DH lines were more similar to Tampar, while SSD lines were more similar to Golf. The same partitioning was found in the observed phenotypic values. Both in the SSD and DH dataset, L(K) did not increase beyond k =1 (Figures S6 and S7), suggesting that there is no population substructure within these datasets. The observed fluctuations in Δk for these datasets are most likely due to noise. PCA also suggested that there is no clear population structure within these datasets, apart from a slight partitioning based on row type as shown in Figure S8.

After the generation of the Golf x Tampar mapping population, all lines were multiplied in 2012 (and 2013) and subsequently sown (and phenotyped) in 2014 and 2016 using seeds harvested during the previous year. This theoretically opens up possibility for selection, as well as outcrossing. However, cultivated barley is almost entirely self-fertilising (~99%) and gene flow is slow (Ritala, Nuutila, Aikasalo, Kauppinen & Tammisola, 2002; Ellstrand, 2003). Also, SSD lines in F_7 should theoretically have attained a homozygosity level of about 0.986, and DH lines should also be almost completely homozygous (Snape & Riggs, 1975). The probability of natural selection taking place during the three to four generations that were grown and harvested before leaf samples were taken for genotyping is therefore very low. The low level of heterozygosity of the mapping population rules at least out that natural selection has had a large, if any noticeable, effect on the population within these three to four generations. Furthermore, by filtering out lines with high levels of heterozygosity before continuation of analysis, any lines subjected to outcrossing are likely to have been removed. The observed levels of heterozygosity in the filtered datasets are all < 0.006, which is sufficiently low to exclude any possibility of outcrossing. The proportion of heterozygote markers in the SSD subset of lines was only slightly larger than that in DH lines. Assuming that these markers are representative for the entire genome, this indicates that both line types have similar levels of heterozygosity, and that the extent of any selection would likely be similar. Moreover, even if natural selection played a negligible role, this would still not explain why SSD lines were similar to Golf, while DH lines were more similar to Tampar. All in all, it appears to be unlikely that the observed differences between DH and SSD lines were (all) generated during multiplication and maintaining of the population in the field. This difference must therefore have occurred during the generation of these lines before they were planted in the field. A study by Ogugu et al. (2014) looked at parental genome contribution (PGC) in maize DH lines derived from six backcrosses, and found that PGC did not always followed the expectations, as the PGC of one parent was higher than expected. On average, PGC is supposed to be approximately 50% for each parent in an F_1 generation. However, due to non-random and random recombination, PGC

can vary in both amount and specific chromosomal content from each parent. This deviation from expected PGC found by Ogugu et al. (2014) is a similar phenomenon to the partitioning of germplasm observed in the Golf x Tampar population investigated in the present study. Although the level of selection that takes place during generation of SSD or DH lines is usually very little (Wricke & Weber, 1986), Ogugu et al. (2014) still suspected artificial selection during the generation of the F_1 generation or the generation of DH lines from the F_1 generation to have contributed to the observed deviations. This may very well have been the case with the Golf x Tampar population as well, and may have occurred as a bottleneck effect during the creation of the SSD lines. However, looking at the frequencies of the Golf and Tampar alleles along chromosomes in the SSD and DH datasets, this appears at least not to be the sole reason for this phenomenon. Instead of the expected allele frequency of 0.5, large fluctuations and deviations from this frequency were observed in not only the SSD dataset but the DH dataset as well (Figure S4). On some chromosomes, alleles frequencies reached as low as about 0.2 and as high as 0.8. Any interval on which one allele has a (much) higher frequency than the other would suggest that some kind of selection must have taken place in favour of the allele with the highest frequency. For the SSD dataset, the plot for chromosome 2H shows that the Tampar allele has a higher frequency along the entire chromosome than the Golf allele. The frequency does however remain not too far from the 0.5 frequency it would be expected to be at, except at the right end of the chromosome where it even reaches a frequency of 0.75 for a short distance. The plot of chromosome 6H for the SSD dataset shows a similar pattern, only more extreme: The allele frequency of Golf is near 0.8 for about 100 cM of the total length of 129 cM of chromosome 6H. These results therefore suggest that selection in favour of the Tampar alleles may have taken place on chromosome 2H, while selection in favour of the Golf alleles may have taken place on chromosome 6H during the generation of the SSD lines. Another possibility is that genetic drift has contributed to the observed allele frequencies. The plots for the DH dataset showed that allele frequencies differed even more from the expected frequency of 0.5 than in the SSD dataset. Especially noteworthy is that only at a few points there is an exchange as to which allele is more frequent than the other. On chromosome 6H this does not happen even once, with the Tampar allele having a higher frequency than the Golf allele along the entire chromosome. Similarly, the frequency of the Golf allele remains much higher than the Tampar allele for almost the entire length of chromosome 1H. On several chromosomes Golf and/or Tampar allele frequencies reach as high as 0.7 or even 0.8, showing clearly that there is no equal parental genome contribution. To conclude, the analysis of allele frequency

along each chromosome appears to suggest that not only must selection and/or genetic drift have taken place on (at least) chromosomes 2H and 6H during the generation of SSD lines, but also that some unusual event such as selection must have occurred during the generation of DH lines as well. It is also noteworthy that no particular pattern in allele frequencies could be observed when looking at the position of the detected QTL. And these results offer therefore no clear solution as to why Golf was more similar to the SSD lines, and Tampar more similar to the DH lines. Analysis of allele frequency further also showed that the total frequency of the Golf and Tampar alleles combined did not always reach 1.0. This may be explained by the fact that genotyping of some markers failed, and that some lines carried alleles for some markers that differed from both the Golf- and Tampar alleles. Although this is possibly the result of a mutation, a more likely explanation is that this is the result of an error during genotyping. Similarly, for some markers, Golf and Tampar carried the same allele, while other lines carried different alleles.

Barley germplasm is often partitioned based on row-type, which is thought to be largely the result of breeding history (Drosse et al., 2014). Since there is no such breeding history in the bi-parental mapping population examined in the present study, no clear partitioning based on row type was therefore expected. However, PCA of the DH and SSD datasets shows a slight partitioning of germplasm based on row type. More striking though, is the association there appears to be between row type and other investigated traits. Some studies have suggested that loci commonly associated with row type, may be linked to loci associated with other traits such as heading (Alqudah et al., 2016). Phenotypic data analysis showed that heading, and maturity, commenced on average earlier in six-rowed lines than two-rowed lines, both in DH and SSD lines, and two-way ANOVA confirmed that this association was significant. This suggests that such a linkage between row type QTL and heading QTL may be present in this population, and may have contributed to the observed partitioning. During QTL analysis, one major QTL on chromosome 2H, for which *Vrs-1* is the most likely candidate gene, explained up to 99.5% of row type variance, while a second closeby (~15 cM) positioned major QTL, for which *HvCEN* is the most likely candidate gene, explained up to 58% of heading. These same QTL were also found using GWAS. Tampar is six-rowed and early flowering and maturing, while Golf is two-rowed and later flowering and maturing. This would suggest that it may be possible that, on average, the alleles causing early flowering and six-rowedness were more often linked together than would be expected if recombination was completely random. The LD plot of chromosome 2H showed that the *HvCEN* and *Vrs-1* loci were indeed significantly linked. In the DH dataset,

r^2 was as high as 0.4 – 0.7. In the SSD and complete dataset, r^2 was lower, but still reaching as high as 0.4 and 0.2 respectively. This would support this hypothesis. It should however be noted that the percentage of six-rowed lines was approximately the same in both the SSD and DH dataset, and that no clear patterns supporting this hypothesis were found using allele frequency analysis.

4.4 Quality of GWAS results and QQ-plots and validation of the mapping approach

Even when population structure and kinship were accounted for using PC and K matrices, QQ-plots showed that observed $-\log_{10}$(p-values) were sometimes inflated or deflated. During the interpretation of results, it should therefore be kept in mind that both Type I and Type II errors may have occurred. QQ-plots from models using a PCA matrix were on average better than or similar to the QQ-plots from models using a Q matrix to account for population structure. As reviewed by Zhao et al. (2007), this is often the case.

Inflation and deflation of $-\log_{10}$(p-values) depends on multiple factors, including population structure, LD structure and number of markers at a specific site, heritability of a trait, population size and number of causal variants (Yang et al., 2011). Since PCA and K matrices were included in the models, population structure was accounted for, which reduced inflation of $-\log_{10}$(p-values). Over-compensation for population structure may result in deflation of $-\log_{10}$(p-values) (Yang et al., 2011), however whether this has also contributed to the deflation observed here, is difficult to say. Thinning of the SSD and complete datasets reduced the maximum number of markers per site. When large numbers of markers are situated at the same site, this can affect PCA and thereby influence GWAS results. And this appeared to be the case here. However, the effects did not appear to be very large, which may be attributed to the fact that the total number of markers left after filtering was not very high to begin with. By thinning the datasets by site, some of the most closely linked markers were randomly removed from the dataset rendering individual tests for individual markers more independent. This resulted in improved QQ-plots, but at the same time introduced the risk of deleting the most significant markers. Another, arguably better (Yang et al, 2011), method that could be implemented, is LD based pruning of the marker set, during which the markers with the highest LD are removed based on for example significance (p-value) of the linkage or where markers in high LD are pruned based on p-value of the marker-trait association. This reduces the risk of accidently removing some of the most significant markers. Since GWAS assumes markers to be independent from each other, high LD in particular may affect and confound results. Should the Golf x Tampar population be

studied in more detail in the future, it may therefore be advisable to examine the option of LD pruning the dataset before continuation of analysis. Increased heritability of a trait has also been associated with increased inflation of $-\log_{10}$(p-values) (Yang et al., 2011). Inflation was highest for row type and heading, which have a high heritability. It is therefore not ruled out that this may have contributed to the observed inflation. Inflation of $-\log_{10}$(p-values) for row type was even higher than for heading, showing extreme deviation from the expected normal distribution of p-values, with p-values being either very high or very low. This may indicate that some markers greatly affect the trait, while other do not. Previous research has shown that row type is primarily regulated by one gene (*Vrs-1*) (Rossini et al., 2014). This might result in markers linked to this gene having very small p-values, while all other markers have extremely high p-values, possibly explaining the observed deviation from normality. Additionally, GWAS assumes a normal distribution of phenotypic values, and is especially useful to study quantitative traits. However, row-type was defined as either two- or six-row, and was thus not normally distributed. This may also have contributed to deviation from a normal distribution of p-values.

Population size is also known to largely affect the results of GWAS and QTL analysis. Larger population size, i.e. larger number of investigated lines or individuals, is associated with increased power to detect QTL, especially those with small to moderate effects. In small populations, p-values are often higher, and the effect of detected QTL may be overestimated compared to larger populations (Beavis, 1998; Jannink & Walsh, 2002). Beavis (1998) suggested that a total of 300 lines or more would be preferred for QTL analysis. The complete dataset of the Golf x Tampar population investigated here consisted of 175 lines, however the SSD and DH datasets consisted of just 109 and 66 lines, respectively. Therefore, it is likely that the high p-values and the deflation of $-\log_{10}$(p-values) observed particularly in the DH dataset, as well as the fact that peaks in the Manhattan plots could less clearly be defined than in the other datasets, can at least partially be explained by the small population size. In addition, the small population is likely at least partially responsible for observed false positives and overestimated effects of QTL. Significance of markers and their effect should therefore be interpreted carefully, especially in the DH dataset and (slightly less so) in the complete and SSD datasets. It is clear that a larger population size would have benefitted the analyses.

4.5 Quality and success of QTL mapping

Comparisons between the DH and SSD maps, showed that the DH map spanned only 614 cM, while the SSD map spanned across 1153.5 cM. SSD lines contain on average twice as many recombinations as DH lines, and this difference in map length is therefore close to the expected difference. In addition, recombination frequencies are genotype dependent, which means that covering the same physical distance can differ in length when measured in genetic distance (cM). However, on average, the barley genome is only about 1100 cM long (IBSC, 2012). It is therefore possible that genetic distances in the SSD map appear longer than they actually are. As a consequence, results obtained using the DH map may be more reliable than those obtained using the SSD map.

When generating linkage maps using a RIL population, the parental lines should be left out of analysis, and only the offspring lines should be used. In the present study however, the parental lines (Golf and Tampar) were also among the lines used to generate the linkage maps for the DH and SSD datasets. This is likely to have influenced the linkage maps, which in turn may have affected QTL mapping results. Since the SSD dataset consists of more lines than the DH dataset, it is likely that this effect is smaller for the results of the SSD dataset than the DH dataset. This should be kept in mind during the interpretation of the results, as it may have reduced the accuracy of the QTL mapping.

4.6 Positioning of detected QTL

The 9k iSelect Barley SNP chip consensus map by Muñoz-Amatriaín et al. (2014) was chosen for positioning of markers during GWAS since it was more recently published than the map for the 9k iSelect Barley SNP chip by Comadran et al. (2012). However, comparison between genetic positions assigned to markers using four different genetic maps (the consensus map by Muñoz-Amatriaín et al. (2014), the iSelect map by Comadran et al. (2012), and the POPSEQ (Mascher et al., 2013) and IBSC_2012 (IBSC, 2012) genetic maps) showed that the genetic positions according to this consensus map differed a lot from the other maps. Although QTL from GWAS were originally reported in cM intervals on the consensus map, QTL were therefore combined or separated based on the genetic distance between QTL according to Comadran, POPSEQ and IBSC_2012 genetic maps as well. GWAS results were also reported in cM in order to facilitate comparison between the results of GWAS and QTL analysis.

4.7 Candidate genes at detected QTL

Potential candidate genes were found for 11 of the 13 QTL found through GWAS and for 15 of the 21 QTL found using QTL analysis. Nine QTL were identified using both GWAS and QTL analysis, however, surprisingly they were not always associated with the same traits. For six different QTL no clear candidate genes were identified, and these are therefore putatively novel QTL, or false positives. Some of the QTL explained high percentages (>5%) of a trait, even though they were not significant for that trait. The small population size may have contributed to the low statistical power in this study. In small populations, p-values are usually higher and the effect of detected QTL may be overestimated compared to larger populations (Beavis, 1998; Jannink & Walsh, 2002). Discussion is therefore mostly limited to significant QTL-trait associations.

4.7.1 Row type QTL

Vrs-1 was the candidate gene for both QTL_2H.5, explaining up to 99.5% of variance, and AM_2H.2, which explained up to 11.8% of variance. Multiple studies have previously identified *Vrs-1* as the primary gene involved in row type determination (Rossini et al., 2014), and these results support those findings. QTL_3H.5 was the only other QTL significantly associated with row type, explaining only 2.5% of the variance. No candidate genes could be identified in the region of this QTL. Since this QTL was only detected in the SSD dataset, explaining only 2.5% and having a low LOD score of 5.23, this could be a false positive result. Especially since it did not show up using GWAS, there is a high possibility that this is the case. If not, this may be a potential novel QTL for row type. A number of other genes determining row type have been mapped and are well described (Rossini et al., 2014). Since genotyping data for row type genes in Golf and Tampar was not available, it was impossible to predict the number of QTL to segregate. Nevertheless, because two-rowed, six-rowed as well as a number of different levels of intermedium phenotypes were observed in this population, more row type QTL than just *Vrs-1* would be expected to be found. However, the exclusion of intermediate row type phenotypes is very probable to have reduced the probability of finding other row type QTL. Additionally, since the mapping population consists of a single cross between just two lines, it is very likely that only a limited number of the known row type QTL were polymorphic, allowing for them to be detected.

4.7.2 Heading and maturity QTL

Candidate genes could be proposed for all but four of the total 22 QTL found to be associated with heading and/or maturity. Only five QTL were significantly associated with both heading and maturity. Since within-year correlation between maturity and heading was strong, more QTL were expected to be associated with both traits. Also unexpected was that only four of the QTL found using QTL analysis were also found using GWAS.

4.7.2.1 *HvCEN* may explain a large percentage of variation in heading and maturity

QTL_2H.3 and the corresponding QTL AM_2H.1 are located in the centromeric region of chromosome 2H, where a number of candidate genes for heading, maturity and other developmental traits are located. This includes *HvCEN*, *HvHOX2*, *HvGID2*, *HDG-2H*, *HvKNOX1* and *HvD11*. Since linkage is extremely high within the proximity of the centromere, it is impossible to determine which gene or genes caused the observed effects. However, since *HvCEN* is known to be very important in flowering time regulation, *HvCEN* may be the most likely candidate gene for these QTL. The *HvCEN* locus, in previous publications also referred to as *earliness per se 2*, *early maturity 6* or *prematurium-C*, has been associated with a large number of traits including, heading, maturity, tiller biomass, tiller grain weight, ear grain number, and plant height (Obsa et al., 2016). *HvCEN* is the barley homolog of *Anthirrinum centroradialis* (*CEN*) and a paralog of *HvFT1*, with which is it proposed to interact (Comadran et al., 2012, Loscos et al., 2014). QTL_2H.3, found through QTL analysis, explained an extremely large percentage of observed variation: Up to 44.5% for heading and up to 20% for maturity, while in the DH dataset it explained up to 58.8% of heading, but was not significantly associated with the other traits. The LOD score for this QTL was higher than for most other QTL. QTL AM_2H.1, found through GWAS, explained up to 8.2% of heading in the DH dataset but up to 19.3% of heading and up to 6.8% of maturity in the SSD dataset. In the complete dataset, this QTL explained up to 8.9% of heading, up to 6.1% of maturity as well as up to 5.6% of height. In all datasets, the Bonferroni-Holm (0.05) threshold was met for each trait in at least one year, which was only the case for two other QTL found through GWAS (AM_2H.2 and AM_7H.1). Thus, these results suggest that genetic variation of *HvCEN* or one or more of the proposed candidate genes may have a very large effect on both (or all three) of these traits in Nordic spring barley. Comadran et al. (2012) suggests that protein sequence variation of *HvCEN* may be sufficient to affect the flowering phenotype. Studying sequence and copy number variation of this gene in Golf and Tampar may therefore be useful to detect the genetic variation associated with the different observed phenotypes.

4.7.2.2 *HvELF3* possibly affects heading and maturity

QTL_1H.2, discovered using QTL analysis, was detected in both the DH and SSD datasets, explaining up to 4.4 and 13.2% of heading respectively. AM_1H.4, the corresponding QTL found using GWAS, was significantly associated with both heading and maturity, explaining up to 6.9 and 7% of variance of these traits respectively. Three candidate genes were found for these QTL. *HvELF3*, which is part of the circadian clock and located close to this QTL, is well known to affect flowering time and is therefore a prime candidate gene (Fauré et al., 2012). Other possible candidate genes include the putative flowering time genes *HvCMF6a* and *HvCMF6b* (Cockram et al., 2012). Since all these genes are closely linked, it is difficult to say for sure which gene(s) are responsible for the observed effects. Sequencing of *HvELF3*, *HvCMF6a* and *HvCMF6b* in the parents would be the next step in order to determine this.

4.7.2.3 *HvFT1* affects heading and maturity in the SSD and complete datasets

HvFT1 was identified as the most likely candidate gene for QTL QTL_7H.1 and AM_7H.1. AM_7H.1, which was found through GWAS, explained up to 12.1% of heading and up to 10.1% of maturity in the SSD dataset, and up to 3.8% of heading and 5.2% of maturity in the complete dataset. It also significantly affected height in 2014 in the SSD dataset, explaining 7.8% of phenotypic variance, but was not found in the DH dataset. QTL_7H.1, found through QTL analysis, explained up to 13.2% of heading and up to 16.2% of maturity in the SSD dataset, but did not significantly affect any trait in the DH dataset. Since *HvFT1* is known to play a crucial role in flowering time regulation (Fauré et al., 2007), it is very likely that this gene is responsible for the phenotypic variation explained by these QTL. It is unclear why these QTL were not found in the DH dataset using either GWAS or QTL analysis, but one possibility is that the size of the DH dataset was too small to allow the detection of this QTL. AM_7H.1 is not only closely positioned to *HvFT1*, but also nearby *HvCO8*, which is also a putative heading gene (Alqudah et al., 2014). Therefore it is possible that *HvFT1* is not the gene responsible for the variance explained by this QTL, but that the effect may be caused by *HvCO8*, or another as of yet unidentified gene. Noteworthy is that this QTL is one of only three QTL detected using GWAS that managed to meet the Bonferroni-Holm (0.05) threshold, which implies that this QTL is unlikely a false positive result, and would deserve to be studied in more detail.

4.7.2.4 Members of the *HvCO* family of putative heading genes likely effect heading and maturity

QTL_1H.1, found using QTL analysis on the SSD dataset, was significantly associated with heading in 2016 and for the average of 2014 and 2016 (AV), explaining up to 5%. However, this QTL was not detected in the DH dataset nor during the GWAS analysis. Found within this QTL is *HvCO9*, also referred to as *HvCMF11*, which is associated with flowering time regulation (Cockram et al., 2012), and although it is surprising that this QTL was not also associated with heading, it is therefore possible that this gene is partially responsible for the observed effects. A second, equally likely candidate gene is *HvGA20ox2*, which was also shown to affect flowering time (Teplyakova et al., 2017). QTL_7H.3 significantly affected maturity in 2016 in the SSD dataset, but was not found in the DH dataset nor using GWAS. The putative flowering time gene *HvCO6* (Griffiths, Dunford, Coupland & Laurie, 2003) is the most likely candidate gene for this QTL. Similarly, the related putative flowering time gene *HvCO8* (Alqudah et al., 2014) is the most likely candidate gene for QTL_7H.2, which was significantly associated with heading, explaining 11.3% of the phenotypic variance. No less than three different CONSTANS like genes, *HvCO2*, *HvCO11* and *HvCO14*, were positioned close to AM_6H.1, which was associated with heading. Since these genes are linked so closely, it is difficult to say which gene or genes caused the variance explained by this QTL. The CONSTANS-like gene *HvCO3* was identified as one of two candidate genes for AM_5H.1, which explained up to 9.2% of heading. In *Arabidopsis*, *AtTFL1* is a key regulator of flowering time and development of the inflorescence meristem (Hanano & Goto, 2011). Although the function of the barley homologue *HvTFL1* is not yet extensively studied, it is likely that it is also involved in flowering time regulation. Since it is positioned close to AM_5H.1, it is also a likely candidate gene for this QTL.

Although their functions may not yet be known, CONSTANS-like genes are suspected to play an important role in flowering time regulation (Alqudah, 2014). And the results of this study do indeed suggest that (some) CONSTANS-like genes are likely to be significantly associated with earliness of heading and flowering, and are therefore likely to play a role in the genetic pathways regulating these stages of development. These genes, and this gene family as a whole do therefore deserve to be researched in more detail.

4.7.2.5 *Vrn-H2* and several putative heading genes likely contribute to variance in heading and maturity

QTL_4H.1 and the corresponding QTL AM_4H.1 were each associated with heading. *Vrn-H2* is a vernalization gene known to play a role in flowering time regulation (Nitcher et al., 2013). It is therefore the most likely candidate gene for this QTL. QTL_2H.1 explained 14.5% of variance in maturity in 2014, but was not significantly associated in 2016, the SSD dataset and no corresponding QTL was found using GWAS either. *HvRAX*, is a member of the redundant family of *regulator of axillary meristems* (*RAX*) genes which encode MYB-like TFs and affect axillary meristem initiation in *Arabidopsis* and in tomato (Keller, Abbot, Moritz & Doermer, 2006). Since axillary meristem initiation is one of the phases the plant must go through to reach maturity, it is therefore possible that this gene has contributed to the observed variance of this trait. However, since this gene and its function have not yet been described extensively in barley, this cannot be said for sure. Also, the fact that this QTL was only found for the year 2014 in the DH dataset using QTL analysis indicates that it is unlikely a major QTL for this trait. QTL_6H.1 significantly affected maturity in both 2014 and 2016, explaining up to 15.6% of variance in the DH dataset. However, the QTL was not found in the SSD dataset or using GWAS. *HvCry1b*, a putative heading time gene (Alqudah et al., 2016) is the most likely candidate gene. However, this gene would be expected to affect heading as well, which was not the case.

4.7.2.6 *Ppd-H1* and *Ppd-H2* are not major determinants of heading and maturity in the Golf x Tampar population.

The photoperiod genes *Ppd-H1* and *Ppd-H2* are known to account for a large percentage of phenotypic variance (Comadran et al., 2012; Kikuchi et al., 2012; Drosse et al., 2014). However, during GWAS neither gene was identified as a possible candidate gene for any of the investigated traits. However, during QTL analysis on the SSD dataset, *Ppd-H1* was identified as potential candidate gene for QTL_2H.2, which significantly affected maturity in 2016, explaining 17.9% of variance. GWAS and QTL analysis can only detect QTL when they are polymorphic. It is therefore possible that Golf and Tampar are monomorphic at the *Ppd-H2* locus, however it is also possible that this locus was polymorphic but that the small population size prevented the detection of this locus.

4.7.2.7 QTL without clear candidate genes

QTL_2H.5, which explained up to 99.5% of row type, was also significantly associated with heading, explaining up to 10% of variance in the DH dataset and up to 6.9% in the SSD dataset. However, no known candidate genes associated with heading are located at this locus. A number of genes of possible interest were identified, however it is unclear which, if any of these genes is responsible for the observed effect on heading. No association between heading and the corresponding QTL AM_2H.2, found using GWAS, was found. Although *Vrs-1*, the candidate gene associated with these QTL that is most likely to be responsible for the observed row type effect, is known to have pleiotropic effects on among others height and tillering, no such effect on heading has yet been described (Alqudah et al., 2016). Consequently, no clear candidate genes were detected that may be responsible for the observed effect on heading and maturity. QTL_3H.4 explained up to 13.2% of heading, and was also associated with height. Candidate genes known to affect height were identified, however no genes known to affect heading were located in the proximity of this QTL, and therefore no candidate genes could be identified. No candidate genes could be identified for QTL_2H.6, QTL_5H.2, AM_1H.1 and AM_8H.1.

4.7.3 QTL affecting height

Using QTL analysis, eight different QTL significantly associated with height were found, two of which were also found using GWAS. In addition, four QTL were only found using GWAS. Candidate genes were found for all but four height QTL. The semi-dwarf gene *Hv20ox2/denso/sdw1* is well known to affect height as well as other traits such as yield and quality (Jia et al., 2009), and was therefore identified as a candidate gene for AM_3H.2, and the corresponding QTL QTL_3H.4. Similarly, a number of possible candidate genes were found for AM_3H.1 and the corresponding QTL_3H.2. *HvBRI1*, *HvGA3ox2*, *HvD2*, *HvHXK7* and *HvD18*, which are all suggested to affect height (Chono et al. 2003; Alqudah et al., 2016), are all positioned close to this QTL. Due to high linkage, it is difficult to say which of these candidate gene(s) caused the observed effects.

QTL_2H.3, which explained large proportions of heading and maturity, also explained up to 12.6% of height in the SSD dataset, but was not significantly associated with height in the DH dataset. A number of tightly linked genes are positioned in the region of this QTL, and it is therefore impossible to deduct which of the genes is responsible for the observed effect on height. However, the candidate genes *HvGID2* (*gibberellin insensitive dwarf2*) and *HvD11* are the most likely candidates, since they are predominantly though to affect height (Davière &

Achard, 2013), unlike some of the other candidate genes for this QTL, which are predominantly involved in flowering time regulation. Yet another QTL for which a candidate gene related to the gibberellin pathway was identified is AM_1H.3, which explained 3.3% of plant height in 2014. A *gibberellin-2 oxidase* gene was found located nearby this QTL. Since gibberellins play an important role in growth regulation (Daviere & Achard, 2013), it is likely that this gene is responsible for the observed effect.

QTL_3H.1, identified using QTL analysis on the DH dataset, was significantly associated with height in both 2014 and 2016, explaining up to 14.2% of variance. *HvSTE1/HvD7* is the barley homologue for *OsSTE1* in rice and *TaSTE* in bread wheat, where these genes are suggested to affect height (Zhang et al., 2012). Although this has yet to be studied in barley, *HvSTE1/HvD7* could still be a candidate gene for QTL_3H.1. Several candidate genes were identified for QTL_5H.1, these were *ari-e*, *HvTPS1*, *HvCPD* and *HvD53*. As described by Alqudah et al. (2016), these are all genes related to plant stature, sugar and heading time. As described by Liu et al. (2014), *ari-e* is a well-known locus that affects height in barley, while the function of the other genes is less clear. All of these genes are located in the centromeric region where linkage is high, and hence, it is not clear which gene(s) cause the effect.

As suggested by Alqudah et al. (2016), it is likely that sugar-related genes such as hexokinases are involved in height regulation. The hexokinase *HvHXK5* was found near QTL AM_1H.2, which explained 7.2% of height in 2016 in the SSD dataset, and may therefore possibly be partially responsible for the reported effect of this QTL on height.

HvFT1 and *HvCO8* were identified as candidate genes for AM_7H.1, which was associated with height as well as heading and maturity. Although these genes are known for their effect on heading and maturity, they are not known to affect height. Many development QTL affect several developmental traits. *HvFT1* is known to not only affect flowering time, but root and shoot development as well (Arifuzzaman et al., 2016). Pleiotropic effects of *HvCO8* on growth are currently unknown. Since no other candidate genes known to be associated with height were found at this QTL, it is possible that *HvFT1* and/or *HvCO8* may have affected height as well, however this cannot be deducted based on these results alone. Another possibility is that this QTL was falsely associated with height.

Although no candidate genes could be identified for QTL_2H.4, QTL_3H.3 and QTL_5H.3, a number of genes of potential interest were identified, including genes associated with plant hormone synthesis and pathways, as well as transcription factors known to play a role in plant development. Although too little is known about these genes to consider them candidate genes,

these results support the possibility of these genes playing a role in developmental processes in barley.

4.7.5 Genes of potential interest for developmental traits

Although candidate genes could not be identified for all QTL, a number of genes of potential interest that were found for some of these loci as well as for QTL for which a candidate gene(s) could be identified. Throughout the range of QTL, certain classes of genes and transcription factors (TFs) appeared numerous times, among which are *FAR1* related sequences, MYB domain proteins and MYB TFs, GRAS family TFs, chlorophyll and photosystem II related proteins and genes related to different plant hormone pathways.

4.7.5.1 *FAR1* related sequences may affect heading and maturity

In *Arabidopsis thaliana*, *FAR1* genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signalling pathway (Hudson, Lisch & Quail, 2003). As summarised by Ma and Li (2018), *FAR1* and its homologue *FHY3* are known to play a crucial role in plant growth and development in *A. thaliana*. They are involved in multiple cellular processes including light signal transduction, photomorphogenesis, circadian clock and flowering time regulation, shoot meristem and floral development and abscisic acid response. Although 12 *FAR1* related sequences (*FRS*) and four *FRS* related factors have so far been identified in *A. thaliana*, not much is known about their physiological and molecular mechanisms. However, multiple *FRS* members have been shown to act as transcription factors that regulate flowering time by repressing or activating a diverse set of genes such as *gigantea* or the circadian clock gene *ELF4*. Other *FRS* member are suspected to repress flowering by affecting the output of the circadian clock. *FRS* and *FRF* genes have not been investigated as much in other species. However, recent studies suggest a role for *FRS* members in panicle and spikelet degeneration in rice (Zhang et al., 2015), and as a photoperiod-dependent flowering time regulator in wheat (Kiseleva, Potokina & Salina, 2017). *FAR1* related sequence proteins were detected within or near QTL QTL_2H.3, QTL_2H.5, QTL_3H.4, AM_2H.1, AM_4H.1, AM_5H.1, AM_7H.1 and AM_8H.1, which were all significantly associated with heading except AM_8H.1, which was associated with maturity. Most of these QTL were also associated with maturity. Although no information could be found on *FRS* and *FRF* genes in barley, it is not unlikely that the barley homologues of these proteins have a similar function. The results presented here suggest that there is a possibility that they may affect heading and maturity.

4.7.5.2 The effect of phytohormones on plant development

Multiple copies of gibberellin receptors, and other proteins related to the gibberellin pathway were found near QTL associated with height, heading and maturity. Some genes involved in the gibberellin pathway are well known, and have been identified as candidate genes for a number of QTL. However, there are also some gibberellin related proteins that have not yet been described in detail. These genes were therefore reported as genes of potential interest. As reviewed by Hedden and Sponsel (2015), gibberellin is a plant hormone known to be involved in the regulation of multiple developmental processes in plants including stem elongation, germination, dormancy, flowering, flower development and leaf and fruit senescence. The large number of detected genes related to the gibberellin pathway suggests that gibberellins do indeed play an important role in plant development in barley. However, since gibberellins interact with many other proteins, factors and plant hormones, it is difficult to get a complete understanding of the gibberellin pathway and function in the plant.

Other plant hormones are also known to control development and maintenance of plant meristems and stem cell systems. Auxin in particular is seemingly involved in the regulation of nearly all stages of growth and development (Enders & Strader, 2015). A number of auxin response factors and other proteins related to the auxin pathway were found in close proximity to 13 different QTL associated with (one or multiple of) all investigated traits. Even though for most of these genes more likely candidate genes were identified, it is still possible that some of these, lesser known, auxin related genes are also responsible in part for some of the observed effects of these QTL.

As reviewed by Iqbal et al. (2017), ethylene is a multifunctional plant hormone that plays a key role in plant growth and reproductive development, as well as senescence. It interacts with a network of different signalling pathways and plant hormones to regulate the transition from the vegetative to the reproductive stage, and depending on the timing, concentration and plant species, it may promote or inhibit developmental processes such as flowering. Ethylene related genes were found close to 11 QTL that were associated with heading, maturity and/or, height. Although the effect of ethylene on these traits have not been studied in detail in barley, and for some of these QTL more likely candidate genes were identified, it is still possible that they were affected by ethylene.

4.7.5.3 MYB domain proteins and transcription factors (TFs) and GRAS family TFs

MYB domain proteins and transcription factors (TFs) are known to play a role in many developmental processes (Ambawat, Sharma, Yadav & Yadav, 2013). A large number of MYB transcription factors was found within or near multiple QTL associated with heading, maturity and height. Although these MYB-like TFs are unlikely to be responsible for the complete effect of these QTL, they may have contributed to the observed effects. However, this is impossible to ascertain with the results presented in the study. Due to the large number of MYB transcription factors found throughout the barley genome, it is a challenging task to isolate those factors with the largest effect, and discover their exact function. Similarly, GRAS family TFs are also known to influence plant development (Hirsch & Oldroyd, 2009). GRAS family TFs were found in the vicinity of five QTL, which were associated mostly with heading and maturity. Although it is impossible to deduct whether the GRAS family TFs contributed to the observed effects, the possibility of these TFs being of any importance should not be disregarded.

4.7.5.4 QTL without candidate genes

The QTL for which no known candidate genes could be identified were likely either putative novel QTL for these traits, or false positive results. Most of these QTL explained only small percentages of phenotypic variance of just one trait, were only just significant and/or were only found in one year, one dataset and using one method, rendering the chances of them being false positive results higher. This might suggest that these QTL are not very important for these traits and therefore do not warrant extensive continued research.

4.8 Comparison of results between methods, datasets and years

Only nine of the discovered QTL were found using both QTL analysis and GWAS. The reported explained percentages of phenotypic variance were higher in the results from QTL analysis than from GWAS. Although some differences between methods were expected, the differences were sometimes striking. For example, the same major row type QTL on chromosome 2H was found using both methods, however using the QTL analysis this QTL was reported to explain up to 99.5% of phenotypic variance, while only a maximum of 11.8% was reported using GWAS. GWAS and QTL analysis using linkage maps are two inherently different methods. While GWAS consists of a series of single t-tests for individual markers, QTL analysis using interval mapping is based on maximum likelihood statistics for QTL being present at certain positions in the intervals between markers along the linkage groups. Furthermore, this method takes recombination between neighbouring markers and the QTL into account, which GWAS does

not consider. With composite interval mapping which is also known as multiple QTL mapping, and was used in the present study, the method also corrects for the effects of other QTL. This causes increased precision and detection power. In summary, QTL analysis using interval mapping is a more powerful method, which may explain the differences between results from GWAS and QTL analysis.

Only four of the QTL found through QTL analysis using the SSD dataset were also found using the DH dataset. Six different QTL were only found using the SSD dataset, while 12 of the QTL found using the DH dataset were not found using the SSD dataset. This is surprising since performing GWAS on the DH data resulted in the detection of far fewer QTL than performing GWAS on the SSD dataset. Performing GWAS on the complete dataset resulted in the detection of the largest number of QTL. Furthermore, the DH dataset is smaller than the SSD dataset, which reduces the statistical power to detect QTL. However, the quality of the DH genetic map generated during QTL analysis was better than the map of the SSD dataset, which may explain this.

QTL found in 2014 were not always found in 2016, or vice versa. This might be explained by the fact that between-year correlation was not always perfect. It is possible that the observed differences between years may be attributed to Genotype by Environment (GxE) factors. In order to properly study GxE interactions, populations should preferably be studied in a variety of different environments or in a number of different years. However, in the present study, phenotypic data was only available for one study site and two years.

5. Conclusions

Analysis of the genetics underlying reproductive, developmental and structural traits such as heading, maturity and lodging is of importance for barley breeders in Iceland and other subarctic and marginal regions that aim to produce early maturing, high yielding and lodging resistant cultivars. QTL- and GWAS analysis are both valuable methods to analyse marker by phenotype associations. This study made use of both methods to examine genetic factors underlying earliness of heading (AGDD), maturity (AGDD), height (cm) and row type.

5.1 Analysis of field data

Analysis of field data showed large variation in phenotypic values for all investigated traits, and large differences between the observed phenotypic values for Golf and Tampar, indicating that the Golf x Tampar population was suited for QTL and GWAS analysis. Line type (DH or SSD) and row type proved to significantly affect phenotype for most traits, and in most years. Differences in average and variation of phenotypes were also observed between years. Between-year correlation was strongest for heading, followed by maturity and height, which suggests that heritability of these traits was higher than for the remaining traits with a weaker between-year correlation coefficient. Investigation of within- year correlations between traits showed that correlations were strongest between heading and maturity while other correlations were weaker. This suggested that partial overlapping of QTL for these traits was likely to occur.

5.2 Analysis of population structure, linkage disequilibrium and quality of mapping

Analysis of field data showed that there are significant phenotypic differences between SSD and DH lines, and between two-rowed and six-rowed lines. This same partitioning was also found when looking at the population structure. Running STRUCTURE ™ on the entire dataset found that Δk was smallest at $k=2$, and PCA also showed two clear distinct groups containing SSD respectively DH lines. Although it was unclear how this partitioning occurred, it is likely that selection and LD have contributed to this partitioning. LD analysis further showed that the mapping population had high LD and slow LD decay. LD decay was slower in the DH dataset than in the SSD dataset. This was most probably the result of the larger number of recombination events that has taken place during generation of the SSD population compared to the DH lines.

Because of the large difference between SSD and DH lines, SSD and DH datasets were analysed separately, which reduced statistical power of the analysis. It is therefore clear that in order to

facilitate better data analysis and increased statistical power during analysis of bi-parental RIL populations, it would be preferable to make only use of either line type instead of both, if circumstances allow. Further, when working with a species with very slow linkage decay, such as for example barley, SSD lines may be preferred since they have faster LD decay than DH lines, which may increase mapping resolution.

5.3 Candidate genes

HvCEN was suggested as a main candidate gene for QTL_2H.3 and AM_2H.1, which explained a large proportion of the observed phenotypic variance of the heading, maturity and height traits. However, since a number of other genes related to development (including *HvHOX2*, *HvGID2*, *HDG-2H*, *HvKNOX1* and *HvD11*) are also positioned in this area, it is possible that one or more of these genes are (also) partially responsible for the observed effect of these QTL. The effect of these QTL alone was often higher than the combined effect of all other QTL for a trait. It is therefore likely that genetic variance of the candidate genes for these QTL is responsible for a large part of the observed increased earliness of Tampar compared to Golf and other Nordic barley varieties. An investigation of sequence variation and copy number variation of the *HvCEN* gene, as well as the other candidate genes, in Golf and Tampar is therefore of special interest. Candidate genes for other, smaller, QTL include *HvELF3*, *HvFT1*, *HvCO2*, *HvCO3*, *HvCO8*, *HvCO9*, *HvCO11*, *HvCO14* and *Vrn-H2*. QTL associated with heading were sometimes also associated with maturity. Some candidate genes for QTL associated with maturity but not with heading include *Ppd-H1*, *HvCry1b*, *HvCO6* and *HvRAX1*. A total of 12 different QTL affecting height were found, with *Hv20ox2/denso/sdw1*, *HvBRI1*, *HvGID2*, *ari-e*, *HvGI* and *HvSTE1/HvD7* being the candidate genes of most interest in this study. Most of these are involved in the gibberellin pathway, highlighting the importance of gibberellins in the regulation of growth. *Vrs-1* was identified as a candidate gene for row type, explaining up to 99.5% of variance. This is in accordance with previous studies that identified *Vrs-1* as the major row type determining gene.

5.4 Future prospects

Results presented in this study showed that although the population consisted of two line types, LD decay was slow and population size was limited, the Golf x Tampar population is still a valuable source of phenotypic and genetic information that can be useful to investigate genetic mechanism underlying developmental traits in Nordic spring barley. Further research based on data from this Golf x Tampar mapping population should focus on sequence and copy number

variation between Golf and Tampar of some of the major candidate genes detected in this study. Especially the candidate genes *HvCEN*, *HvELF3* and *HvFT1* deserve to be studied further to investigate in more detail the effect of these genes on developmental traits in the Golf x Tampar mapping population. Increased knowledge on the genetic variation, pathways and molecular mechanisms of developmental traits such as heading, maturity, height and lodging will facilitate molecular breeding of barley cultivars adapted to grow in the subarctic climate conditions found for example on Iceland, and continued research on this matter is therefore of high importance.

6. References

Aastveit, K. *Yield components of some high yielding and some low yielding Nordic spring barley varieties.* Unpublished data.

Allaby, R.G. (2015). Barley domestication: The end of a central dogma? *Genome Biology, 16*, 176.

Alqudah, A.M., Sharma, R., Pasam, R.K., Graner, A., Kilian, B. & Schnurbusch, T. (2014) Genetic dissection of photoperiod response based on GWAS of pre-anthesis phase duration in spring barley. *PLOS ONE, 9*(11): e113120.

Alqudah, A.M., Koppolu, R., Wolde, G.M., Graner, A. & Schnurbusch, T. (2016). The genetic architecture of barley plant stature. *Frontiers in Genetics, 7*, 117.

Alqudah, A.M. & Schnurbusch, T. (2017). Heading date is not flowering time in spring barley. *Frontiers in Plant Science, 8*, 896.

Ambawat, S., Sharma, P., Yadav, N. R. & Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: An overview. *Physiology and Molecular Biology of Plants, 19*(3), 307–321.

Arifuzzaman, M., Günal, S., Bungartz, A., Muzammil, S., Afsharyan, N.P., Léon, J. & Naz, A.A. (2016). Genetic mapping reveals broader role of *Vrn-H3* gene in root and shoot development beyond heading in barley. *PLoS ONE, 11*(7), e0158718.

Badr, A., Müller, K., Schäfer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H. et al. (2000). On the origin and domestication history of barley (*Hordeum vulgare*). *Molecular Biology and Evolution, 17*(4), 499-510.

Bauer, A., Black, A.L., Frank, A.B. & Vasey, E.H. (1992). Agronomic characteristics of spring barley in the Northern Great Plains. *North Dakota State University Agricultural Experiment Station Bulletin No. 523*, 47.

Bauer, A., Fanning, C., Enz, J.W. & Eberlein, C.V. (1984). Use of growing-degree days to determine spring wheat growth stages. *North Dakota Cooperative Extensive Services*. EB-37. Fargo, ND.

Beavis, W.D. (1998). QTL analyses: Power, precision and accuracy. In Paterson, A.H. (Ed.), *Molecular dissection of complex traits* (pp. 145-162). Boca Raton: CRC Press.

Bengtsson, T., The PPP Barley Consortium, Manninen, O., Jahoor, A. & Orabi, J. (2017). Genetic diversity, population structure and linkage disequilibrium in Nordic spring barley (*Hordeum vulgare* L. subsp. *vulgare*). *Genetic Resources and Crop Evolution, 64*(8), 2021-2033.

Bergþórsson, P., Björnsson, H., Dýrmundsson, O., Gudmundsson, B., Helgadóttir, A. & Jónmundsson, J.V. (1987). The effect of climatic variations on agriculture in Iceland. In Parry, M.L., Carter, T.R. & Konijn, N.T. (Eds.), *The impact of climatic variations on agriculture, Vol I: Assessments in cool temperature and cold regions* (pp. 383-09). Dordrecht: Kluwer Academic Publishers Group.

Blattner, F.R. (2009). Progress in phylogenetic analysis and a new infrageneric classification of the barley genus *Hordeum* (*Poaceae*: *Triticeae*). *Breeding Science, 59*(5), 471-480.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635.

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

Caierão, E., (2006). Effect of induced lodging on grain yield and quality of brewing barley. *Crop Breeding and Applied Biotechnology, 6*, 215-221.

Campoli, C., Drosse, B., Searle, I., Coupland, G. & von Korff, M. (2012a). Functional characterisation of *HvCO1*, the barley (*Hordeum vulgare*) flowering time ortholog of *CONSTANS*. *The Plant Journal*, *69*(5), 868-880.

Campoli, C., Shtaya, M., Davis, S.J. & von Korff, M. (2012b). 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 Biology*, *12*(1), 1-15.

Cantalapiedra, C.P., Boudiar, R., Casas, A.M., Igartua, E. & Contreras-Moreira, B. (2015). BARLEYMAP: Physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. *Molecular Breeding, 35*, 1-11.

Carlson, C.S., Eberle, M.A., Kruglyak, L. & Nickerson, D.A. (2004). Mapping complex disease loci in whole-genome association studies. *Nature, 429*, 446-452.

Casao, M.C., Karsai, I., Igartua, E., Gracia, M.P., Veisz, O. & Casas, A.M. (2011). Adaptation of barley to mild winters: A role for *PPDH2*. *BMC Plant Biology*, *11*(1), 164-177.

Chan, E.K.F., Rowe, H.C. & Kliebenstein, D.J. (2010). Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics, 185*(3), 991-1007.

Chandler, P.M. & Harding, C.A. (2013). 'Overgrowth' mutants in barley and wheat: New alleles and phenotypes of the 'Green Revolution' *DELLA* gene. *Journal of Experimental Botany, 64*(6), 1603-1613.

Chono, M., Honda, I., Zeniya, H., Yoneyama, K., Saisho. D., Takeda, K. et al. (2003). Semidwarf phenotype of barley *uzu* results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. *Plant Physiology, 133*(3), 1209-1219.

Cockram,J., White, J., Zuluaga, D.L., Smith, D., Comadran, J., Macaulay, M. et al. (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences, 107*(50) 21611-21616.

Cockram, J, Thiel, T., Steuernagel, B., Stein, N., Taudien, S., Bailey, P.C. & O'Sullivan, D.M. (2012). Genome dynamics explain the evolution of flowering time *CCT* domain gene families in the *poaceae*. *PLOS ONE, 7*(9), e45307.

Comadran, J., Kilian, B., Russel, J., Ramsay, L., Stein, N., Ganal, M. et al. (2012). Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics, 44*, 1388-1392.

Crossa, J., Burgueño, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S.A., Lillemo, M. et al. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics*, *177*(3), 1889-1913.

Dai, F., Wang, X., Zhang, X., Chen, Z., Nevo, E., Jin, G. et al. (2018). Assembly and analysis of a *qingke* reference genome demonstrate its close genetic relation to modern cultivated barley. *Plant Biotechnology Journal, 16*, 760-770.

Davière, J.M. & Achard, P. (2013). Gibberellin signaling in plants. *Development, 140*(6), 1147-51.

Dolezel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysak, M.A., Nardi, L. & Obermayer, R. (1998). Plant genome size estimation by flow cytometry: Inter-laboratory comparison. *Annals of Botany, 82*, 17-26.

Drosse, B., Campoli, C., Mulki, A. & von Korff, M. (2014). The world importance of barley and challenges for further improvement. In Kumlehn, J. & Stein, N. (Eds.), *Biotechnological approaches to Barley Improvement, Biotechnology in Agriculture and Forestry 69* (pp. 81-99). Berlin Heidelberg: Springer Verlag.

Dunford, R.P., Griffiths, S., Christodoulou, V. & Laurie, D.A. (2005). Characterization of a barley (*Hordeum vulgare* L.) homologue of the *Arabidopsis* flowering time regulator *GIGANTEA. Theoretical Applied Genetics, 110*(5), 925-931.

Earl, D.A. & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*(2), 359-361.

Ellstrand, N.C. (2003). *Dangerous liaisons? When cultivated plants mate with their wild relative.* Baltimore: The John Hopkins University Press.

Enders, T.A. & Strader, L.C. (2015). Auxin activity: Past, present, and future. *American Journal of Botany, 102*(2), 180–196.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology, 14*(8), 2611-2620.

FAOSTAT (2014). Food and Agricultural Organization of the United Nations. http://faostat.fao.org. © FAO Statistics division.

Fauré, S., Higgins, J., Turner, A. & Laurie, D.A. (2007). The *FLOWERING LOCUS T*-like gene family in barley (*Hordeum vulgare*). *Genetics, 176* (1), 599-609.

Fauré, S., Turner, A.S., Gruszka, D., Christodoulou, V., Davis, S.J., von Korff, M. & Laurie, D.A. (2012). Mutation at the circadian clock gene *EARLY MATURITY 8* adapts domesticated barley (*Hordeum vulgare*) to short growing seasons. *Proceedings of the National Academy of Sciences of the United States of America, 109*(21), 8328-8333.

Fischbeck, G. (2003). Diversification through breeding. In von Bothmer, R., van Hintum, T, Knüpffer H. & Sato, K (Eds.), *Diversity in barley* (Hordeum vulgare) (pp. 29-54). Amsterdam: Elsevier BV.

Franckowiak, J.D. & Lundqvist, U. (2002). New and revised barley genetic stock description. *Barley Genetics Newsletter, 32*, 49-137.

Ganal, M.W., Altmann, T. & Roder, M.S. (2009). SNP identification in crop plants. *Current Opinion in Plant Biology, 12*(2), 211-217.

Ganal, M.W., Polley, A., Graner, E.M., Plieske, J., Wieseke, R., Luerssen, H. & Durstewitz, G. (2012). Large SNP arras for genotyping in crop plants. *Journal of Biosciences, 37*(5), 821-828.

Griffiths, S., Dunford, R. P., Coupland, G. & Laurie, D. A. (2003). The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis*. *Plant Physiology, 131*(4), 1855–1867.

Gupta, P.K., Kulwal, P.L., Jaiswal, V. (2014). Association mapping in crop plants: Opportunities and challenges. *Advances in Genetics, 85*, 524-535.

Gyenis, L., Yun, S.J., Smith, K.P., Steffenson, B.J., Bossolini, E. & Muehlbauer, G.J. (2007). Genetic architecture of quantitative trait loci associated with morphological and agronomic trait differences in a wild by cultivated barley cross. *Genome, 50*, 714-723.

Halldór Björnsson, Árný E. Sveinbjörnsdóttir, Anna K. Daníelsdóttir, Árni Snorrason, Bjarni D. Sigurðsson, Einar Sveinbjörnsson et al. (2008). *Hnattrænar loftslagsbreytingar og áhrif þeirra á Íslandi*. *Skýrsla vísindanefndar um loftslagsbreytingar*. Reykjavík: Ministry for the Environment.

Hamblin, M.T., Close, T.J., Bhat, P.R., Chao, S., Kling, J.G., Abraham, K.J. et al. (2010). Population structure and linkage disequilibrium in US barley germplasm: Implications for association mapping. *Crop Science, 50*(2), 556–566.

Hanano, S. & Goto, K. (2011). *Arabidopsis TERMINAL FLOWER1* is involved in the regulation of flowering time and inflorescence development through transcriptional repression. *The Plant Cell, 23*(9), 3172–3184.

Hedden, P. & Sponsel, V. (2015). A century of gibberellin research. *Journal of Plant Growth Regulation, 34*, 740–760.

Helgi Hallgrímsson & Guðríður Gyða Eyjólfsdóttir, (2004). Íslenskt sveppatal I. Smásveppir (Checklist of Icelandic fungi I. Microfungi.). *Fjölrit Náttúrufræðistofnunar, 45*, 1-189.

Hemming, M.N., Fieg, S., James Peacock, W., Dennis, E.S. & Trevaskis, B. (2009). Regions associated with repression of the barley (*Hordeum vulgare*) *VERNALIZATION1* gene are not required for cold induction. *Molecular Genetics and Genomics, 282*(5), 107-117.

Hemming, M.N., Peacock, W.J., Dennis, E. S. & Trevaskis, B. (2008). Low-temperature and daylength cues are integrated to regulate *FLOWERING LOCUS T* in barley. *Plant Physiology, 147*(5), 355–366.

Hill, W.H. & Weir, B.S. (1988). Variances and covariances of squared linkage disequilibria in finite population. *Theoretical Population Biology, 33*(1), 54-78.

Hilmarsson, H.S., Göransson, M., Lillemo, M., Kristjánsdóttir, Þ.A., Hermannsson, J. & Hallsson, J.H. (2017). An overview of barley breeding and variety trials in Iceland in 1987-2014. *Icelandic Agricultural Sciences, 30*, 13-28.

Hirsch, S. & Oldroyd, G. E. (2009). GRAS-domain transcription factors that regulate plant development. *Plant Signaling & Behavior, 4*(8), 698–700.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P. et al. (2015). MAGIC populations in crops: Current status and future prospects. *Theoretical and Applied Genetics, 128*(6), 999-1017.

Hudson, M.E., Lisch, D.R. & Quail, P.H. (2003). The *FHY3* and *FAR1* genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. *The Plant Journal, 34*, 453-471.

Humphreys, D.G. & Knox, R.E. (2015) Doubled haploid breeding in cereals. In Al-Khayri, J., Jain S., Johnson D. (Eds.), *Advances in plant breeding strategies: Breeding,*

biotechnology and molecular tools (pp. 241-290). Cham: Springer International Publishing.

Intellecta ehf. (2009). *Kornrækt á Íslandi – tækifæri til framtiðar*. Report commissioned by the Ministry of Fisheries and Agriculture. Reykjavík: Ministry of Fisheries and Agriculture.

IBSC (International Barley Sequencing Consortium), Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P. et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, *491*(7426), 711-716.

Iqbal, N., Khan, N.A., Ferrante, A., Trivellini, A., Francini, A. & Khan, M. (2017). Ethylene role in plant growth, development and senescence: Interaction with other phytohormones. *Frontiers in plant science*, *8*, 475.

Jannink, J.L. & Walsh, B. (2002). Association mapping in plant populations. In Kang, M.S. (Ed.), *Quantitative Genetics Genomics and Plant Breeding* (pp. 59-68). Wallingford, UK: CAB International.

Jia, Q., Hang, J., Westcott, S., Zhang, X., Bellgard, M., Lance, R. & Li, C. (2009). GA-20 oxidase as a candidate for the semidwarf gene *sdw1/denso* in barley. *Functional and Integrative Genomics*, *9*(2), 255-262.

Jónatan Hermannsson (1993). Kornrækt á Íslandi. In *Ráðunautafundur* (*1993*). Reykjavík: Búnaðarfélag Íslands, Rannsóknarstofnun Landbúnaðarins.

Jónatan Hermannsson (2004). Sjúkdómar í byggi. In *Fræðaþing landbúnaðarins 2004* (pp. 178-184). Reykjavík: Bændasamtök Íslands, Landbúnaðarháskólinn á Hvanneyri, Landgræðsla Ríkisins, Rannsóknarstofnun Landbúnaðarins, Skógrækt Ríkisins.

Keller, T., Abbott, J., Moritz, T. & Doerner, P. (2006). *Arabidopsis REGULATOR OF AXILLARY MERISTEMS1* controls a leaf axil stem cell niche and modulates vegetative development. *The Plant Cell*, *18*(3), 598–611.

Khush, G.S. (2001). Green revolution: The way forward. *Nature Reviews Genetics*, *2*(10), 815-822.

Kikuchi, R., Kawahigashi, H., Ando, T., Tonooka, T. & Handa, H. (2009). Molecular and functional characterization of PEBP genes in barley reveal the diversification of their roles in flowering. *Plant Physiology*, *149*(3), 1341-1353.

Kikuchi, R., Kawahigashi, H., Oshima, M., Ando, T. & Handa, H. (2012). The differential expression of *HvCO9*, a member of the *CONSTANS*-like gene family, contributes to the control of flowering under short-day conditions in barley. *Journal of Experimental Botany*, *63*(2), 773-784.

Kiseleva, A.A., Potokina, E.K. & Salina, E.A. (2017). Features of *Ppd-B1* expression regulation and their impact on the flowering time of wheat near-isogenic lines. *BMC Plant Biology*, *17*(Suppl 1), 172.

Komatsuda, T. & Mano, Y. (2002). Molecular mapping of the *intermedium spike-c* (*int-c*) and *non-brittle rachis 1* (*btr1*) loci in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, *105*, 85-90.

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D. et al. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proceedings of the National Academy of Sciences of the USA*, *104***,** 1424-1429.

Korte, A. & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods, 9*, 29.

Kraakman, A.T.W., Niks, R.E., van den Berg, P.M., Stam, P. & van Eeuwijk, F.A. (2004). Linkage disequilibrium mapping of yield and yield stability in modern spring barley sultivars. *Genetics, 168*(1), 435-446.

Kuczyńska, A., Surma, M., Adamski, T., Mikołajczak, K., Krystkowiak, K. & Ogrodowicz, P. (2013). Effects of the semi-dwarfing *sdw1/denso* gene in barley. *Journal of applied genetics, 54*(4), 381-90.

Kumlehn, J. & Stein, N. (Eds.) (2014), *Biotechnological approaches to Barley Improvement, Biotechnology in Agriculture and Forestry 69*. Berlin Heidelberg: Springer Verlag.

Künzel, G., Korzun, L. & Meister, A. (2000). Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. *Genetics, 154,* 397-412.

Laurie, D.A., Pratchett, N., Snape, J.W. & Bezant, J.H. (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.

Li, C. & Dubcovski, J. (2008). Wheat FT protein regulates *VRN1* transcription through interaction with *FDL2*. *The Plant Journal, 55*(4), 543-554.

Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y.M., Todhunter, R.J. et al. (2014). Enrichment of statistical power for genome-wide association studies. *BMC Biology, 12*(73).

Li, H., Chen, G. & Yan, W. (2015). Molecular characterization of barley 3H semi-dwarf genes. *PLoS ONE, 10*(3), e0120558.

Lillemo, M., Reitan, L. & Bjørnstad, Å. (2010). Increasing impact of plant breeding on barley yields in central Norway from 1946 to 2008. *Plant Breeding, 129,* 484-490.

Liu, H., Bayer, M., Druka, A., Russell, J. R., Hackett, C. A., Poland, J. et al. (2014). An evaluation of genotyping by sequencing (GBS) to map the *Breviaristatum-e* (*ari-e*) locus in cultivated barley. *BMC Genomics, 15*, 104.

Loscos, J., Igartua, E., Contreras-Moreira, B., Gracia, M.P. & Casas, A.M. (2014). *HvFT1* polymorphism and effect-survey of barley germplasm and expression analysis. *Frontiers in plant science, 5*, 251.

Lundqvist, U. (2009). Eighty years of Scandinavian barley mutation genetics and breeding. In Shu, Q.U. (Ed.), *Induced mutations in the genomics era* (pp. 39-43). Rome: FAO.

Ma, L. & Li, G. (2018). FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) family proteins in *Arabidopsis* growth and development. *Frontiers in Plant Science, 9,* 692.

Mammadov, J., Aggarwal, R., Buyyarapu, R. & Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *International Journal of Plant Genomics, 2012*, 1-11.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T. et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature, 544*(7651), 427-433.

Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K. et al. (2013). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *The Plant Journal, 76,* 718-727.

Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E. et al. (2015). Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics, 16*, 290.

Maurer, A., Draba, V. & Pillen, K. (2016). Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. *Journal of experimental botany, 67*(8), 2507-18.

Mayer, K.F.X., Martis, M., Hedley, P.E., Šimková, H., Liu, H., Morris, J.A. & Stein, N. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *The Plant Cell, 23*(4), 1249-1263.

Mickelson, H.R. & Rasmusson, D.C. (1994). Genes for short stature in barley. *Crop Science, 34*(5), 180-1183.

Mikołajczak, K., Kuczyńska, A., Krajewski, P., Sawikowska, A., Surma, M., Ogrodowicz, P. et al. (2017). Quantitative trait loci for plant height in Maresi × CamB barley population and their associations with yield-related traits under different water regimes. *Journal of Applied Genetics, 58*(1), 23-35.

Muñoz-Amatriaín, M., Cuesta-Marcos, A., Endelman, J.B., Comadran, J., Bonman, J.M., Bockelman, H.E. et al. (2014). The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. *PloS one, 9*(4), e94688.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E. & Buckler, E.S. (2009). Association mapping: Critical considerations shift from genotyping to experimental design. *The Plant Cell, 21*(8), 2194-2202.

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

Nice, L.M., Steffenson, B.J., Thomas, B., Horsley, R.D., Smith, K.P. & Muehlbauer, G.J. (2017). Mapping agronomic traits in a wild barley advanced backcross–nested association mapping population. *Crop Science, 57*, 1199-1210.

Nitcher, R., Distelfeld, A., Tan, C., Yan, L. & Dubcovsky, J. (2013). Increased copy number at the *HvFT1* locus is associated with accelerated flowering time in barley. *Molecular genetics and genomics, 288*(5-6), 261-75.

NordGen (2017). *PPP Pre-breeding.* Accessed 10. February 2018 via https://www.nordgen.org/en/plants/ppp-pre-breeding/.

Nurminiemi, M. (1995). *Environmental stability of Nordic barley materials*. Non-published Doctoral Thesis, Agricultural University of Norway, Ås.

Nurminiemi, M., Bjørnstad, Å. & Rognli, O.A. (1996). Yield stability and adaptation of Nordic barleys. *Euphytica, 92*(1-2), 191-202.

Obsa, B.T., Eglinton, J., Coventry, S., March, T., Langridge, P. & Fleury, D. (2016). Genetic analysis of developmental and adaptive traits in three doubled haploid populations of barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics, 129*(6), 1139-1151.

Ogugo, V., Semagn, K., Beyene, Y., Runo, S., Olsen, M. & Warburton, M.L. (2014). Parental genome contribution in maize DH lines derived from six backcross populations using genotyping by sequencing. *Euphytica, 202*(1), 129-139.

Pasam, R.K., Sharma, R., Malosetti, M., van Eeuwijk, F.A., Haseneyer, G., Kilian, B. & Graner, A. (2012). Genome-wide association studies for agronomical traits in a worldwide spring barley collection. *BMC Plant Biology, 12*, 16.

Poets, A.M., Fang, Z., Clegg, M.T. & Morrell, P.L. (2015). Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biology, 16*, 173.

Pollard D.A. (2012). Design and construction of recombinant inbred lines. In Rifkin S.A. (Eds.), *Quantitative Trait Loci (QTL). Methods in Molecular Biology (Methods and Protocols), vol 871*. New York: Springer Science + Business Media.

Pourkheirandish, M. & Komatsuda, T. (2007). The importance of barley genetics and domestication in a global perspective. *Annals of botany, 100*(5), 999-1008.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*, 904-909.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*, 945-959.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Rafalski, A, (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current opinions in Plant Biology, 5*(2), 94-100.

Rajkumara, S., (2008). Lodging in cereals. *Agricultural Reviews, 29*(1), 55-60.

Ramsay, L., Comadran, J., Druka, A., Marshall, D.F., Thomas, W.T.B., Macaulay, M. et al. (2011). *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. *Nature Genetics, 43*, 169-172.

Ramsay, L. (2014). Modulation of meiotic recombination. In Kumlehn, J. & Stein, N. (Eds.), *Biotechnological approaches to Barley Improvement, Biotechnology in Agriculture and Forestry 69* (pp. 311-329). Berlin Heidelberg: Springer Verlag.

Reykdal, Ó., Sveinsson, S., Dalmannsdóttir, S., Martin, P., Gerðinum, J.I., Kananagh, V. et al. (2016). Northern cereals – New opportunities. *Matís report.* ISSN 1670-7192.

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.

Rossini, L., Okagaki, R., Druka, A. & Muehlbauer, G.J. (2014). Shoot and inflorescence architecture. In Kumlehn, J. & Stein, N. (Eds.), *Biotechnological approaches to Barley Improvement, Biotechnology in Agriculture and Forestry 69* (pp. 3-19). Berlin Heidelberg: Springer Verlag.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Ruge-Wehling, B. & Wehling, P. (2014). The secondary gene pool of barley (*Hordeum bulbosum*): Gene introgression and homoeologous recombination. In Kumlehn, J. &

Stein, N. (Eds.), *Biotechnological approaches to Barley Improvement, Biotechnology in Agriculture and Forestry 69* (pp.331-343). Berlin Heidelberg: Springer Verlag.

Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J. & Itoh, T. (2016). Improvement of barley genome annotations by deciphering the Haruna Nijo genome. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *23*(1), 21-28.

Schulte, D., Close, T.J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G. et al. (2009). The International Barley Sequencing Consortium - At the threshold of efficient access to the barley genome. *Plant physiology*, *149*(1), 142-147.

Shimada, S., Ogawa, T., Kitagawa, S., Suzuki, T., Ikari, C., Shitsukawa, N. & Murai, K. (2009). A genetic network of flowering-time genes in wheat leaves, in which an *APETALA1/FRUITFULL*-like gene, *VRN1*, is upstream of *FLOWERING LOCUS T*. *The Plant Journal*, *58*(4), 668-681.

Shin-Young, H., Sangmin, L., Pil, J.S., Moon-Sik, Y. & Chung-Mo, P. (2010). Identification and molecular characterization of a *Brachipodium distachyon GIGANTEA* gene: Functional conservation in monocot and dicot plants. *Plant Molecular Biol*ogy, *72*(4-5), 485-497.

Shitsukawa, N., Ikari, C., Shimada, S., Kitagawa, S., Sakamoto, K., Saito, H. & Murai, K. (2007). The einkorn wheat (*Triticum monococcum*) mutant, maintained vegetative phase, is caused by a deletion in the VRN1 gene. *Genes & Genetic Systems*, *82*(2), 167-170.

Sigurbjörnsson, B. (2014). Research on small grains in support of a short-lived renaissance in cereal production in Iceland in the 1960s and its recent revival. Rit LbhÍ nr. 52.

Slatkin, M. (2008). Linkage disequilibrium – Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics, 9*, 477:485.

Snape, J.W. & Riggs, T.J. (1975). Genetical consequences of single seed descent in the breeding of self-pollinating crops. *Heredity, 35*, 11-219.

Somers, D.J., Isaac, P. & Edwards, K. (2004). A high density micro-satellite consensus map for bread wheat (*Triticum aestivum*, L). *Theoretical and Applied Genetics, 109*(6), 1105-1114.

Statistics Norway (2016a). Area used for grain and oil seed. Table 04607.

Statistics Norway (2016b). Grain. Yields. 1000 tonnes. Table 07479.

Stefansson, T.S., Serenius, M. & Hallsson, J.H. (2012). The genetic diversity of Icelandic populations of two barley leaf pathogens, *Rhynchosporium commune* and *Pyrenophora teres*. *European Journal of Plant Pathology*, *134*(1), 167-180.

Surma, M., Adamski, T., Kaczmarek, Z. & Czajka, S. (2006). Phenotypic distribution of barley SSD lines and doubled haploids derived from F1 and F2 hybrids. *Euphytica, 149*, 19–25

Szucs, P., Karsai, I., von Zitzewitz, J., Meszaros, K., Cooper, L.L., Gu, Y.Q. et al. (2006) Positional relationships between photoperiod response QTL and photoreceptor and vernalization genes in barley. *Theoretical and Applied Genetics, 112*, 1277-1285.

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F. et al. (2016). GAPIT Version 2: An enhanced integrated tool for genomic association and prediction. *The Plant Journal*, *9*(2).

Teplyakova, S., Lebedeva, M., Ivanova, N., Horeva, V., Voytsutskaya, N., Kovaleva, O. & Potokina, E. (2017). Impact of the 7-bp deletion in *HvGA20ox2* gene on agronomic important traits in barley (*Hordeum vulgare* L.). BMC Plant Biology, 17(Suppl 1), 181.

The James Hutton Institute (2012). *Barley iSelect SNP chip*. Accessed May 2018 at http://bioinf. hutton.ac.uk/iselect/app/.

Turner, A., Beales, J., Fauré, S., Dunford, R.P. & Laurie, D.A. (2005). The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science, 310*(5750), 1031-1034.

Turner, S.D. (2014). Qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. biorXiv. DOI: 10.1101/005165.

van Ooijen, J.W. (2006). JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

van Ooijen, J.W. (2009). MapQTL ® 6, Software for the mapping of quantitative trait loci in experimental ppulations of diploid. Kyazma B.V., Wageningen, Netherlands.

Verstegen, H., Köneke, O., Korzun, V. & von Broock, R. (2014). The world importance of barley and challenges for further improvement. In Kumlehn, J. & Stein, N. (Eds.), *Biotechnological approaches to Barley Improvement, Biotechnology in Agriculture and Forestry 69* (pp. 3-19). Berlin Heidelberg: Springer Verlag.

von Bothmer, R.V., Jacobsen, N., Baden, C., Jørgensen, R.B. & Linde-Laursen, I. (1995). *An ecogeographical study of the genus Hordeum. Systematic and ecogeographic studies on crop genepools 7* (2nd edition.). Rome: International Plant Genetic Resources Institute.

Voorrips, R. (2001). MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity, 93*(1), 77-78.

Vos, P.G., Paulo, M.J., Voorrips, R.E., Visser, R.G.F., van Eck, H.J. & van Eeuwijk, F.A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theoretical and Applied Genetics, 130*(1), 123-135.

Wonneberger, R., Ficke, A. and Lillemo, M. (2017). Identification of quantitative trait loci associated with resistance to net form net blotch in a collection of Nordic barley germplasm. *Theoretical and Applied Genetics, 130*(10), 2025-2043.

Wricke, G. & Weber, W.E. (1986). *Quantitative genetics and selection in plant breeding*. Berlin: Walter de Gruyten & Co.

Xiao, Y., Liu, H., Wu, L., Warburton, M. & Yan, J. (2017). Genome-wide association studies in maize: Praise and Stargaze. *Molecular Plant, 10*, 359-374.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Dubcovsky, J. (2006). The wheat and barley vernalization gene *VRN3* is an orthologue of FT. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(51), 19581-19586.

Yan, L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P. et al. (2004). The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science*, *303*(5664), 1640-1644.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T. & Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene *VRN1*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 6263-6268.

Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J et al. (2011). Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics*, *19*(7), 807–812.

Yuo, T., Yamashita, Y., Kanamori, H., Matsumoto, T., Lundqvist, U., Sato, K. et al. (2012). A SHORT INTERNODES (SHI) family transcription factor gene regulates awn elongation and pistil morphology in barley. *Journal of Experimental Botany*, 63(14), 5223–5232.

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

Zakhrabekova, S., Gough, S.P., Braumann, I., Müller, A.H., Lundqvist, J., Ahmann, K. et al. (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*(11), 4326-31.

Zhang, J. (2003). Inheritance of plant height and allelism tests of the dwarfing genes in Chinese barley. *Plant Breeding, 122*, 112–115.

Zhang, W., Zhang, .Qiao, Q., Jing, Q., Zhao, G., Jing, R. et al. (2012). Cloning and haplotype analysis of TaSTE, which is associated with plant height in bread wheat (*Triticum aestivum* L.). *Molecular Breeding, 31*(1), 47-57.

Zhang S., Xu F., Zhang Y., Lin J., Song C. & Fang X. (2015). Fine mapping and candidate gene analysis of a novel *PANICLE AND SPIKELET DEGENERATION* gene in rice. *Euphytica, 206* 793–803.

Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C. et al. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS Genetics, 3*(1), e4.

Zhou, H., Muehlbauer, G. & Steffenson, B. (2012). Population structure and linkage disequilibrium in elite barley breeding germplasm from the United States. *Journal of Zhejiang University. Science. B, 13*(6), 438-451.

Þórdís Anna Kristjánsdóttir (Ed.) (2013). Jarðræktarrannsóknir 2013. *Rit Lbhí, 44*, 30.

Þórdís Anna Kristjánsdóttir (Ed.) (2014). Jarðræktarrannsóknir 2013. *Rit Lbhí, 47*, 37.

Þórdís Anna Kristjánsdóttir (Ed.) (2015). Jarðræktarrannsóknir 2014. *Rit Lbhí, 60*, 31-33.

Þórdís Anna Kristjánsdóttir (Ed.) (2017). Jarðræktarrannsóknir 2016. *Rit LbhÍ, 76*, 34.

7. Supplementary material

Figure S1. Left: Total monthly precipitation (mm) in April-October 2014 and 2016. Right: Average monthly temperature (°C) in April-October 2014 and 2016 (Based on Þórdís Anna Kristjánsdóttir, 2015; 2017).

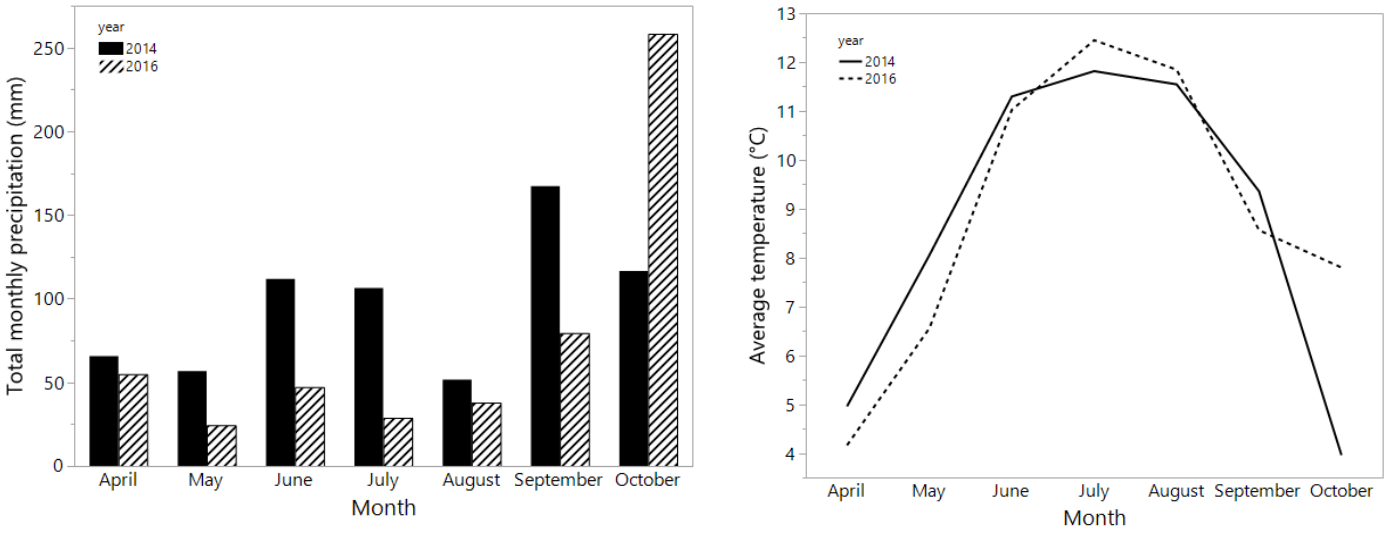


Figure S2. Experimental field design for 2014 and 2016 with 12 rows and 30 columns. The red/black checkered pattern differentiates the 12 blocks per repetition.

Compartment size 1,2 m^2 – Small blocks with each 15 compartments

1	72	63	96	151	7	103	154	124	121	152	97	34	127	64	134
2	73	37	163	105	164	95	2	66	4	40	100	69	158	67	93
3	133	39	71	162	15	11	122	41	123	157	65	99	161	1	130
4	74	132	155	44	12	125	14	75	92	156	91	31	94	8	36
5	126	129	32	104	6	68	153	128	101	135	3	131	9	13	43
6	70	102	165	35	61	160	98	62	33	5	45	10	38	42	159
7	145	109	57	139	170	50	23	147	106	30	136	25	56	172	180
8	90	171	166	144	114	108	29	86	107	115	142	148	52	167	54
9	143	169	78	76	60	80	79	177	178	89	117	84	116	88	119
10	59	118	53	146	175	179	174	110	17	137	87	149	82	26	21
11	22	173	120	19	112	18	24	176	168	77	47	113	81	51	141
12	85	150	140	58	27	83	48	49	138	20	111	16	55	46	28
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Repetition I

1	152	135	131	104	151	124	126	127	70	10	63	61	164	9	153
2	13	130	128	69	155	73	38	11	66	32	1	35	34	154	42
3	97	75	157	67	72	96	95	158	93	43	159	44	40	71	14
4	99	64	74	162	100	133	31	45	161	7	123	65	165	102	160
5	129	39	68	92	121	6	3	36	41	91	103	156	125	15	5
6	2	163	94	12	122	134	4	37	33	105	101	132	98	8	62
7	171	147	172	168	118	50	179	87	52	180	84	48	119	146	138
8	175	88	17	111	107	148	109	169	140	112	106	21	82	149	57
9	174	176	137	53	110	56	173	142	80	116	85	83	78	81	30
10	23	141	58	113	166	86	76	22	150	144	77	117	143	108	139
11	79	16	136	167	29	54	18	114	28	178	60	55	177	20	145
12	27	170	59	47	51	90	26	25	46	120	24	49	19	115	89
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Repetition II

N ↓

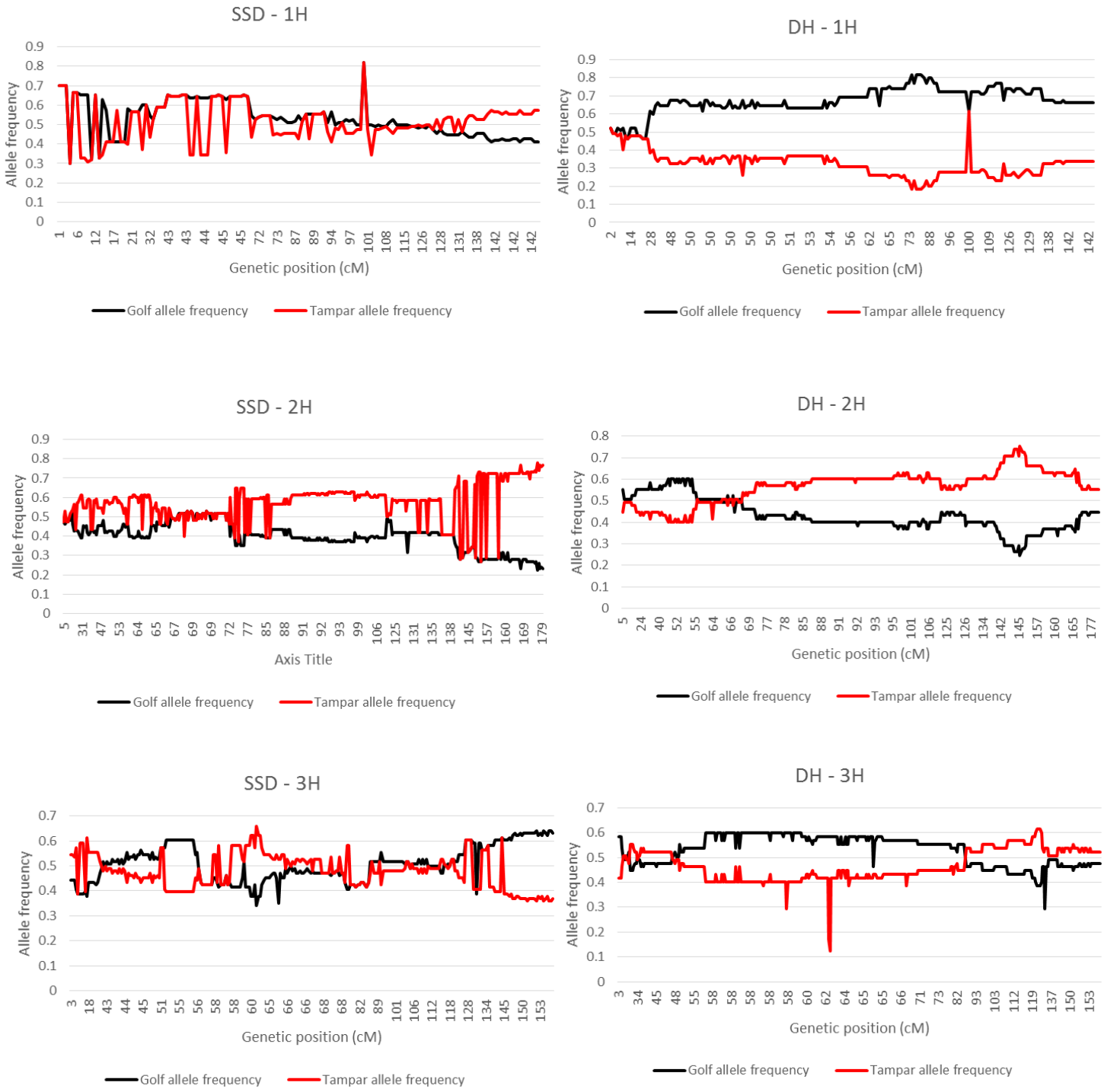
Figure S3. Table showing definitions of the traits discussed in this study.

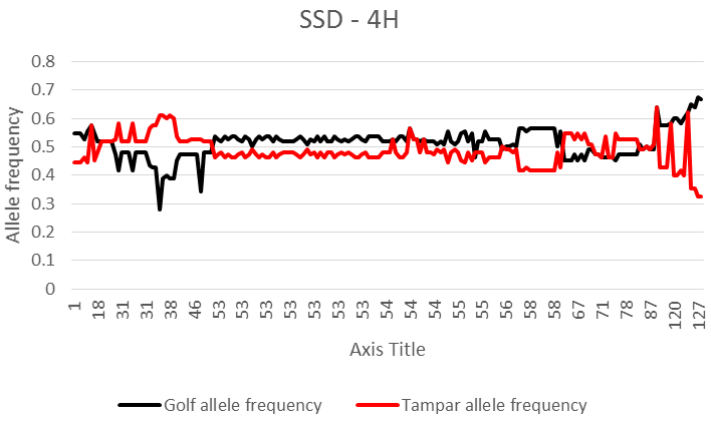
Trait	Description of trait	Method of measurement	Additional information
Heading date	Date at which 100% of the spike is showing in 50% of plants	a) Days from sowing b) AGDD*	2014: Date at which 50% of the spike is showing in 50% of plants + 3 days*
Maturity date	Date at which the peduncle is yellow in 50% of plants	a) Days from sowing b) AGDD*	
Height	Straw height from soil to spike	cm	Data for 2016 obtained 9.August
Row Type	Spike morphology: six row, two row or intermedium		

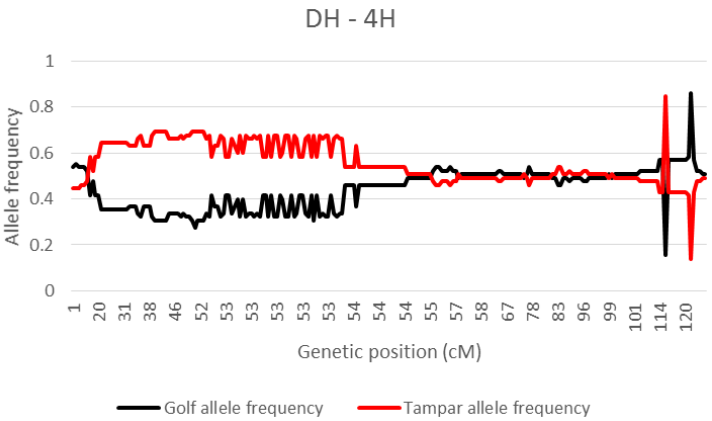
* AGDD = accumulated growing degree days

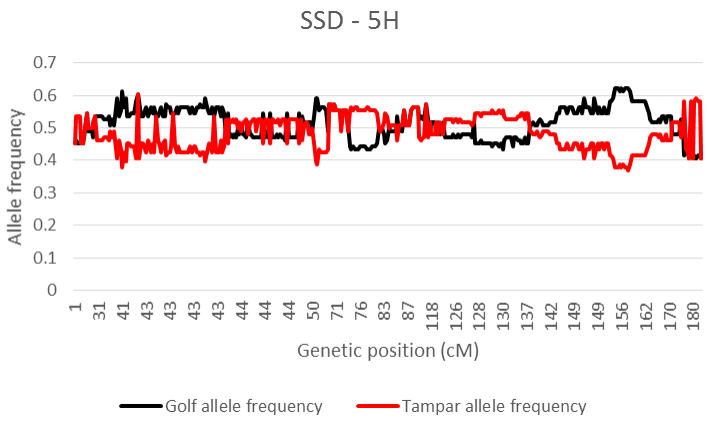
** On average, it takes 3 days for spikes that are 50% visible to reach full emergence (Jónatan Hermansson, personal communications, 2016).

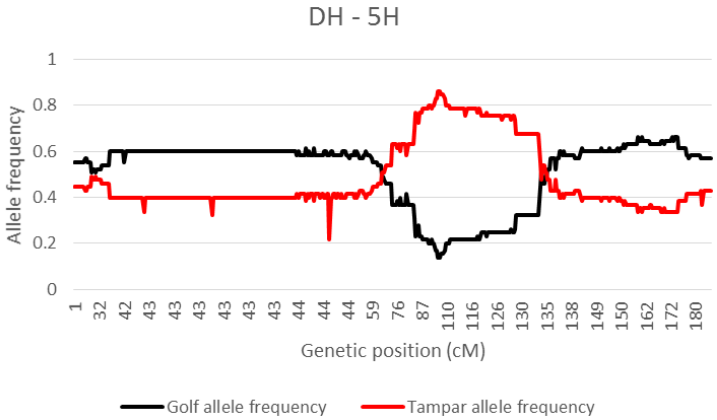
Figure S4. Graphs showing the frequency of the Golf respectively Tampar alleles along each chromosome in the SSD and DH datasets. Markers are positioned according to the consensus map by Muñoz-Amatriaín et al. (2014).

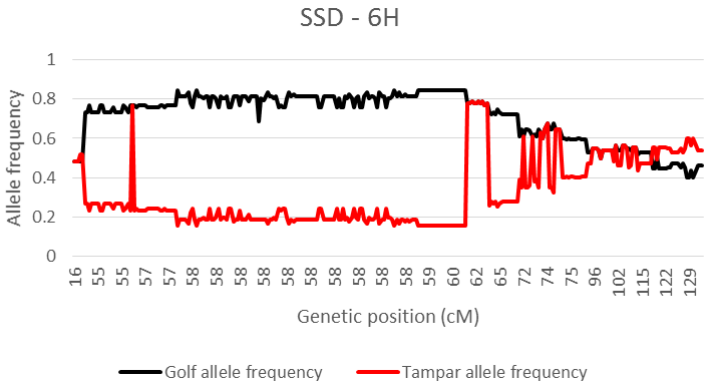


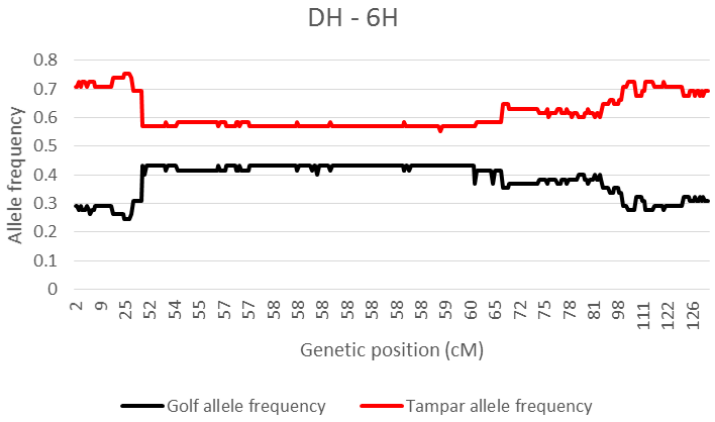

SSD - 4H
Allele frequency
Axis Title
Golf allele frequency
Tampar allele frequency

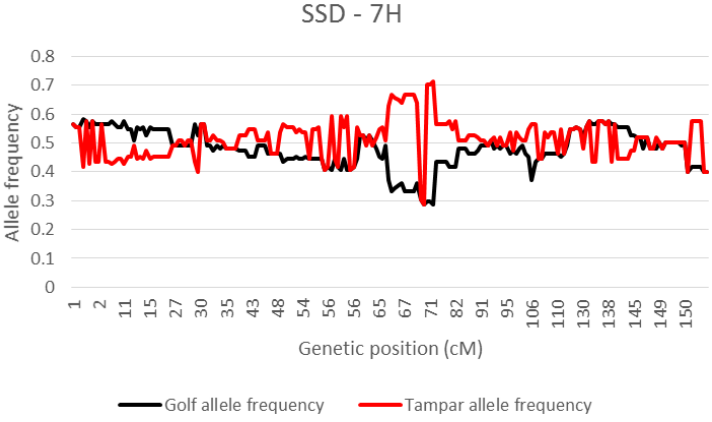

DH - 4H
Allele frequency
Genetic position (cM)
Golf allele frequency
Tampar allele frequency


SSD - 5H
Allele frequency
Genetic position (cM)
Golf allele frequency
Tampar allele frequency


DH - 5H
Allele frequency
Genetic position (cM)
Golf allele frequency
Tampar allele frequency


SSD - 6H
Allele frequency
Genetic position (cM)
Golf allele frequency
Tampar allele frequency


DH - 6H
Allele frequency
Genetic position (cM)
Golf allele frequency
Tampar allele frequency


SSD - 7H
Allele frequency
Genetic position (cM)
Golf allele frequency
Tampar allele frequency

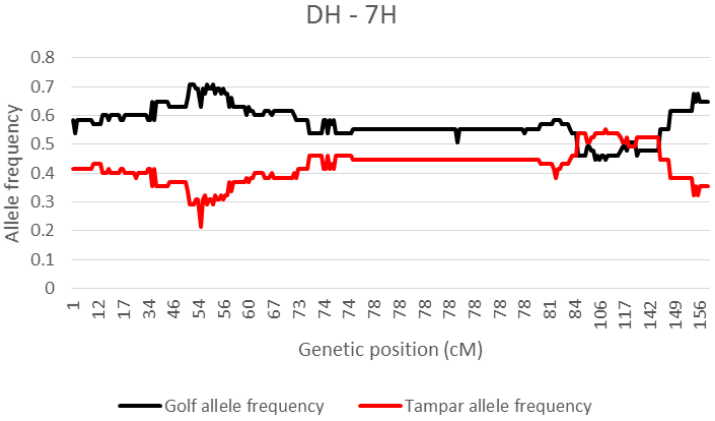

DH - 7H
Allele frequency
Genetic position (cM)
Golf allele frequency
Tampar allele frequency

Figure S5. Exploration of the optimal number of genetic subpopulations (k) in the complete mapping population. **Left:** Log probability of the data (LnP(D)) for each k between 1 and 12. **Right:** Δk values as a function of k.

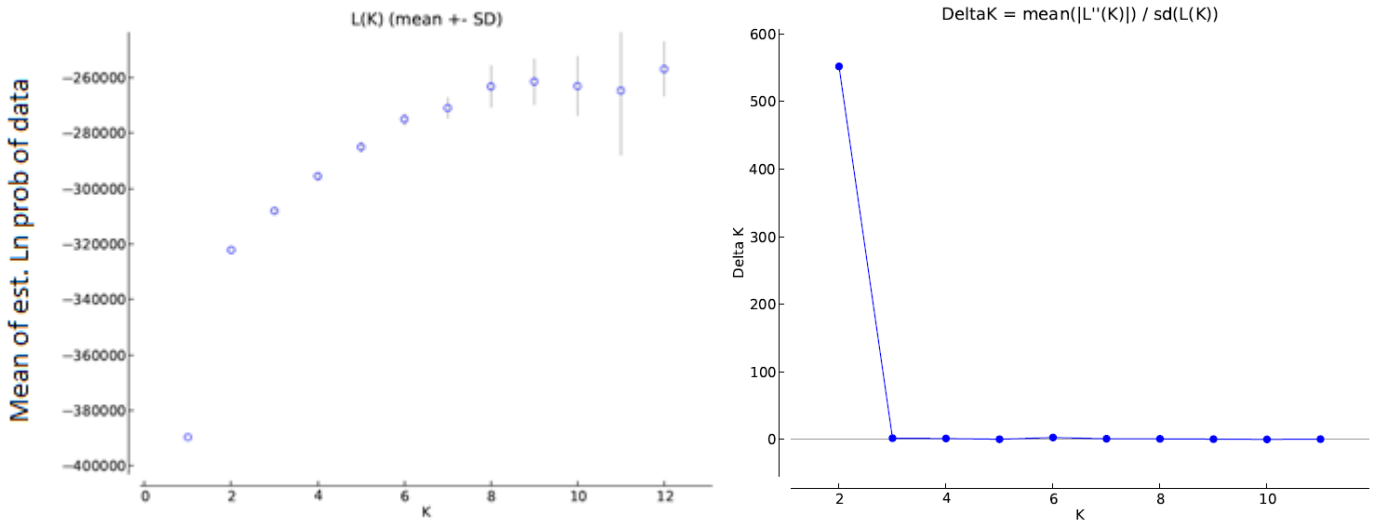


Figure S6. Exploration of the optimal number of genetic subpopulations (k) in the SSD dataset. **Left:** Log probability of the data (LnP(D)) for each k between 1 and 12. **Right:** Δk values calculated as a function of k.

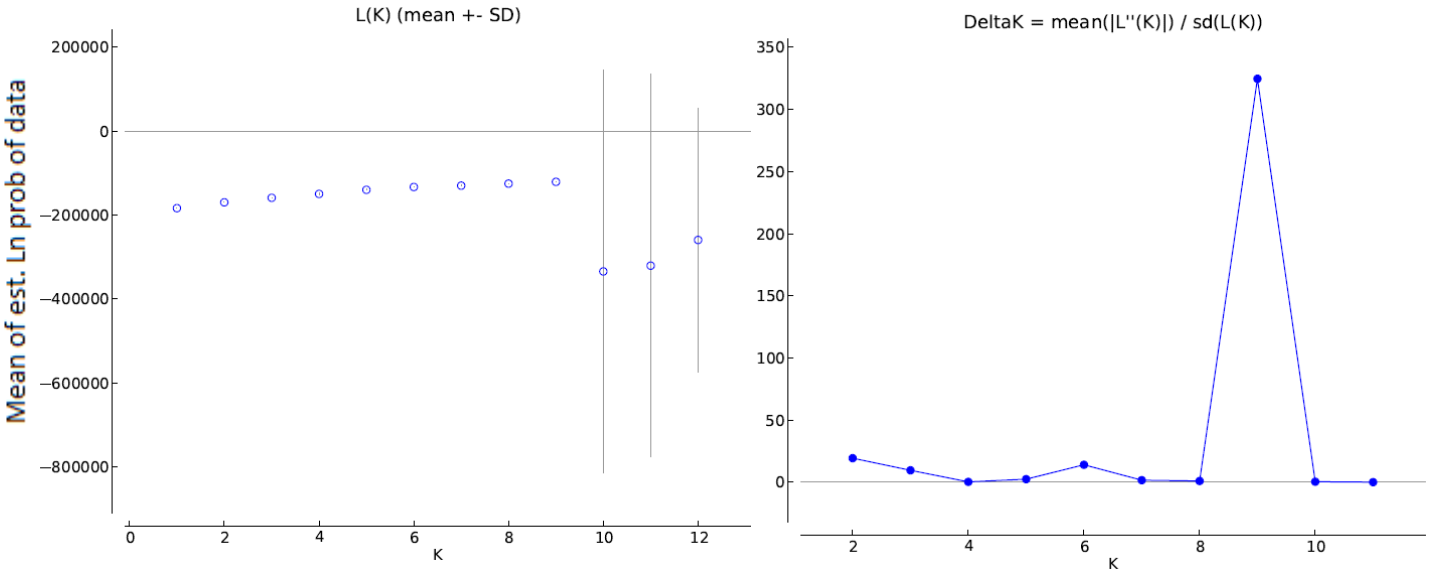


Figure S7. Exploration of the optimal number of genetic subpopulations (k) in the DH dataset. **Left:** Log probability of the data (LnP(D)) for each k between 1 and 12. **Right:** Δk values as a function of k.

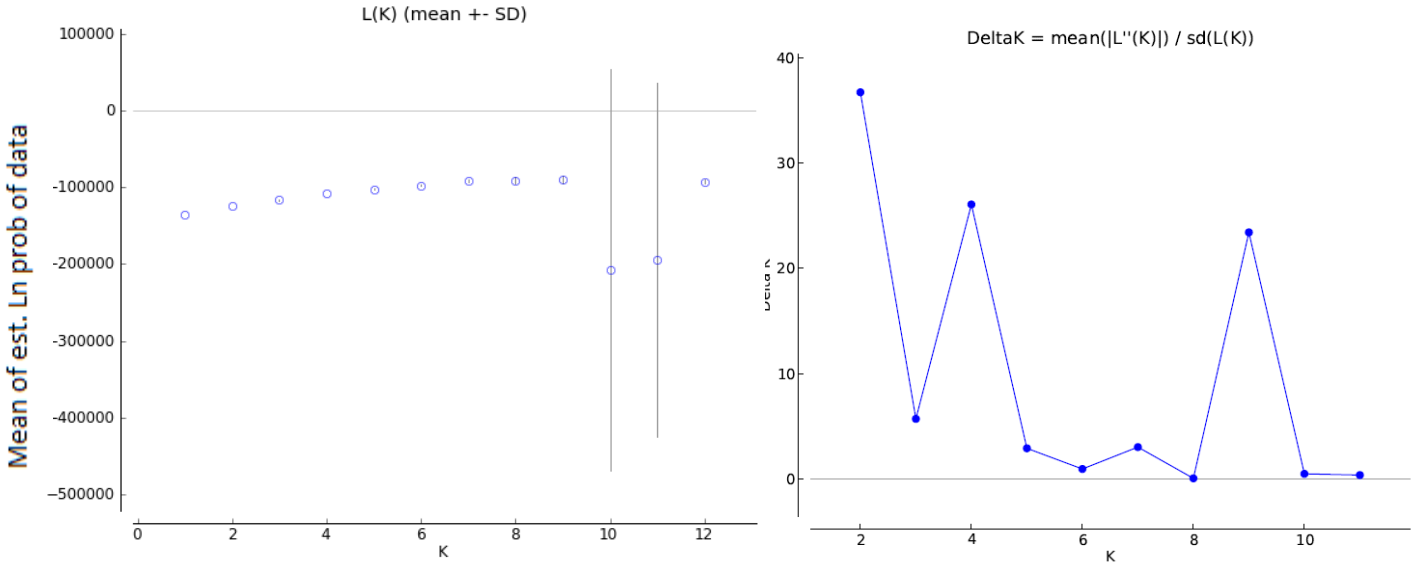


Figure S8. Results of Principal Component Analysis (PCA). **Left:** plot showing the variance explained by the first two PCs in the DH dataset. **Right:** plot showing the variance explained by the first two PCs in the SSD dataset. Accessions are coloured based on row type.

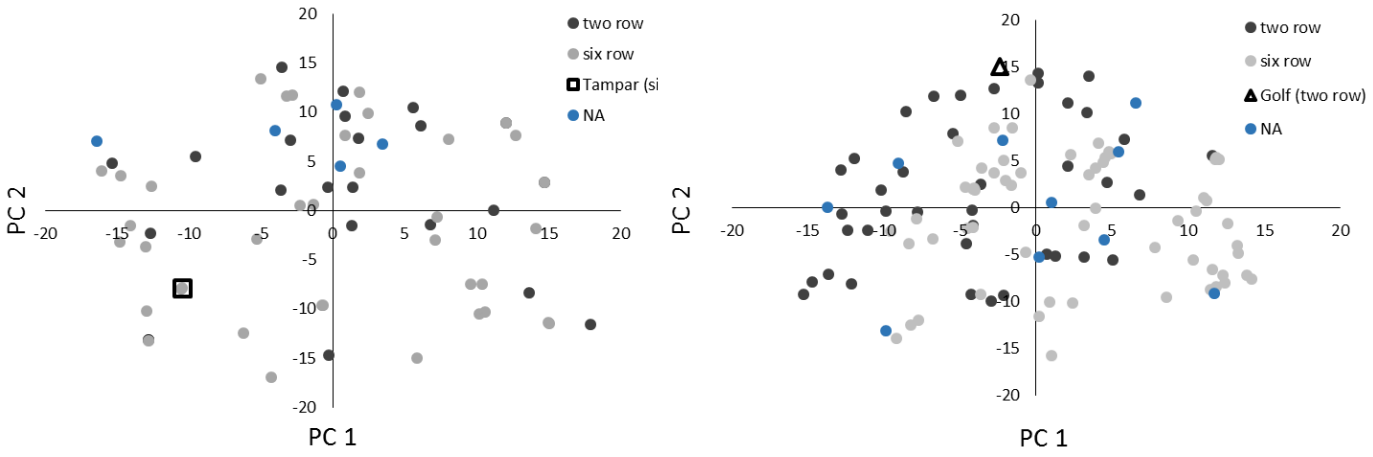
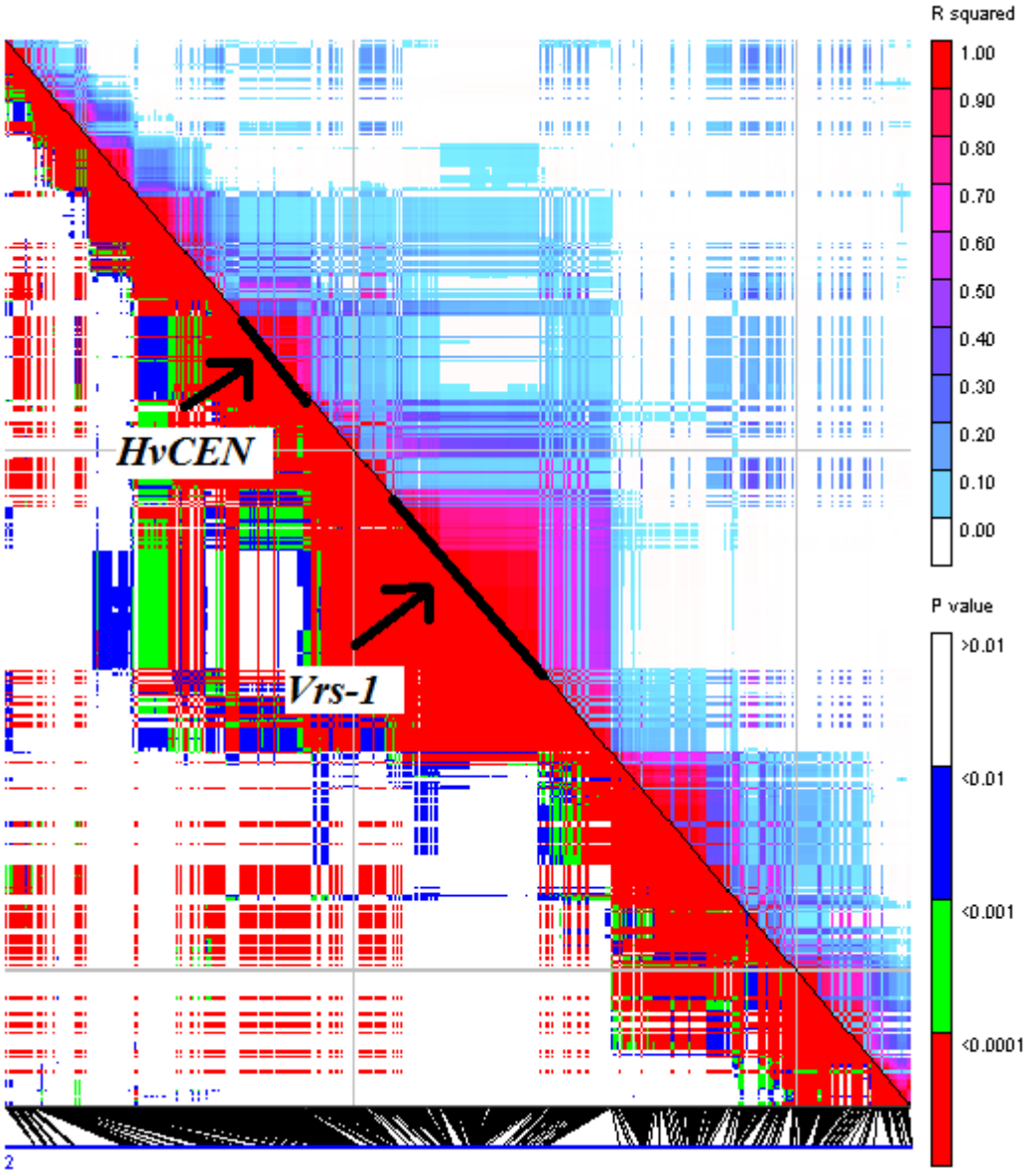
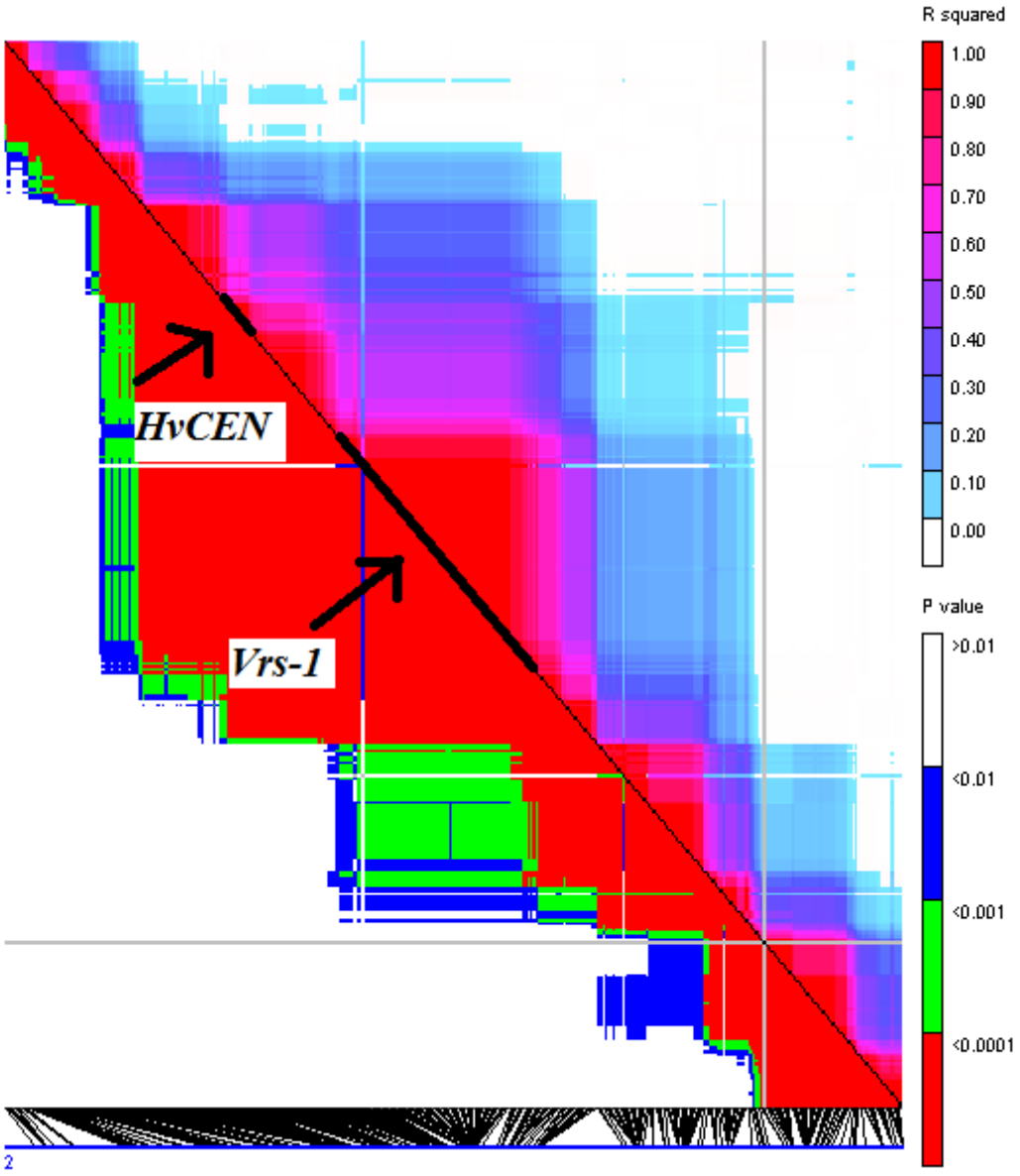


Figure S9. LD plot of chromsome 2H in A) the complete, B) the DH and C) the SSD dataset. The uper triangle shows R^2 and the lower triangle shows the P-value. Bold lines indicate the interval at which markers are linked to the gene.

A)



B)


R squared
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
P value
>0.01
<0.01
<0.001
<0.0001
HvCEN
Vrs-1
2

C)

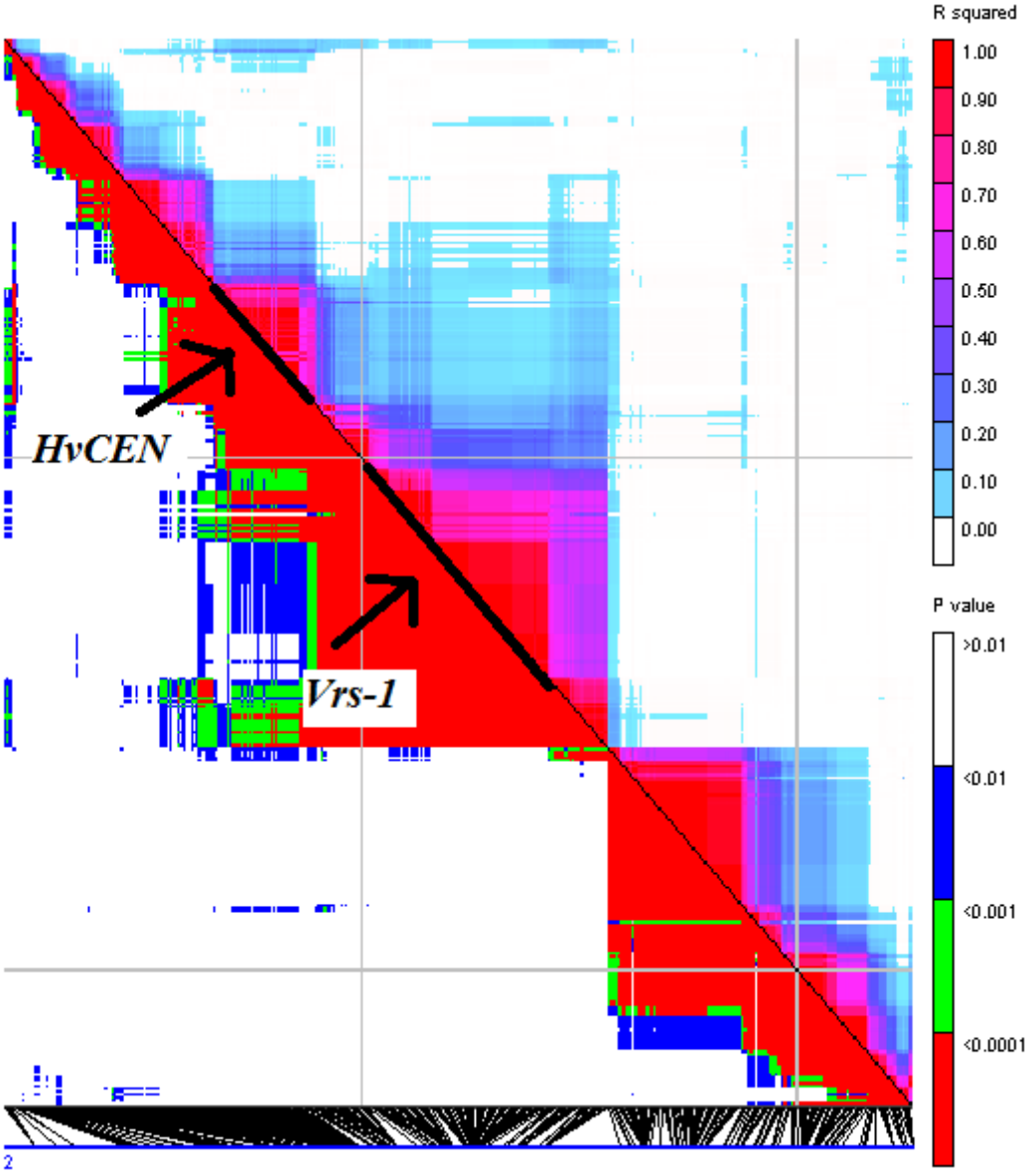
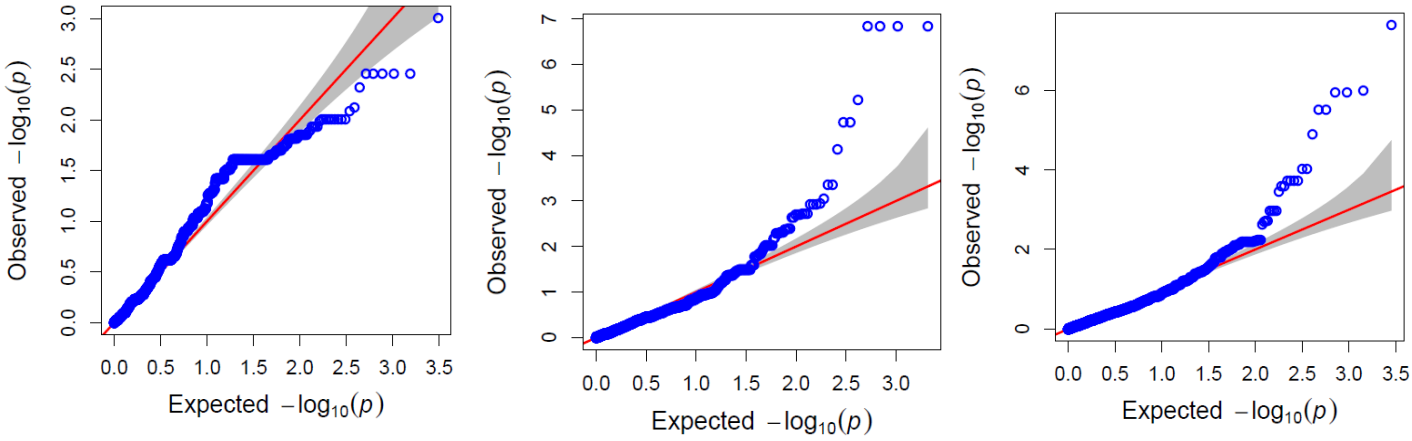
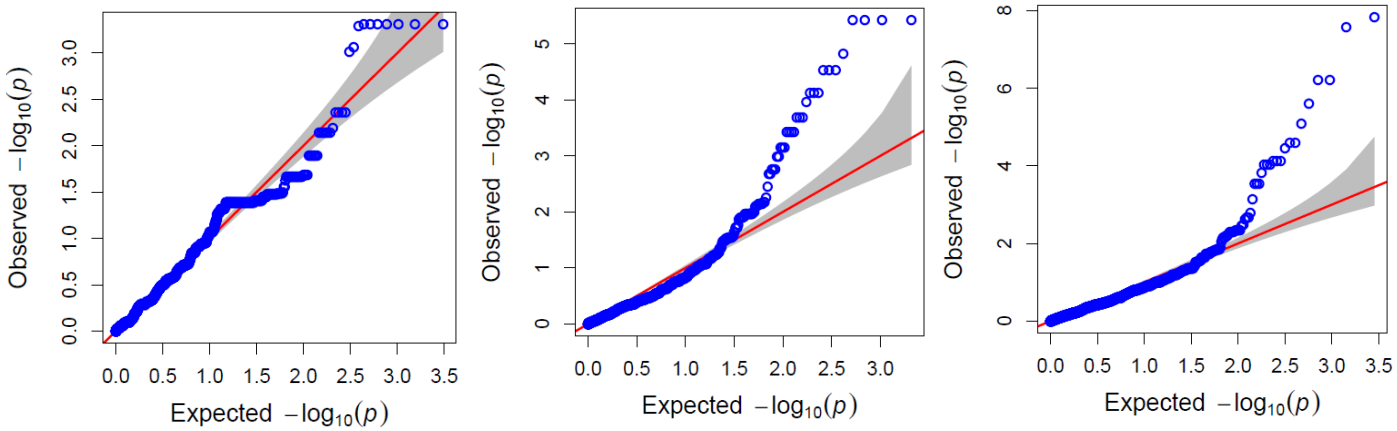

R squared
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
P value
>0.01
<0.01
<0.001
<0.0001
HvCEN
Vrs-1
2

Figure S10. QQ-plots for each trait, in each tested subset of lines.

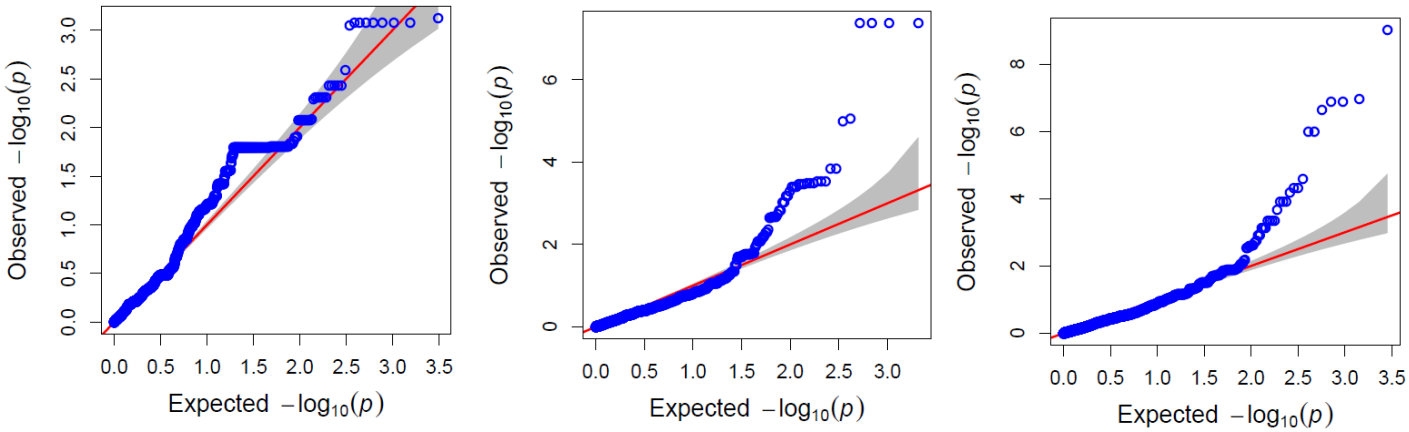
A). Heading (AGDD) (2014). From left to right: Plots for the DH, SSD and complete datasets.



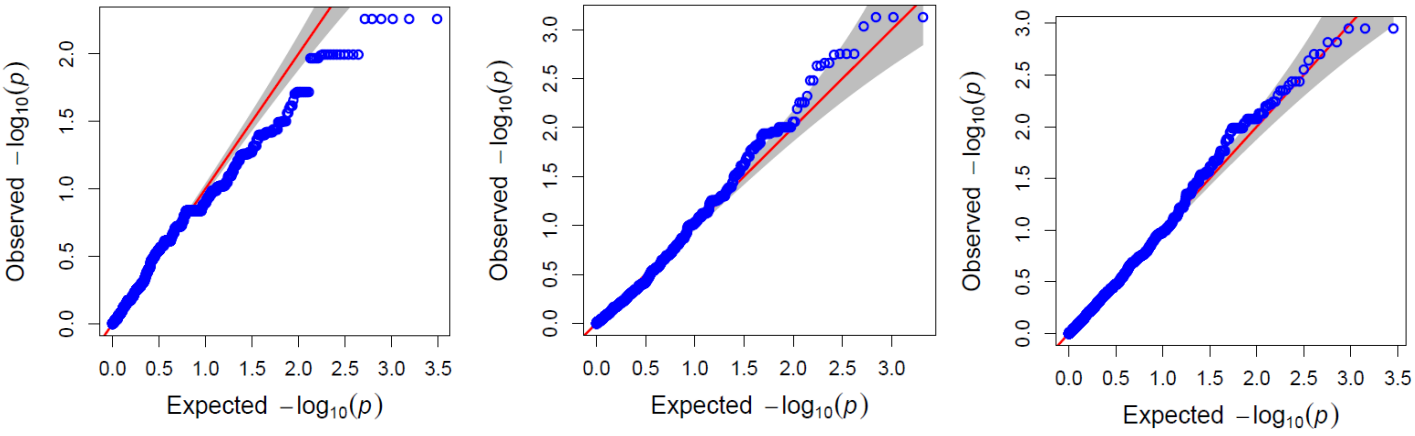
B). Heading (AGDD) (2016). From left to right: Plots for the DH, SSD and complete datasets.



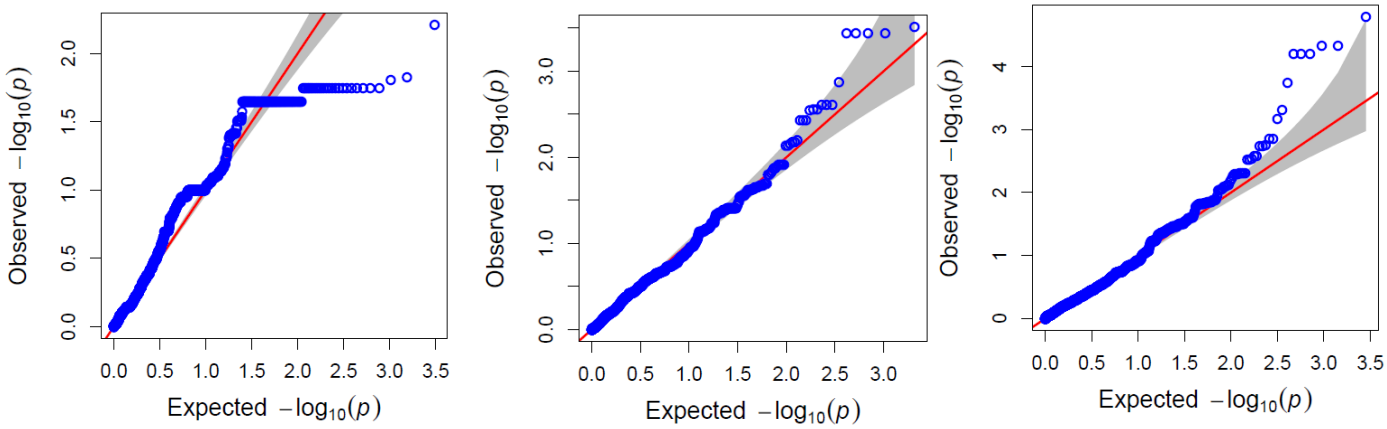
C). Heading (AGDD) (AV). From left to right: Plots for the DH, SSD and complete datasets.



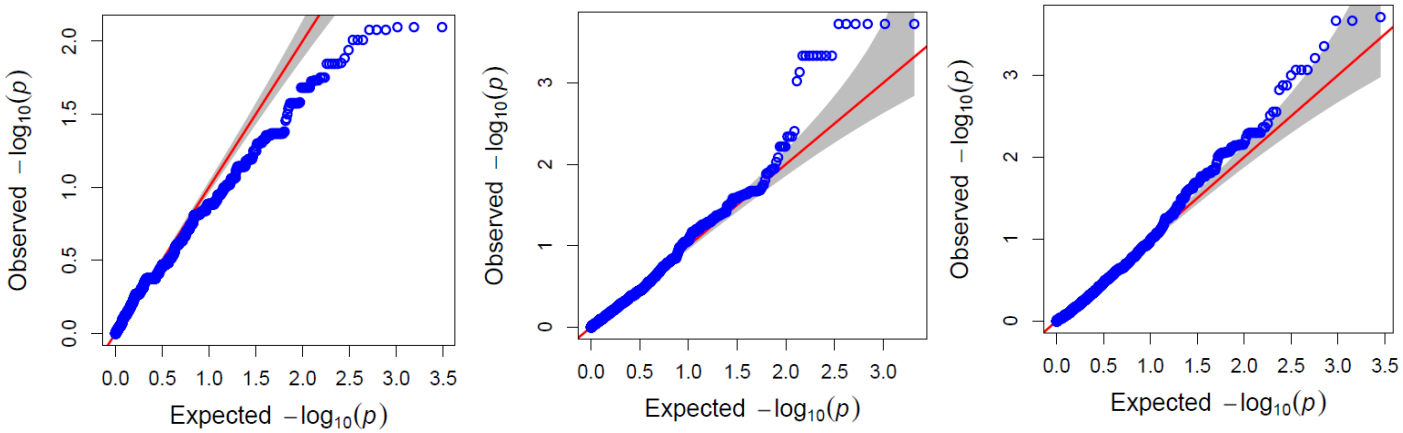
D). Maturity (AGDD) (2014). From left to right: Plots for the DH, SSD and complete datasets.



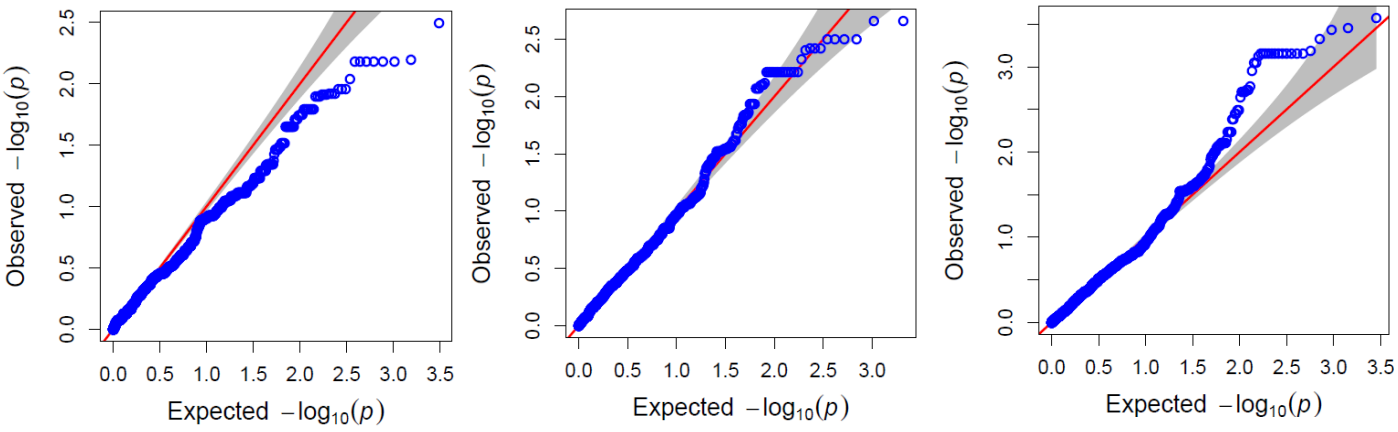
E). Maturity (AGDD) (2016). From left to right: Plots for the DH, SSD and complete datasets.



F). Height (2014). From left to right: Plots for the DH, SSD and complete datasets.



G). Height (2016). From left to right: Plots for the DH, SSD and complete datasets.



H). Row type. From left to right: Plots for the DH, SSD and complete datasets.

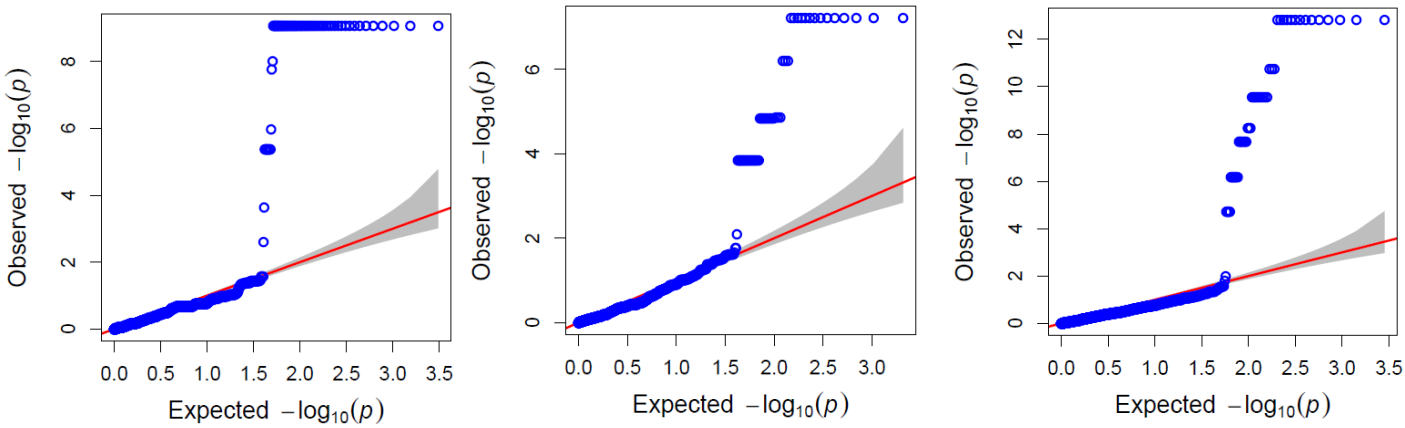


Figure S11. Position of QTL detected using GWAS according to the consensus map by Muñoz-Amatriaín et al. (2014), the genetic map by Comadran et al. (2012), the POPSEQ map (Mascher et al., 2013) and the IBSC_2012 map (IBSC, 2012).

Combined QTL	QTL	Significantly associated traits*	Chr.	Consensus map range (cM)***		Comadran map range (cM)***		POPSEQ map range (cM)***		IBSC_2012 map range (cM)***	
AM_1H.1	*AM_ALL_1H.1*	HD(16)	1H	5.68	5.68	5	5	4.96	4.96	4.96	4.96
	AM_ALL_8H.6	HD(16)	1H	-	-	-	-	6.02	6.02	6.02	6.02
AM_1H.2	*AM_SSD_1H.1*	HT(16)	1H	88.82	89.34	86.8	87	86.54	86.54	86.76	86.83
AM_1H.3	*AM_ALL_1H.2*	HT(14)	1H	133	133	122	122	122	122	119.69	119.69
AM_1H.4	*AM_SSD_1H.2*	HD(14, 16, AV), MT(16)	1H	142.7	142.7	132	132	-	-	132.37	132.37
	AM_ALL_1H.3	HD(14, 16, AV)	1H	142.7	142.7	133	133	132.1	132.5	132.37	132.51
AM_2H.1	*AM_SSD_2H.1*	HD(14, 16, AV)	2H	69	74.49	57.4	62.5	58.64	63.6	56.59	62.76
	AM_DH_2H.1	HD(16, AV)	2H	69	69.6	57.4	59.5	58.64	58.92	56.37	59.49
	AM_ALL_2H.1	HD(14, 16, AV), MT(14, 16)	2H	69	74.5	57.4	62.7	58.07	63.6	56.52	62.76
	AM_SSD_8H.2	HD(14, 16, AV)	2H	-	-	-	-	62.46	62.46	-	-
	AM_ALL_8H.5	HD(16, AV)	2H	-	-	-	-	62.46	62.46	-	-
AM_2H.2	*AM_DH_2H.2*	RT	2H	85.19	99	74.2	86.6	74.08	86.76	74.29	86.61
	AM_ALL_2H.2	RT	2H	86.4	95.2	75.2	82.8	75.21	82.08	75.21	86.86
	AM_SSD_2H.2	RT	2H	86.44	95.24	75.2	82.8	75.21	82.08	75.21	82.51
	AM_SSD_8H.4	RT	2H	-	-	-	-	76.91	77.09	70.90	80.03
	AM_DH_8H.1	RT	3H	-	-	-	-	76.91	77.09	77.9	80.03
AM_3H.1	*AM_ALL_8H.3*	HT(16)	3H	-	-	-	-	45.4	49.72	46.18	51.2
	AM_ALL_3H.1	HT(14, 16)	3H	55.4	61.9	46.2	55.1	45.82	54.53	46.18	55.1
	AM_SSD_3H.1	HT(14)	3H	58.31	61.89	55.1	55.1	52.03	54.53	51.2	55.1
	AM_SSD_8H.1	HT(14)	3H	-	-	-	-	51.2	52.03	51.2	53.26
AM_3H.2	*AM_ALL_3H.2*	HT(16)	3H	117.6	122.1	116	118	109.8	117.6	112.71	118.2
	AM_DH_3H.1	HT(16)	3H	122.1	122.1	118	118	117.6	117.6	118.20	118.20
AM_4H.1	*AM_ALL_4H.1*	HD(14)	4H	117.1	117.1	111	111	111.1	111.1	111.33	111.33
AM_5H.1	*AM_DH_5H.1*	HD(14)	5H	45	45	46.7	46.7	46.32	46.32	46.74	46.74
AM_6H.1	*AM_DH_6H.1*	HD(16, AV)	6H	74.2	74.2	67.9	67.9	-	-	67.92	67.92
AM_7H.1	*AM_ALL_7H.1*	HD(14, 16, AV), MT(14,16)	7H	35.01	45.14	29.8	38.8	29.96	40.65	29.82	38.81
	AM_SSD_7H.1	HD(14,16, AV), MT(14, 16), HT(14)	7H	34.74	45.14	32.8	38.8	26.71	40.65	32.05	38.81
	AM_SSD_8H.3	HD(14, 16, AV), MT(14, 16)	7H	-	-	-	-	28.98	28.98	30.56	34.45
	AM_ALL_8H.1	MT(16)	7H	-	-	-	-	28.98	28.98	30.56	34.35
AM_8H.1	*AM_ALL_8H.4*	MT(14)	5H**	-	-	-	-	-	-	584990735	584990735

* HD= Heading date, MT= maturity, HT= height, RT= row type

** Morex_2016 physical map, via T3/barley triticeae toolbox (https://triticeaetoolbox.org)

*** Range of markers assoicated with the QTL for which a position was available.

Figure S12. Significant markers associated with each QTL found using GWAS. Continuation of table of next page.

Combined QTL	QTLs	Significant markers
AM_1H.1	*AM_ALL_1H.1*	SCRI_RS_232577
	AM_ALL_8H.6	SCRI_RS_198544
AM_1H.2	*AM_SSD_1H.1*	BOPA1_4578-899, BOPA1_3087-1763, BOPA1_3201-603, BOPA2_12_30546, SCRI_RS_17486
AM_1H.3	*AM_ALL_1H.2*	BOPA2_12_10693
AM_1H.4	*AM_SSD_1H.2*	SCRI_RS_143790, SCRI_RS_4928, BOPA1_13095-187
	AM_ALL_1H.3	SCRI_RS_143790, SCRI_RS_4928, BOPA1_13095-187, BOPA1_4057-2114, SCRI_RS_117283, SCRI_RS_150935
AM_2H.1	*AM_SSD_2H.1*	BOPA1_ABC17685-1-4-365, BOPA1_1826-229, BOPA1_2719-672, BOPA1_3813-1629, BOPA1_1917-848, BOPA1_1940-567, BOPA1_217-677, BOPA1_2895-1064, BOPA1_6510-1430, BOPA1_6911-866, SCRI_RS_4802,
	AM_DH_2H.1	BOPA1_ABC17685-1-4-365, BOPA2_12_30042, SCRI_RS_117951, SCRI_RS_175300, BOPA1_1826-229, BOPA1_2719-672, BOPA1_3813-1629, BOPA1_7489-442, SCRI_RS_144776
	AM_ALL_2H.1	BOPA1_ABC17685-1-4-365, BOPA2_12_30042, SCRI_RS_117951, BOPA1_1826-229, BOPA1_1826-229, BOPA1_2719-672, BOPA1_2719-672, BOPA1_3813-1629, SCRI_RS_127347, SCRI_RS_144776, SCRI_RS_1502, SCRI_RS_165574, BOPA1_6804-1197, SCRI_RS_162413, SCRI_RS_202786, SCRI_RS_153531, BOPA1_4434-804
	AM_SSD_8H.2	SCRI_RS_156063
	AM_ALL_8H.5	SCRI_RS_156063, SCRI_RS_34413, SCRI_RS_215063
AM_2H.2	*AM_DH_2H.2*	SCRI_RS_133539, SCRI_RS_156323, SCRI_RS_129857, SCRI_RS_221795, SCRI_RS_222844, SCRI_RS_239742, BOPA1_7549-782, SCRI_RS_139193, SCRI_RS_166540, SCRI_RS_200291, SCRI_RS_2961, SCRI_RS_154398, SCRI_RS_157347, SCRI_RS_134812, BOPA1_4629-162, BOPA1_6117-1507, BOPA1_ABC04861-2-1-334, BOPA1_ABC09163-1-3-313, SCRI_RS_235860 SCRI_RS_4930, BOPA1_2322-462, BOPA1_2371-950, SCRI_RS_91810, BOPA2_12_30896, BOPA2_12_30897, SCRI_RS_196853, SCRI_RS_88704, SCRI_RS_137263, BOPA1_ConsensusGBS0705-1, BOPA2_12_31205, SCRI_RS_13565, SCRI_RS_192398, SCRI_RS_221886, SCRI_RS_3376, SCRI_RS_180028, BOPA1_3469-1152, SCRI_RS_10670, SCRI_RS_128449, SCRI_RS_156090, SCRI_RS_159462, SCRI_RS_160616, SCRI_RS_160833, SCRI_RS_172648, SCRI_RS_185505, SCRI_RS_198601, SCRI_RS_198603, SCRI_RS_211281, SCRI_RS_235221, SCRI_RS_179213, SCRI_RS_188339, SCRI_RS_208760, BOPA2_12_10936, BOPA2_12_31424, BOPA1_8632-1809, BOPA2_12_10969, SCRI_RS_119261, SCRI_RS_155456, SCRI_RS_157236, SCRI_RS_162798, SCRI_RS_174800, SCRI_RS_176114, SCRI_RS_182039, SCRI_RS_183984, SCRI_RS_211894, SCRI_RS_237481, BOPA1_6024-1095, SCRI_RS_138463, SCRI_RS_219074
	AM_ALL_2H.2	SCRI_RS_129857, SCRI_RS_221795, SCRI_RS_222844, SCRI_RS_239742, BOPA1_7549-782, SCRI_RS_166540, SCRI_RS_200291, SCRI_RS_154398, SCRI_RS_157347, SCRI_RS_134812, BOPA1_4629-162, BOPA1_6117-1507, BOPA1_ABC04861-2-1-334, BOPA1_ABC09163-1-3-313, SCRI_RS_235860, SCRI_RS_4930, BOPA1_2322-462, BOPA1_2371-950, SCRI_RS_91810, BOPA2_12_30896, BOPA2_12_30897, SCRI_RS_196853, SCRI_RS_88704, SCRI_RS_137263, BOPA1_ConsensusGBS0705-1, BOPA2_12_31205, SCRI_RS_13565, SCRI_RS_192398, SCRI_RS_221886, SCRI_RS_180028, SCRI_RS_128449, SCRI_RS_156090, SCRI_RS_159462, SCRI_RS_160616, SCRI_RS_179213, SCRI_RS_188339, SCRI_RS_208760, BOPA2_12_10936, BOPA2_12_31424, BOPA1_8632-1809, BOPA2_12_10969, SCRI_RS_119261, SCRI_RS_155456, BOPA1_6024-1095, SCRI_RS_138463, SCRI_RS_219074
	AM_SSD_2H.2	SCRI_RS_129857, SCRI_RS_221795, SCRI_RS_222844, SCRI_RS_239742, BOPA1_7549-782, SCRI_RS_166540, SCRI_RS_200291, SCRI_RS_154398, SCRI_RS_157347, SCRI_RS_134812, BOPA1_4629-162, BOPA1_6117-1507, BOPA1_ABC04861-2-1-334, BOPA1_ABC09163-1-3-313, SCRI_RS_235860, SCRI_RS_4930, BOPA1_2322-462, BOPA1_2371-950, SCRI_RS_91810, BOPA2_12_30896, BOPA2_12_30897, SCRI_RS_196853, SCRI_RS_88704, SCRI_RS_137263, BOPA1_ConsensusGBS0705-1, BOPA2_12_31205, SCRI_RS_13565, SCRI_RS_192398, SCRI_RS_221886, SCRI_RS_180028, BOPA1_3469-1152, SCRI_RS_10670, SCRI_RS_128449, SCRI_RS_156090, SCRI_RS_179213, SCRI_RS_188339, SCRI_RS_208760, BOPA2_12_10936, BOPA2_12_31424, BOPA1_8632-1809, BOPA2_12_10969, SCRI_RS_119261, SCRI_RS_155456, BOPA1_6024-1095, SCRI_RS_138463, SCRI_RS_219074
	AM_DH_8H.1	SCRI_RS_133327, SCRI_RS_150663, SCRI_RS_154203
	AM_SSD_8H.4	SCRI_RS_133327, SCRI_RS_150663, SCRI_RS_154203

AM_3H.1	*AM_ALL_8H.3*	SCRI_RS_200297, SCRI_RS_235443, SCRI_RS_239787, SCRI_RS_153735, BK_08, BOPA2_12_20037, SCRI_RS_126627
	AM_ALL_3H.1	BOPA1_6354-1193, SCRI_RS_182479, SCRI_RS_185596, SCRI_RS_218235, SCRI_RS_230096, BOPA2_12_31122, BOPA1_6460-355, SCRI_RS_189045, SCRI_RS_219247, SCRI_RS_175073, BOPA1_5038-1035, SCRI_RS_152172, SCRI_RS_186102, SCRI_RS_222102, SCRI_RS_173186, BOPA1_2335-1614, BOPA1_ABC06381-1-5-73
	AM_SSD_3H.1	BOPA1_10353-119, BOPA1_1630-1150, BOPA1_2288-126, BOPA1_2588-229, SCRI_RS_173186
	AM_SSD_8H.1	SCRI_RS_6877, SCRI_RS_209249, SCRI_RS_212864, SCRI_RS_219551, SCRI_RS_239403, SCRI_RS_221644, SCRI_RS_181360, SCRI_RS_190876, SCRI_RS_160464, SCRI_RS_138918, SCRI_RS_126246
AM_3H.2	*AM_ALL_3H.2*	BOPA2_12_30972, SCRI_RS_146012, SCRI_RS_167755
	AM_DH_3H.1	SCRI_RS_167755
AM_4H.1	*AM_ALL_4H.1*	BOPA1_4160-1365, SCRI_RS_107762
AM_5H.1	*AM_DH_5H.1*	BOPA1_12045-83
AM_6H.1	*AM_DH_6H.1*	SCRI_RS_9620
AM_7H.1	*AM_ALL_7H.1*	SCRI_RS_138457, SCRI_RS_138461, SCRI_RS_160602, SCRI_RS_236580, SCRI_RS_153154, BOPA1_12701-485, BOPA2_12_30893, BOPA2_12_30894, BOPA2_12_30895, BOPA1_1213-1959, SCRI_RS_155061, SCRI_RS_179937, BOPA2_12_31305, SCRI_RS_209176
	AM_SSD_7H.1	BOPA2_12_30083, BOPA1_12701-485, BOPA2_12_30893, BOPA2_12_30894, BOPA2_12_30895, BOPA1_1213-1959, SCRI_RS_155061, SCRI_RS_179937, BOPA2_12_31305, SCRI_RS_209176, SCRI_RS_204256
	AM_SSD_8H.5	SCRI_RS_140831, SCRI_RS_220780, SCRI_RS_165545, BK_05
	AM_ALL_8H.1	BK_05, SCRI_RS_140831, SCRI_RS_220780, SCRI_RS_139762
AM_8H.1	*AM_ALL_8H.4*	BOPA1_1213-1959, SCRI_RS_155061, SCRI_RS_179937, BOPA2_12_31305, SCRI_RS_209176

Figure S13. Genetic maps for the DH dataset generated using Joinmap and visualised using MapChart.

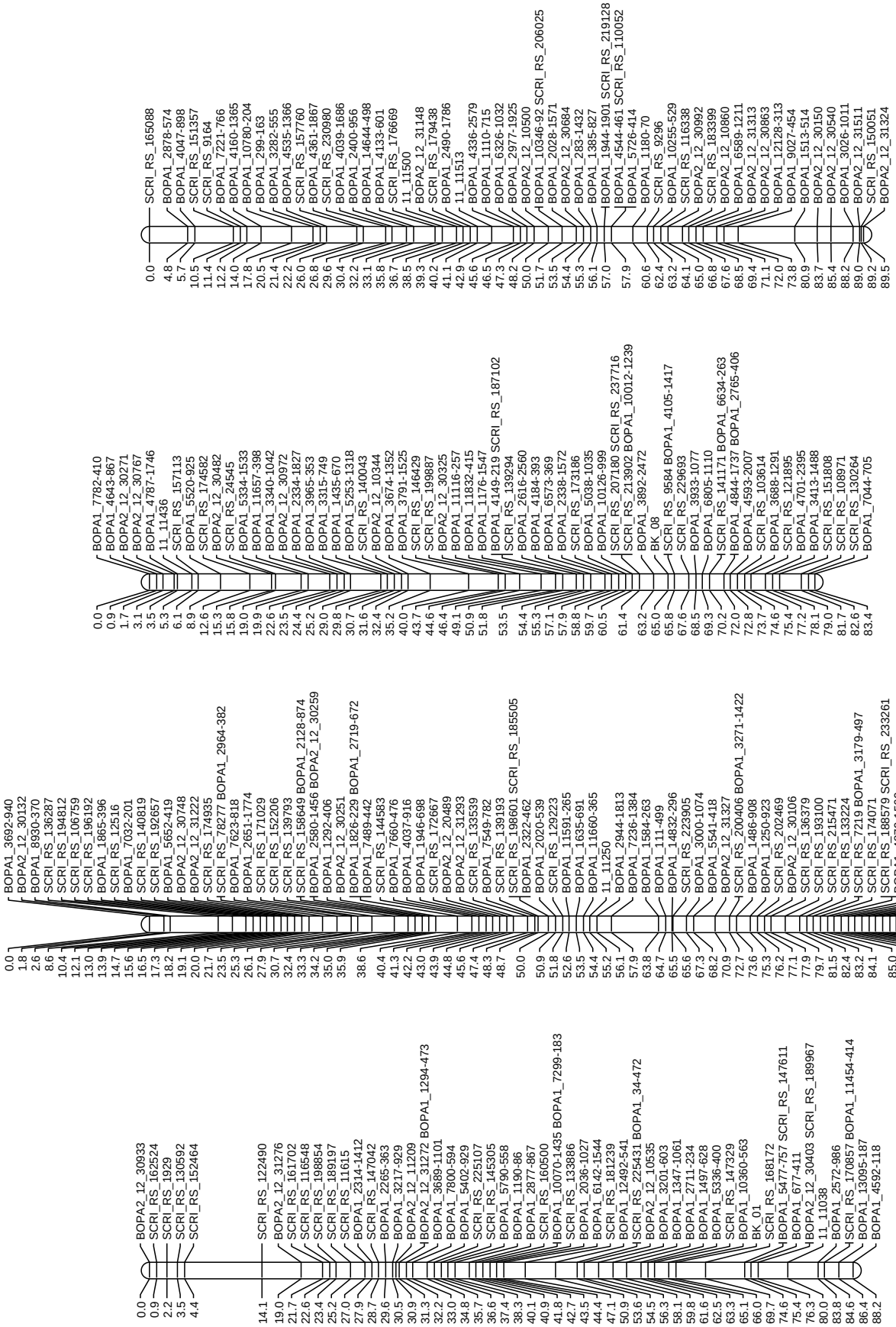


Figure S13. – continued.

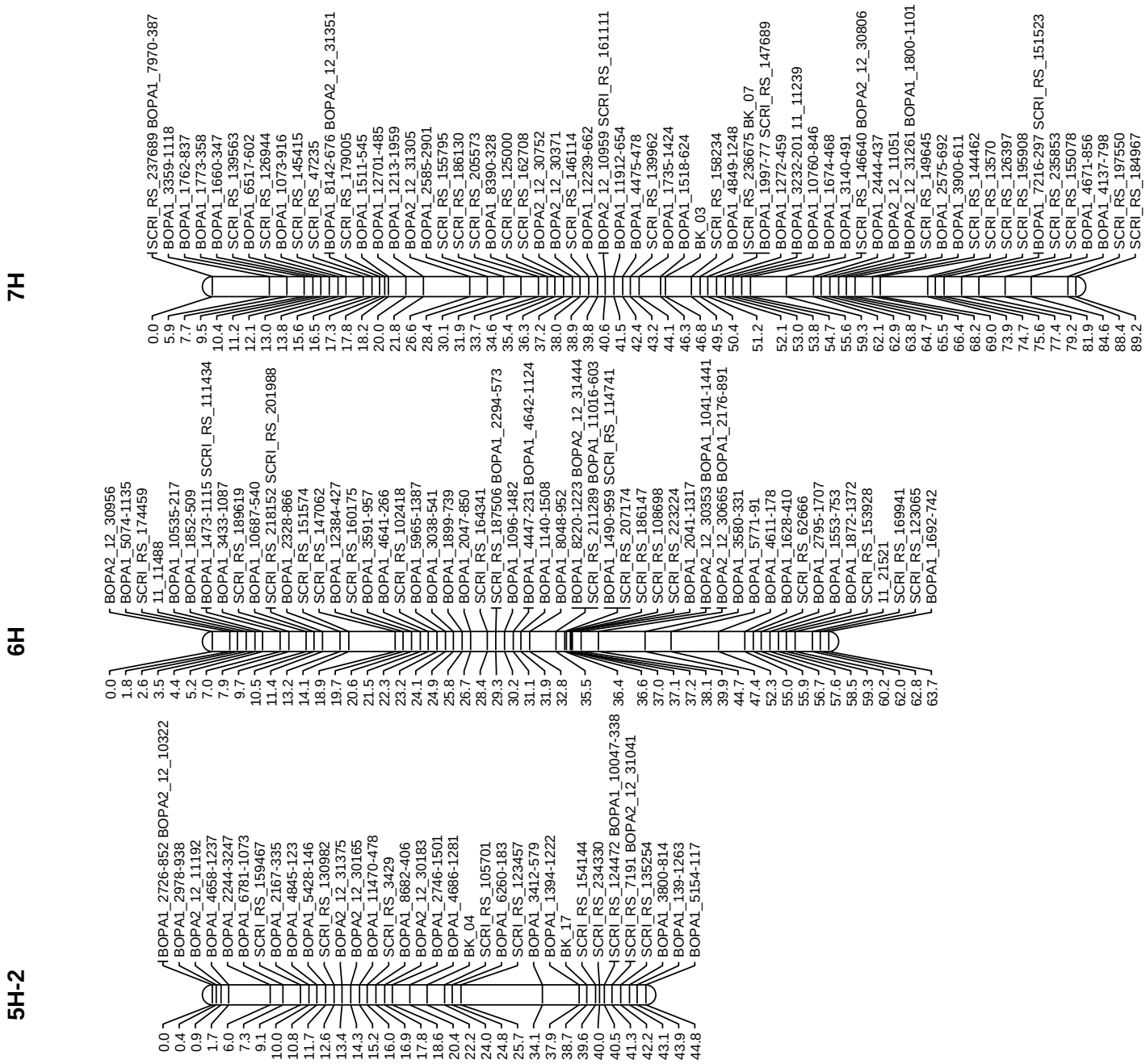


Figure S14. Genetic maps for the SSD dataset generated using Joinmap and visualised using MapChart.

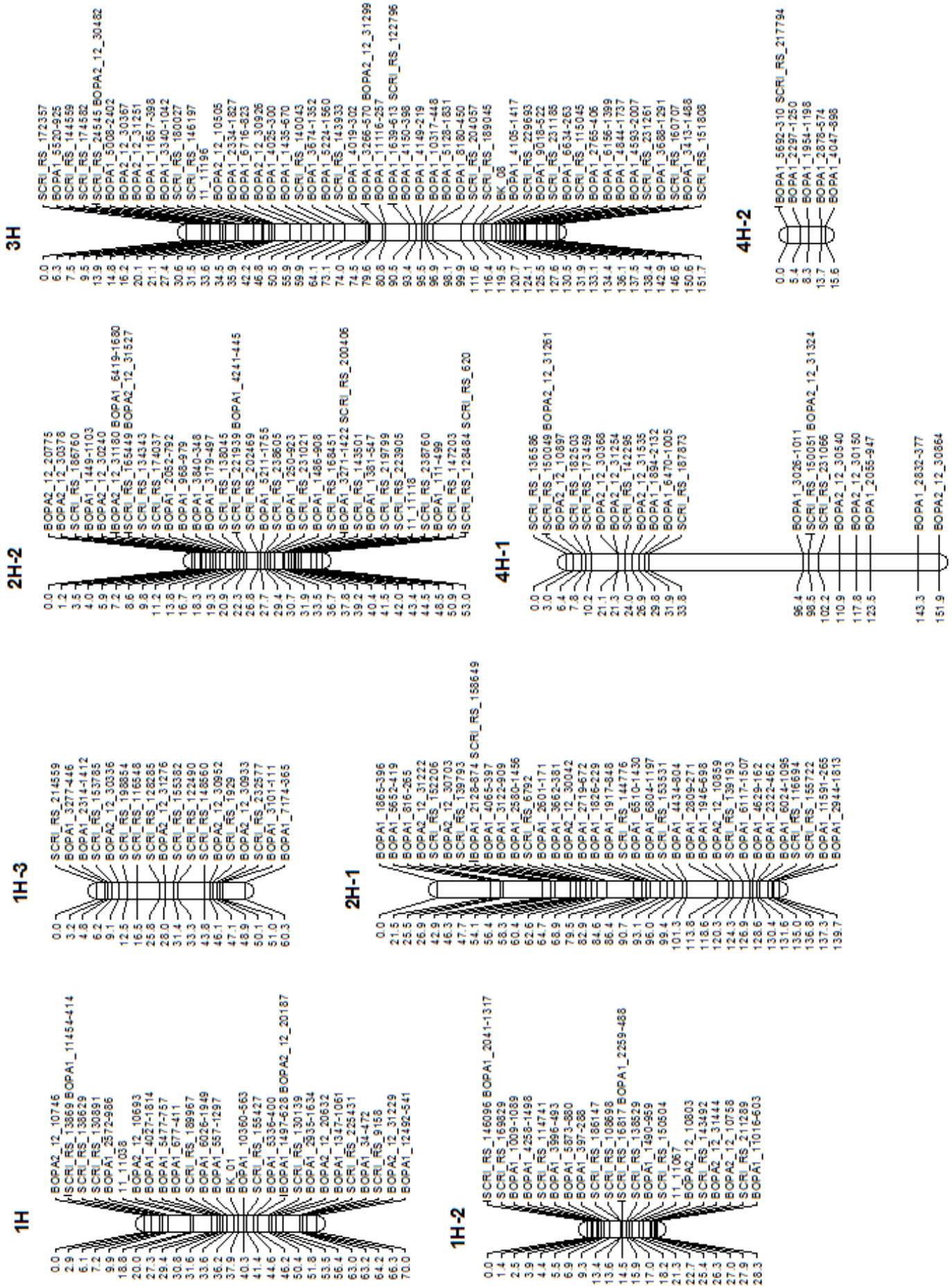


Figure S14. – Continued.

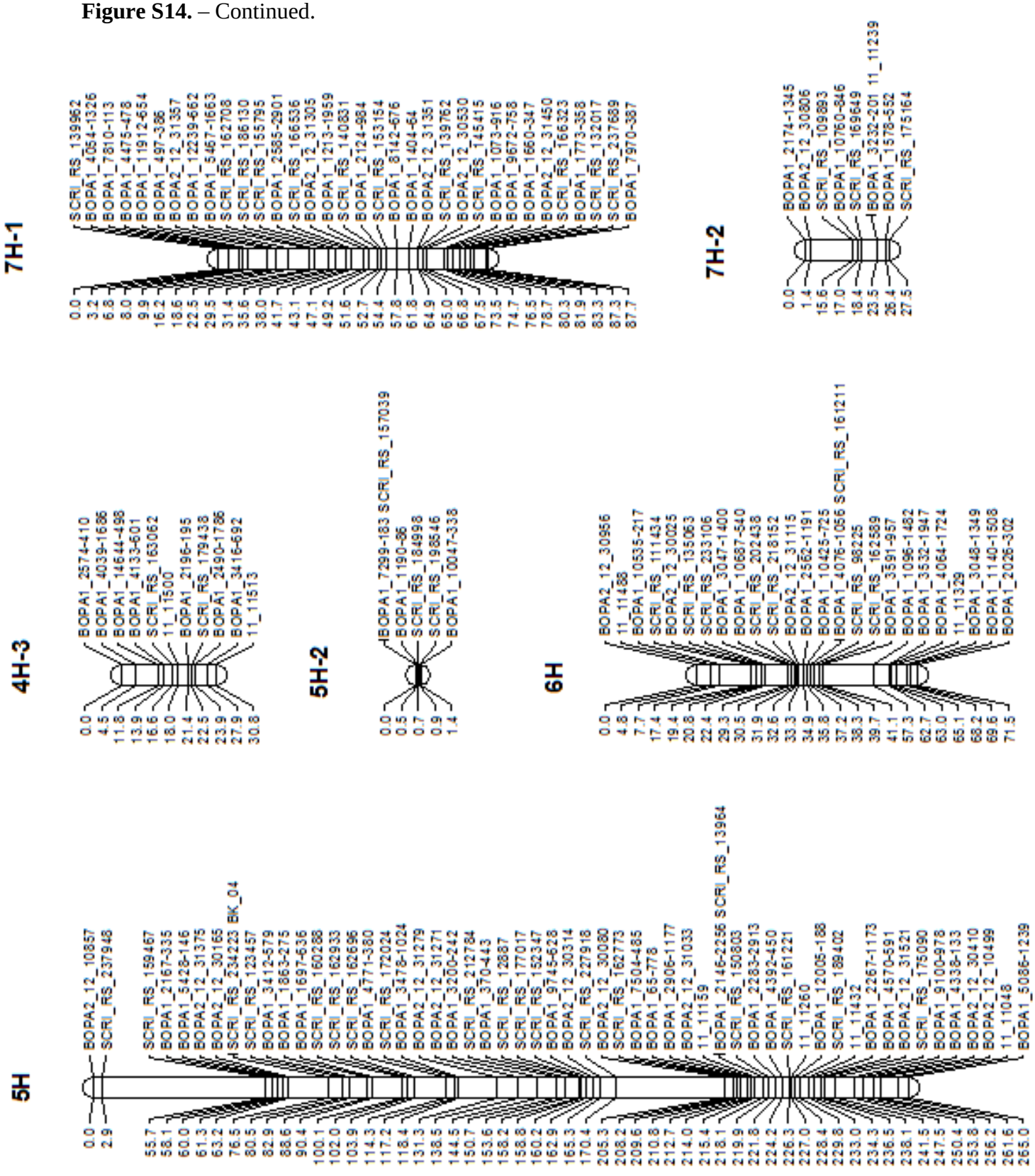


Figure S15. Position of QTL detected using QTL analysis according to the consensus map by Muñoz-Amatriaín et al. (2014), the genetic map by Comadran et al. (2012), the POPSEQ map (Mascher et al., 2013) and the IBSC_2012 map (IBSC, 2012).

Combined QTL	QTLs	Chr.	Consensus map range (cM)**		Comadran map range (cM)**		POPSEQ map range (cM)**		IBSC_2012 map range (cM)**	
QTL_1H.1	*QTL_SSD_1H-2.1*	1H	64.9	66.2	61.5	62.8	61.5	61.5	61.0	62.8
QTL_1H.2	*QTL_SSD_1H.1*	1H	142.7	142.7	132.5	132.5	131.9	131.9	128.1	132.5
	QTL_DH_1H.1	1H	142.7	142.7	131.9	132.7	-	-	132.4	132.7
QTL_2H.1	*QTL_DH_2H.1*	2H	-	-	-	-	-	-	11.6	11.6
QTL_2H.2	*QTL_DH_2H.2*	2H	31.9	40.2	19.1	23.8	25.3	25.3	18.9	23.6
QTL_2H.3	*QTL_DH_2H.3*	2H	66.1	69.6	58.8	58.8	53.8	59.4	54.3	58.8
	QTL_SSD_2H.1	2H	69.2	69.6	57.4	57.4	58.1	58.7	56.5	57.4
QTL_2H.4	*QTL_DH_2H.4*	2H	81.3	81.3	70.8	70.8	70.8	71.0	70.8	71.0
QTL_2H.5	*QTL_DH_2H.5*	2H	85.2	92.8	74.2	81.5	74.1	81.8	74.3	81.5
	QTL_SSD_2H.2	2H	88.0	98.4	79.9	86.0	82.1	86.8	76.7	86.0
QTL_2H.6	*QTL_DH_2H.6*	2H	100.9	100.9	88.5	88.5	88.5	88.5	88.5	88.5
QTL_3H.1	*QTL_DH_3H.4*	3H	13.5	14.2	8.3	8.8	8.4	10.6	8.3	11.5
QTL_3H.2	*QTL_SSD_3H.2*	3H	58.0	58.0	49.3	49.3	49.3	49.3	49.3	49.3
	QTL_DH_3H.3	3H	59.9	60.8	53.3	54.5	52.8	53.3	53.3	54.2
QTL_3H.3	*QTL_DH_3H.2*	3H	91.4	94.0	88.8	90.9	90.2	90.2	90.7	90.7
QTL_3H.4	*QTL_DH_3H.1*	3H	113.4	119.8	109.5	115.9	108.9	116.4	108.9	108.9
QTL_3H.5	*QTL_SSD_3H.1*	3H	128.5	128.5	126.1	126.7	126.1	126.7	126.1	126.7
QTL_4H.1	*QTL_DH_4H.1*	4H	104.2	117.1	111.3	111.3	103.8	111.1	100.6	111.3
QTL_5H.1	*QTL_SSD_5H.1*	5H	43.9	43.9	44.0	44.0	44.0	44.0	44.1	44.1
QTL_5H.2	*QTL_DH_5H-1.2*	5H-1	80.5	80.5	85.6	85.6	84.8	84.8	85.6	85.6
QTL_5H.3	*QTL_DH_5H-1.1*	5H-1	95.3	95.7	-	-	97.4	97.4	98.6	98.6
QTL_6H.1	*QTL_DH_6H.1*	6H	65.4	65.4	60.6	60.6	-	-	60.2	60.2
QTL_7H.1	*QTL_SSD_7H.2*	7H	43.1	43.1	37.6	37.6	29.0	37.6	32.8	37.6
QTL_7H.2	*QTL_SSD_7H.1*	7H	46.7	48.7	40.0	40.7	-	-	40.5	40.5
QTL_7H.3	*QTL_SSD_7H-2.1*	7H	126.5	126.5	-	-	119.8	119.8	118.3	120.0

* Range on the Golf x Tampar genetic map made using the DH or SSD dataset

** Position range of markers associated with the QTL for which a position was available.

Figure S16. Significant markers associated with each QTL found using QTL analysis.

Combined QTL	Significant markers
QTL_1H.1	BOPA1_7299-183, SCRI_RS_157039
QTL_1H.2	BOPA2_12_10746, SCRI_RS_13869, BOPA1_13095-187, BOPA1_4592-118
QTL_2H.1	SCRI_RS_136287
QTL_2H.2	BOPA1_1865-396, SCRI_RS_12516
QTL_2H.3	BOPA1_1292-406, BOPA2_12_30251, BOPA1_1826-229, BOPA1_2719-672, BOPA1_7489-442, BOPA1_2719-672, BOPA1_1826-229, BOPA1_1917-848
QTL_2H.4	BOPA1_1946-698, SCRI_RS_172667
QTL_2H.5	SCRI_RS_133539, BOPA1_7549-782, SCRI_RS_139193, SCRI_RS_198601, SCRI_RS_185505, BOPA1_2322-462, BOPA1_4629-162, BOPA1_2322-462, BOPA1_6024-1095, SCRI_RS_116694
QTL_2H.6	SCRI_RS_129223, BOPA1_11591-265
QTL_3H.1	SCRI_RS_108971, SCRI_RS_130264, BOPA1_7044-705
QTL_3H.2	SCRI_RS_204057, BOPA1_5038-1035, BOPA1_10126-999
QTL_3H.3	BOPA2_12_10344, BOPA1_3674-1352
QTL_3H.4	BOPA1_2334-1827, BOPA1_3965-353, BOPA2_12_30972
QTL_3H.5	BOPA1_5008-2402, BOPA2_12_30367
QTL_4H.1	BOPA1_4160-1365, BOPA1_10780-204
QTL_5H.1	BOPA1_12005-188
QTL_5H.2	SCRI_RS_140054
QTL_5H.3	BOPA1_3478-1024, BOPA1_9084-713
QTL_6H.1	BOPA1_8220-1223
QTL_7H.1	BOPA1_1213-1959, SCRI_RS_140831
QTL_7H.2	BOPA1_2585-2901, SCRI_RS_166536
QTL_7H.3	BOPA1_1894-2132, BOPA2_12_31535