Utilizing a wild barley nested association mapping (NAM) population for genome-wide association studies

Dissertation

zur Erlangung des Doktorgrades

der Agrarwissenschaften (Dr. agr.)

der

Naturwissenschaftlichen Fakultät III

Agrar- und Ernährungswissenschaften,

Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Herrn Andreas Maurer

Geb. am 04.04.1986 in Würzburg

1. Gutachter: Prof. Dr. Klaus Pillen
2. Gutachter: Prof. Dr. Rod Snowdon

Verteidigung am: 19. Juni 2017

Contents

General introduction 1

Barley domestication 1

Flowering time regulation in barley 2

DNA markers and QTL analysis 2

Association mapping 4

Multi-parental association mapping designs 5

Strategies to identify favorable exotic alleles 6

Objectives 9

Chapter 1) Modelling the genetic architecture of flowering time control in barley through nested association mapping[1] 10

Chapter 2) Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping[2] 23

Chapter 3) Estimating parent-specific QTL effects through cumulating linked identity-by-state SNP effects in multiparental populations[3] 35

General discussion 44

Suitability of HEB-25 for GWAS 44

Evaluation of allelic diversity and usefulness of wild alleles 45

Characteristics of HEB-25 and alternative multi-parental populations 49

Summary 52

Zusammenfassung 55

References 58

---

[1] Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics 16: 290.

[2] Maurer A, Draba V, Pillen K (2016) Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. J Exp Bot 67: 2507-2518.

[3] Maurer A, Sannemann W, Leon J, Pillen K (2016) Estimating parent-specific QTL effects through cumulating linked identity-by-state SNP effects in multiparental populations. Heredity. doi: 10.1038/hdy.2016.121

Abbreviations

AB	advanced backcross
AFLP	Amplified Fragment Length Polymorphism
BIL	backcross inbred line
CIM	composite interval mapping
DH	doubled haploid
DNA	deoxyribonucleic acid
FAO	Food and Agriculture Organization of the United Nations
GWAS	genome-wide association study
HEB	Halle exotic barley
HIF	heterogeneous inbred family
IL	introgression line
JLAM	joint linkage association mapping
LD	linkage disequilibrium
LSMEANS	least squares means
MAGIC	multi-parent advanced generation inter-cross
MAS	marker-assisted selection
MTA	marker trait association
NAM	nested association mapping
PCR	polymerase chain reaction
QTL	quantitative trait locus/loci
RFLP	Restriction Fragment Length Polymorphism
RIL	recombinant inbred line
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat

General introduction

Barley (*Hordeum vulgare* ssp. *vulgare*) is one of the most important cereals worldwide. With a total acreage of 49.4 million hectares and an average grain yield of 2.9 t/ha it has been the world's fourth-most grown cereal in 2014 (FAO 2016). Due to its high adaptability and stress tolerance (Munns and Tester 2008; Nevo et al. 2012) barley can be successfully grown in more extreme regions than any other cereal crop (Baik et al. 2011). Thus, barley represents a globally important crop, especially in harsh environments.

Most of the harvested barley grains serve as animal feed. Moreover, barley malt provides the basis for beverages like beer and whisky. In recent years barley retrieved importance also in human nutrition, since its valuable nutrient composition may support reducing the risk for several widespread diseases like type II diabetes and cardiovascular disease (Baik et al. 2011).

Barley domestication

The domestication of barley began at least 10,000 years ago in the Fertile Crescent (Zohary et al. 2012). However, barley domestication has been complex and difficult to retrace. In this regard several further domestication centers have been proposed (Negassa 1985; Molina-Cano et al. 1999; Badr et al. 2000; Morrell and Clegg 2007; Dai et al. 2012). Although there is a wide range of potential domestication centers there is no doubt that wild barley (*Hordeum vulgare* ssp. *spontaneum*) is the direct progenitor of the cultivated form. As a typical characteristic wild barley grains disperse from the spike at maturity to ensure effective grain spreading. The loss of this function due to two independent deletion events at *Btr1* or *Btr2* can be seen as the most important event during barley domestication since this enabled effective harvesting and thereby increasing yields (Pourkheirandish et al. 2015). Cultivated barley can be separated into two-rowed and six-rowed spike types. While in two-rowed barley only the central spikelet of a rachis node is fertile, in six-rowed barley also the two lateral spikelets are fertile, as a result of a mutated *VRS1* gene

(Komatsuda et al. 2007). The latter represented another crucial step in barley domestication by tripling the yield potential (Sakuma et al. 2011).

Flowering time regulation in barley

Besides the discrimination of row-type barley can furthermore be separated in winter and spring types, depending on the requirement of vernalization to initiate the reproductive phase. Ninety-eight percent of wild barleys represent winter barley (Cockram et al. 2011) that needs to be exposed to lower temperatures over a certain time to initiate flowering. In several genomic studies the genetic mechanisms underlying the development of different barley types could be revealed during the last decades. Spring barley lacks the vernalization requirement due to a natural deletion of the flowering repressor gene *Vrn-H2* (Yan et al. 2004; von Zitzewitz et al. 2005). In winter barley *Vrn-H1* (Yan et al. 2003) captures the cold signal (Oliver et al. 2013) by *cis*-regulatory elements in its promotor region (Alonso-Peral et al. 2011). The resulting up-regulation of *Vrn-H1* then causes a deactivation of *Vrn-H2*, which in turn promotes flowering through the floral integrator gene *Vrn-H3/HvFT1* (Yan et al. 2006), likely corresponding to the mobile signal 'florigen' (Chailakhyan 1937). Besides vernalization, also day length (photoperiod) determines flowering time in barley. In this regard *Ppd-H1* and *Ppd-H2* promote flowering under long days (Turner et al. 2005) and short days (Kikuchi et al. 2009), respectively. In addition to photoperiod and vernalization, light quality (Nishida et al. 2013; Pankin et al. 2014) circadian rhythms (Campoli et al. 2012; Faure et al. 2012; Zakhrabekova et al. 2012; Campoli et al. 2013; Calixto et al. 2015), phytohormones like gibberellic acid (GA) (Boden et al. 2014; Jia et al. 2015) and cytokinins (Mrízová et al. 2013) also contribute to the induction of flowering in barley. Most of these factors interact and the responsive genes can be integrated in a complex pathway. Unravelling this pathway can be supported by comparing with the model species *Arabidopsis* (Blümel et al. 2015).

DNA markers and QTL analysis

A first step towards being able to identify the above mentioned genes was the creation and screening of stable genetic mutants that were initially induced through

X-rays in the 1920s (Lundqvist and Franckowiak 2003). With the advent of molecular marker techniques in the early 1980s another crucial step to systematically attribute specific phenotypes to certain genotypic characteristics were made. The first DNA marker system was called RFLP (Restriction Fragment Length Polymorphism, Botstein et al. (1980)) and is able to detect polymorphisms that are based on different DNA fragment sizes resulting from restriction enzyme digestion. After the invention of the PCR (polymerase chain reaction) technique a multitude of other marker systems evolved. Simple Sequence Repeat markers (SSRs, Weber and May (1989)) and Amplified Fragment Length Polymorphism markers (AFLPs, Vos et al. (1995)) dominated the field in the 1990's until the highly abundant Single Nucleotide Polymorphism markers (SNPs, Landegren et al. (1998)) emerged. SNPs allowed the systematical characterization of thousands of markers without the need of gel-based assays, which was an important step towards high-throughput genotyping (Gupta et al. 2001). Nowadays, so-called SNP chips are available that enable genotyping of ten thousands of SNPs at the same time in a single assay.

Several methods have been developed to detect marker trait associations (MTAs) and to define quantitative trait loci (QTL) in the genome. QTL are genomic regions that concertedly affect a quantitative trait. According to Bernardo (2010) conducting QTL studies to find genomic regions that influence the desired trait basically comprises the following steps: 1) creating a segregating population; 2) genotyping the population with molecular markers; 3) phenotyping the population for traits of interest; and 4) applying statistical procedures to find markers linked to QTL.

Classical segregating populations are created by crossing two parents that differ for the trait of interest and subsequently selfing the F_1. The resulting F_2 population segregates at each marker locus in an expected ratio of 1 : 2 : 1 (homozygous parent A : heterozygous : homozygous parent B). It can be directly used for QTL studies. However, also enhanced populations are well-established for QTL mapping. Doubled haploids (DHs) are created by *in vitro* regeneration of plants out of F_1 anthers/microspores and have the advantage that they are homozygous throughout the genome and, thus, represent a stable population for years of trials. In contrast, a recombinant inbred line (RIL) population, resulting from subsequent

selfing of F_2 plants, still segregates at certain loci, depending on the selfing generation. However, this harbors great potential for further fine-mapping approaches of QTL via utilization of heterogeneous inbred families (HIFs) (Tuinstra et al. 1997). HIFs are created through selfing a rare RIL that is heterozygous for the marker of interest. As a result a typical segregation of 1 : 2 : 1 at this locus is obtained, while the remaining genome is almost fixed and therefore not disturbing the analysis. In a similar way, also backcross inbred lines (BILs) (Lin et al. 1998) can be utilized. These result from repeated selfing of a backcross (BC_1) population, which is a product of crossing F_1 plants back to one of the parents.

The basic concept of being able to identify a genomic region that affects a certain trait is that molecular markers are genetically linked to the causative gene. They are in linkage disequilibrium (LD) and inherited together. Initially, MTAs were defined by simple approaches in single-factor models. Here, a simple mean difference between two marker alleles was used to characterize differences with a *t*-test. However, if the number of available markers is low it might be that the detection of a relevant MTA is impeded. Lander and Botstein (1989) described interval mapping as a method that was able to map the position of a QTL between two markers. This method was enhanced by Zeng (1994), who combined interval mapping with multiple regression by fitting other molecular markers outside the interval in the model. This so-called composite interval mapping (CIM) method enabled higher precision of QTL mapping by accounting for other QTL in the background.

Association mapping

Due to the complex and laborious process of developing a suitable mapping population for QTL mapping and the emergence of affordable large SNP chips, covering the whole genome with high density, a new approach of finding MTAs evolved. The so-called (genome-wide) association mapping does not require an artificially designed mapping population. Instead, it uses the diversity that is present in nature, for instance by selecting a collection of highly diverse accessions of the desired species in a world-wide scale. This way it makes use of historical recombination events that occurred during evolution (Soto-Cerda and Cloutier 2012) enabling a higher potential to find allelic diversity in the population and

increasing mapping resolution. These advantages, however, are accompanied by lower marker efficiency and reduced statistical power compared to QTL mapping (Flint-Garcia et al. 2003). Another major constraint of association mapping is that undetected 'cryptic' population structure may lead to spurious, false-positive, associations (Pritchard et al. 2000). Yu et al. (2006) developed a method that is able to account for this problem by integrating measures of relatedness among individuals in a linear mixed model. Nevertheless, doubts remain about the method, since genetic background (i. e. other causative loci) is a more critical factor than population structure (Vilhjálmsson and Nordborg 2013). Therefore, Segura et al. (2012) developed a model integrating genetic background and relatedness in a linear mixed model when testing a marker's association with a specific trait.

Multi-parental association mapping designs

In parallel to the development of sophisticated models for association mapping, also enhanced mapping populations have been developed as a way to overcome the known problems. One example is the so-called nested association mapping (NAM) design that was first developed in maize by Yu et al. (2008). In NAM, a diverse set of founder lines is independently crossed with a reference parent. Out of these crosses RIL families are generated through several rounds of selfing. It represents a mixture of classical linkage mapping and association mapping and combines advantages of both approaches (Yu et al. 2008). Due to the mating design, both historical LD (within short segments of chromosomes) and LD resulting from recombination in the designed population can be exploited (Guo et al. 2013). Due to recombination also population structure is eliminated, since spurious long range LD on a chromosome is diminished (Guo et al. 2010). The first NAM population in maize (US-NAM) has successfully been applied to dissect, for instance, the genetic architecture of flowering time (Buckler et al. 2009), leaf architecture (Tian et al. 2011) and plant height (Peiffer et al. 2014) with high precision. Besides US-NAM, further NAM populations in maize (CN-NAM, Li et al. (2013)), sorghum (Jordan et al. 2011) and wheat (Bajgain et al. 2016; Li et al. 2016b) have been developed. In barley, currently two different forms of NAM populations are available (Maurer et al. 2015; Nice et al. 2016), which are described in more detail later.

Another multi-parental concept to combine allele richness, high resolution and statistical power to detect MTAs is called multi-parent advanced generation inter-cross (MAGIC, Cavanagh et al. (2008)). It also consists of several diverse founders that are intermated. In the basic concept eight founders are combined over three generations of single crosses, double crosses and a final quadruple cross so that ultimately all resulting individuals harbor different mixes of genome segments from all eight founders. In contrast to a NAM population no family sub-structure is available in MAGIC, which impedes an exact tracing back of founder alleles in the final MAGIC lines. In barley, Sannemann et al. (2015) developed a MAGIC population out of one elite barley cultivar (Barke) and seven old landraces that are considered to be the founders of German barley breeding. Further MAGIC populations are available for the cereal crops maize (Dell'Acqua et al. 2015), wheat (Huang et al. 2012; Mackay et al. 2014; Milner et al. 2015; Thépot et al. 2015) and rice (Bandillo et al. 2013; Meng et al. 2016), as well as faba bean (Sallam and Martsch 2015), tomato (Pascual et al. 2015) and cotton (Li et al. 2016a).

Liller et al. (2017) developed a barley population that represents a mixture of NAM and MAGIC. Here, four different founders have been crossed with the recurrent parent Morex and subsequently F_1s were backcrossed once with Morex. BC_1 generations were then intercrossed with each other at every possible combination, resulting in six independent subpopulations.

Strategies to identify favorable exotic alleles

One of the most important challenges of future agricultural research is to find ways to achieve higher yields on a reduced area of available arable land in times of climate change and global population growth. Plant breeding is supposed to play a major role to meet this target. Domestication and breeding efforts in the past centuries led to a severe depletion of the allele richness of important crop species, known as the bottleneck effect (Tanksley and McCouch 1997). A recent study revealed a ~50% reduction of genetic diversity in domesticated compared to wild barley (Pankin et al. 2016). It is obvious that this causes a reduced adaptability to changing climates. Stagnating yields of barley, especially in Southern Europe, might be a direct consequence of this allele depletion (Dawson et al. 2015). Therefore, there is need

to replenish the elite breeding material with new exotic alleles to cope with upcoming agronomic challenges (Ellis et al. 2000; Zamir 2001; Gur and Zamir 2004; Zamir 2008; Ma et al. 2012; McCouch et al. 2013; Zhang et al. 2017).

Since exotic material harbors a multitude of undesired negative effects, direct crosses with elite material are no option for a fast development of superior breeding lines. Even if a favorable exotic allele can be crossed into a well-adapted background there are plenty of linked unfavorable alleles introgressed at the same time (so-called linkage drag). For this purpose advanced backcross (AB)-QTL analysis has been established (Tanksley and Nelson 1996). Here, several rounds of backcrossing with the elite parent are performed, while at the same time negative selection reduces the frequency of disadvantageous donor alleles. After identifying favorable QTL linkage maps can assist to select lines for breeding that contain recombinant chromosomes, which break linkage drag (Tanksley and Nelson 1996). AB-QTL analysis has successfully been applied to barley (Pillen et al. 2003; Pillen et al. 2004; von Korff et al. 2005; von Korff et al. 2006; von Korff et al. 2008; Saal et al. 2011).

Introgression lines (ILs) represent another concept similar to AB-QTL. Here the aim is that the whole donor genome is represented in a so-called IL library. This library consists of a multitude of lines each carrying only a single segment of the respective donor (Zamir 2001). This is achieved by recurrent backcrossing with the elite parent and simultaneous marker-assisted selection (MAS). In a relatively simple statistical approach a specific phenotype can then be attributed to a single defined chromosomal region, which is for instance achieved by a Dunnett multiple comparison test (Dunnett 1955). Here, least squares means (LSMEANS) of each IL are compared with the recurrent parent as a control. In case of significance the difference can then be explained by the introgression, since the remaining genome is shared across the IL and the elite parent control. This also increases the ability to statistically identify small phenotypic effects, since epistasis is largely removed (Zamir 2001). The IL concept has been established in tomato by Eshed et al. (1992) and has been applied to a multitude of plant species. In barley Schmalenbach et al. (2008) developed a set of ILs by crossing the Israeli wild barley accession ISR42-8 and the German elite cultivar Scarlett. This so-called S42IL population has been subjected to a multitude of trials and several QTL for agronomic traits could be defined (Schmalenbach et al. 2008; Schmalenbach et al. 2009; Schmalenbach and

Pillen 2009; Wang et al. 2010; Schmalenbach et al. 2011; Hoffmann et al. 2012; Schnaithmann and Pillen 2013; Honsdorf et al. 2014a; Honsdorf et al. 2014b; Naz et al. 2014; Reuscher et al. 2016). ILs carrying an advantageous QTL allele can be used directly for breeding, since they already harbor a magnitude of desirable elite genome and therefore also reduce potential sterility problems (Zamir 2001).

In principle, also NAM populations can serve as a useful tool to evaluate exotic germplasm. Therefore a multitude of exotic accessions can be crossed with the recurrent elite parent. However, disadvantageous exotic alleles might complicate mechanical processing (in barley mainly caused by brittleness and bad threshability) and impede direct usability for plant breeders. Therefore, to get rid of these problems backcrossing with the recurrent parent or selection against unfavorable phenotypes is performed. In the classical NAM concept of Yu et al. (2008) no backcross step was included, leading to an exotic allele frequency of 50% in each RIL. In barley, there are currently only backcross-based NAM populations available. Nice et al. (2016) for instance developed a so-called Advanced Backcross (AB)-NAM population, where each F_1 plant has been backcrossed twice with the recurrent elite barley cultivar Rasmusson. This leads to an exotic allele frequency of 12.5% for each BIL.

The present thesis is based on another backcross-derived NAM population. The so-called Halle Exotic Barley 25 (HEB-25) population is the result of initial crosses of 25 highly divergent wild barley accessions (hereafter called donors) with the recurrent elite barley cultivar Barke. The donors were selected to maximize the genetic diversity in HEB-25 and originate from Afghanistan, Iran, Iraq, Israel, Lebanon, Syria and Turkey (*Hordeum vulgare* ssp. *spontaneum*) as well as Tibet, China, (*Hordeum vulgare* ssp. *agriocrithon* (Åberg)). F_1 plants were backcrossed once with Barke, resulting in an exotic allele frequency of 25%. After three subsequent selfing steps the final population consists of 1,420 BC_1S_3 lines, subdivided into families of 22-75 individual BILs.

Objectives

In the present thesis the NAM population HEB-25 was utilized for the first time to detect QTL for important agronomic traits via genome-wide association studies (GWAS). This way the potential of exotic alleles for future barley breeding should be defined. Since there has been no preliminary experience about how to handle a backcross-based NAM population this can be seen as an iterative process where different models for GWAS were applied. The work was split in the following sections:

1. Establishing a model to reliably detect QTL for flowering time.
   This trait is well-studied and many genes are already known so that it is perfectly suited as a proof of concept and to evaluate the power of GWAS in HEB-25. Furthermore, since flowering time is of major importance with regard to climate change, the question should be answered whether there is useful variation among wild barley to improve future breeding progress (Chapter 1).

2. Dissecting the genetic control of flowering time in more detail to fine-tune plant development with regard to yield components.
   Since not only flowering time itself but also other developmental stages like maturity have a major impact on final grain yield the question was whether different developmental stages are under independent genetic control. This should be achieved by comparing GWAS results of several developmental traits and showing their impact on the yield component thousand grain weight. Finally, conclusions should be drawn about an optimal strategy for timing of developmental phases to increase thousand grain weight (Chapter 2).

3. Comparing methods to estimate donor-specific QTL allele effects.
   Since HEB-25 is a multi-parental population, theoretically up to 26 different alleles could be present at one QTL. The arising question was whether this issue can be addressed in GWAS without losing statistical power (Chapter 3).

Chapter 1

Modelling the genetic architecture of flowering time control in barley through nested association mapping[1]

The aim of the presented study was to establish a model to reliably detect QTL for flowering time in HEB-25.

Flowering time is a well-studied trait and many genes are already known so that it is perfectly suited as a proof of concept and to evaluate the power of GWAS in HEB-25. In addition, a direct comparison with the genetic architecture of the allogamous species maize was possible, which has been investigated in the US-NAM population (Buckler et al. 2009). Furthermore, since flowering time is of major importance with regard to climate change, the question should be answered whether there is useful variation among wild barley to improve future breeding.

[1] Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics 16: 290.

Maurer *et al. BMC Genomics* (2015) 16:290
DOI 10.1186/s12864-015-1459-7

BMC Genomics

RESEARCH ARTICLE Open Access

Modelling the genetic architecture of flowering time control in barley through nested association mapping

Andreas Maurer[1†], Vera Draba[1,2†], Yong Jiang[3], Florian Schnaithmann[1], Rajiv Sharma[3,4], Erika Schumann[1], Benjamin Kilian[3,5], Jochen Christoph Reif[3] and Klaus Pillen[1*]

Abstract

Background: Barley, globally the fourth most important cereal, provides food and beverages for humans and feed for animal husbandry. Maximizing grain yield under varying climate conditions largely depends on the optimal timing of flowering. Therefore, regulation of flowering time is of extraordinary importance to meet future food and feed demands. We developed the first barley nested association mapping (NAM) population, HEB-25, by crossing 25 wild barleys with one elite barley cultivar, and used it to dissect the genetic architecture of flowering time.

Results: Upon cultivation of 1,420 lines in multi-field trials and applying a genome-wide association study, eight major quantitative trait loci (QTL) were identified as main determinants to control flowering time in barley. These QTL accounted for 64% of the cross-validated proportion of explained genotypic variance (p_G). The strongest single QTL effect corresponded to the known photoperiod response gene *Ppd-H1*. After sequencing the causative part of *Ppd-H1*, we differentiated twelve haplotypes in HEB-25, whereof the strongest exotic haplotype accelerated flowering time by 11 days compared to the elite barley haplotype. Applying a whole genome prediction model including main effects and epistatic interactions allowed predicting flowering time with an unmatched accuracy of 77% of cross-validated p_G.

Conclusions: The elaborated causal models represent a fundamental step to explain flowering time in barley. In addition, our study confirms that the exotic biodiversity present in HEB-25 is a valuable toolbox to dissect the genetic architecture of important agronomic traits and to replenish the elite barley breeding pool with favorable, trait-improving exotic alleles.

Keywords: Barley, Wild barley, Nested association mapping (NAM), Flowering time, Genome-wide association study (GWAS), Quantitative trait locus (QTL), Genomic prediction, Epistasis, Haplotypes

Background

Barley is among the oldest crop species human civilization was built on. Approximately 10,500 years ago, barley was domesticated in the Fertile Crescent [1,2], presumably followed by additional independent domestication events in East Asia [3,4]. Domestication and subsequent genetic selection led to gene erosion in most crop species' gene pools [5,6], threatening future breeding advances. Utilizing the untapped biodiversity, present in wild progenitors is one promising approach to replenish the elite breeding pool with new favorable alleles [6-13]. The enriched diversity may be pivotal to boost the rate of genetic improvement and to cope with the anticipated enhanced effects of biotic and abiotic stresses due to climate change.

In this regard, time of flowering is expected to play a major role in future crop improvement. It is a key trait for the successful completion of a plant's life cycle and, therefore, it has a strong impact on grain yield [14]. Flowering time indicates the transition from vegetative to reproductive stage, which is mainly influenced by environmental cues like day length (photoperiod) and prolonged exposure to cold temperatures (vernalization). In

* Correspondence: klaus.pillen@landw.uni-halle.de
†Equal contributors
[1]Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle Wittenberg, Betty-Heimann-Str. 3, 06120 Halle, Germany
Full list of author information is available at the end of the article

BioMed Central

© 2015 Maurer et al.; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

barley, the day length determining light signal is transmitted from a circadian clock oscillator, with *Ppd-H1,* a *PSEUDO-RESPONSE REGULATOR 7 (PRR7)* gene, in its center [15]. Under long day condition, *Ppd-H1,* through mediation of *CONSTANS (CO),* promotes the expression of the floral inducer *Vrn-H3,* a homolog of the *Arabidopsis thaliana FLOWERING LOCUS T (FT)* gene [16]. On the other hand, *Vrn-H2,* a zinc-finger *CONSTANS*, *CO*-like and *TOC1 (CCT)*-domain protein (*ZCCT1*) acts as a repressor of *Vrn-H3* [17]. *Vrn-H2*, in turn, is repressed by *Vrn-H1,* an *APETALA1* family *MADS*-box transcription factor [18], which is up-regulated during vernalization. Thus, after vernalization, the repression of *Vrn-H3* is abolished and flowering is induced. Based on its vernalization requirement, winter barley and spring barley can be distinguished. Spring barley lacks the vernalization requirement due to a deletion of *Vrn-H2* [19].

Besides photoperiod and vernalization, there are also genetic mechanisms acting independently of environmental cues, so-called earliness *per se* [20]. Although several key regulatory cereal genes of flowering time were characterized and finally cloned during the last two decades, still little is known about the genetic architecture underlying flowering time regulation in temperate cereals, as compared to the model species *A. thaliana* [14,21-23]. So far, a holistic explanation of flowering time in a segregating germplasm population and the accurate prediction of a plant's time of flowering, based on the combined action and interaction of major genes, is still not fully achieved in cereal species. Furthermore, it is reported that wild barley accessions possess a rich reservoir of beneficial alleles controlling flowering time [7,24,25].

Nested association mapping (NAM) emerged as a multi-parental mapping design to investigate genomic regions with unprecedented genetic resolution and allelic variation by combining the advantages of linkage analysis and association mapping [26]. Hence, it facilitates the elucidation of a trait's genetic architecture via genome-wide association study (GWAS). Until now, the NAM design was applied to the allogamous species maize and sorghum [26,27]. NAM populations for autogamous species like barley or wheat have not been developed, yet. In maize, the genetic dissection of various agronomic traits, including flowering time, has been investigated [28-34]. However, it was not possible to completely dissect the genetic architecture of flowering time in maize due to its complex inheritance and the multitude of involved small effect QTL. We developed the first NAM population in the autogamous species barley, termed 'Halle Exotic Barley 25' (HEB-25). The population results from initial crosses between the spring barley elite cultivar Barke (*Hordeum vulgare* ssp. *vulgare, Hv*) and 25 highly divergent exotic barley accessions, contributing an ideal instrument to study biodiversity. The exotic donors comprise 24 wild barley accessions of *H. vulgare* ssp. *spontaneum (Hsp),* the progenitor of domesticated barley, and one Tibetan *H. vulgare* ssp. *agriocrithon (Hag)* accession. Barke was selected since it was also used as a parent of a barley high-resolution mapping population [35] and as a genetic stock for mutation screening [36]. The exotic donors were selected from Badr *et al.* [37] to represent a substantial part of the genetic diversity that is present across the Fertile Crescent, where barley domestication occurred. To generate the nested population, F_1 plants were backcrossed to Barke and, subsequently, selfed three times (Figure 1). In

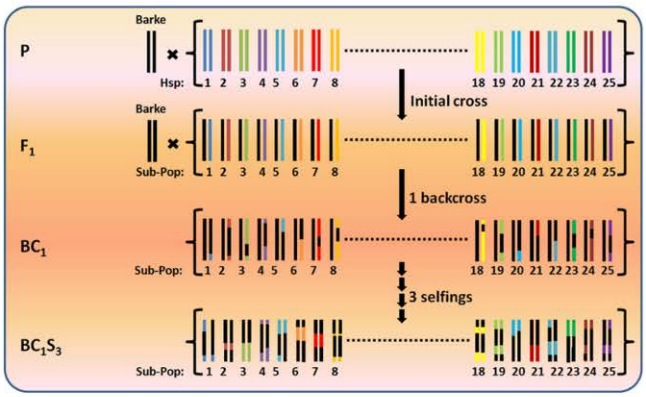

Figure 1 Development of the nested association mapping population HEB-25. HEB-25 is made of 25 families with 1,420 NAM lines in BC_1S_3. Per NAM line, one chromosome pair is illustrated as a double bar. Black and colored bars represent chromosome segments originating from Barke and the exotic donor accessions, respectively. At each SNP locus, HEB-25 is expected to segregate into 71.875% homozygous Barke, 6.25% heterozygous and 21.875% homozygous donor genotypes.

total, HEB-25 consists of 1,420 BC_1S_3 lines, divided into 25 HEB families of up to 75 lines per family (Additional file 1).

In the present study we investigated the genetic architecture of flowering time in barley. For this purpose, the NAM population HEB-25 was grown from 2011 to 2013 in multi-field trials to gather data on flowering time. By combining these data with high-density SNP marker information via genome-wide association studies and genomic prediction models, we could show that flowering time in barley mainly depends on a low number of large-effect QTL and epistatic interactions.

Results and discussion

Characterization of HEB-25

The inheritance of parental segments across the genomes of the 1,420 HEB lines was characterized through genotyping 5,709 informative, genic single nucleotide polymorphism (SNP) markers [35]. Marker saturation was high with an average genetic distance of 0.17 cM and a maximum gap of 11.1 cM between adjacent markers. Linkage disequilibrium (LD) among the 26 parents decayed rapidly (Additional file 2) enabling a high mapping resolution [26]. The SNP data revealed a low degree of genetic similarity between Barke and the donors, ranging from 40 to 54% (Additional file 1). Parents and the HEB-25 population could be clearly separated in a principal component analysis (PCA) (Additional file 3). Also, HEB families could be ordered in the PCA based on their geographical origin. These findings point to the high genetic diversity that is present among HEB-25 and its parents.

Diversity in HEB-25 was also visible phenotypically. During the seasons 2011 through 2013, HEB-25 was cultivated at the Halle University Experimental Field Station to collect flowering time data. The HEB lines flowered on average 68.1 days after sowing with a range from 51.0 to 98.9 days and a standard deviation of 6.5 days (Additional files 4 and 5). The broad variation in flowering time, covering almost 50 days among the 1,420 HEB lines, and a high heritability of 91.6%, as well as the genetic properties of the NAM population provided an excellent starting point to study the genetic architecture of flowering time through GWAS.

Genome-wide association study

For GWAS, we initially applied the multiple linear regression Model-B with step-wise selection of cofactors, as outlined in Liu *et al.* [38]. Model-B was found most suitable to study traits across multiple related families [39], where a family effect and additional SNPs, selected as cofactors, are included in the model. GWAS revealed eight highly significant major QTL regions controlling flowering time with $P_{BON\text{-}HOLM} < 1.0$ E-10 (Table 1, Figure 2, Additional files 6 and 7). Testing the combined explanatory power of the single peak markers of the eight major QTL revealed a cross-validated explained proportion of genotypic variance (p_G, [40]) of 64% (Figure 3). To check if genetic relatedness, as reported elsewhere [41], affects this parameter in HEB-25, we also estimated p_G for different sets of eight randomly chosen SNPs, excluding regions with significant QTL. However, since the cross-validated explained p_G remained low with an average of 8%, we conclude that genetic relatedness between individual lines does not play a major role in HEB-25. This emphasizes the power and precision of QTL detection in HEB-25, which may be a combined effect of the low extent of LD and the particular mating design, resulting in an elevated rate of chromosomal recombination. Thus, flowering time of barley can be reliably predicted based on

Table 1 List of eight major QTL controlling flowering time in HEB-25

QTL	Chr[a]	Pos[b]	Range[b]	Peak marker[c]	No. Seg. Fam.[d]	$P^{e}_{BON\text{-}HOLM}$	p^{f}_{G}	CV Freq.[g]	Effect[h]	CG[i]
QFt.HEB25-1b	1H	128.3	128.0-128.3	SCRI_RS_150786	25	2.41E-18	0.01	68	−1.4	*HvELF3* [46,47]
QFt.HEB25-2b	2H	23.0	16.8-23.8	BK_16	24	3.39E-130	0.36	100	−9.5	*Ppd-H1* [15]
QFt.HEB25-2c	2H	57.4	56.4-58.1	BOPA2_12_30265	25	2.25E-42	0.05	84	−3.0	*HvCEN* [35]
QFt.HEB25-3c	3H	108.4	107.8-109.2	BOPA1_ABC07496_ pHv1343_02	23	2.62E-62	0.04	83	−3.1	*denso* [45]
QFt.HEB25-4a	4H	3.5	3.5	BOPA2_12_31458	24	5.08E-15	0.05	82	3.2	
QFt.HEB25-4e	4H	113.4	113.4-114.3	SCRI_RS_216897	24	4.58E-17	0.02	100	2.2	*Vrn-H2* [17]
QFt.HEB25-5d	5H	125.5	125.5-125.8	BOPA1_4795_782	24	2.31E-33	0.06	60	3.8	*Vrn-H1* [18]
QFt.HEB25-7a	7H	34.3	25.9-34.3	BOPA2_12_30895	23	6.04E-69	0.07	100	4.1	*Vrn-H3* [16]

[a]Barley chromosome on which the QTL was determined.
[b]Genetic position of the peak marker and range of the QTL in cM, based on Comadran *et al.* [35].
[c]Marker of the QTL with the highest significance (peak marker).
[d]Number of families, in which peak marker is segregating.
[e]Significance of the peak marker, expressed as $P_{BON\text{-}HOLM}$.
[f]Cross-validated proportion of explained genotypic variance of peak marker.
[g]Frequency of significant detections of the peak marker in 100 five-fold cross-validation runs.
[h]Difference between the wild genotype and the cultivated genotype in days until flowering. Early flowering effects of exotic alleles are indicated in red.
[i]Candidate gene, potentially explaining the QTL effect with reference.

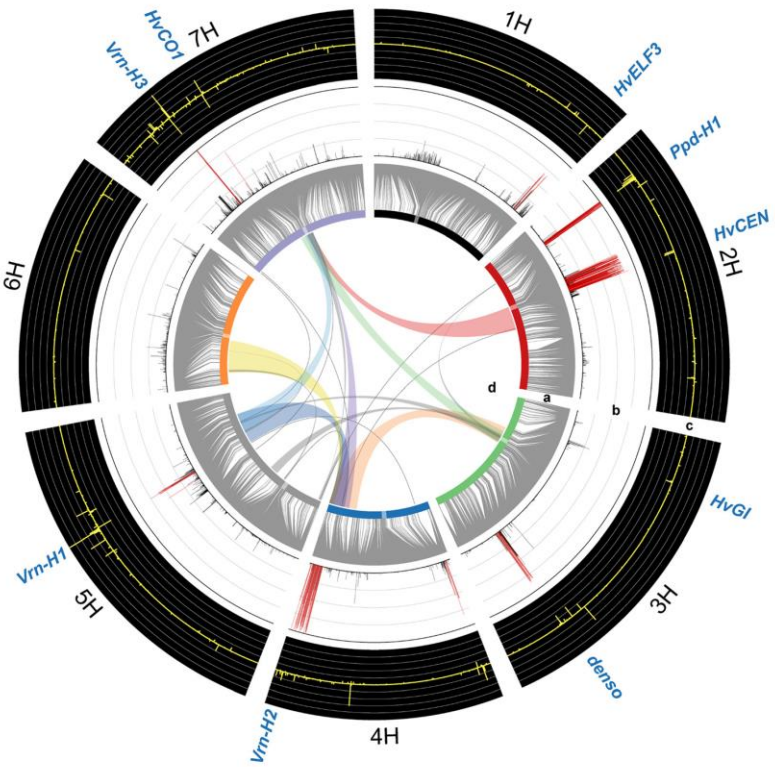

Figure 2 Genetic architecture of flowering time in HEB-25. Barley chromosomes are indicated as colored bars on the inner circle, centromeres are highlighted as transparent boxes. **a)** Grey connector lines represent the genetic position of SNPs on the chromosomes. **b)** Frequency of QTL detection in 100 cross-validation runs via GWAS (0 to 100, grid line spacing: 25); markers with > 50 detections are colored in red. **c)** Additive effect of the SNP obtained from the BayesCπ genomic prediction model. **d)** Links in the center of the circle represent significant ($P_{BON\text{-}HOLM} < 0.05$) di-genic interactions between SNP markers via GWAS. Clusters of significant SNP interactions are indicated by different colors. Position of candidate genes, potentially explaining major effects and epistatic effects, correspond to Table 1 and are indicated in blue outside the circle.

eight major QTL. This finding is in contrast to flowering time regulation in the allogamous species maize and sorghum, where only small effect minor QTL were detected [28,42]. The most significant association in HEB-25 ($P_{BON\text{-}HOLM}$ = 3.4 E-130) was observed on the short arm of chromosome 2H and explained a p_G of 36%. This SNP is directly located within *Ppd-H1*, the major determinant of photoperiod response in barley under long day condition [15]. Seven further genomic regions of extraordinary high significance were detected on chromosome arms 1HL, 2HS, 3HL, 4HS, 4HL, 5HL, and 7HS. All except one QTL (4HS) could be assigned to known flowering time genes (Table 1, Figure 2 and Additional file 7). Besides *Ppd-H1*, also the vernalization genes *Vrn-H1* and *Vrn-H2*, as well as the floral inducer *Vrn-H3* and its putative paralog *HvCEN* [43] exhibited highly significant effects. In addition, we could confirm the importance of gibberellic acid (GA) in flowering time regulation [44] through detection of *denso* [45] and *HvELF3* [46,47] as two further major QTL. Both genes are shown to be involved in GA biosynthesis [45,48]. So far, only the QTL on 4HS could not been referenced. This QTL, thus, remains a subject for further genetic investigations.

The eight major QTL were located with high genetic precision, with four QTL restricted to confidence intervals of less than 0.9 cM (Table 1). In cases where gene-specific SNPs were available (i.e. *Ppd-H1* and *Vrn-H3*), exactly those SNPs revealed the highest significance within the respective QTL window (Additional file 6). The exotic alleles at *Ppd-H1* and *Vrn-H3* revealed the strongest effects, accelerating flowering time by 9.5 days and delaying flowering time by 4.1 days, respectively. The drastic effects of single QTL outline the high potential of introducing wild barley alleles from HEB-25 in

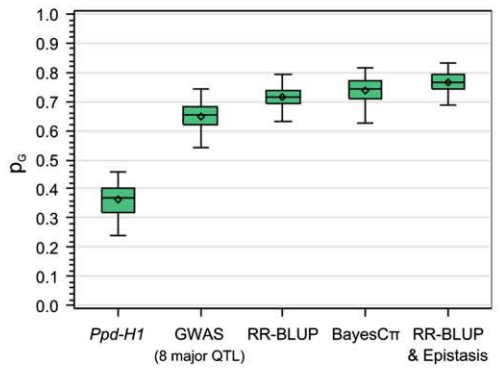

Figure 3 Cross-validated proportion of explained genotypic variance (p_G) of different applied models. The box-whisker plots depict the variation of explained genotypic variance after 100 cross-validations. The tested QTL models are **(i)** the single SNP locus *Ppd-H1* (Mean $p_G = 0.36$), **(ii)** GWAS with peak markers, representing the eight major QTL indicated in Table 1 (Mean $p_G = 0.64$), **(iii)** the whole genome ridge regression best linear unbiased prediction (RR-BLUP, mean $p_G = 0.71$), **(iv)** the BayesCπ prediction (Mean $p_G = 0.74$), and **(v)** RR-BLUP including epistasis (Mean $p_G = 0.77$).

order to adapt flowering time to environmental requirements and to enhance biodiversity in the elite barley breeding pool.

Ppd-H1 haplotype study

As we used bi-allelic SNP markers, additive effects were estimated across the NAM population. Theoretically, there may be up to 26 different alleles present at each locus in HEB-25. Thus, distinct alleles that show contrasting effects between families potentially escaped detection in our SNP-based GWAS. Contrasting effects are illustrated in Figure 4 and, in detail, in Additional file 6. For instance, SNPs at position 46.2 cM on chromosome 3H, which are tightly linked to *HvGI* [49], revealed opposing effects across HEB families. We tested the potential to integrate SNP haplotypes in the GWAS model for *Ppd-H1,* which exhibited the largest p_G. After re-sequencing the last two exons and three introns of *Ppd-H1,* twelve haplotypes could be distinguished (Additional file 1). All *Hsp* donor haplotypes at *Ppd-H1* showed a significantly reduced flowering time (Additional file 8 and Figure 5), where a maximum reduction of flowering time was associated with H-6 (−11.1 days compared to elite barley haplotype H-2). Only the *Hag* haplotype H-45 did not differ from H-2. This finding confirms the presence of haplotype-specific effects present in HEB-25. Consequently, we expect the existence of further haplotype effects for other candidate genes controlling flowering time. The haplotype-based *Ppd-H1* model resulted in a slight increase of the cross-validated explained p_G from 36% to 38%. This finding implies that modelling haplotype-specific effects for a substantial portion of the barley gene space may result in an improved prediction of flowering time in HEB-25. However, a genome-wide re-sequencing of HEB-25 lines will be required to identify and distinguish those haplotypes.

Applying genomic prediction models

To check whether we could further elucidate the genetic architecture of barley flowering time we applied genomic prediction models that considered all markers simultaneously. Genomic prediction evolved in animal breeding as a tool to predict a phenotype based on modelling a large set of SNP data [50]. It is used for selection of improved genotypes based on estimated genomic breeding values. Applying RR-BLUP [51] and BayesCπ [52] models, we could further increase the cross-validated explained p_G to 71% and 74%, respectively (Figure 3). These findings are in agreement with comparisons of multiple linear regression and genomic prediction of traits in bi-parental plant populations [53]. However, our p_G values substantially exceed the prediction accuracies of genomic prediction models reported in comparable studies [54-56], underlining the tremendous predictive power of HEB-25. Interestingly, compared to GWAS, only a few additional loci had non-zero effects in the BayesCπ model, indicating that flowering time is indeed mainly controlled by the eight major loci detected via GWAS.

We assume that important reasons for the slightly higher explained p_G of genomic prediction compared to GWAS are that minor QTL effects and marginally existing genetic relatedness [55,57] among HEB lines may be better modeled in the first case. Furthermore, modeling all makers simultaneously enables a better prediction of flowering time due to the estimation of family-specific QTL effects. This is indicated by the occurrence of opposing additive effects between HEB families alongside tightly linked SNPs (Figure 2 and Additional file 6).

A model including epistasis to maximize the cross-validated explained p_G

A final increase of the cross-validated explained p_G to an extraordinary high level of 77% was achieved by including di-genic epistatic interactions between significant main effect SNPs in the RR-BLUP model. This finding indicates that epistasis explains a portion of the 'missing heritability' [58] of flowering time regulation in barley, whereas in maize it does not [28]. The term 'missing heritability' is highly debated in quantitative genetics and refers to the observation that the explained genotypic variance of combined marker effects is usually lower than the heritability of the trait. Epistatic interactions between candidate genes may point to functional relationships and genetic

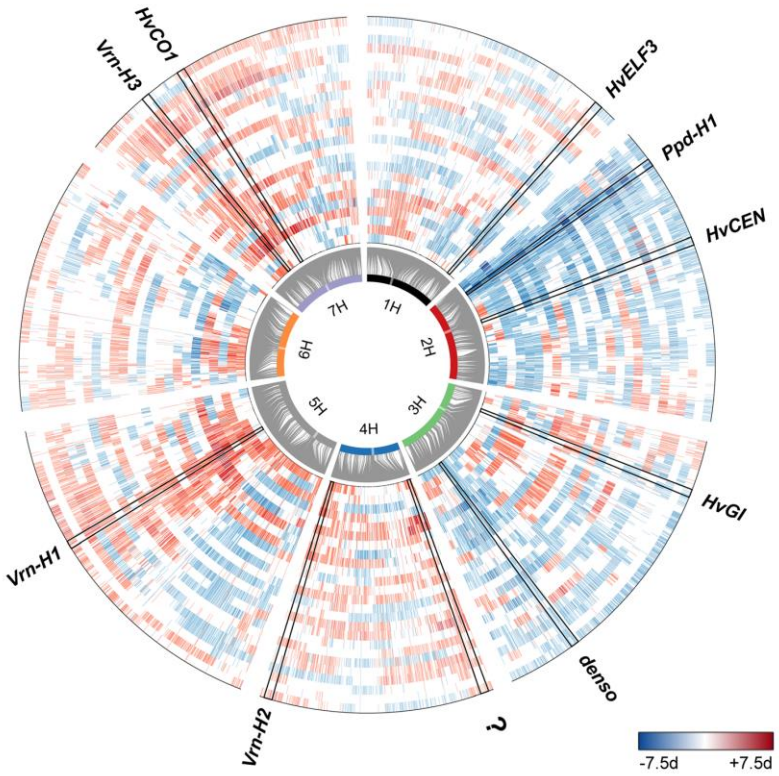

Figure 4 Visualization of family-wise SNP effects. Barley chromosomes are indicated as inner circle of colored bars, centromeres are highlighted as transparent boxes. Grey connector lines represent the genetic position of SNPs on chromosomes. Each track displays one HEB family (F01 – F25, from inside to outside). The heatmap indicates the difference in days between the donor and Barke genotype. Blue and red colors specify early and late flowering, respectively, caused by the donor genotype. White color indicates no SNP effect or SNPs monomorphic in the respective family. Candidate genes (Table 1) are indicated outside the circle. Black frames highlight their family-specific effects as indicated in Additional file 6.

networks [59]. Our findings indicate that the flowering time genes *HvGI*, *Vrn-H2*, *Vrn-H1* and *HvCO1* [60] on chromosomes 3H, 4H, 5H and 7H, respectively, are probably major players of di-genic epistatic interactions in HEB-25. All four genes potentially interact with each other as well as with further genes on additional chromosomes (Figure 2 and Additional file 9). These observations are in agreement with independent studies in barley and *A. thaliana* where these interacting genes were placed in a day length and temperature depending signaling network that controls flowering time [14,21-23]. It is, thus, likely that the observed interaction between the chromosomal regions in HEB-25 may be a function of the mentioned flowering time genes. As an example we refer to the potential interaction found between *Vrn-H1* and *Vrn-H2*. Epistatic interactions between these loci were already reported [17,61,62] and support the model that *Vrn-H2* is a long-day suppressor of flowering, that is itself suppressed by *Vrn-H1* following vernalization [63]. Barke is a spring type barley cultivar that lacks the vernalization requirement due to a deletion of *Vrn-H2* [19]. Hence, our findings may indicate that the epistatic interaction found between the two regions on chromosomes 4H and 5H is based on the presence (exotic allele) or absence (Barke allele) of *Vrn-H2*, the target of *Vrn-H1*. In general, the epistatic interactions detected in HEB-25 may provide hints for the presence of so far unknown functional networks of genes, which assist in fine-tuning flowering time in barley. Studies with knock out lines of these genes may be used to validate the observed interaction effects.

Conclusions

The first barley NAM population HEB-25 provides great opportunities for future research and breeding. The genetic constitution of HEB-25 allows to carry out detailed studies on the genetic architecture of important agronomic traits, as exemplified by flowering time. The present study substantiated that flowering time in barley is primarily

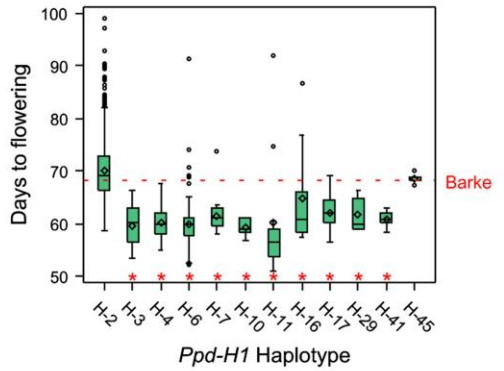

Figure 5 Box-whisker plots of flowering time BLUEs for *Ppd-H1* haplotypes. Green box-whisker-plots display the distribution of flowering time BLUEs of all HEB lines carrying the respective haplotype. Horizontal lines and diamonds indicate median and mean, respectively, for each haplotype. The extension of vertical lines indicates minimum and maximum observations, excluding outliers, which are indicated as circles. The red dotted horizontal line indicates the BLUE of cultivar Barke (68.2 days). H-2 represents the haplotype of the Barke genotype present in HEB lines. All haplotypes except H-45 differ significantly ($P < 0.05$) from H-2, as indicated by red asterisks. Further information to haplotypes is given in Additional files 1 and 8.

determined by large-effect QTL and epistatic interactions. This finding is in contrast to flowering time regulation in the allogamous species maize and sorghum, where only small effect minor QTL were detected [28,42], indicating that the mating system may control the genetic architecture of adaptive traits [28].

In future, the NAM population HEB-25 will be utilized in two directions: On the one hand, HEB-25 may support elucidating the genetic architecture of quantitatively inherited agronomic traits, ultimately resulting in cloning of yet unknown causal genes. On the other hand, HEB-25 will be exploited by breeders to enhance biodiversity of the elite barley gene pool. This may occur through introgression of favorable wild alleles with the aim to sustainably increase yield and stress tolerances against disadvantageous climate conditions like drought, heat and salt.

Methods

Development of the NAM population

The development of the NAM population 'Halle Exotic Barley 25' (HEB-25) was initiated in 2007 conducting crosses between the spring barley cultivar Barke (*Hordeum vulgare* ssp. *vulgare*) and 25 highly divergent exotic wild barley accessions. The latter were used as pollen donors. Twenty-four accessions, originating from Afghanistan, Iran, Iraq, Israel, Lebanon, Turkey, and Syria (*Hordeum vulgare* ssp. *spontaneum*), were selected to maximize the genetic diversity in HEB-25. One further accession, HID380, originating from Tibet, China, was classified as *Hordeum vulgare* ssp. *agriocrithon* (Åberg). F_1 plants of the initial crosses were backcrossed with Barke as the female parent. Twenty BC_1 plants per cross were subsequently selfed three times, using the single seed descent (SSD) technique to generate the next generations. The resulting BC_1S_3 generation consists of 1,420 individual lines, classified in 25 HEB families with 22 to 75 individual lines per family (Additional file 1). Subsequently, each HEB line was bulk propagated until $BC_1S_{3:6}$ to achieve sufficient seed numbers for field testing. No artificial selection was carried out during the development of HEB-25.

Collecting single nucleotide polymorphism (SNP) data

SNP genotype data were collected at TraitGenetics, Gatersleben, Germany, for all 1,420 individual BC_1S_3 lines and their corresponding parents with the barley Infinium iSelect 9k chip consisting of 7,864 SNPs [35]. At each locus, three genotypes were differentiated, with an expected BC_1S_3 segregation ratio of 0.71875 : 0.0625 : 0.21875 for homozygous recipient (i.e. Barke), heterozygous and homozygous donor genotypes, respectively. In total, 1,027 monomorphic SNPs and 1,128 SNPs with high failure rates (i.e. no call in >10% of HEB lines) were excluded from the dataset, resulting in 5,709 informative SNPs for further analyses.

Extraction of genomic DNA

DNA was extracted from leaf tissue of 1,420 single founder HEB plants in generation BC_1S_3. The subsequent seed propagation of HEB lines was based on these founder HEB plants. For Barke and the wild barley accessions leaf material from three to four plants was used to create bulked samples. The plants were cultivated in a glasshouse and 50 to 100 mg of leaf material was harvested for each sample. DNA was extracted according to the manufacturer's protocol, using the BioSprint 96 DNA Plant Kit and a BioSprint work station (Qiagen, Hilden, Germany), and finally dissolved in distilled water at approximately 50 ng/μl.

SNP mapping

The chromosomal positions of 3,391 out of 5,709 SNPs were taken from Comadran *et al.* [35]. The remaining SNPs were fitted next to the mapped SNPs applying chi-square tests of independence. Each non-mapped SNP was compared to each mapped SNP based on genotype segregation across all HEB lines. If two SNPs segregated completely independent from each other, i.e. in case of no linkage disequilibrium (LD), one expects to find all possible genotype combinations according to the product of their single locus genotype frequencies. However, in case of tight linkage, there should be a significant deviation from the expected genotype combination frequencies due to

reduced recombination between these markers. Consequently, a high chi-square statistic and a low *P*-value likely indicate a tight linkage. Therefore, we assigned the position of the SNP with the lowest *P*-value (minimum: $P < 0.001$) to the non-mapped SNP under investigation. If there were more than one SNP with the same *P*-value, the position of the unmapped SNP was defined as the average of the minimum and the maximum position of the respective markers. In this way, all except six of the non-mapped SNPs were placed into the Comadran map.

SNP calling

The differentiation of the HEB genotypes was based on an identity-by-state approach. Based on parental genotype information, the exotic allele could be specified in each segregating family. Thus, HEB lines that showed a homozygous exotic genotype were assigned a value of 2 and HEB lines that showed a homozygous Barke genotype were assigned a value of 0. Consequently, heterozygous HEB lines were assigned a value of 1. If a SNP was monomorphic in one HEB family but polymorphic in a second family, lines of the first HEB family were assigned a genotype value of 0 to keep a full genotype data set, which is a pre-requisite for the subsequent multiple regression analysis. For the same reason, missing genotypes were estimated applying the mean imputation (MNI) approach [64]. For this, each missing SNP value was replaced with the mean of the non-missing values of that SNP in the respective HEB family. Quantitative SNP genotypes were subsequently used for multiple regression analysis.

Evaluation of genetic diversity

SAS 9.4 Software (SAS Institute Inc., Cary, NC, USA) was used to evaluate genetic diversity among parents and progenies of the HEB-25 population. Genetic similarities (GS) between HEB lines and their parents and among HEB lines were calculated with *Proc Distance*, based on a simple matching comparison between the three possible genotype states across all informative SNPs. In addition, we performed principal component analysis (PCA) using R [65]. First we applied PCA for the 26 parents (the cultivar Barke and 25 wild donors) based on the SNP matrix. The first two PCs explained 51.9 and 4.8% of the variation. Then, all progenies of HEB-25 were projected to the space spanned by the two PCs (Additional file 3) as outlined in detail elsewhere [66].

Linkage disequilibrium (LD)

LD was calculated as r^2 [67] between all mapped SNPs with the software package TASSEL [68]. For this purpose, heterozygous genotypes and SNPs with a minor allele frequency < 0.05 were excluded. LD was calculated across the 26 parents of HEB-25. LD decay across intra-chromosomal SNPs was displayed by plotting r^2 between SNP pairs against their genetic distance. A second-degree smoothed loess curve [69] was fitted in SAS with *Proc Loess*. The population-specific baseline r^2 was defined as the 95% percentile of the distribution of r^2 for unlinked markers [70]. LD decay was defined as the distance, at which this baseline crossed the loess curve.

Ppd-H1 haplotype definition

For sequencing of the *Ppd-H1* locus on chromosome 2HS we used the following primers: PP05 (forward) 5′-GTGCAAAGCATAATATCAGTGTCC-3′ and PP04 (reverse) 5′-GGCCAAAGACACAAGAATCAG-3′. These primers amplify the last two exons and three introns of *Ppd-H1* covering the CCT domain that contains SNP22, the causal SNP of *Ppd-H1* [15]. Identical sequences were grouped into haplotypes. A detailed description of the sequencing is given in Jakob *et al.* [71].

HEB-25 field trials

Between 2011 and 2013, three field trials were conducted at the 'Kühnfeld Experimental Station' of the University of Halle to gather phenotype data on flowering time. In 2011, the field trial was conducted with selfed progenies of BC_1S_3 lines (so-called $BC_1S_{3:4}$). Sowing occurred in single to five row plots with a length of 1.50 m and a distance of 0.20 m between rows. The number of rows per HEB line and the position inside the field trial depended on the number of available $BC_1S_{3:4}$ seeds. Lines with seed numbers lower than ten were planted in plots with a length of 0.50 m. In 2012 and 2013, the field trials were conducted with the selfed progenies in $BC_1S_{3:5}$ and $BC_1S_{3:6}$, respectively. Two replications per HEB line, arranged in two randomized complete blocks, were cultivated in 2012 and 2013. The plots consisted of two rows (30 seeds each) with a length of 1.50 m and a distance of 0.20 m between rows. All field trials were sown in spring between March and April with fertilization and pest management following local practice.

Phenotypic data

The occurrence of flowering time was recorded as days after sowing, when the first awns were visible (BBCH49 [72]) for 50% of all plants of a plot. We performed a one-step phenotypic data analysis with SAS, using a linear mixed model with effects for genotype (i.e. 1,420 HEB lines), environment (i.e. 3 years) and interaction of genotype and environment. To estimate variance components, all effects were assumed to be random. Broad-sense heritability (h^2) was estimated on an entry-mean basis. Best linear unbiased estimates (BLUEs) of flowering time were calculated for each genotype assuming fixed genotype effects.

Genome-wide association study (GWAS)

For GWAS, we applied Model-B as outlined in detail by Liu *et al.* [38]. This model was found most suitable to carry out GWAS with multiple families [39]. It is based on multiple regression considering an SNP effect and a family effect in addition to cofactors, which control both population structure and genetic background [39]. Cofactor selection was carried out by applying *Proc Glmselect* in SAS and minimizing the Schwarz Bayesian Criterion [73]. The genome-wide scan for presence of marker-trait associations was implemented in the statistical software R [65], excluding cofactors linked closer than 1 cM to the SNP under investigation. The Bonferroni–Holm procedure [74] was used to adjust marker-trait associations for multiple testing. Significant marker main effects were accepted with $P_{BON\text{-}HOLM} < 0.05$. Additive effects for each SNP were estimated based on regression across but also within families. Significant marker trait associations were grouped to a singled QTL if the significant SNPs were linked by less than 5 cM and revealed additive effects of the same direction, i.e. both exotic alleles increased or decreased flowering time. In addition, a two-dimensional epistasis scan was carried out to identify pairwise marker interactions. For this, the GWAS Model-B was extended to cover a second main SNP effect and the interaction effect between the two SNPs.

Haplotype-based association mapping for *Ppd-H1*

A haplotype-based association mapping test was implemented in HEB-25 to test for effects of haplotypes at *Ppd-H1*. We used the same GWAS procedure with cofactor selection as mentioned before. However, bi-allelic SNPs covering the region of *Ppd-H1* were replaced by a qualitative variable containing the defined *Ppd-H1* haplotype. BLUEs were determined for each haplotype. Subsequently, pairwise comparisons between all haplotype BLUEs were performed using the Tukey-Kramer [75] multiple comparison test.

Genomic prediction

Based on BLUEs of the 1,420 HEB genotypes, two approaches for genomic prediction were applied considering additive effects: ridge regression best linear unbiased prediction (RR-BLUP [51]) and BayesCπ [52]. All statistical procedures for genomic prediction approaches were executed using R. We briefly describe the two models in the following.

Let n be the number of genotypes, m be the number of markers and l be the number of environments. The RR-BLUP model has the form $y = 1_n\mu + Xg + e$, where y is the vector of BLUEs of flowering time scores for all HEB genotypes across environments, 1_n denotes the vector of 1's, μ is the overall mean, g is the vector of marker effects (for SNP markers, allele effects), X is the corresponding design matrix and e is the residual term. In the model we assumed that $g \sim N\left(0, \sigma_g^2\right)$, $e \sim N\left(0,\ \sigma_e^2\right)$, where $\sigma_g^2 = \sigma_G^2/m$ for SNP markers and $\sigma_e^2 = \sigma_R^2/l$. Here σ_G^2 and σ_R^2 are the genotypic and residual variance components obtained in the mixed model in the phenotypic data analysis. The penalty parameter is $\lambda = \left(\sigma_R^2/l\right)/\left(\sigma_G^2/m\right)$. The estimation of marker effects is then given by the mixed model equations [76].

The basic model of BayesC π is the same as RR-BLUP. However, all parameters are treated as random variables in a Bayesian framework. First, we defined the prior distributions as $g \sim N\left(0, \sigma_g^2\right)$, $e \sim N\left(0,\ \sigma_e^2\right)$. The prior of μ is a constant. The prior distribution of σ_g^2 is assumed to be zero with probability π and a scaled inverse chi-squared distribution with probability (1-π). The probability π is a random variable whose prior distribution is uniform on the interval [0,1]. The prior distribution of σ_e^2 is also scaled inverse chi-squared. A Gibbs sampler algorithm was then implemented to infer all the parameters in the model. It was run for 10,000 cycles and the first 1,000 cycles were discarded as burn in. The samples of g from all later cycles were averaged to obtain estimates of the marker effects.

Cross-validation for additive models

The accuracy of the prediction of flowering time by GWAS and the two genomic prediction approaches were evaluated using five-fold cross-validations [77]. In each run of cross-validation, the estimation set included 80% of HEB lines, randomly selected per HEB family, while the remaining 20% of HEB lines were assigned to build the test set. For GWAS, we performed an association mapping scan within the estimation set and recorded the detected significant markers. To determine the cross-validated proportion of explained genotypic variance (p_G), we estimated the effects of the significant peak markers within the estimation set and predicted the genotypic value of the lines in the test set [40]. We then calculated the cross-validated p_G as the squared Pearson product–moment correlation between predicted and observed genotypic values in the test set standardized with the heritability. The mean p_G in 100 cross-validation runs (20 times five-fold cross-validations) was taken as the final record. In addition, the number of significances for each SNP was cumulated across all runs and is referred to as QTL detection rate.

For genomic prediction we estimated the effects for all markers using the estimation set and predicted the genotypic value of the lines in the test set. The cross-validated p_G was calculated as in GWAS.

Exploiting additive and additive times additive epistatic effects in genomic prediction

We extended the RR-BLUP based on main effects to model also epistasis for markers with significant main effects in the GWAS. The model is $y = 1_n\mu + \sum_{i=1}^{n} X_i g_i + \sum_{j<l}(X_j \cdot X_l) f_{jl} + e$, where y is the vector of BLUEs of flowering time for all HEB genotypes, 1_n denotes the vector of 1's, μ is the overall mean, g_i is the main additive effect of the i-th marker, X_i is the vector of marker indices, f_{jl} is the epistatic effects of the j- and the l-th marker, X_j X_l is the point-wise product of the two vectors X_j and X_l, and e is the vector of residual terms. Note that in the third term of the right hand side of the formula, the sum is not taken over all pairs of markers but only pairs of markers exhibiting a significant additive effect in the GWAS study performed previously. Hence, in different cross-validation runs, different pairs of markers were considered in the model. The model assumptions are similar to the usual RR-BLUP, except treating additive and epistatic effects separately. We assumed $g_i \sim N\left(0, \sigma_g^2\right)$, $f_{jl} \sim N\left(0, \sigma_f^2\right)$, where $\sigma_g^2 = p_G\sigma_G^2/m$, $\sigma_f^2 = (1-p_G)\sigma_G^2/p$. Here P_G is the cross-validated proportion of explained genotypic variance for genomic prediction, obtained previously by the RR-BLUP, only considering additive effects, m is the number of markers, p is the number of pairs of markers having significant additive effect. Therefore the penalty parameter λ is different for additive and epistatic effects.

Using the above extended model, for each cross-validation run we estimated the additive effects of all markers and epistatic effects of all pairs of markers exhibiting significant additive effects in GWAS using the estimation set. Then we predicted genotypic values of the lines in the test set and calculated the p_G in the same way as outlined above.

Availability of supporting data

Raw data, including data on SNPs, *Ppd-H1* haplotypes and GWAS, and all other supporting data are provided as additional files.

Additional files

Additional file 1: Genetic constitution of HEB-25: classification of families and donors. Tabular overview of the genetic constitution of HEB-25, classifying the 25 families and donors and containing the *Ppd-H1* haplotypes.

Additional file 2: LD decay of intra-chromosomal markers among HEB-25 parents. Figure showing the LD decay of intra-chromosomal markers among HEB-25 parents by plotting r^2 against the genetic marker distance.

Additional file 3: Principal component analysis for HEB-25 and its parents. Figure showing the relatedness of HEB-25 lines by plotting of the first two principal components of a principal component analysis for HEB-25 and its parents.

Additional file 4: Distribution of flowering time. Figure showing the frequency distribution of flowering time BLUEs across three field trials and illustrating contrasting phenotypes in the field.

Additional file 5: Phenotype and genotype data for HEB-25. Table listing the complete phenotype and genotype data of HEB-25 underlying this study as well as marker information.

Additional file 6: Estimates of single marker GWAS and genomic prediction effects across HEB-25 and within individual HEB families. Table listing the results of GWAS and genomic prediction across HEB-25 and within individual HEB families.

Additional file 7: GWAS Manhattan plot for flowering time. Figure displaying the GWAS results through plotting the significance and effects of markers in a Manhattan plot.

Additional file 8: *Ppd-H1* haplotype comparison. Two tables contrasting the different *Ppd-H1* haplotype effects by comparison of their BLUEs.

Additional file 9: Significant epistatic interactions via GWAS. Table listing all significant ($P_{BON\text{-}HOLM} < 0.05$) epistatic interactions between SNPs that were obtained via GWAS.

Abbreviations

BC_1: Progeny of F_1 after back-crossing with Barke; BC_1S_3: Progeny of BC_1 after three rounds of selfing; $BC_1S_{3:x}$: Progeny of BC_1S_3 individual after x-3 rounds of bulk propagation; BLUE: Best linear unbiased estimate; cM: Centimorgan; F_1: First generation after initial crosses; GA: Gibberellic acid; GS: Genetic similarity; GWAS: Genome-wide association study; *Hag*: *Hordeum agriocrithon*; HEB: Halle exotic barley; HID: *Hordeum* identity; *Hsp*: *Hordeum spontaneum*; *Hv*: *Hordeum vulgare*; LD: Linkage disequilibrium; MNI: Mean imputation; NAM: Nested association mapping; PC: Principal component; PCA: Principal component analysis; p_G: Proportion of explained genotypic variance; QTL: Quantitative trait locus/loci; RR-BLUP: Ridge regression best linear unbiased prediction; SNP: Single nucleotide polymorphism; SSD: Single seed descent.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AM planned and conducted the field trials in 2012 and 2013, analyzed the genotype and phenotype data, carried out the GWAS, created the figures, and drafted the manuscript. VD was involved in the development of the HEB-25 population, gathered the genotype data, and planned and conducted the field trial in 2011. YJ performed the genomic prediction and cross-validation approaches. FS was involved in the development of the HEB-25 population. RS performed the re-sequencing of *Ppd-H1*. ES supervised the field trials. BK supervised the *Ppd-H1* re-sequencing. JCR supervised the genomic prediction and cross-validation approaches. KP acquired the funding, supervised the development of the HEB-25 population and all analyses, and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft (DFG), Bonn (projects Pi339/7-1 and Pi 339/7-2), and the Interdisciplinary Centre for Crop Plant Research (IZN), Halle. We are grateful to Roswitha Ende, Jana Müglitz, Diana Rarisch, Helga Sängerlaub, Brigitte Schröder, Bernd Kollmorgen and various student assistants for excellent technical help and to TraitGenetics GmbH, Gatersleben, Germany, for genotyping HEB-25 with the Infinium iSelect 9k SNP chip.

Author details

[1]Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle Wittenberg, Betty-Heimann-Str. 3, 06120 Halle, Germany. [2]Interdisciplinary Center for Crop Plant Research (IZN), Betty-Heimann-Str. 3, 06120 Halle, Germany. [3]Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Stadt Seeland, OT Gatersleben, Germany. [4]Current address: University of Dundee at the James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK. [5]Current address: Bayer CropScience NV, Technologiepark 38, 9052 Ghent, Belgium.

Received: 12 October 2014 Accepted: 9 March 2015
Published online: 12 April 2015

References

1. Sakuma S, Salomon B, Komatsuda T. The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. Plant Cell Physiol. 2011;52:738–49.
2. Zohary D, Hopf M, Weiss E. Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. 4th ed. Oxford: Oxford University Press; 2012.
3. Morrell PL, Clegg MT. Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. Proc Natl Acad Sci U S A. 2007;104:3289–94.
4. Dai F, Nevo E, Wu D, Comadran J, Zhou M, Qiu L, et al. Tibet is one of the centers of domestication of cultivated barley. Proc Natl Acad Sci U S A. 2012;109:16969–73.
5. Tanksley SD, McCouch SR. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. Science. 1997;277:1063–6.
6. Zamir D. Improving plant breeding with exotic genetic libraries. Nat Rev Genet. 2001;2:983–9.
7. Ellis RP, Forster BP, Robinson D, Handley LL, Gordon DC, Russell JR, et al. Wild barley: a source of genes for crop improvement in the 21st century? J Exp Bot. 2000;51:9–17.
8. Pillen K, Zacharias A, Leon J. Advanced backcross QTL analysis in barley (Hordeum vulgare L.). Theor Appl Genet. 2003;107:340–52.
9. Wang G, Schmalenbach I, von Korff M, Leon J, Kilian B, Rode J, et al. Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC(2)DH population and a set of wild barley introgression lines. Theor Appl Genet. 2010;120:1559–74.
10. Schmalenbach I, March TJ, Bringezu T, Waugh R, Pillen K. High-Resolution Genotyping of Wild Barley Introgression Lines and Fine-Mapping of the Threshability Locus thresh-1 Using the Illumina GoldenGate Assay. G3 (Bethesda). 2011;1:187–96.
11. Ma X, Li C, Wang A, Duan R, Jiao G, Nevo E, et al. Genetic diversity of wild barley (Hordeum vulgare ssp. spontaneum) and its utilization for barley improvement. Sci Cold Arid Reg. 2012;4:453–61.
12. McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, et al. Agriculture: Feeding the future. Nature. 2013;499:23–4.
13. Schnaithmann F, Kopahnke D, Pillen K. A first step toward the development of a barley NAM population and its utilization to detect QTLs conferring leaf rust seedling resistance. Theor Appl Genet. 2014;127:1513–25.
14. Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, et al. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J Exp Bot. 2007;58:1231–44.
15. Turner A, Beales J, Faure S, Dunford RP, Laurie DA. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science. 2005;310:1031–4.
16. Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, et al. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proc Natl Acad Sci U S A. 2006;103:19581–6.
17. Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, et al. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science. 2004;303:1640–4.
18. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci U S A. 2003;100:6263–8.
19. von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, et al. Molecular and structural characterization of barley vernalization genes. Plant Mol Biol. 2005;59:449–67.
20. Laurie D, Pratchett N, Snape J, Bezant J. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (Hordeum vulgare L.) cross. Genome. 1995;38:575–85.
21. Distelfeld A, Li C, Dubcovsky J. Regulation of flowering in temperate cereals. Curr Opin Plant Biol. 2009;12:178–84.
22. Jung C, Mueller AE. Flowering time control and applications in plant breeding. Trends Plant Sci. 2009;14:563–73.
23. Milec Z, Valárik M, Bartoš J, Šafář J. Can a late bloomer become an early bird? Tools for flowering time adjustment. Biotechnol Adv. 2014;32:200–14.
24. Cockram J, Hones H, O'Sullivan DM. Genetic variation at flowering time loci in wild and cultivated barley. Plant Genet Resour. 2011;9:264–7.
25. Nevo E, Fu Y-B, Pavlicek T, Khalifa S, Tavasi M, Beiles A. Evolution of wild cereals during 28 years of global warming in Israel. Proc Natl Acad Sci U S A. 2012;109:3412–5.
26. Yu J, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. Genetics. 2008;178:539–51.
27. Jordan D, Mace E, Cruickshank A, Hunt C, Henzell R. Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. Crop Sci. 2011;51:1444–57.
28. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The Genetic Architecture of Maize Flowering Time. Science. 2009;325:714–8.
29. Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet. 2011;43:163–8.
30. Poland JA, Bradbury PJ, Buckler ES, Nelson RJ. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci U S A. 2011;108:6893–8.
31. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet. 2011;43:159–62.
32. Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, et al. Genetic Architecture of Maize Kernel Composition in the Nested Association Mapping and Inbred Association Panels. Plant Physiol. 2012;158:824–34.
33. Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, et al. The genetic architecture of maize height. Genetics. 2014;196:1337–56.
34. Wallace J, Larsson S, Buckler E. Entering the second century of maize quantitative genetics. Heredity (Edinb). 2014;112:30–8.
35. Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganal M, et al. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet. 2012;44:1388–92.
36. Gottwald S, Bauer P, Komatsuda T, Lundqvist U, Stein N. TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene HvHox1. BMC Res Notes. 2009;2:258.
37. Badr A, Muller K, Schafer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, et al. On the origin and domestication history of barley (Hordeum vulgare). Mol Biol Evol. 2000;17:499–510.
38. Liu W, Gowda M, Steinhoff J, Maurer HP, Würschum T, Longin CFH, et al. Association mapping in an elite maize breeding population. Theor Appl Genet. 2011;123:847–58.
39. Würschum T, Liu W, Gowda M, Maurer H, Fischer S, Schechert A, et al. Comparison of biometrical models for joint linkage association mapping. Heredity (Edinb). 2012;108:332–40.
40. Utz HF, Melchinger AE, Schön CC. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics. 2000;154:1839–49.
41. Gowda M, Zhao Y, Würschum T, Longin CF, Miedaner T, Ebmeyer E, et al. Relatedness severely impacts accuracy of marker-assisted selection for disease resistance in hybrid wheat. Heredity (Edinb). 2014;112:552–61.
42. Mace E, Hunt C, Jordan D. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. Theor Appl Genet. 2013;126:1377–95.
43. Loscos J, Igartua E, Contreras-Moreira B, Gracia MP, Casas AM. HvFT1 polymorphism and effect – survey of barley germplasm and expression analysis. Front Plant Sci. 2014;5:251.
44. Mutasa-Göttgens E, Hedden P. Gibberellin as a factor in floral regulatory networks. J Exp Bot. 2009;60:1979–89.
45. Jia Q, Zhang J, Westcott S, Zhang X-Q, Bellgard M, Lance R, et al. GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. Funct Integr Genomics. 2009;9:255–62.
46. Faure S, Turner AS, Gruszka D, Christodoulou V, Davis SJ, von Korff M, et al. Mutation at the circadian clock gene EARLY MATURITY 8 adapts domesticated barley (Hordeum vulgare) to short growing seasons. Proc Natl Acad Sci U S A. 2012;109:8328–33.
47. Zakhrabekova S, Gough SP, Braumann I, Muller AH, Lundqvist J, Ahmann K, et al. Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. Proc Natl Acad Sci U S A. 2012;109:4326–31.
48. Boden SA, Weiss D, Ross JJ, Davies NW, Trevaskis B, Chandler PM, et al. EARLY FLOWERING3 Regulates Flowering in Spring Barley by Mediating

Gibberellin Production and FLOWERING LOCUS T Expression. Plant Cell. 2014;26:1557–69.
49. Dunford RP, Griffiths S, Christodoulou V, Laurie DA. Characterisation of a barley (Hordeum vulgare L.) homologue of the Arabidopsis flowering time regulator GIGANTEA. Theor Appl Genet. 2005;110:925–31.
50. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics. 2001;157:1819–29.
51. Whittaker JC, Thompson R, Denham MC. Marker-assisted selection using ridge regression. Genet Res. 2000;75:249–52.
52. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the Bayesian alphabet for genomic selection. BMC Bioinf. 2011;12:186.
53. Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet. 2009;120:151–61.
54. Guo ZG, Tucker DM, Lu JW, Kishore V, Gay G. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor Appl Genet. 2012;124:261–75.
55. Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, Guo B, et al. The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet. 2014;127:749–62.
56. Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity (Edinb). 2014;112:638–45.
57. Habier D, Fernando R, Dekkers J. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;177:2389–97.
58. Mackay TF. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet. 2013;15:22–33.
59. von Korff M, Léon J, Pillen K. Detection of epistatic interactions between exotic alleles introgressed from wild barley (H. vulgare ssp. spontaneum). Theor Appl Genet. 2010;121:1455–64.
60. Griffiths S, Dunford RP, Coupland G, Laurie DA. The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant Physiol. 2003;131:1855–67.
61. Tranquilli G, Dubcovsky J. Epistatic interaction between vernalization genes Vrn-Am1 and Vrn-Am2 in diploid wheat. J Hered. 2000;91:304–6.
62. Szűcs P, Skinner JS, Karsai I, Cuesta-Marcos A, Haggard KG, Corey AE, et al. Validation of the VRN-H2/VRN-H1 epistatic model in barley reveals that intron length variation in VRN-H1 may account for a continuum of vernalization sensitivity. Mol Genet Genomics. 2007;277:249–61.
63. Chen A, Dubcovsky J. Wheat TILLING mutants show that the vernalization gene VRN1 down-regulates the flowering repressor VRN2 in leaves but is not essential for flowering. PLoS Genet. 2012;8:e1003134.
64. Rutkoski JE, Poland J, Jannink J-L, Sorrells ME. Imputation of Unordered Markers and the Impact on Genomic Selection Accuracy. G3 (Bethesda). 2013;3:427–39.
65. R Development Core Team (2010) R: A language and environment for statistical computing, R Foundation for Statistical Computing. http://www.R-project.org
66. Xia X, Reif J, Melchinger A, Frisch M, Hoisington D, Beck D, et al. Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers. Crop Sci. 2005;45:2573–82.
67. Weir BS. Genetic Data Analysis II. Sunderland, MA: Sinauer Associates, Inc.; 1996.
68. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633–5.
69. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74:829–36.
70. Breseghello F, Sorrells ME. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics. 2006;172:1165–77.
71. Jakob SS, Rödder D, Engler JO, Shaaf S, Özkan H, Blattner FR, et al. Evolutionary History of Wild Barley (Hordeum vulgare subsp. spontaneum) Analyzed Using Multilocus Sequence Data and Paleodistribution Modeling. Genome Biol Evol. 2014;6:685–702.
72. Lancashire PD, Bleiholder H, Boom TVD, Langelüddeke P, Stauss R, Weber E, et al. A uniform decimal code for growth stages of crops and weeds. Ann Appl Biol. 1991;119:561–601.
73. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–4.
74. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6:65–70.
75. Kramer CY. Extension of multiple range tests to group means with unequal numbers of replications. Biometrics. 1956;12:307–10.
76. Henderson CR. Applications of Linear Models in Animal Breeding. Guelph: University of Guelph; 1984.
77. Hjorth JU. Computer intensive statistical methods: Validation, model selection, and bootstrap. London: Chapman & Hall/CRC; 1993.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Chapter 2

Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping[2]

After the initial investigation of flowering time in HEB-25 (Chapter 1) the aim of the following study was to dissect the genetic control of flowering time in more detail to fine-tune plant development with regard to yield components.

Since not only flowering time itself but also other developmental stages like maturity have a major impact on final grain yield the question was whether different developmental stages are under independent genetic control or if they are co-regulated by the same major genes. This should be achieved by comparing GWAS results of several developmental traits and showing their impact on the yield component thousand grain weight. Finally, conclusions should be drawn about an optimal strategy for the timing of developmental phases to increase thousand grain weight.

[2] Maurer A, Draba V, Pillen K (2016) Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. J Exp Bot 67: 2507-2518.

as being two major determinants of flowering time. *Vrn-H3* (Yan *et al.*, 2006) is the key gene controlling flower initiation in barley. It is an orthologue of the Arabidopsis *FT* (*FLOWERING LOCUS T*) gene. The FT protein moves from leaves to the shoot apical meristem, as postulated decades ago for the mobile signal 'florigen' (Chailakhyan, 1937). Its function as a promoter of flowering is assumed to be preserved across different plant species (Turck *et al.*, 2008).

The key vernalization genes *Vrn-H1* (Yan *et al.*, 2003) and *Vrn-H2* (Yan *et al.*, 2004) have a major impact on *Vrn-H3* expression. *Vrn-H2* was determined to be a repressor of *Vrn-H3*, which prevents flowering under long days before vernalization. *Vrn-H1* responds to low temperatures (Oliver *et al.*, 2013) as a result of *cis*-regulatory elements in its promotor region (Alonso-Peral *et al.*, 2011). It is up-regulated after vernalization and, thus, promotes flowering through direct binding to the promoters of *Vrn-H2* (repression) and *Vrn-H3* (activation) (Deng *et al.*, 2015). A differentiation between winter and spring barley can be made based on their response to vernalization. The latter lacks the vernalization requirement as a result of a natural deletion of *Vrn-H2* (von Zitzewitz *et al.*, 2005).

Flowering is furthermore promoted by *Ppd-H1* under long days (Turner *et al.*, 2005) and by *Ppd-H2* under short days (Kikuchi *et al.*, 2009). In addition to photoperiod and vernalization, light quality (Nishida *et al.*, 2013; Pankin *et al.*, 2014), circadian rhythms (Campoli *et al.*, 2012; Faure *et al.*, 2012; Zakhrabekova *et al.*, 2012; Campoli *et al.*, 2013; Calixto *et al.*, 2015), and phytohormones like gibberellic acid (GA) (Jia *et al.*, 2009; Jia *et al.*, 2011; Boden *et al.*, 2014) and cytokinins (Mrízová *et al.*, 2013) also contribute to the induction of flowering in barley. Many loci that were previously classified as earliness *per se* genes that act independently from external signals (Laurie *et al.*, 1995) were recently assigned to those classes. The complexity of floral networks (Blümel *et al.*, 2015) is caused by the interplay between numerous genes and external signals. It is worth taking a closer look at specific developmental subphases to help unravel these networks.

The life cycle of barley consists of several subphases. The most basic differentiation divides the life of a barley plant into vegetative, reproductive and grain-filling phases (Sreenivasulu and Schnurbusch, 2012). Vegetative plant organs develop during the vegetative phase. The reproductive phase starts with the initiation of spikelets, which develop over time. This phase is terminated with anthesis resulting in the onset of the grain-filling phase. The length of different preanthesis subphases has been shown to be under genetic control (Kernich *et al.*, 1997; Borras *et al.*, 2009; Borras-Gelonch *et al.*, 2010; Borras-Gelonch *et al.*, 2012; Alqudah *et al.*, 2014) and impacts yield-related traits (Miralles *et al.*, 2000; Alqudah and Schnurbusch, 2014, 2015). The postanthesis phase is also assumed to have a major impact on yield by determining the time frame for grain-filling (Evans and Wardlaw, 1976; Egli, 2004), which controls the yield component grain weight (Distelfeld *et al.*, 2014). Although the timing and duration of developmental phases is important for a plant's yield potential, most genomic studies dealing with the regulation of plant development and its impact on yield focus solely on flowering time as the only developmental parameter.

Nested association mapping (NAM) has been shown to be a valuable tool in the dissection of the genetic architecture of many traits in maize, sorghum, and barley (Buckler *et al.*, 2009; Jordan *et al.*, 2011; Maurer *et al.*, 2015). A NAM population is the result of wide crosses of highly diverse donor genotypes with a recurrent elite cultivar, followed by several rounds of selfing. It combines the advantages of association mapping (i.e. high allele richness and mapping resolution) and linkage mapping (i.e. high statistical power). It also provides excellent opportunities to evaluate the performance of untapped wild alleles in an elite background as a result of its exclusive mating design (Ogut *et al.*, 2015).

The aim of this study was to characterize the genetic architecture of barley development. To do this, we used the NAM population HEB-25 (Maurer *et al.*, 2015) consisting of 1420 highly divergent BC_1S_3 lines. We also wanted to shed more light on the flowering time pathway in barley by comparing the impact flowering time genes had on different developmental subphases. In addition to time to flowering (HEA), we investigated time to shooting (SHO), duration of the shoot elongation phase (SEL), duration of the ripening phase (RIP) and time to maturity (MAT). In order to gain insight into additional physiological functions of flowering time genes and their impact on yield formation, we also investigated plant height (HEI) and thousand grain weight (TGW, Table 1). Furthermore, we looked at whether there was useful variation present in the wild barley germplasm that could be used to fine-tune specific developmental phases in order to increase yield potential in future barley breeding programmes.

Materials and methods

Plant material

The NAM population HEB-25 (Maurer *et al.*, 2015), consisting of 1420 individual BC_1S_3 lines in 25 wild-barley-derived subfamilies, was used in this study. HEB-25 was the result of initial crosses between the spring barley cultivar Barke (*Hordeum vulgare* ssp. *vulgare*) and 25 highly divergent exotic wild barley accessions (*Hordeum vulgare* ssp. *spontaneum* and *agriocrithon*). F_1 plants of the initial crosses were backcrossed with Barke. For detailed information about the population design, see Maurer *et al.* (2015).

Collecting single nucleotide polymorphism data

Single nucleotide polymorphism (SNP) genotype data were collected at TraitGenetics, in Gatersleben, Germany, for all 1420 individual BC_1S_3 lines and their corresponding parents. The barley Infinium iSelect 9K chip consisted of 7864 SNPs (Comadran *et al.*, 2012). The genotype data were processed and stored as indicated in Maurer *et al.* (2015) and 5709 informative SNPs, which met the quality criteria, could be utilized in this study. An identity-by-state approach was used to differentiate between the HEB genotypes. Based on parental genotype information, the exotic allele could be specified in each segregating family, and homozygous exotic genotypes were assigned a value of 2. HEB lines that showed a homozygous Barke genotype were assigned a value of 0. Consequently, heterozygous HEB lines were assigned a value of 1. Numbers can therefore be interpreted as a quantitative variable representing the dose of the wild allele.

HEB-25 field trials

Between 2011 and 2014, four field trials were conducted at the 'Kühnfeld Experimental Station' of Martin Luther University Halle-Wittenberg (51°29′46.47″N; 11°59′41.81″E) to gather phenotype data. In 2011, the field trial was conducted with selfed progenies of BC_1S_3 lines (so-called $BC_1S_{3:4}$), arranged as a single randomized block. The majority of plots (92%) included one or two rows per HEB line whereas the remaining 8% of plots included three to five rows, depending on the number of available $BC_1S_{3:4}$ seeds. Plots in 2011 had a length of 1.50 m, a distance of 0.20 m between rows, and were separated by 0.50 m to reduce competition between plots. In 2012 and 2013, the field trials were conducted with the selfed progenies in $BC_1S_{3:5}$ and $BC_1S_{3:6}$, respectively. Two replications per HEB line, arranged in two randomized complete blocks, were cultivated in 2012 and 2013. The plots consisted of two rows (30 seeds each) with a length of 1.50 m, a distance of 0.20 m between rows, and a spacing of 0.50 m between plots. In 2014, $BC_1S_{3:7}$ seeds were sown in a single randomized block. The plots consisted of two rows (50 seeds each) with a length of 1.50 m, a distance of 0.20 m between rows and a spacing of 0.50 m between plots. Barke was integrated as a check line in all trials. All field trials were sown in spring, between March and April, with fertilization and pest management carried out according to local practice. No additional fertilizer was applied in 2014.

Phenotypic data

Table 1 shows a list of all of the investigated traits and a description of their measurements and the years studied. This information is supplemented insofar as all developmental traits were recorded both as 'days from sowing' and 'growing degree days' (GDD), which were calculated with a base temperature of 0 °C in accordance with equation (1) in McMaster and Wilhelm (1997). Thus, the mean daily temperatures of all of the days with a mean temperature above 0 °C were cumulated.

Statistical analyses

We performed a one-step phenotypic data analysis for all traits with SAS 9.4 (SAS Institute Inc., Cary, NC, USA), based on a linear mixed model (PROC MIXED) with effects for genotype (i.e. 1420 HEB lines), environment (i.e. 4 years) and interaction of genotype and environment. To estimate variance components, all effects were assumed to be random (PROC VARCOMP). Heritabilities across years were calculated as: $h^2 = \frac{V_G}{V_G + \frac{V_{GY}}{y} + \frac{V_R}{yr}}$, where V_G, V_{GY} and V_R represent the genotypic components, genotype × year, and error variance components, respectively. The terms y and r indicate the number of years and replicates, respectively. Best linear unbiased estimates (BLUEs) were calculated with PROC MIXED for each genotype assuming fixed genotype effects. BLUEs were used to calculate Pearson's correlation coefficients (r) with PROC CORR.

Genome-wide association study

We applied Model B on trait BLUEs as outlined in detail by Liu *et al.* (2011). This model was found to be most suited for carrying out genome-wide association study (GWAS) with multiple families (Würschum *et al.*, 2012) and has already been shown to work properly in HEB-25 (Maurer *et al.*, 2015). It is based on multiple regression, taking into account a quantitative SNP effect and a qualitative family effect in addition to quantitative cofactors that control both population structure and genetic background (Würschum *et al.*, 2012). Cofactor selection was carried out on this model and included all SNPs simultaneously by applying PROC GLMSELECT in SAS and minimizing the Schwarz Bayesian Criterion (Schwarz, 1978). PROC GLM was used to perform the genome-wide scan for the presence of marker–trait associations. Cofactors that were linked closer than 1 cM to the SNP under investigation were excluded. The Bonferroni–Holm method (Holm, 1979) was used to adjust marker–trait associations for multiple testing. Significant marker effects were accepted with $P_{BON\text{-}HOLM}<0.05$. The proportion of the phenotypic variance explained by a marker was determined by estimating R^2 after modelling the marker solely in a linear model. Additive effects for each SNP were taken as the regression coefficient of the SNP directly from the GWAS model. Family-specific effects were calculated for all markers based on a simple linear model, including a general family term and the marker effect as nested within the family.

To increase the robustness of the method, the entire procedure was applied 200 times on random subsamples of the full dataset. Each subsample included 80% of the lines, randomly selected per HEB family. We recorded the significant ($P_{BON\text{-}HOLM}<0.05$) markers detected, which is referred to as the detection rate. Markers that were detected in at least 10% of subsamples were accepted as putative QTLs, following Ogut *et al.* (2015). Significant markers were merged into a single QTL if they were linked by less than 4 cM. Additive effects, $P_{BON\text{-}HOLM}$ values and R^2 were averaged across all runs, in which the respective marker was significant. In order to evaluate the explained phenotypic variance, the unbiased estimator R^2_{adj} (Draper and Smith, 1981) was determined for each subsample by simultaneously modelling all of the significant markers in a linear model. In order to determine the predictive ability R^2_{pred}, the estimated additive effects of each subsample were used to predict the phenotypic value of the remaining

Table 1. *List of evaluated traits*

Abbreviation	Trait	Unit	Method of measurement	Years studied
SHO	Shooting	days	Number of days from sowing until first node palpable at least 1 cm above the tillering node for 50% of all plants of a plot (BBCH 31; Lancashire *et al.*, 1991)	2011–2014
SEL	Shoot elongation phase	GDD	Time from SHO to HEA	2011–2014
HEA	Flowering	days	Number of days from sowing until first awns visible (BBCH 49; Lancashire *et al.*, 1991) for 50% of all plants of a plot	2011–2014
RIP	Ripening phase	days	Time from HEA to MAT	2012–2014
MAT	Maturity	days	Number of days from sowing until hard dough: grain content solid and fingernail impression held (BBCH 87; Lancashire *et al.*, 1991) for 50% of all plants of a plot	2012–2014
TGW	Thousand grain weight	g	Calculated after harvest by use of MARVIN seed analyser (GTA Sensorik GmbH, Neubrandenburg, Germany) based on a 200 seeds sample of each plot. Before, seeds were cleaned and damaged seeds were sorted out	2011–2013
HEI	Plant height	cm	Recorded at maturity as the distance from ground to tip of the erected ear (without awns), taken as an average across the ears of a plot	2011–2013

20% of the lines. We then calculated R^2_{pred} to be the squared Pearson product–moment correlation between predicted and observed phenotypic values. The means of R^2_{adj} and R^2_{pred}, measured over 200 runs, were ultimately recorded as the final values.

We used the BARLEYMAP pipeline (Cantalapiedra *et al.*, 2015) to identify potential candidate genes to explain the QTLs. BARLEYMAP enables markers to be mapped and gene sequences to be aligned against sequence-enriched genetic (Mascher *et al.*, 2013) and physical frameworks (International Barley Genome Sequencing Consortium, 2012). This represents a very precise way to find positional coincidence of QTLs with putative candidate genes. The results are presented in Supplementary Table S1 at *JXB* online. We also compared our wild allele effects to those reported in a barley BC_2DH population and a set of wild barley introgression lines (Wang *et al.*, 2010). Here several known flowering time genes were sequenced in wild and cultivated barley. Since the wild allele effect was also produced directly from a cross of wild barley and cultivated barley, we assume that similar effects indicate the same candidate genes.

Results and discussion

Phenotypes

A broad variation in phenotypes of HEB-25 lines was observed for all traits across and within years. This resulted in high coefficients of variation (Supplementary Table S2). After calculating BLUEs, which were corrected for year effects, we observed large differences of more than 100% between the most extreme HEB lines for each trait except MAT (Table 2 and Supplementary Fig. S1). For instance, a difference of 51.4 days between the earliest and the latest genotypes could be observed for HEA, and likewise TGW varied between 19.4 g and 60.2 g. The lowest coefficient of variation was obtained for the trait MAT, where the most extreme genotypes nevertheless showed a difference of 33.4 days.

Phenotypic correlations

We calculated Pearson's correlation coefficients (r) in order to gain basic insights into the relationships between the different phases of plant development, HEI, and their influence on TGW. We observed very high correlations (0.88–0.93) between SHO, HEA, and MAT (Table 3). This indicates that early shooting lines also tend to be early for other stages. Another peculiarity is that in the case of RIP correlation coefficients were negative for all other developmental stages. Together with the relatively low coefficient of variation for MAT, this indicates that the time of maturity may be predetermined or limited to some extent by environmental factors such as heat and drought. As both SEL and RIP were calculated to be the difference between two other stages, we can draw conclusions about their main determinant. Following this, HEA had the greatest impact on the duration of SEL and RIP, since its correlation coefficients outperformed those of SHO and MAT, respectively. The duration of SEL greatly impacted HEI (r=0.45), indicating that the occurrence of (semi-)dwarf plants was based more on a shortened period of SEL than on a reduced growth rate. No correlation with TGW could be observed for SEL. TGW was positively correlated with RIP (r=0.37), which may be due to an extension of grain-filling, where starch is being stored in the grains (Distelfeld *et al.*, 2014). Interestingly, HEI was also positively correlated with TGW (r=0.31).

Table 2. *Descriptive statistics for best linear unbiased estimates (BLUEs) and heritabilities across all environments*

Trait[a]	N[b]	Mean[c]	SD[d]	Min[e]	Max[f]	CV%[g]	h²[h]
SHO	1422	53.3	5.6	38.6	82.6	10.4	0.93
SEL	1422	237.8	42.6	108.9	396.0	17.9	0.75
HEA	1422	67.9	6.3	50.4	101.9	9.2	0.94
RIP	1420	32.7	2.6	19.2	40.5	7.9	0.81
MAT	1420	101.3	4.5	88.5	121.9	4.5	0.91
TGW	1422	46.5	5.0	19.4	60.2	10.8	0.57
HEI	1420	63.9	8.5	41.0	100.5	13.3	0.88

[a] Trait abbreviations are given in Table 1.
[b] Number of observations (genotypes).
[c] Arithmetic mean.
[d] Standard deviation.
[e] Minimum.
[f] Maximum.
[g] Coefficient of variation (%).
[h] Heritability.

Table 3. *Pearson's correlation coefficients (r)*

	SEL	HEA	RIP	MAT	TGW	HEI
SHO	**0.32**	**0.92**	**–0.67**	**0.88**	**–0.38**	–0.01
SEL		**0.66**	**–0.60**	**0.57**	–0.07	**0.45**
HEA			**–0.79**	**0.93**	**–0.35**	**0.17**
RIP				**–0.54**	**0.37**	**–0.19**
MAT					**–0.28**	**0.13**
TGW						**0.31**

Bold values indicate significant correlations at P<0.0001. Trait abbreviations are given in Table 1.

Heritabilities

Heritabilities for all investigated traits were calculated over 3–4 years. We observed heritabilities >0.5 for all traits (Table 2). Heritabilities for the developmental traits SHO, HEA, and MAT were almost identical, regardless of whether they were measured in days or growing degree days (GDD) (Supplementary Table S3). However, days outperformed GDD (0.81 vs 0.68) in the case of RIP, while GDD outperformed days (0.75 vs 0.69) in the case of SEL. Thus, for SEL, GDD rather than days offers a better estimate for the relative time needed to fulfil this stage, since plant growth rate and, hence, plant development is based on the interplay between different temperature-dependent biochemical processes (Atwell *et al.*, 1999; Atkin and Tjoelker, 2003). This is of particular importance, especially when dealing with data from different years and when taking into account that the onset of SEL differs greatly (from 38.6 to 82.6 days after sowing) between individual lines in the HEB-25 population. In contrast, the use of days instead of GDD resulted in a higher heritability for RIP. This observation may be attributed to the fact that plant development

does not benefit from higher temperatures if a certain temperature threshold is met or the plant has reached a critical physiological state. McMaster and Smika (1988) compared temperature-dependent and -independent models to predict growth stages in winter wheat and pointed out that the best model for predicting a developmental stage varied depending on the respective stage. Therefore, we decided to concentrate on GDD for SEL and days for all other developmental traits in our analyses.

GWAS

We conducted GWAS for each trait in order to further analyse the above-mentioned correlations between traits and to elucidate which QTLs are responsible for controling trait variation in HEB-25. We applied a multiple linear regression model, including a population main effect and selected markers as cofactors, to account for genetic background and relatedness. This model was recently shown to perform best in HEB-25 (Maurer *et al.*, 2015).

We were able to detect numerous associated genomic regions for all of the traits studied using our GWAS method (Fig. 1 and Supplementary Table S4). A total of 89 QTLs could be defined (Table 4 and Supplementary Table S5). Most QTLs were shared by multiple traits. However, we could also detect trait-specific QTLs for all of the seven traits (Supplementary Fig. S2). We obtained broad variation for wild allele QTLs, with an increase or decrease in

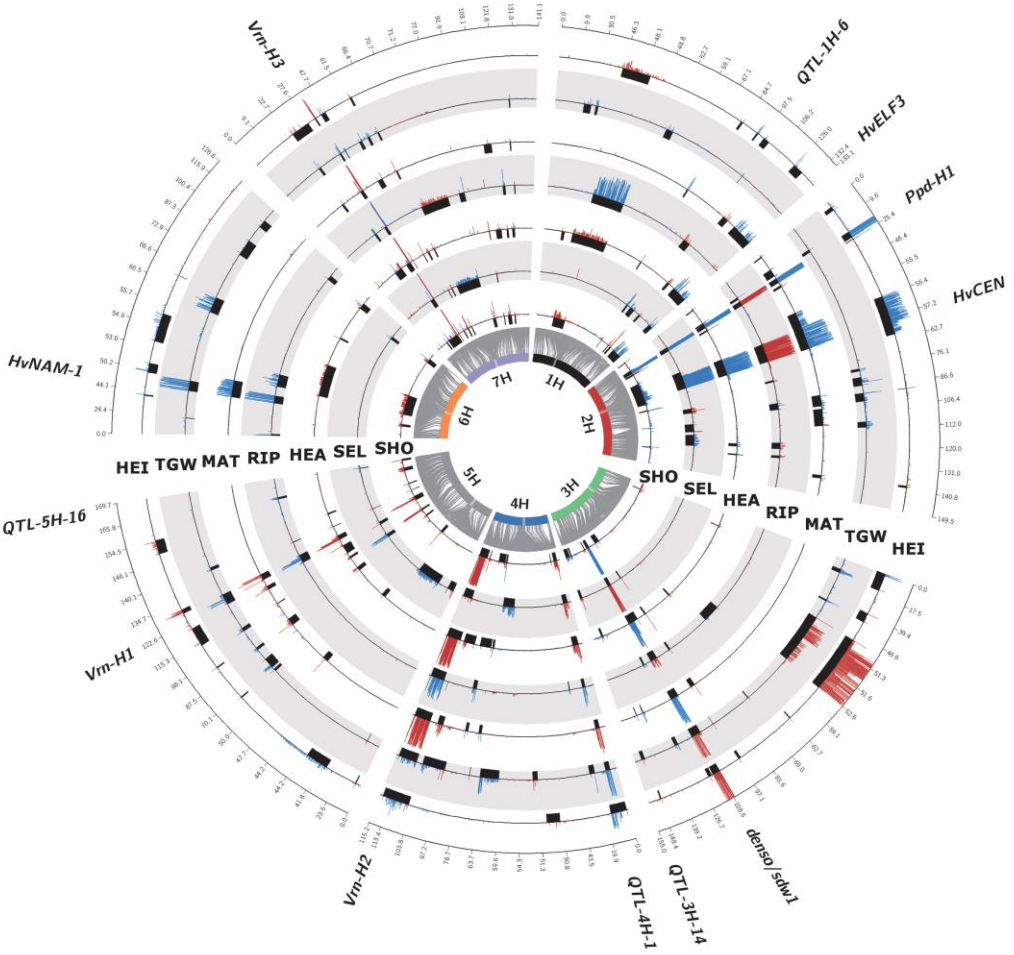

Fig. 1. Comparison of GWAS results across developmental traits, thousand grain weight and plant height. Barley chromosomes are indicated as coloured bars on the inner circle, and centromeres are highlighted as transparent boxes. Grey connector lines represent the genetic position of SNPs on the chromosomes, which is given in centimorgans on the outer circle. Each track represents one trait, and these are (from inside to outside) SHO, SEL, HEA, RIP, MAT, TGW and HEI. Trait abbreviations are given in Table 1. Black boxes indicate the QTL positions. The height of histogram bars above represent the detection rate across 200 repeated random subsamples. The blue and red colours of the bars indicate trait-reducing and trait-increasing effects, respectively, exerted by exotic QTL alleles. Candidate genes of major QTLs are indicated outside the circle.

trait values compared with the Barke control allele. In this regard, we were able to identify 10 QTLs where the wild allele increased TGW by up to 6.6 g (Supplementary Table S5). Modelling all significant markers of one trait simultaneously resulted in explained percentages of phenotypic variance (R^2_{adj}) ranging from 63.6% (TGW) to 82.3% (HEI). The fact that R^2_{pred} values, resulting from a prediction of phenotypes of an independent sample, were also comparatively high (Table 4) confirms the robustness of the method and indicates that a large fraction of phenotypic variation can be explained by genotypic variation.

Below, we present eight common QTLs (starting with chromosome 1H in ascending order) and discuss their relevance in controlling the five developmental traits SHO, SEL, HEA, RIP, and MAT. Then, in order to draw conclusions about their potential QTL function, we study their effects on HEI and TGW. Since HEB-25 enables high genetic precision in mapping of QTLs (Maurer *et al.*, 2015), we obtained strong positional coincidence with plausible candidate genes (Supplementary Table S1). We therefore directly refer to these candidate genes in the headings of the subsections. The results of these eight major QTLs are summarized in Table 5 and illustrated in Fig. 2. Finally, we discuss the developmental phase-specific QTLs that were obtained in this study.

QTL-1H-10 (HvELF3)

We observed a QTL that showed significant effects for all seven traits studied except for RIP and TGW. The QTL is located close to the telomere of chromosome arm 1HL (QTL-1H-10). The exotic allele at this locus is associated with a slight acceleration of plant development. The time to reach each stage was shortened by 2–3 days in contrast to the allele of the spring barley cultivar Barke. This QTL region harbours the earliness-inducing *eam8*/*mat-a* locus (Franckowiak *et al.*, 1997) and may correspond to *HvELF3* (*EARLY FLOWERING3*), which is orthologous to the Arabidopsis circadian clock gene *ELF3* (Faure *et al.*, 2012; Zakhrabekova *et al.*, 2012). *HvELF3* was recently shown to influence flowering by regulating GA production in barley (Boden *et al.*, 2014).

Table 4. *Number of QTLs and total explained phenotypic variance*

Trait[a]	QTLs (*n*)[b]	R^2_{adj} (%)[c]	R^2_{pred} (%)[d]
SHO	49	81.6	59.9
SEL	28	65.9	42.5
HEA	43	79.0	55.5
RIP	30	66.6	37.5
MAT	32	76.9	56.2
TGW	43	63.6	42.8
HEI	37	82.3	45.7
No. of unique QTLs	89		

[a] Trait abbreviations are given in Table 1.
[b] Number of QTLs detected for the respective trait.
[c] Mean explained phenotypic variance by GWAS.
[d] Mean ability to predict phenotypes of an independent sample.

QTL-2H-4 (Ppd-H1)

QTL-2H-4 exerted significant effects on all developmental traits and HEI, thereby explaining up to 34% of the phenotypic variance (R^2). The most significant SNP marker (i_BK_12) is directly located within the *Ppd-H1* gene, which is the main determinant of response to long day conditions in barley (Turner *et al.*, 2005). Most wild barley accessions were shown to carry a highly responsive *Ppd-H1* allele, accelerating development under long day conditions (Cockram *et al.*, 2011). In contrast, Barke, like most spring barley from Northern Europe, carries an allele that exhibits a reduced response to long days. In our study, the wild allele led to a strong acceleration of all developmental phases by up to 9.3 days, except for RIP, which was delayed by up to 3 days. In contrast, HEI was reduced by 7.3 cm when the wild *Ppd-H1* allele was present. After reaching the reproductive stage, relatively more energy is put into reproductive organs. This leads to reduced vegetative growth. Thus, the wild allele at *Ppd-H1* may accelerate floral development at the expense of growth and biomass production.

QTL-2H-7 (HvCEN)

We observed a QTL, located next to the centromere of chromosome 2H, that showed significant effects on all traits except TGW. Comadran *et al.* (2012) identified *HvCEN* (*CENTRORADIALIS*), a homologue of Arabidopsis *TFL1* (*TERMINAL FLOWER1*), as the possible gene behind this locus. Recently, Loscos *et al.* (2014) found evidence that *HvCEN* plays a central role in the induction of flowering in barley. The effects of *HvCEN* on all developmental phases were similar to the effects of *Ppd-H1*. However, the scale of the effects was clearly reduced. At this point we should note that the *HvCEN* effect is presumably not due to genetic linkage to *Ppd-H1*, since *Ppd-H1* was selected as a cofactor in the GWAS model for calculating the *HvCEN* association.

QTL-3H-9 (denso/sdw1)

QTL-3H-9 was detected for all five developmental traits. This QTL may correspond to the *denso*/*sdw1* locus, causing a semi-dwarf phenotype. *HvGA20ox2*, coding a GA-20 oxidase enzyme, is a candidate gene for explaining its function (Jia *et al.*, 2009; Jia *et al.*, 2011). The *denso* allele is commonly present in modern European malting barley cultivars like Barke (Jia *et al.*, 2009). Therefore, it is not surprising that we also observe a strong effect on HEI for this locus, explaining 41% of phenotypic variance (R^2). The presence of the wild allele increased HEI by up to 12.3 cm. This QTL simultaneously affected all developmental traits and TGW. The wild allele reduced the time required to reach SHO, HEA and MAT by 5.7, 4.3 and 4.0 days, respectively. At the same time it delayed the time between these stages (SEL and RIP). Furthermore, the wild allele increased TGW by up to 4.5 g. As the absolute effects steadily diminished throughout the developmental stages (Table 5), we presume that *denso*/*sdw1* plays a major role in the very early stages of development. All in all, the influence of *denso*/*sdw1* on all investigated traits underlines

Table 5. *Major developmental QTLs and their impact on further traits*

QTL	Chr[a]	Peak marker[b]	cM interval[c] From	cM interval[c] Until	SHO[d]	SEL	HEA	RIP	MAT	TGW	HEI	Candidate gene/locus with reference	
QTL-1H-10	1H	i_SCRI_RS_150786	128.0	133.1	**–1.9**	**–10.2**	**–2.3**		**–3.4**		**–3.2**	*HvELF3*	(Faure *et al.*, 2012; Zakhrabekova *et al.*, 2012)
QTL-2H-4	2H	i_BK_12	22.2	23.8	**-7.4**	**–46.7**	**–9.3**	2.8	**–6.8**		**–7.3**	*Ppd-H1*	(Turner *et al.*, 2005)
QTL-2H-7	2H	i_12_30265	53.3	60.8	**–1.9**	**–39.8**	**–3.6**	1.9	**–2.6**		**–4.7**	*HvCEN*	(Comadran *et al.*, 2012)
QTL-3H-9	3H	i_11_11172	103.8	109.8	**–5.7**	44.9	**–4.3**	0.9	**–4.0**	4.5	12.3	*denso/sdw1*	(Jia *et al.*, 2009)
QTL-4H-1	4H	i_12_31458	0.6	14.9	2.0	15.4	2.5	**–0.7**	2.0	**–3.5**	**–2.6**		
QTL-4H-9	4H	i_SCRI_RS_216897	110.2	114.3	7.8	11.3	8.2	**–3.1**	5.5	**–1.5**	4.2	*Vrn-H2*	(Yan *et al.*, 2004)
QTL-5H-10	5H	i_11_10783	122.4	128.5	7.9		8.5	**–3.5**	5.2	6.6		*Vrn-H1*	(Yan *et al.*, 2003)
QTL-7H-3	7H	i_12_30895	29.8	34.3	3.4	36.8	5.7	**–2.1**	3.5		2.5	*Vrn-H3*	(Yan *et al.*, 2006)

Negative values are indicated in bold. Blank cells indicate that the respective QTL was not detected for the trait. Trait abbreviations are given in Table 1. For a complete table of all QTLs, see Table S5.
[a] Chromosome on which the QTL was detected.
[b] iSelect name of marker with the highest significance for HEA.
[c] Genetic interval (cM) with lower and upper threshold of QTL, based on the map of Maurer *et al.* (2015).
[d] Most extreme effect (absolute difference of homozygous wild genotype and homozygous cultivated genotype) of all significant SNPs in respective QTL interval.

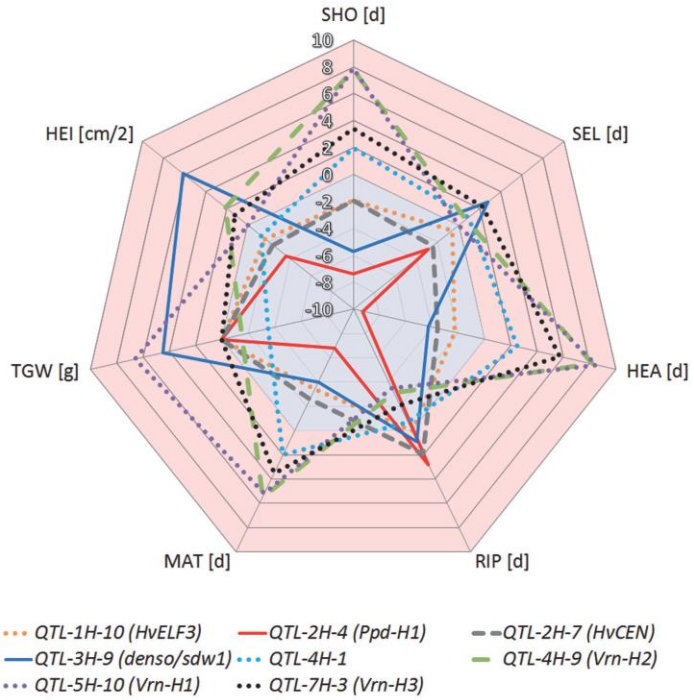

Fig. 2. Spider diagram of major QTL effects across traits. Different traits are represented by corner points of the net. Trait abbreviations are given in Table 1. Effects of wild alleles are indicated by differentially shaped lines for the respective QTL. The blue-shaded and red-shaded backgrounds of the spider net indicate trait-reducing and trait-increasing effects, respectively, exerted by exotic QTL alleles. To enable a comparison of traits within the same scale, values of SEL have been divided by 16.3, which represents the equivalent of one GDD to one day during SEL. Values of HEI have been divided by 2.

its complex and important role in plant physiology throughout a plant's life cycle—a known feature of GA (Taiz and Zeiger, 2007). In addition to the importance of GA in floral regulatory networks (Mutasa-Göttgens and Hedden, 2009), elevated levels of GA also play a role in the delay of senescence, as shown for pea (Wang *et al.*, 2007). Furthermore, GA levels were shown to be crucial for endosperm differentiation and for the growth of barley grains (Weier *et al.*, 2014).

QTL-4H-1

QTL-4H-1 is another ubiquitous QTL that has recently been identified as being a major QTL for flowering time in barley (Maurer *et al.*, 2015). The position of the most significant peak markers differs between 3.5 and 14.9 cM depending on the investigated trait. Nevertheless we assume that both peaks correspond to the same QTL due to a lack of marker

density in this region. This QTL has also been found for TGW and HEI, underscoring its crucial role in plant physiology. This QTL delayed all developmental stages by approximately 2 days, except for RIP, which was accelerated by approximately 2 days. Simultaneously, TGW was reduced by 3.5 g and HEI was reduced by 2.6 cm. To date, no clear candidate gene for this QTL has been found. However, after aligning the sequences of significant markers in this region against sequence-enriched genetic and physical frameworks via BARLEYMAP (Cantalapiedra *et al.*, 2015) a *CCT* (*CONSTANS*, *CO*-like, and *TOC1*) domain gene was identified in this region (Genbank accession number AK354746). Many known flowering time regulators have *CCT* domains (Cockram *et al.*, 2012). Furthermore, the *LOG* (*LONELY GUY*) gene (MLOC_45038.2) could be assigned to this region. *LOG* encodes a cytokinin-activating enzyme that is required to maintain meristem activity. Its loss of function causes pre-mature termination of the shoot meristem in rice development (Kurakawa *et al.*, 2007).

*QTL-4H-9 (*Vrn-H2*)*

QTL-4H-9, which may harbour *Vrn-H2*, the main repressor of flowering in vernalization-dependent barley (Yan *et al.*, 2004), was reliably detected for SHO, HEA, RIP and MAT. The wild allele delayed the respective developmental stages by up to 8 days, whereby RIP was shortened by up to 3 days. However, the high variation of marker effects among significant SNPs indicates that effects vary greatly depending on the segregating families of the respective SNPs (Supplementary Table S4). This locus is naturally deleted in spring barley cultivars like Barke. The result is that no vernalization is required to induce flowering (von Zitzewitz *et al.*, 2005). This explains why the alleles from wild barley, which are predominantly winter types (Cockram *et al.*, 2011), delayed flowering in our spring-sown field trials. The fact that this QTL was infrequently detected for the traits SEL, TGW and HEI leads us to assume that the impact of *Vrn-H2* on these traits may be diminished or biased by family-specific effects.

*QTL-5H-10 (*Vrn-H1*)*

This QTL was shared between SHO (+7.9 d), HEA (+8.5 d), RIP (–3.5 d) and MAT (+5.2 d) and may correspond to the *Vrn-H1* locus with *HvBM5A* (*MADS-box 5A*) being a candidate gene (von Zitzewitz *et al.*, 2005). *Vrn-H1* is known to be involved in the vernalization pathway of flowering time regulation by responding to the low temperatures required for vernalization (Oliver *et al.*, 2013). *HvPhyC* (*PHYTOCHROME C*) is another gene that affects flowering time (Nishida *et al.*, 2013; Pankin *et al.*, 2014). It is closely linked to *HvBM5A*, which makes it difficult to distinguish between their effects. In addition, a QTL for HEI (QTL-5H-11) has been defined approximately 3 cM distal from the position of *HvBM5A* and *HvPhyC*. This further complicates the interpretation of the QTL effects in this genomic region. Again, as already shown for *Vrn-H2*, the high variation of marker effects points to family-specific effects for this locus. An extraordinarily high variation exists for the trait TGW, which is caused by a single SNP marker (i_11_11090) that is only polymorphic in HEB-F08. The wild allele caused an estimated increase in TGW by 6.6 g, whereas all other significant markers in the *Vrn-H1* region reduced TGW by 1.5–1.8 g. Interestingly, there is low variation for the locus affecting HEI, located 3 cM distal from *Vrn-H1*. This indicates that this locus might be independent of *Vrn-H1*.

*QTL-7H-3 (*Vrn-H3*)*

The most significant SNP marker at this QTL (i_12_30895) is directly located within *Vrn-H3*. The *Vrn-H3* locus in barley has been shown to correspond to *HvFT1*, the orthologue of Arabidopsis *FT* (Yan *et al.*, 2006; Faure *et al.*, 2007). This gene plays a central role in the flowering pathways as it is involved in the switch from vegetative to reproductive growth under long day conditions (Turck *et al.*, 2008). In our population, significant effects of this QTL were observed for all developmental traits and HEI. On average, the wild allele in the *Vrn-H3* region delayed all developmental phases by 3–6 days, except RIP, which was shortened by approximately 2 days. However, ample variation in marker effects exists for all developmental traits (Supplementary Table S4). For instance we observed effects for SHO ranging from –3.3 to +3.4 days depending on the families in which the respective marker segregates. This clearly shows the presence of family-specific effects, most likely due to functionally different haplotypes of *Vrn-H3* among the HEB-25 donor accessions. However, for the trait HEI (increased by 2.5 cm) the effect of the wild allele was relatively constant. Interestingly, gene-specific markers of *Vrn-H3* did not impact TGW. Instead, the QTL for TGW was defined at around 4 cM proximal of *Vrn-H3* at 38.8 cM (QTL-7H-4) causing a reduction of TGW by up to 2.2 g.

Developmental phase-specific QTLs

In addition to the majority of QTLs that were shared across developmental traits, we also found certain QTLs that impacted a specific stage of plant development. These QTLs are a potential source for fine-tuning the phases of plant development and unravelling the physiological pathways.

QTLs affecting early development

QTLs that affect early development are characterized by their influence on SHO and SEL. SEL is strongly related to the phase from awn primordium to tipping, which was shown to be the most decisive developmental phase for spikelet survival (Alqudah and Schnurbusch, 2014) and leaf growth rate (Alqudah and Schnurbusch, 2015) in barley. The length of this phase is therefore assumed to play a key role in determining a plant's yield potential.

We found three striking QTLs (QTL-1H-6, QTL-3H-14 and QTL-5H-16) that exhibited exclusive effects on SHO and SEL. In the case of QTL-3H-14, SHO was accelerated by approximately 2 days and TGW increased by 1.3 g. In the case of QTL-1H-6 and QTL-5H-16, SHO was delayed by 1–2 days

without any effect on TGW. These QTLs were not shown to affect HEA since the effect on SEL was contrary to the effect on SHO. This therefore compensated for the effect of SHO in the subsequent developmental phases. Thus, we assume that these QTLs play a role in very early plant development by affecting the time to reach SHO and the duration of SEL.

We also observed QTLs specific to the SEL trait. For instance, QTL-2H-10 extended SEL by 19.7 GDD, while QTL-4H-5, QTL-5H-2 and QTL-5H-7 shortened SEL by 12.2, 12.1 and 18.0 GDD, respectively. Interestingly, some of the above-mentioned QTLs also affected TGW and HEI in addition to affecting SEL. In the case of QTL-5H-2, the shortened period of SEL correlates with reduced HEI. Although QTL-4H-5 and QTL-2H-10 show contrasting effects on SEL, they co-localized with QTLs that decrease TGW. In the case of QTL-2H-10 this effect is likely caused by the six-rowed spike locus *Vrs1* (Komatsuda *et al.*, 2007) originating from the *Hordeum vulgare* ssp. *agriocrithon* donor of HEB-F24 in our population. Six-rowed spikes generally show decreased TGW due to a higher grain number. This causes single grains to compete for assimilates. Interestingly, *Vrs1* also seems to extend SEL. This corroborates the observation that six-rowed barley has a higher leaf area and leaf dry weight than two-rowed barley, since leaf biomass is mainly produced during SEL (Alqudah and Schnurbusch, 2015).

QTLs affecting late development

QTLs that affect late development are thought to influence RIP and MAT. Following Davies and Gan (2012), whole plant senescence is initiated through a physiological transition at flowering. This leads to a remobilization of nutrients to the developing seeds (Distelfeld *et al.*, 2014). In the present study, RIP represents the period of whole plant senescence and grain-filling. Thus, it is considered to be highly important in determining yield, particularly as a result of the grain weight component (Distelfeld *et al.*, 2014). Egli (2004) also emphasized that the duration of this period played an important role in grain crop yields.

The most obvious QTL to specifically affect late development was detected in the centromeric region of chromosome 6H (QTL-6H-4). In addition to affecting RIP and MAT, this QTL also significantly affected TGW and HEI. *HvNAM-1* (*NO APICAL MERISTEM-1*), coding for an NAC transcription factor (Distelfeld *et al.*, 2008), is a promising candidate for this locus. It plays an important role in inducing senescence (Distelfeld *et al.*, 2014). Moreover, it has been shown that its wheat homologue *NAM-B1* influences the remobilization process of nitrogen, iron and zinc in the developing grain (Uauy *et al.*, 2006; Waters *et al.*, 2009). In our study the wild allele accelerated senescence, which shortened RIP by approximately 1 day and consequently led to earlier MAT (–3.4 days). Simultaneously, TGW was reduced by 2.6 g and HEI was reduced by 4.5 cm. Interestingly, there is a second QTL for RIP (QTL-6H-5) that is closely linked to *HvNAM-1*. It mainly affects early stages by delaying SHO and HEA and shortening RIP. This behaviour points to the candidate gene *HvGR-RBP1* (*GLYCINE-RICH RNA BINDING PROTEIN*), which is consistent with the findings of other groups (Jukanti *et al.*, 2008; Streitner *et al.*, 2008; Mason *et al.*, 2014). This is the first time in barley genomics that we have been able to clearly distinguish this adjacent locus from *HvNAM-1*.

Comparison of shared QTLs

In previous sections we mentioned many QTLs that simultaneously affected several developmental traits. Comparing their effects across traits provides insights into the different modes of action of the respective QTLs (Fig. 1 and Table 5).

In general, most QTLs affecting HEA also have a similar effect on SHO. This is also reflected by the high correlation of r=0.92 (Table 3). However, comparing the absolute effects on both stages, we can state that the effect on HEA is generally more pronounced. This is reflected by an increased SEL and indicates that the effect may accumulate over time. However, the *denso/sdw1* locus is a striking exception to this rule. Although both time to SHO and HEA are reduced, it simultaneously extends the time needed for SEL. This indicates that the effect on SHO is stronger than on HEA at *denso/sdw1*. We therefore assume that *denso/sdw1* has the biggest impact in the very early developmental stages and that its influence decreases during subsequent developmental phases.

Almost half of all developmental QTLs in our study simultaneously showed significant effects on HEI. However, the direction in which HEI was influenced differed between QTLs, even if they shared the same tendencies with regard to developmental traits. This is interesting when it comes to separating general developmental effects from purely reproduction-promoting effects. In most cases, HEI increased as the duration of SEL increased and vice versa. This is already indicated by the positive correlation of both traits (r=0.45). However, in the case of QTL-4H-1 and *Vrn-H2*, the prolonged duration of SEL is accompanied by a reduction in HEI. Since these QTLs retard development and simultaneously reduce HEI, we assume that their gene products may generally hamper plant development.

Relevance of developmental subphases with regard to TGW

We observed numerous QTLs that were significantly associated with the developmental subphases SEL and RIP. We observed a general influence on TGW for the duration of grain-filling, i.e. RIP; as RIP increased, TGW increased as well. However, we could not find any direct correlations between SEL and TGW. There were hints that the interplay between RIP and SEL may be of interest. In general, we observed a negative correlation between both traits with r=–0.6. Consequently, a shortened SEL would cause an extended RIP. This would then cause a longer period of grain-filling and thus increase TGW. The time of physiological maturity (MAT) is largely predetermined by environmental factors, as discussed above. Therefore, increasing the duration of RIP leads to earlier flowering, and very likely also to a shorter period of SEL. Thus, if the duration of RIP is extended to increase TGW, this will result in a reduction in

SEL. This lowers the amount of biomass that would be able to supply the developing grains with assimilates (Alqudah and Schnurbusch, 2015). Normally this reduction would not have serious consequences since plants produce excess vegetative mass (Egli, 1998). However, if the effect is severe (e.g. in the case of *Ppd-H1* in our study), the beneficial effect on TGW may be compensated. We therefore presume that increasing TGW is based on a trade-off between an extended RIP and a shortened SEL. However, one has to keep in mind that yield formation is based on a complex interplay between several yield components. To scrutinize our findings, we therefore suggest studying all yield components simultaneously, i.e. tiller number, number of grains per spike and grain weight, as well as total grain yield, in future HEB-25 field experiments. In addition, spike photosynthesis may also require analysis, since its role in grain-filling has recently been highlighted (Sanchez-Bragado *et al.*, 2014).

Beneficial wild germplasm present in HEB-25

One major advantage of using HEB-25 is the ability to directly evaluate the value of wild barley alleles in a cultivated background. Thus, we were able to identify 10 QTLs where the wild allele increased TGW by up to 6.6 g compared with the elite Barke allele. These wild QTL alleles are promising candidates for improving future barley grain weight once they are introduced into barley breeding programmes. However, the optimum strategy of adjusting plant developmental subphases to increase grain weight may vary depending on the ecogeographic region in which barley is grown. As latitudes increase, alleles conferring a decelerated plant development would be favourable for making use of the longer and cooler season for grain-filling. For barley cultivated in lower latitudes, early flowering would be favourable, for instance, to escape early season terminal drought (Comadran *et al.*, 2012). Therefore, it is currently not possible to make a general statement about the usefulness of specific plant development QTL effects. However, as we see that there is tremendous effect variation for every trait, we conclude that HEB-25 harbours a multitude of beneficial plant development QTL alleles for different ecogeographic regions.

In this study, the most prominent example for putative advantageous effects of wild alleles is the *denso/sdw1* QTL on chromosome 3H. Barke carries the dwarfing *denso* allele, which was introduced in barley breeding programmes during the 'Green Revolution'. It reduces plant height and improves resistance to lodging, whereby also yield is increased (Jia *et al.*, 2009). The wild allele extends SEL and RIP, and increases TGW, compared with the cultivated Barke allele. Unfortunately, these beneficial effects are accompanied by the unfavourable effect of increasing HEI. However, we observed lines (e.g. HEB_25_050) in HEB-25 that carry the wild allele at the *denso/sdw1* locus and nevertheless demonstrate high agronomic performance. This is most likely due to introgression of additional wild alleles, which appear to compensate for the effect of increasing HEI (Supplementary Table S6). We thus encourage breeders to integrate wild germplasm into their breeding programmes. This way, the elite barley gene pool can be replenished with new favourable alleles to overcome future agricultural challenges.

Supplementary data

Supplementary data are available at *JXB* online.

Figure S1. Frequency distributions of BLUEs for all investigated traits, plotted as histograms.

Figure S2. Venn diagrams indicating the number of shared QTLs across traits.

Table S1. Results of BARLEYMAP alignments to detect QTL candidate genes.

Table S2. Descriptive statistics of all investigated traits, grouped by years.

Table S3. Heritabilities (h^2) of all investigated traits, including the comparison of days and GDD for developmental traits.

Table S4. Tabular overview of all results gathered from GWAS, along with family-specific effect estimates for each SNP.

Table S5. Overview of all QTLs and their effects, with position, estimated effect on the different traits, and plausible candidate genes.

Table S6. Raw phenotype data and BLUEs of the investigated traits across years and blocks, along with information about the allelic state at the *denso/sdw1* locus.

Acknowledgements

This work was supported by the German Research Foundation (DFG) via the priority program 1530: Flowering time control—from natural variation to crop improvement (Grant Pi339/7-1), via ERA-CAPS (Grant Pi339/8-1) and via the Interdisciplinary Centre for Crop Plant Research (IZN), Halle. We are grateful to Roswitha Ende, Jana Müglitz, Diana Rarisch, Helga Sängerlaub, Brigitte Schröder, Bernd Kollmorgen, Markus Hinz and various student assistants for excellent technical help and to TraitGenetics GmbH, Gatersleben, Germany, for genotyping HEB-25 with the Infinium iSelect 9k SNP chip.

References

Alonso-Peral MM, Oliver SN, Casao MC, Greenup AA, Trevaskis B. 2011. The promoter of the cereal VERNALIZATION1 gene is sufficient for transcriptional induction by prolonged cold. PLoS ONE **6,** e29456.

Alqudah AM, Schnurbusch T. 2014. Awn primordium to tipping is the most decisive developmental phase for spikelet survival in barley. Functional Plant Biology **41,** 424–436.

Alqudah AM, Schnurbusch T. 2015. Barley leaf area and leaf growth rates are maximized during the pre-anthesis phase. Agronomy **5,** 107–129.

Alqudah AM, Sharma R, Pasam RK, Graner A, Kilian B, Schnurbusch T. 2014. Genetic dissection of photoperiod response based on GWAS of pre-anthesis phase duration in spring barley. PLoS ONE **9,** e113120.

Atkin OK, Tjoelker MG. 2003. Thermal acclimation and the dynamic response of plant respiration to temperature. Trends in Plant Science **8,** 343–351.

Atwell BJ, Kriedemann PE, Turnbull CGN. 1999. *Plants in action: adaptation in nature, performance in cultivation*. South Yarra, Victoria: Macmillan Education Australia.

Blümel M, Dally N, Jung C. 2015. Flowering time regulation in crops—what did we learn from Arabidopsis? Current Opinion in Biotechnology **32,** 121–129.

Boden SA, Weiss D, Ross JJ, Davies NW, Trevaskis B, Chandler PM, Swain SM. 2014. EARLY FLOWERING3 regulates flowering in spring barley by mediating gibberellin production and FLOWERING LOCUS T expression. *The Plant Cell* **26,** 1557–1569.

Borras-Gelonch G, Denti M, Thomas WTB, Romagosa I. 2012. Genetic control of pre-heading phases in the Steptoe x Morex barley population under different conditions of photoperiod and temperature. Euphytica **183,** 303–321.

Borras-Gelonch G, Slafer GA, Casas AM, van Eeuwijk F, Romagosa I. 2010. Genetic control of pre-heading phases and other traits related to development in a double-haploid barley (Hordeum vulgare L.) population. Field Crops Research **119,** 36–47.

Borras G, Romagosa I, van Eeuwijk F, Slafer GA. 2009. Genetic variability in duration of pre-heading phases and relationships with leaf appearance and tillering dynamics in a barley population. Field Crops Research **113,** 95–104.

Buckler ES, Holland JB, Bradbury PJ, *et al*. 2009. The genetic architecture of maize flowering time. Science **325,** 714–718.

Calixto CPG, Waugh R, Brown JWS. 2015. Evolutionary relationships among barley and Arabidopsis core circadian clock and clock-associated genes. Journal of Molecular Evolution **80,** 108–119.

Campoli C, Pankin A, Drosse B, Casao CM, Davis SJ, Korff M. 2013. HvLUX1 is a candidate gene underlying the early maturity 10 locus in barley: phylogeny, diversity, and interactions with the circadian clock and photoperiodic pathways. New Phytologist **199,** 1045–1059.

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

Cantalapiedra CP, Boudiar R, Casas AM, 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.

Chailakhyan MK. 1937. *Gormonal'naya teoriya razvitiya rastenii* (Hormonal theory of plant development) . Moscow: Akademii Nauk SSSR.

Cockram J, Jones H, O'Sullivan DM. 2011. Genetic variation at flowering time loci in wild and cultivated barley. Plant Genetic Resources: Characterization and Utilization **9,** 264–267.

Cockram J, Thiel T, Steuernagel B, Stein N, Taudien S, Bailey PC, O'Sullivan DM. 2012. Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. PLoS ONE **7,** e45307.

Comadran J, Kilian B, Russell J, *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.

Davies P, Gan S. 2012. Towards an integrated view of monocarpic plant senescence. Russian Journal of Plant Physiology **59,** 467–478.

Deng W, Casao MC, Wang P, Sato K, Hayes PM, Finnegan EJ, Trevaskis B. 2015. Direct links between the vernalization response and other key traits of cereal crops. Nature Communications **6,** 5882.

Distelfeld A, Avni R, Fischer AM. 2014. Senescence, nutrient remobilization, and yield in wheat and barley. Journal of Experimental Botany **65,** 3783–3798.

Distelfeld A, Korol A, Dubcovsky J, Uauy C, Blake T, Fahima T. 2008. Colinearity between the barley grain protein content (GPC) QTL on chromosome arm 6HS and the wheat Gpc-B1 region. Molecular Breeding **22,** 25–38.

Draper N, Smith H. 1981. *Applied regression analysis* . New York: John Wiley.

Egli DB. 1998. *Seed biology and the yield of grain crops* . Wallingford, Oxon: CAB International.

Egli DB. 2004. Seed-fill duration and yield of grain crops. Advances in Agronomy **83,** 243–279.

Evans LT, Wardlaw IF. 1976. Aspects of the comparative physilogy of grain yield in cereals. Advances in Agronomy **28,** 301–359.

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

Faure S, Turner AS, Gruszka D, Christodoulou V, Davis SJ, von Korff M, Laurie DA. 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,** 8328–8333.

Franckowiak J, Lundqvist U, Konishi T, Gallagher L. 1997. Description of stock number: BGS 214. Barley Genetics Newsletter **26,** 213–215.

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

International Barley Genome Sequencing Consortium. 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature **491,** 711–716.

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

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

Jordan D, Mace E, Cruickshank A, Hunt C, Henzell R. 2011. Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. Crop Science **51,** 1444–1457.

Jukanti AK, Heidlebaugh NM, Parrott DL, Fischer IA, McInnerney K, Fischer AM. 2008. Comparative transcriptome profiling of near-isogenic barley (Hordeum vulgare) lines differing in the allelic state of a major grain protein content locus identifies genes with possible roles in leaf senescence and nitrogen reallocation. New Phytologist **177,** 333–349.

Kernich GC, Halloran GM, Flood RG. 1997. Variation in duration of pre-anthesis phases of development in barley (Hordeum vulgare). Australian Journal of Agricultural Research **48,** 59–66.

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

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A. 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 United States of America **104,** 1424–1429.

Kurakawa T, Ueda N, Maekawa M, Kobayashi K, Kojima M, Nagato Y, Sakakibara H, Kyozuka J. 2007. Direct control of shoot meristem activity by a cytokinin-activating enzyme. Nature **445,** 652–655.

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

Laurie D, Pratchett N, Snape J, Bezant J. 1995. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (Hordeum vulgare L.) cross. Genome **38,** 575–585.

Liu W, Gowda M, Steinhoff J, Maurer HP, Würschum T, Longin CFH, Cossic F, Reif JC. 2011. Association mapping in an elite maize breeding population. Theoretical and Applied Genetics **123,** 847–858.

Loscos J, Igartua E, Contreras-Moreira B, Gracia MP, Casas AM. 2014. HvFT1 polymorphism and effect – survey of barley germplasm and expression analysis. Frontiers in Plant Science **5,** 251.

Mascher M, Muehlbauer GJ, Rokhsar DS, *et al*. 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). The Plant Journal **76,** 718–727.

Mason KE, Tripet BP, Parrott D, Fischer AM, Copié V. 2014. 1H, 13C, 15N backbone and side chain NMR resonance assignments for the N-terminal RNA recognition motif of the HvGR-RBP1 protein involved in the regulation of barley (Hordeum vulgare L.) senescence. Biomolecular NMR Assignments **8,** 149–153.

Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K. 2015. Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics **16,** 290.

McMaster GS, Smika DE. 1988. Estimation and evaluation of winter wheat phenology in the central Great Plains. Agricultural and Forest Meteorology **43,** 1–18.

McMaster GS, Wilhelm W. 1997. Growing degree-days: one equation, two interpretations. Agricultural and Forest Meteorology **87,** 291–300.

Miralles DJ, Richards RA, Slafer GA. 2000. Duration of the stem elongation period influences the number of fertile florets in wheat and barley. Functional Plant Biology **27,** 931–940.

Mrízová K, Jiskrová E, Vyroubalová Š, Novák O, Ohnoutková L, Pospíšilová H, Frébort I, Harwood WA, Galuszka P. 2013. Overexpression of cytokinin dehydrogenase genes in barley (Hordeum vulgare cv. Golden Promise) fundamentally affects morphology and fertility. PLoS ONE **8,** e79029.

Mutasa-Göttgens E, Hedden P. 2009. Gibberellin as a factor in floral regulatory networks. Journal of Experimental Botany **60,** 1979–1989.

Nishida H, Ishihara D, Ishii M, Kaneko T, Kawahigashi H, Akashi Y, Saisho D, Tanaka K, Handa H, Takeda K. 2013. Phytochrome C is a key factor controlling long-day flowering in barley. Plant Physiology **163,** 804–814.

Ogut F, Bian Y, Bradbury P, Holland J. 2015. Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. Heredity **114,** 1–12.

Oliver SN, Deng W, Casao MC, Trevaskis B. 2013. Low temperatures induce rapid changes in chromatin state and transcript levels of the cereal VERNALIZATION1 gene. Journal of Experimental Botany **64,** 2413–2422.

Pankin A, Campoli C, Dong X, Kilian B, Sharma R, Himmelbach A, Saini R, Davis SJ, Stein N, Schneeberger K. 2014. Mapping-by-sequencing identifies HvPHYTOCHROME C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. Genetics **198,** 383–396.

Sanchez-Bragado R, Molero G, Reynolds MP, Araus JL. 2014. Relative contribution of shoot and ear photosynthesis to grain filling in wheat under good agronomical conditions assessed by differential organ δ13C. Journal of Experimental Botany **65,** 5401–5413.

Schwarz G. 1978. Estimating the dimension of a model. The Annals of Statistics **6,** 461–464.

Sreenivasulu N, Schnurbusch T. 2012. A genetic playground for enhancing grain number in cereals. Trends in Plant Science **17,** 91–101.

Streitner C, Danisman S, Wehrle F, Schöning JC, Alfano JR, Staiger D. 2008. The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in Arabidopsis thaliana. The Plant Journal **56,** 239–250.

Taiz L, Zeiger E. 2007. *Plant physiology* . Heidelberg: Spektrum Akademischer Verlag/Springer.

Turck F, Fornara F, Coupland G. 2008. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annual Review of Plant Biology **59,** 573–594.

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

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. 2006. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science **314,** 1298–1301.

von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, Casas A, Chen TH, Hayes PM, Skinner JS. 2005. Molecular and structural characterization of barley vernalization genes. Plant Molecular Biology **59,** 449–467.

Wang D-Y, Li Q, Cui K-M, Zhu Y-X. 2007. Gibberellin is involved in the regulation of cell death-mediated apical senescence in G2 pea. Journal of Integrative Plant Biology **49,** 1627–1633.

Wang G, Schmalenbach I, von Korff M, Leon J, Kilian B, Rode J, Pillen K. 2010. Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC(2)DH population and a set of wild barley introgression lines. Theoretical and Applied Genetics **120,** 1559–1574.

Waters BM, Uauy C, Dubcovsky J, Grusak MA. 2009. Wheat (Triticum aestivum) NAM proteins regulate the translocation of iron, zinc, and nitrogen compounds from vegetative tissues to grain. Journal of Experimental Botany **60,** 4263–4274.

Weier D, Thiel J, Kohl S, Tarkowská D, Strnad M, Schaarschmidt S, Weschke W, Weber H, Hause B. 2014. Gibberellin-to-abscisic acid balances govern development and differentiation of the nucellar projection of barley grains. Journal of Experimental Botany **65,** 5291–5304.

Würschum T, Liu W, Gowda M, Maurer H, Fischer S, Schechert A, Reif J. 2012. Comparison of biometrical models for joint linkage association mapping. Heredity **108,** 332–340.

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

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J. 2004. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science **303,** 1640–1644.

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,** 6263–6268.

Zakhrabekova S, Gough SP, Braumann I, *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,** 4326–4331.

Chapter 3

Estimating parent-specific QTL effects through cumulating linked identity-by-state SNP effects in multiparental populations[3]

The models used in Chapter 1 and Chapter 2 mainly focused on the estimation of main effects to detect QTL. However, since HEB-25 is a multi-parental population, theoretically up to 26 different alleles could be present at each QTL. Therefore, the aim of the following study was to compare different methods to estimate donor-specific QTL allele effects. The arising question was whether this issue can be addressed in GWAS without losing statistical power.

[3] Maurer A, Sannemann W, Leon J, Pillen K (2016) Estimating parent-specific QTL effects through cumulating linked identity-by-state SNP effects in multiparental populations. Heredity. doi: 10.1038/hdy.2016.121

parent. Nevertheless, this SNP could have been inherited from another parent sharing the same IBS allele state as the reference parent. In this case they would show different IBD states, although they share the same IBS allele.

Here, we derive IBD from IBS genotype data in a wild barley NAM population. Subsequently, we test whether parent-specific IBD calling of SNPs is superior over classical biallelic IBS calling in GWAS. For this, we apply GWAS to four agronomic traits of increasing complexity (grain colour, grain threshability, flowering time and thousand grain weight). Subsequently, we develop a novel approach to model parent-specific quantitative trait locus (QTL) effects in the wild barley NAM population HEB-25 (Maurer *et al.*, 2015) without the need of IBD information or modelling haplotypes *a priori*. To test the method's accuracy we perform extensive simulation studies with varying trait complexities. We then show that this approach is also suitable to model parent-specific QTL effects in a barley MAGIC population (Sannemann *et al.*, 2015).

MATERIALS AND METHODS

Plant material

The NAM population HEB-25 (Maurer *et al.*, 2015), consisting of 1420 individual BC1S3 lines in 25 wild barley-derived families, was used in this study. HEB-25 is the result of initial crosses between the spring barley cultivar Barke (*Hordeum vulgare* ssp. *vulgare*) and 25 highly divergent exotic wild barley accessions (*H. vulgare* ssp. *spontaneum* and *H. vulgare* ssp. *agriocrithon*), hereafter referred to as donors. F1 plants of the initial crosses were backcrossed with Barke. For detailed information about the population design, see Maurer *et al.* (2015).

The barley MAGIC population consists of 533 doubled haploid lines, created through intermating eight founder genotypes of barley breeding in Germany. For more details about this population, see Sannemann *et al.* (2015).

Phenotype data

Four major agronomic traits were investigated in this study. Phenotype data were collected in field trials in Halle, Germany (51°29′46.47′N; 11°59′41.81′E) during the seasons 2011–2014. Briefly, the complete population was grown in double rows following a randomised complete block design in two replications. For details on the experimental setup, see Maurer *et al.* (2016). Flowering time (heading (HEA)) and thousand grain weight (TGW) data were taken from Maurer *et al.* (2016). Grain threshability (THR) was manually scored on a scale from 1 (difficult to thresh) to 9 (easy to thresh), after threshing mature spikes using a home-made rotating threshing drum. In addition, data on grain colour (GrCol), manually scored as 1 (light) or 9 (dark) after visual assessment, were scored. These traits were selected because we assumed that they are controlled by few (GrCol, THR) or many (HEA, TGW) genes.

Genotype data

All 1420 BC1S3 lines and their corresponding parents were genotyped with the barley Infinium iSelect 9K chip (Illumina Inc., San Diego, CA, USA) (Maurer *et al.*, 2015), consisting of 7864 SNP markers as reported in Comadran *et al.* (2012). SNP markers that did not meet the quality criteria (polymorphic in at least one HEB family, $<10\%$ failure rate, $<12.5\%$ heterozygous calls) were removed from the data set. A total of 305 markers were removed as they revealed the exact segregation among all HEB lines as a twin marker, indicating that they were in complete linkage disequilibrium (LD). Only one of the twin markers was kept, resulting in a total set of 5398 remaining markers. Out of these markers, 4861 segregated in less than 25 families, and 448 thereof segregated only in a single family.

Defining IBS and IBD matrices

Polymorphic SNP alleles originating from Barke or the wild barley donors of the NAM population HEB-25 are easily distinguishable by state. Based on the Barke reference genotype, the wild barley allele can be specified in each segregating family. To set up the IBS matrix the state of the homozygous Barke allele was coded as 0, whereas HEB lines that showed a homozygous wild barley genotype were assigned a value of 2. Consequently, heterozygous HEB lines were assigned a value of 1. If a SNP was monomorphic in one HEB family but polymorphic in a second family, lines of the first HEB family were assigned a genotype value of 0, as their state is not different from the Barke allele. Gaps resulting from missing genotypes (0.6% of all data points) were filled with the mean of polymorphic flanking markers, based on the map of Maurer *et al.* (2015). This way a complete genotype data set (IBS) was retained that is required to carry out the following multiple regression GWAS.

To convert the IBS matrix into an IBD matrix, we first replaced each marker value that was monomorphic in a HEB family by an empty value. Then, the resulting gaps (44.9% of all data points) were filled with the mean of the next polymorphic flanking markers of this gap. This way we can distinguish whether the allele is inherited from the recurrent parent Barke or a wild donor across the whole NAM population. The newly assigned IBD value reflects the marker's probability of being inherited from the wild barley donor.

Both matrices are available as Supplementary Figure S1 and Additional File S1.

Models used for genome-wide association mapping

We used two different multiple linear regression models to conduct genome-wide association mapping on best linear unbiased estimates of each HEB line trait performance. The best linear unbiased estimates were obtained from a linear mixed model with effects for genotype, environment and interaction of genotype and environment.

Model 'IBS-M' corresponds to Model-A of Liu *et al.* (2011), where SNP markers are included as main effects using the quantitative IBS genotype matrix scores.

$$\text{Model 'IBS-M'}: \quad Y = \mu + \Sigma \text{SNP}_{\text{IBS}} + \varepsilon$$

This model showed the highest predictive power and detected the highest number of QTLs when compared with other joint linkage association mapping models (Würschum *et al.*, 2012). Model 'IBD-M × F' models the SNP markers as interaction effect with the HEB family. This model is based on the quantitative IBD genotype matrix scores.

$$\text{Model 'IBD-M} \times \text{F'}: \quad Y = \mu + \Sigma(\text{SNP}_{\text{IBD}} \times \text{Fam}) + \varepsilon$$

The analyses were carried out with SAS 9.4 Software (SAS Institute Inc., Cary, NC, USA) using Proc GLMSELECT. This procedure selects the best model out of a set of predefined possible factors. In our case, all SNPs were initially defined as possible factors. Significant SNPs were then determined by stepwise forward–backward regression. SNPs were allowed to enter or leave the model at each step until the Schwarz Bayesian criterion (Schwarz, 1978) could not be reduced further. SNPs included in the final model are hereafter referred to as significant SNPs. The total number of significant SNPs included in the final model was recorded. A SNP effect estimate can be interpreted as the allele substitution effect (α) and represents the regression coefficient of the respective SNP in the final model. Note that all significant SNP effect estimates are modelled at the same time in the final model.

Cross-validation

A fivefold cross-validation was run 20 times to increase the robustness of the results. For this, 100 subsets were extracted out of the total phenotypic data. Each subset consisted of 80% randomly chosen HEB lines per family. This set was used as the training set to define significant markers and to estimate their effects, whereas the remaining 20% of lines were used as the validation set. The phenotypes of the validation set lines were predicted based on marker effects estimated in the training set. Prediction ability (R^2val) was then calculated as the squared Pearson product–moment correlation between the observed and predicted phenotypes of the validation set, whereas R^2train represents the model fit of the training set.

To define QTL regions, we calculated a SNP marker's detection rate as the number of times, out of 100, it was included in the final model. Robust major QTLs were defined if they were detected more than 20 times in IBD-M × F.

Cumulating SNPs to estimate parent-specific QTL effects

To estimate a parent-specific QTL effect from model 'IBS-M' we cumulated significant SNP marker effects. First, a peak marker for each expected parent-

specific QTL was selected from model 'IBD-M × F' where the number of model inclusions across all 100 cross-validation runs was maximised. Each peak marker was placed central in a 26 cM interval (that resembles the mean introgression size in HEB-25) to look for significant SNPs in this region. Then, 'IBS-M' SNP effect estimates of all markers within this interval were cumulated for each of the 25 donors, following $\sum_{i=1}^{n} SNP(donor)_i * \alpha_i$, where i iterates through all significant SNPs (n) in the respective QTL interval. $SNP\ (donor)_i$ represents the quantitative IBS donor genotype (that is, 0 vs 2) of the i-th significant SNP and α_i denotes the SNP effect estimate of this SNP obtained from model 'IBS-M'. As SNPs show different IBS segregation patterns across the donors of HEB families (Supplementary Table S1), a different cumulated effect was obtained for each donor. This procedure was conducted within each of the 100 cross-validation runs and the mean of them was taken as the final parent-specific QTL effect estimate. For an illustration of the workflow and an example see Supplementary Figure S2.

Finally, we determined cumulation precision that is the correlation of the cumulated parent-specific effects with the effects obtained from model 'IBD-M × F'. We chose this comparison to check our hypothesis of whether multiple SNPs clustering in specific regions are able to represent parent-specific effects.

To test its general transferability to other multiparental populations, we applied the same approach to a barley MAGIC population comprising 533 doubled haploid lines (Sannemann *et al.*, 2015). We used the above described model 'IBS-M' to derive parent-specific QTL effects for flowering time from 4550 biallelic SNP markers. Peak markers were chosen based on minimum P-values and a window of 20 cM was used to cumulate QTL effects on flowering time. We chose this window size, as here LD fell below the population-specific critical value of 0.021 (Breseghello and Sorrells, 2006; Sannemann *et al.*, 2015). We then compared the estimated parent-specific QTL effects with haplotype-based QTL estimates obtained from modelling parental haplotypes with R/mpMap, as presented in Sannemann *et al.* (2015).

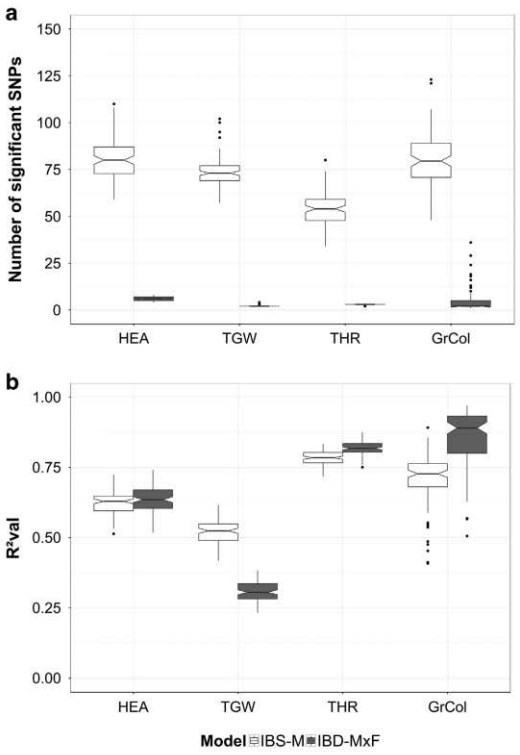

Figure 1 Variation of number of significant SNPs (**a**) and variation of prediction ability (R^2val) (**b**) in the barley NAM population. Box plots indicate the distribution across 100 cross-validation runs. Empty and filled boxes represent models 'IBS-M' and 'IBD-M × F', respectively. Traits are separated by columns.

Simulation studies

We performed simulation studies to further check the suitability of the investigated models and the 'cumulation method'. For this purpose we used our existing real genotype matrices and simulated different QTLs for an artificial trait. We created scenarios that differed in the number of estimated QTLs (1, 3 and 8) and the amount of noise added to the phenotypes to decrease heritability. QTL positions were defined by picking single random SNPs from the IBD genotype matrix throughout the genome. The SNP that was selected for the one QTL scenario was also one of the three SNPs in the three QTL scenario and these three QTLs were part of the eight SNP scenario. The SNP genotypes of the eight simulated QTLs were removed from both genotype matrices and not used in the further analyses. The trait mean was set to 50. Parent-specific allele substitution (α) effects could take defined values (−5, −3, 0, 1 and 2) and were randomly assigned to families (Supplementary Table S2). To add noise to the phenotype data an error term was added that was defined as a normally distributed value ($\mu = 0$, $\sigma = 1$), multiplied by 0 (no noise), 3 (medium noise: noise moderate compared with simulated effect sizes) or 6 (high noise: noise may be bigger than simulated effect sizes). The same training and test sets as described above have been used to scan for significant associations and to estimate prediction abilities. The obtained QTL positions and parent-specific effect estimates were compared with the truly simulated data and the rate of false positives (that is, significant SNPs that did not match the respective QTL interval) and the power to detect QTL precisely (that is, at least one significant marker in a 5 cM interval surrounding the QTL) have been defined for each of the 100 cross-validation runs. In addition, different window sizes (2–40 cM) to cumulate SNP effects in model IBS-M were tested to determine the optimum window for SNP effects to be cumulated. For this purpose, cumulation precision (that is, correlation of cumulated and true parent-specific QTL effects) and the mean difference from the true effect (that is, absolute difference of the cumulated effect and the true parent-specific QTL effect, averaged across parents) have been determined.

RESULTS

QTL detection

In general, a considerably lower number of significant markers was detected by 'IBD-M × F' than by 'IBS-M', irrespective of the trait. For instance, on average 80 and 6 significant markers for HEA were detected by model 'IBS-M' and model 'IBD-M × F', respectively (Figure 1a and Table 1). All significant markers detected by model 'IBD-M × F' were also detected by model 'IBS-M', irrespective of the trait (Figure 2).

Prediction ability

The prediction ability estimates (R^2val) of 'IBS-M' and 'IBD-M × F' were on comparable levels for the traits HEA and THR. Model 'IBD-M × F' showed the highest predction abilites for the assumed mono-/oligogenic traits GrCol and THR. However, for TGW, 'IBS-M' predicted phenotypes better than 'IBD-M × F' (Figure 1b and Table 1). All comparisons of 'IBS-M' and 'IBD-M × F' were significant at $P < 0.001$ after applying one-way analysis of variance, except for HEA (Supplementary Table S3).

Cumulating SNPs to estimate parent-specific QTL effects

As already mentioned above, model 'IBS-M' detected much more significantly associated SNPs than model 'IBD-M × F' (Figure 1a and Table 1). In Figure 2, the contrast in detection rate of QTL regions is visualised. The comparison indicates that model 'IBD-M × F' detects major QTLs as single associations, whereas numerous significant markers from model 'IBS-M' cluster in these major QTL regions. Based on the observed differences, we wondered whether these SNP

clusters from model 'IBS-M' were able to reflect parent-specific effects obtained from model 'IBD-M × F'. Therefore, we cumulated SNP effect estimates (taken from model 'IBS-M') surrounding the 12 clear QTL peaks obtained from model 'IBD-M × F' within a window of 26 cM, representing the mean introgression size in HEB-25 (Supplementary Figure S3).

To estimate the cumulation precision as a measure of appropriateness of the method we correlated the averaged cumulated QTL effects with the average IBD-M × F effect estimate for each QTL. Cumulation precision ranged from 0.26 (HEA, 3H-107.8 cM) to 0.96 (GrCol, 1H-116.8 cM) with a mean of 0.65 (Table 2 and Figure 3). In addition, the mean number of significant SNPs per QTL interval was recorded. The mean number ranged from 2.6 (HEA, 3H-107.8 cM) to 24.1 (GrCol, 1H-116.8 cM) and was positively correlated ($r = 0.69$) with the above-mentioned cumulation precision (Table 2).

Table 1 Comparison of mean R^2 and mean number of significant SNPs across 100 cross-validation runs, calculated for two models and four traits

Trait	*R^2val*		*R^2train*		*Number of significant SNPs*	
	'IBS-M'	*'IBD-M × F'*	*'IBS-M'*	*'IBD-M × F'*	*'IBS-M'*	*'IBD-M × F'*
HEA	0.62	0.63	0.85	0.76	80	6
TGW	0.52	0.31	0.77	0.40	74	2
THR	0.78	0.82	0.89	0.86	54	3
GrCol	0.71	0.85	0.93	0.95	80	5
Mean	0.66	0.65	0.86	0.74	72	4

Abbreviations: GrCol, grain colour; HEA, heading; IBD, identical by descent; IBS, identical by state; R^2train, model fit of the training set; R^2val, prediction ability; SNP, single-nucleotide polymorphism; TGW, thousand grain weight; THR, threshability.
Traits studied: HEA, TGW, THR and GrCo; Models applied: only marker main effects based on IBS ('IBS-M') and only marker-by-family effects based on IBD genotypes ('IBD-M × F'). The last row indicates the mean across traits.

When applying the method to a barley MAGIC population to estimate parent-specific QTL effects for flowering time and comparing them with the haplotype-specific QTL effects presented in Sannemann *et al.* (2015), we observed a mean cumulation precision of 0.60, ranging from 0.26 (QFT.MAGIC.HA-3H.a) to 0.98 (QFT.MAGIC.HA-7H.a, Table 2).

Simulation studies

In a designed simulation study we modelled specific scenarios of possible trait architectures in our NAM population to check the performance of model IBS-M and IBD-M × F and the general suitability of the cumulation method. These scenarios differed for the number of QTLs, the size of QTL effects and the background noise of modelled phenotype values.

Generally, both models detected simulated QTLs with high precision, that is, a significant marker was detected in a 5 cM interval surrounding the true QTL position (Table 3). However, with increasing noise added to the phenotype values, QTL detection was decreased. In particular, model IBD-M × F was not able to detect any QTL if eight QTLs with high background noise were modelled. In contrast, model 'IBS-M' could detect all modelled QTL with high precision even at higher background noise. The aforementioned clustering of significant SNPs surrounding the QTLs in model IBS-M was also observed in the simulations (Supplementary Figure S4). The rate of false positive associations that were not part of the respective QTL interval was low for model IBD-M × F, ranging from 0.00 to 0.31, whereas for model IBS-M, rates from 0.46 to 0.83 were obtained (Table 3). Prediction abilities exceeded 0.5 in all scenarios when no noise was added to the phenotypes and decreased with

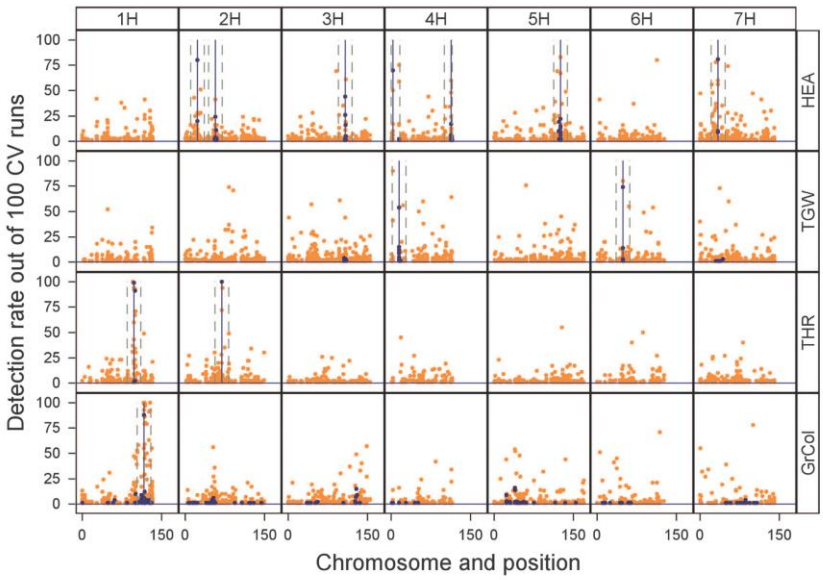

Figure 2 Comparison of detection rates of significant markers across the genome between models 'IBS-M' and 'IBD-M × F' in the barley NAM population. Each of the four rows represents a different trait. The height of the peaks indicates the number of significant effects detected per SNP marker out of 100 cross-validation runs, ordered by genetic position according to the map of Maurer *et al.* (2015). Orange dots represent IBS-M, and blue dots IBD-M × F. Major QTLs detected by IBD-M × F are indicated by vertical lines. Dashed vertical lines indicate the 26 cM window used for cumulation.

Table 2 Cumulation precision (*r*) of major QTLs located in barley NAM and MAGIC populations

Population[a]	*Trait*	*QTL*	*Cumulation precision (r)*[b]	*No. of cumulated SNPs*[c]	*CV*[d]
NAM	HEA	2H-23 cM	0.40	4.8	0.25
	HEA	2H-57 cM	0.60	3.2	0.40
	HEA	3H-107.8 cM	0.26	2.6	0.45
	HEA	4H-3.5 cM	0.68	2.7	0.71
	HEA	4H-113.4 cM	0.69	3.9	1.41
	HEA	5H-125.5 cM	0.90	7.7	0.66
	HEA	7H-34.3 cM	0.70	8.1	0.87
	TGW	4H-14.9 cM	0.36	2.8	0.75
	TGW	6H-49.1 cM	0.58	3.6	0.61
	THR	1H-97.9 cM	0.93	10.8	0.32
	THR	2H-69.3 cM	0.75	5.3	0.43
	GrCol	1H-116.8 cM	0.96	24.1	2.85
	Correlation with cumulation precision[e]			0.69	0.44
MAGIC	FT	QFT.MAGIC.HA-2H.a	0.81	2	6.35
	FT	QFT.MAGIC.HA-3H.a	0.26	1	6.14
	FT	QFT.MAGIC.HA-3H.b	0.51	2	6.51
	FT	QFT.MAGIC.HA-4H.a	0.37	3	3.11
	FT	QFT.MAGIC.HA-5H.a	0.42	2	2.99
	FT	QFT.MAGIC.HA-5H.b	0.89	1	6.82
	FT	QFT.MAGIC.HA-7H.a	0.98	4	6.38
	Correlation with cumulation precision			0.32	0.53

Abbreviations: CV, coefficient of variation; FT, flowering time; GrCol, grain colour; HEA, heading; MAGIC, multiparent advanced generation inter-cross; NAM, nested association mapping; QTL, quantitative trait locus; SNP, single-nucleotide polymorphism; TGW, thousand grain weight; THR, threshability.
[a]NAM: barley population HEB-25 (Maurer *et al*., 2015); MAGIC of barley (Sannemann *et al*., 2015).
[b]Correlation coefficient of QTL effect, obtained from 'IBD-M×F' (NAM) and parental allelic means, obtained from the haplotype approach (MAGIC), respectively, versus QTL effect, estimated by cumulating nearby SNP effects from 'IBS-M'. Means: 0.67 (NAM); 0.60 (MAGIC).
[c]Number of significant SNPs that were cumulated in an interval surrounding the QTL (NAM: 26 cM, MAGIC: 20 cM). For NAM, the mean across 100 cross-validation runs is shown.
[d]CV of all parent-specific QTL effects obtained from model 'IBD-M×F' (NAM) and the haplotype approach (MAGIC), respectively.
[e]Pearson's correlation coefficient (*r*) between cumulation precision for NAM and MAGIC population, respectively.

increasing number of simulated QTLs and noise (Table 3). Prediction abilities of model IBD-M×F were higher than those of IBS-M, except when IBD-M×F failed to detect any QTL.

Applying the cumulation method to estimate parent-specific effects from IBS-M revealed high accordance with the truly estimated effects. Cumulation precision increased with increasing window size of included SNP effects and reached a plateau at ~22 cM (Supplementary Figure S5). At this position a cumulation precision of 0.94 was obtained if one QTL was simulated and no noise was added. In case of eight simulated QTLs with high noise, a cumulation precision of 0.54 was obtained. The mean difference of cumulated effects and the truly simulated effects decreased with increasing window size. At a window size of 26 cM, a mean difference of 0.6 was obtained if one QTL was simulated and no noise was added.

DISCUSSION

QTL detection

In general, both models reliably detected simulated QTLs with high precision (Table 3 and Supplementary Figure S4). Expectedly, with increasing number of simulated QTLs and increasing noise, QTL detection power was decreased, especially when model IBD-M×F was used. The number of significant markers was higher for 'IBS-M' (Table 1, 72 on average in the real data set) than for 'IBD-M×F' (4). One reason to explain the higher number of significant SNPs in 'IBS-M' is clustering of SNPs in major QTL regions. However, if we look at Figure 2 and Figure S4, we see that additional genomic regions of significant SNPs are present in 'IBS-M'. On one side, a substantial part of them might be false positive associations, a fact that has also been pointed out in our simulation studies. According to the results obtained therein, up to 83% of associated SNPs were false positives. This is a known issue if the number of available markers exceeds the number of phenotypes to explain, leading to overfitting of the model. However, this problem in defining true associations can be overcome by cross-validation of the results and counting the number of significances across several runs (Valdar *et al*., 2009; Würschum *et al*., 2012). If we look at the detection rates across 100 runs, we clearly observe the highest peaks at the positions of true QTLs, both in the real data set (Figure 2) and the simulation studies (Supplementary Figure S4). Besides false positive associations, however, some regions in the real data set might correspond to known QTLs as, for example, the known flowering time gene *HvELF3* at 128 cM on 1H (Figure 2). For this locus the weakest effect out of eight major HEA QTLs was observed in Maurer *et al*. (2015). Most likely, 'IBD-M×F' is not able to detect it as smaller subgroups are used for the scan of marker trait associations when only interaction effects are modelled. Similar observations were made by Ogut *et al*. (2015) in the maize NAM population. The authors observed that for small effect QTL, a joint-family model was able to detect them more reliably than a single-family model. Therefore, model 'IBD-M×F' seems to be able to detect predominantly QTLs with strong effects. This makes it suitable to separate useful and valuable major QTLs, explaining a high amount of variance, from minor low-impact QTLs. This finding might be of particular interest for plant breeders.

Prediction abilities

Comparing prediction abilities of the different models enabled us to gain insight into the reliability of estimated QTL effects. We used QTL effects of the training population, which consisted of 80% of randomly chosen lines per family, to predict the phenotypes of the remaining 20% and repeated this procedure 100 times to make it more robust.

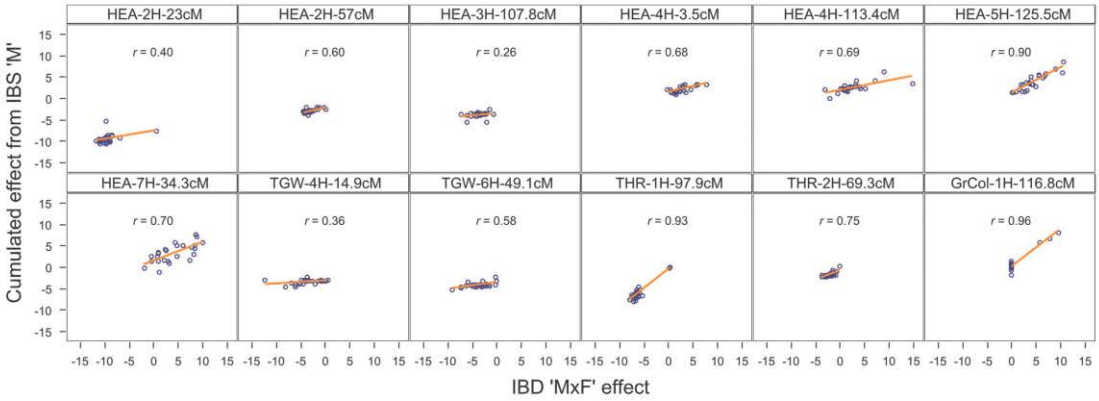

Figure 3 Scatter plots of the cumulated SNP effects from model 'IBS-M' against the M × F effects obtained from model 'IBD-M × F' for 12 major QTLs in the barley NAM population. Each dot represents the estimated QTL effect of one exotic HEB donor. A linear regression line (in orange) and Pearson's correlation coefficients (*r*), indicating the cumulation precision, are given for each QTL. Axis values represent the absolute difference between exotic and cultivated (that is, Barke) QTL alleles.

Table 3 Quality parameters of IBS-M and IBD-M × F in different simulation scenarios

			No noise	*Medium noise*	*High noise*
IBS-M	QTL detection power[a]	1 QTL	1.00	0.48	0.01
		3 QTLs	0.99	0.78	0.37
		8 QTLs	0.91	0.74	0.45
		Mean	0.94	0.72	0.39
	False positives[b]	1 QTL	0.63	0.75	0.83
		3 QTLs	0.54	0.57	0.73
		8 QTLs	0.46	0.48	0.55
		Mean	0.55	0.60	0.70
	Prediction ability[c]	1 QTL	0.65	0.13	0.03
		3 QTLs	0.64	0.26	0.08
		8 QTLs	0.55	0.28	0.08
		Mean	0.61	0.22	0.07
IBD-M × F	QTL detection power	1 QTL	1.00	1.00	0.07
		3 QTLs	1.00	1.00	0.34
		8 QTLs	1.00	0.96	0.00
		Mean	1.00	0.98	0.09
	False positives	1 QTL	0.31	0.00	0.00
		3 QTLs	0.22	0.00	0.00
		8 QTLs	0.07	0.00	n/a
		Mean	0.20	0.00	0.00
	Prediction ability	1 QTL	0.97	0.35	0.00
		3 QTLs	0.96	0.54	0.13
		8 QTLs	0.95	0.68	0.00
		Mean	0.96	0.52	0.04

Abbreviations: IBD-M × F, identical-by-descent marker-by-family effects; IBS-M, identical-by-state marker main effects; QTL, quantitative trait locus.
All values are averaged across 100 simulated cross-validation runs.
[a]QTL detection power is defined as the model's ability to precisely detect the simulated QTL within a 5 cM window surrounding the true position.
[b]False positive associations are all detected significant associations that were outside the interval of cumulated single-nucleotide polymorphisms (SNPs; 26 cM) surrounding a QTL.
[c]Prediction ability represents the correlation coefficient of predicted phenotypes (based on SNP effects obtained in the training set) and observed phenotypes in the test set.

It appears that models 'IBS-M' and 'IBD-M × F' possess a similar power to predict phenotypes of HEA and THR, however there were substantial differences for TGW and GrCol (Table 1). Prediction abilities of polygenic traits like HEA and TGW showed no increase if a parent-specific effect was modelled. Most likely, the reason for this observation is that wild donor-specific QTL effects for HEA are predominantly pointing to the same direction compared with the elite allele of the reference parent Barke (Supplementary Table S4), as a consequence of domestication (Cockram *et al.*, 2011). In particular, it was shown that in the HEB-25 population at the *Ppd-H1* gene, which revealed a major impact on HEA by explaining 36% of genotypic variance, 24 out of 25 wild donors carry alleles with almost identical effects (Maurer *et al.*, 2015). Therefore, modelling M × F effects might not be able to substantially improve the model fit. For TGW, we observed a clearly reduced R^2val for 'IBD-M × F'. This might illustrate that with increasing trait complexity and, thus, decreasing heritability, the modelling of parent-specific marker effects impedes detection of relevant QTLs and diminishes reliable effect estimation. This is confirmed by our simulation study where QTL detection power and prediction ability decreased in scenarios with higher trait complexity, represented by more simulated QTLs and higher amount of noise. In contrast, oligo- or monogenic traits like THR and GrCol benefitted from modelling of parent-specific marker effects. This was most prominent for the trait GrCol, a trait that is only segregating in three families (F-06, F-16, F-24) and no more than 46 lines in total are showing the dark-grained phenotype. Under this circumstance, model 'IBS-M' was able to reach a prediction ability of 0.71, whereas model 'IBD-M × F' reached a prediction ability of 0.85. This outlines the potential to increase prediction ability when parent-specific effects are modelled. Interestingly, although 'IBS-M' only assumes marker main effects and no causative SNP for grain colour is available in our marker set, a remarkable prediction ability of 0.71 was observed. This led to the hypothesis that multiple IBS markers can account for parent-specific effects.

Cumulation method enables realistic modelling of parent-specific effects

To check whether linked SNP markers are suited to reflect parent-specific QTL effects, we cumulated SNP effects surrounding the peak marker of each QTL. Thereby, we focussed on strong QTLs that were detected by 'IBD-M × F' and compared the estimated M × F effects with the parent-specific effects derived by cumulation of 'IBS-M' estimates. We used a window of 26 cM, with the peak marker in its centre, to scan for significant IBS markers. This window size turned

out to be reasonable to capture enough SNPs to maximise the prediction ability of parent-specific QTL effects in simulation studies (Supplementary Figure S5). This window size also reflects the mean introgression size in the HEB-25 population. Thus, markers in this window are often inherited together. The idea behind cumulating their effects is that they are estimated at the same time in the final model and each of the corresponding IBS markers segregates in different families. Therefore, if a marker is not segregating in a particular family (that is, genotype score = 0 = Barke), its effect does not contribute to the cumulated effect (that is, 0 × SNP effect = 0), whereas others do (that is, 2 × SNP effect ≠ 0). Consequently, by combining all markers surrounding a QTL, a specific effect for each parent is estimable, based on the combination of differently segregating significant SNPs. This way, we estimated the parent-specific effects for 12 major QTLs.

Giraud *et al.* (2014) followed a similar approach when comparing QTL effects derived from models assuming ancestral haplotypes with QTL effects gathered from cumulating closely linked single-marker effects. In their case the cumulated effects of two markers already reflected the effect of the respective haplotypes at two different QTLs with high precision.

To estimate how reliable the cumulation of 'IBS-M'-derived SNP effects is able to predict the parent-specific QTL effect, we correlated the cumulated estimates with the M × F effect obtained from model 'IBD-M × F'. We chose this comparison as these M × F effect estimates are very robust and presumably give the best insight into the true parent effect. We observed high positive correlation coefficients for most of the QTLs, ranging from 0.26 (HEA, 3H-107.8 cM) to 0.96 (GrCol, 1H-116.8 cM) with a mean of 0.65 (Figure 3). This shows that cumulating SNP main effects within a QTL region is suitable to estimate a parent-specific effect. Especially for oligo- or monogenic traits (THR, GrCol), we observed extremely high cumulation precision, indicating that the method is of special appropriateness if background noise from other QTLs is low.

The presence of parent-specific QTL effects in HEB-25 was first indicated in Maurer *et al.* (2015), where resequencing of *Ppd-H1* clearly revealed the presence of different haplotypes and consequential parent-specific effects. This resulted in one specific haplotype, originating from HEB family F-24, that showed no difference in HEA as compared with the Barke haplotype. In our study, this could also be observed when looking at the 'IBD-M × F' estimate of this QTL for F-24. However, the method of cumulating SNP effect estimates from model 'IBS-M' failed to detect this fact. We raised the question of why in this obvious case the method seems to fail. One reason could be that in this QTL region there are no F-24-specific SNPs available that could account for the allelic effect. However, in close proximity to the *Ppd-H1* gene (BK_12-BK_16, 23.0 cM), there are three tightly linked SNPs available that solely segregate in F-24 (BOPA1_5880_2547 at 23.2 cM, SCRI_RS_182270 at 24.9 cM and SCRI_RS_115892 at 25.4 cM). When checking their linkage in more detail we recognised that there is recombination between them and Ppd-H1 in seven cases, whereas only in four cases they are inherited together (Additional File S1). Estimating a compensating effect for one of these SNPs that could fine-tune the Ppd-H1 effect will therefore not work. Thus, the GWAS procedure is not able to take any of the F-24-specific SNPs into account to optimise the model. We ran 'IBS-M' again on the whole data set and excluded those seven lines that showed recombination. As expected, BOPA1_5880_2547 now became significant and, consequently, the cumulation method allowed obtaining a more realistic parent-specific effect estimate (Supplementary Table S5).

In addition, for other flowering QTLs, which were described in HEB-25, we could corroborate the presence of parent-specific effects. For instance, the three vernalisation genes *Vrn-H1*, *Vrn-H2* and *Vrn-H3* were supposed to show a parent-specific effect pattern (Maurer *et al.*, 2015, 2016). In this study, we were also able to estimate the parent-specific effects of these QTLs by cumulating SNP effects. We could show that there is plenty of diversification for these vernalisation loci available, depending on the origin of the respective donor. For instance, we observed extremely different parent-specific HEA effects of +8.5 days in F-09 and +1.3 days in family F-19 at *Vrn-H1* locus (Supplementary Table S4).

As we do not know the true QTL effects in our real data set, the correlation of the cumulated effects with the IBD-M × F effects might not really represent an adequate measure of appropriateness of the method. Therefore, we also applied the cumulation method to the simulated data set, where we determined exact parent-specific effects. As a result we obtained a high cumulation precision of 0.94 for the case where one QTL was simulated and no noise was modelled (Supplementary Figure S5). Even for eight simulated QTLs with high noise, a cumulation precision of 0.54 was obtained. At the same time, the mean difference of the cumulated effect and the truly simulated parent-specific effects was low (0.6), indicating the appropriateness of the method. This is in particular remarkable as the simulated parent-specific QTL effects were randomly assigned to donors, that is, closely related donors could have opposing effects and *vice versa*.

Applying the cumulation method to a barley MAGIC population

Besides the general suitability of the cumulation method in a NAM population, we checked whether this also works for MAGIC populations. Therefore, we took raw data on flowering time from an eight-way barley MAGIC population (Sannemann *et al.*, 2015) and applied model 'IBS-M'. Compared with both GWAS approaches presented in Sannemann *et al.* (2015), model 'IBS-M' detected more QTLs, while keeping all QTLs detected before (Supplementary Table S6). Furthermore, total R^2 increased to 74.8%. When using the effect estimates of all significant SNPs to predict the phenotypes of the eight parents, we obtained high accordance ($r = 0.85$, Figure 4 and Supplementary Table S6), indicating the model's general suitability. Then, we applied the cumulation method and compared our estimates with the estimates obtained from the haplotype approach, published in Sannemann *et al.* (2015), that is based on founder haplotype

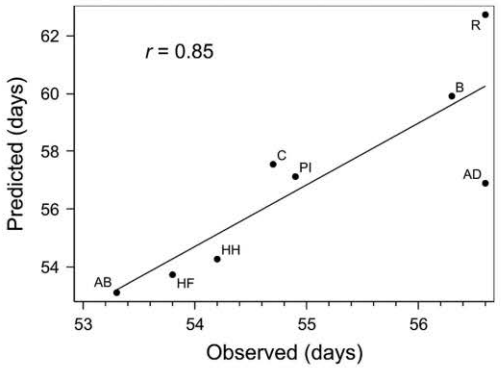

Figure 4 Correlation of observed and predicted flowering time of the eight MAGIC founders. Observed phenotypes are presented in Sannemann *et al.* (2015), and predicted phenotypes are based on the effect estimates of all significant SNPs obtained in model 'IBS-M'. Founder lines are abbreviated: *AB*, Ack. Bavaria; *AD*, Ack. Danubia; *B*, Barke; *C*, Criewener 403; *HF*, Heils Franken; *HH*, Heines Hanna; *PI*, Pflugs Intensiv; *R*, Ragusa.

probabilities calculated with R/mpMap (Huang and George, 2011). On average, we observed a correlation of 0.60 between our estimates and those obtained from the haplotype approach (Table 2). For QFT.MAGIC.HA-7H.a, which explains 28.5% of the variance for heading in the MAGIC population, we reached a correlation of 0.98. By using our method it was also possible to estimate a QTL effect for the MAGIC parent 'Pflugs Intensiv'/'Criewener 403', where the haplotype approach used in Sannemann *et al.* (2015) failed. The given results demonstrate the potential of applying our cumulation method to MAGIC populations to estimate parent-specific QTL effects without the need of haplotype or IBD information.

Prerequisites and characteristics of the cumulation method

The method's success depends on two major prerequisites. First, map positions of all investigated markers must be known. The more accurate the map is, the higher is the chance to differentiate effects reliably. Second, there must be LD present in the population. The fact that SNPs are inherited together because of genetic linkage enables merging these SNP effects to define a parent-specific effect. LD in a multiparental population can be seen as a function of the LD that is present among the parents and the specific population design. In the NAM population HEB-25, there is low LD present among the parents (Maurer *et al.*, 2015) and F1 plants were backcrossed to the recurrent parent Barke. Because of reduced recombination after backcrossing, the introgressed wild barley segments are relatively large, allowing many SNPs to be included in a QTL-surrounding window. However, with an increasing number of differentially segregating SNPs we expect that a smaller window may be sufficient for cumulation of QTL effects.

Using model 'IBS-M' and deriving the parent-specific effect estimate out of it instead of modelling IBD or haplotype effects has several benefits that should be highlighted. (1) More QTLs are detectable compared with a model containing only M × F effects modelled as IBD. This is easily visible in Figure 2 and Supplementary Figure S4, where we see multiple additional SNP peaks for model 'IBS-M' compared with model 'IBD-M × F'. This is most likely because of the fact that in 'IBD-M × F' smaller subgroups are used for the scan of marker trait associations that impede detection of minor QTLs. (2) The grouping of SNP effects is not restricted to 25 families as in models assuming a M × F effect in a NAM population. Because of the different segregation patterns of SNPs, much more information can be gathered by specific combinations of these SNPs. This results in the definition of phenotypic clusters rather than parent-specific haplotypes. Especially, if IBD cannot clearly be derived from pedigree information, as it is for instance the case in MAGIC populations, this method represents an excellent way to model allelic effects originating from different parents independent of IBD information. (3) Haplotypes are not defined *a priori* based on SNP profiles of parental lines like, for example, in clusthaplo (Leroux *et al.*, 2014) or R/mpMap (Huang and George, 2011). Instead, our method represents a more functional approach that is not solely based on ancestral relationships. This enables to track down beneficial genetic variation in a more practical manner.

Besides the above-mentioned beneficial aspects, there are also some limitations that one has to take into account when applying the cumulation method. First, the method is not able to separate effects from tightly linked QTLs, at least not within the selected genetic interval of SNPs being cumulated. Another fact is that the method's success seems to decrease with increasing trait complexity. Strong QTLs with ample allelic variation are still reliably represented by the cumulation method, but one has to be cautious in interpreting parent-specific effects defined for minor QTLs. Furthermore, the effects estimated by the cumulation method are in general less extreme and show lower variation across parents than IBD-based methods do. Another critical point is the high number of false positive associations detected by the model. Probably, they cause the low prediction ability of IBS-M when background noise increased. However, the parent-specific effects obtained via the cumulation method nevertheless turned out to be clearly correlated to the true QTL effects in the simulations. Therefore, we strongly recommend running several cross-validation runs to identify the most reliable QTL positions before cumulation.

To sum up, the comparison of a family-based method (IBD-M × F) with a method assuming general SNP effects (IBS-M) revealed a slight advantage in prediction ability of IBD-M × F, especially for highly heritable traits. However, IBS-M turned out to be superior for traits with lower heritabilities. The idea of cumulating genetically linked SNP effects from model 'IBS-M' provided a novel approach to reconstruct parent-specific QTL effects. This method proved to be applicable to NAM and MAGIC types of multiparental populations even if no IBD information is available. At present, there seems to be the tendency that both haplotype-based linkage models and single-marker association models should be used in a complementary way for QTL detection in multiparental populations (Lorenz *et al.*, 2010; Kump *et al.*, 2011; Tian *et al.*, 2011; Bardol *et al.*, 2013). Our method represents an intermediate path, combining a high QTL detection rate with the possibility to predict parental QTL effects under a reduced computational load. In future, we assume that the cumulation method will benefit from a massive increase in available SNP genotype data that can enhance the precision of this method, for instance by utilising SNP information from exome capture sequencing (Mascher *et al.*, 2013) or increased sizes of SNP chips.

Data archiving

All relevant data are available as supplementary files at Heredity's website or are taken from published articles (Maurer *et al.*, 2015; Sannemann *et al.*, 2015). Additional files containing genotype and phenotype data used as input as well as the obtained GWAS results are available from the Dryad Digital Repository http://dx.doi.org/10.5061/dryad.36rm1.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation (DFG) via priority program 1530: Flowering time control — from natural variation to crop improvement (Grant Pi339/7-1) and via The European Research Area Network for Coordinating Action in Plant Sciences (ERA-CAPS, Grant Pi339/8-1). We are grateful to Roswitha Ende, Jana Müglitz, Diana Rarisch, Helga Sängerlaub, Brigitte Schröder, Bernd Kollmorgen, Markus Hinz and various student assistants for excellent technical assistance and to TraitGenetics GmbH, Gatersleben, Germany, for genotyping HEB-25 with the Infinium iSelect 9k SNP chip. Furthermore, we are grateful to Vera Draba and Stefanie Pencs for providing phenotype data.

Bandillo N, Raghavan C, Muyco PA, Sevilla MA, Lobina IT, Dilla-Ermita CJ *et al.* (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice (NY)* **6**: 11.

Bardol N, Ventelon M, Mangin B, Jasson S, Loywick V, Couton F *et al.* (2013). Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (Zea mays L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. *Theor Appl Genet* **126**: 2717–2736.

Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L *et al.* (2013). Intraspecific variation of recombination rate in maize. *Genome Biol* **14**: R103.
Breseghello F, Sorrells ME (2006). Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. *Genetics* **172**: 1165–1177.
Cockram J, Jones H, O'Sullivan DM (2011). Genetic variation at flowering time loci in wild and cultivated barley. *Plant Genet Resour* **9**: 264–267.
Comadran J, Kilian B, Russell 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. *Nat Genet* **44**: 1388–1392.
Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G *et al.* (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome Biol* **16**: 1–23.
Ding K, Zhou K, Zhang J, Knight J, Zhang X, Shen Y (2005). The effect of haplotype-block definitions on inference of haplotype-block structure and htSNPs selection. *Mol Biol Evol* **22**: 148–159.
Giraud H, Lehermeier C, Bauer E, Falque M, Segura V, Bauland C *et al.* (2014). Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic QTL for hybrid performance in the flint and dent heterotic groups of maize. *Genetics* **198**: 1717–1734.
Huang BE, George AW (2011). R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics* **27**: 727–729.
Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK *et al.* (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J* **10**: 826–839.
Jordan D, Mace E, Cruickshank A, Hunt C, Henzell R (2011). Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Sci* **51**: 1444–1457.
Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA *et al.* (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat Genet* **43**: 163–168.
Leroux D, Rahmani A, Jasson S, Ventelon M, Louis F, Moreau L *et al.* (2014). Clusthaplo: a plug-in for MCQTL to enhance QTL detection using ancestral alleles in multi-cross design. *Theor Appl Genet* **127**: 921–933.
Liu W, Gowda M, Steinhoff J, Maurer HP, Würschum T, Longin CFH *et al.* (2011). Association mapping in an elite maize breeding population. *Theor Appl Genet* **123**: 847–858.
Lorenz AJ, Hamblin MT, Jannink J-L (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One* **5**: e14079.
Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N *et al.* (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3 (Bethesda)* **4**: 1603–1610.
Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D *et al.* (2013). Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *Plant J* **76**: 494–505.
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. *J Exp Bot* **67**: 2507–2518.
Milner SG, Maccaferri M, Huang BE, Mantovani P, Massi A, Frascaroli E *et al.* (2015). A multiparental cross population for mapping QTL for agronomic traits in durum wheat (Triticum turgidum ssp. durum). *Plant Biotechnol J* **14**: 735–748.
Nice LM, Steffenson BJ, Brown-Guedira GL, Akhunov ED, Liu C, Kono TJ *et al.* (2016). Development and genetic characterization of an advanced backcross-nested association mapping (AB-NAM) population of wild x cultivated barley. *Genetics* **203**: 1453–1467.
Ogut F, Bian Y, Bradbury P, Holland J (2015). Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. *Heredity* **114**: 1–12.
Powell JE, Visscher PM, Goddard ME (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet* **11**: 800–805.
Sannemann W, Huang BE, Mathew B, Léon J (2015). Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Mol Breed* **35**: 1–16.
Schwarz G (1978). Estimating the dimension of a model. *Ann Stat* **6**: 461–464.
Thépot S, Restoux G, Goldringer I, Gouache D, Mackay I, Enjalbert J (2015). Efficiently tracking selection in a multiparental population: the case of earliness in wheat. *Genetics* **199**: 609–623.
Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S *et al.* (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* **43**: 159–162.
Valdar W, Holmes CC, Mott R, Flint J (2009). Mapping in structured populations by resample model averaging. *Genetics* **182**: 1263–1277.
Würschum T, Liu W, Gowda M, Maurer H, Fischer S, Schechert A *et al.* (2012). Comparison of biometrical models for joint linkage association mapping. *Heredity* **108**: 332–340.
Yu J, Holland JB, McMullen MD, Buckler ES (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**: 539–551.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

© The Author(s) 2016

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)

General discussion

Suitability of HEB-25 for GWAS

HEB-25 is a novel population that has never been used to conduct GWAS before. Therefore, there was the need to identify methods that are able to reliably detect MTAs and explain the genetic architecture of traits. Previous work already describes the advantages of NAM populations in general – i. e. it combines the power of linkage mapping with the resolution of association mapping (Yu et al. 2008). For this kind of population joint linkage association mapping (JLAM) is an appropriate choice. Würschum et al. (2012) compared several different models for the detection of QTL in a joint set of bi-parental sugar beet breeding populations, a set-up that is comparable to a NAM population. The authors concluded that Model-B, a multiple linear regression model containing cofactors and a population effect to account for genetic background, effectively controlled population structure and possessed a high predictive power. Therefore, this model was initially chosen for application in HEB-25 (Maurer et al. 2015; Maurer et al. 2016a). Although in preliminary work a linear mixed model was applied on a subset of HEB-25 (so-called HEB-5, Schnaithmann et al. (2014)), Model-B was used as it clearly outperformed the former model with regard to QTL detection power and computational load (data not shown). The increased QTL detection power might be attributed to the fact that Model-B contains cofactors to effectively account for genetic background in a multiple regression framework. Cofactors are other markers that explain part of the variance of a trait and may for instance be identified via forward-backward selection in multiple regression models. They are included in the model when testing a marker's association. In classical CIM this basic concept is also implemented and superior to classical interval mapping where no cofactors are included (Zeng 1994). According to Vilhjálmsson and Nordborg (2013) controlling for genetic background is also a major issue in GWAS and Würschum and Kraft (2015) showed that incorporating cofactors in a multi-locus model combined sufficient control of the false-positive rate with high QTL detection and predictive power.

As a proof of concept the trait flowering time has been investigated (Maurer et al. 2015). Flowering time is a highly heritable and well-investigated trait across

species (Jung and Müller 2009; Hill and Li 2016) and a lot could be learned from the model species *Arabidopsis* (Blümel et al. 2015). In barley the main genes involved in the regulation of flowering time have been identified throughout the last decades. Briefly, *Vrn-H1* and *Vrn-H2* are the main determinants of vernalization requirement and *Pph-H1* and *Ppd-H2* control responsiveness to day length. Flowering is promoted through the floral integrator gene *Vrn-H3/HvFT1*. In Maurer et al. (2015) eight major QTL that explained 64% of the cross-validated proportion of explained genotypic variance were reliably identified. Seven out of these QTL corresponded to known genes, namely *HvELF3* (Faure et al. 2012; Zakhrabekova et al. 2012), *Ppd-H1* (Turner et al. 2005), *HvCEN* (Comadran et al. 2012), *HvGA20ox2* (Jia et al. 2015), *Vrn-H2* (Yan et al. 2004), *Vrn-H1* (Yan et al. 2003) and *Vrn-H3/HvFT1* (Yan et al. 2006). Gene-specific SNP markers for *Ppd-H1* (BK_12, BK_13, BK_14, BK_15, BK_16, BOPA2_12_30871, BOPA2_12_30872) and *Vrn-H3/HvFT1* (BOPA2_12_30893, BOPA2_12_30894, BOPA2_12_30895, BK_05) were available on the Infinium iSelect 9K chip (Illumina Inc., San Diego, CA, USA) used for genotyping. Notably, exactly those SNPs were identified as the peak markers of the corresponding QTL regions. Besides the known genes one further QTL on chromosome 4H could be reliably detected, for which no candidate gene has been described yet. Altogether, this outlines the high potential of conducting GWAS in HEB-25 for analyzing the genetic architecture of important agronomic traits.

Evaluation of allelic diversity and usefulness of wild alleles

Not only the question of which QTL regulate a specific trait, but also the question whether there is useful allelic variation available to enhance modern barley cultivars is important. With regard to phenology there is a growing demand for adaptation to environmental threats like drought and heat periods caused by climate change (Rötter et al. 2015). This way the optimal timing of flowering directly affects grain yield as it needs to occur during specific seasons to avoid abiotic and biotic stresses (Hill and Li 2016). Counteracting the trend that almost all modern barley cultivars flower in a relatively narrow space of time, replenishing the elite barley genepool with favorable exotic alleles may be a promising way to meet future food and feed demands.

Within the framework of GWAS also QTL effects were estimated to obtain a measure for the potential value of a wild allele in regard to fine-tuning flowering time of elite material. It turned out that flowering time is severely accelerated by approximately 10 days on average if a wild *Ppd-H1* allele is present. In contrast, at the major vernalization loci *Vrn-H1* and *Vrn-H2* on average a delayed flowering time of 2-4 days could be observed. This is generally attributed to the fact that the reference parent Barke as a spring barley is not responsive to day length and needs no vernalization, whereas wild barley can predominately be seen as a winter barley (Cockram et al. 2011). However, one cannot assume that all 25 wild donors carry the same allele exerting the same effect at a given locus. This was investigated in detail in a case study, where *Ppd-H1* was re-sequenced to determine haplotypes for the donors of HEB-25. Calculating the haplotype effects for this gene revealed strong phenotypic differences between the haplotype of the wild barley accession HID_380 (H-45) and the remaining wild barley haplotypes (Maurer et al. 2015). Interestingly, this accession is classified as *Hordeum vulgare* ssp. *agriocrithon* (Åberg), originating from Tibet, China, which underlines its potentially different domestication history (Dai et al. 2012; Ren et al. 2013). Since re-sequencing of every QTL allele was no feasible option another method to estimate donor-specific QTL allele effects should be defined. This was one aim of Maurer et al. (2016b). With the models introduced therein it was possible to estimate donor-specific QTL effects in a computationally effective manner. This way, for most flowering time QTL allelic differences could be shown. Especially at the vernalization loci *Vrn-H1* and *Vrn-H2*, but also at the *Vrn-H3/HvFT1* locus there is much allelic variation available with some alleles even showing opposing effects.

It is difficult to give general suggestions about the usefulness of wild alleles to fine-tune flowering time in future plant breeding. The optimal timing of flowering varies largely, depending on the environment. While early flowering varieties are favored in Southern Europe and Australia to escape drought (Comadran et al. 2012; Hill and Li 2016), in Northern Europe late flowering cultivars are preferred to make use of the prolonged vegetation period (Cockram et al. 2007). Another fact that further complicates the design of an optimal flowering cultivar is that QTL × environment interaction has to be considered. *Ppd-H1* for instance shows a completely different effect on flowering time when lines carrying the wild allele were grown close to the

equator, where an average wild allele effect of approximately +2 days on flowering time compared to the Barke allele was observed (Saade et al. 2016).

Since the field experiments that built the base for the present thesis had all been conducted at the same location over several years conclusions about a local ideotype can be drawn. Since not only flowering time itself but also other growth stages play a role to increase yield in cereals (Alqudah and Schnurbusch 2014; Semenov et al. 2014) further key parameters of plant development (shooting and maturity) as well as their impact on plant height and thousand grain weight have been investigated in HEB-25 in Maurer et al. (2016a).

It turned out that the eight previously detected major flowering time QTL also have major impacts on shooting and maturity. Most of them also showed effects on plant height and thousand grain weight, suggesting that plant development can have a direct impact on them. However, there were some peculiarities of specific QTL that need to be highlighted. First, not all QTL exerted the same effect on all three developmental stages. Consequently, if the effect on one stage is stronger than on the following stage also the time span between them is affected. Generally, for most QTL the effect on flowering time was more pronounced than on shooting, indicating that the effect somehow accumulates over time. This was true for all major QTL except *denso/sdw1*. Here, the effect on shooting was stronger than on flowering, which in turn caused an extended shoot elongation period. This lead to an increased plant height, which is a known feature of the wild allele at this locus and the reason why short-straw alleles at this locus have been established in breeding to increase yields due to higher lodging resistance (Jia et al. 2009). However, in case of the newly defined major QTL for flowering time, QTL-4H-1 (Maurer et al. 2015), an extended shoot elongation period resulted in reduced plant height. The next outcome of Maurer et al. (2016a) was that flowering time is negatively correlated with maturity; i. e. early flowering causes an extended period of ripening while late flowering results in a shortened ripening period. Since grain filling occurs during this period its duration can directly affect thousand grain weight (Distelfeld et al. 2014). The negative correlation of flowering time and maturity might be explained by an environmental stress that triggers the plant to finish its life cycle earlier than the genetic constitution of the plant would allow. When looking at the average temperature and precipitation patterns across the field trial years (2011-2014) a

period of drought and heat occurred approximately between 80 and 90 days after sowing (Figure 1). It might be that this period caused the plants to prematurely initiate finishing their life cycle due to external factors.

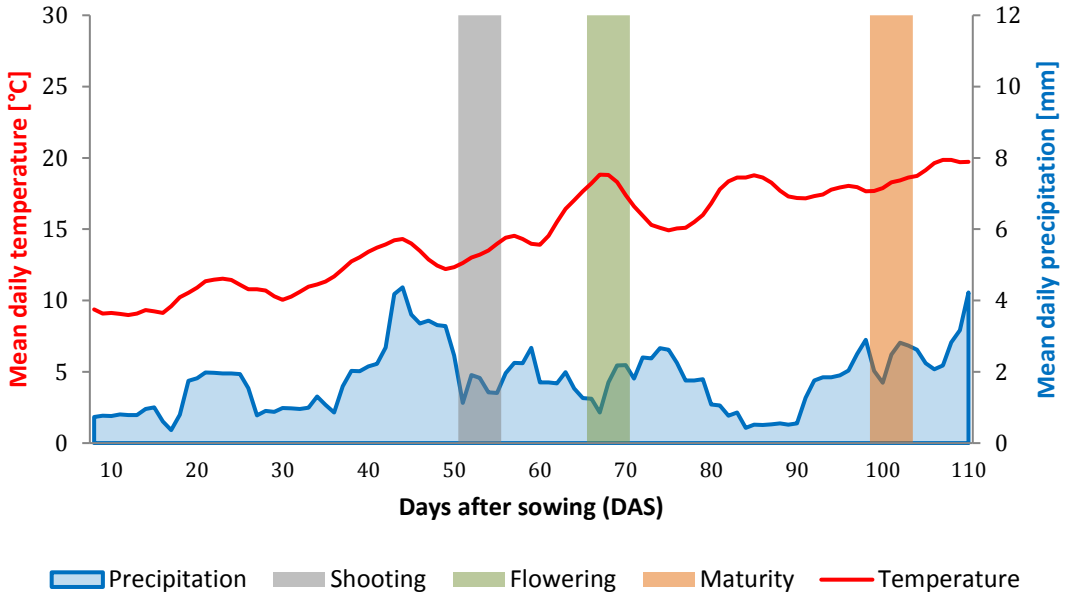

Figure 1) Average temperature and precipitation across 2011-2014 in Halle. Shown are the mean daily temperature (red) and precipitation (blue) averaged across 2011-2014. The values are presented as moving mean of eight days. Colored boxes indicate the population mean of HEB-25 for shooting, flowering and maturity.

Gooding et al. (2003) came to the conclusion that drought and increased temperature before the end of grain filling shortened the grain filling period and, thus, reduced grain yield in wheat. In barley, drought, particularly when combined with high temperature, caused large reductions in grain weight under field conditions (Savin and Nicolas 1996). Preponing flowering in barley might therefore be a suitable strategy to increase yield in drought-prone regions like Halle. Introgressing wild alleles can be a way to achieve this adaptation in future barley breeding. However, one has to take into account that earlier flowering might also influence other yield components like tillering and spikelet formation (Sreenivasulu and Schnurbusch 2012) and thereby alter yield structure of plants. In this context Alqudah and Schnurbusch (2014) revealed that awn primordium to tipping is the most decisive developmental phase for spikelet survival in barley. Since earlier flowering also impacts the duration of this phase one has to be cautious not to

shorten this period severely. Determining the optimal flowering time can therefore be seen as a trade-off between early flowering and a long shoot elongation phase. The wild allele at the *denso/sdw1* locus might therefore be a promising choice to fine-tune plant development, since it combines early flowering with a prolonged shoot elongation phase, which is however accompanied by an increased plant height potentially leading to negative effects like lodging.

Characteristics of HEB-25 and alternative multi-parental populations

As shown in the previous chapters HEB-25 represents a powerful GWAS tool combined with the possibility to retrieve useful exotic alleles. In the following section its properties are discussed and compared to classical mapping concepts as well as other multi-parental population designs.

As already shown before eight major flowering time QTL were identified that jointly account for 64% of the cross-validated proportion of explained genotypic variance in a multiple linear regression model (Model-B). In addition to Model-B the analysis was conducted by applying whole genome prediction models, originally developed for genomic selection in animal breeding (Meuwissen et al. 2001). Instead of including only a subset of markers to control for genetic background genomic prediction incorporates all available markers to detect QTL and estimate their effects. By applying BayesCπ (Habier et al. 2011) the cross-validated proportion of explained genotypic variance could be further increased to 74%, which is comparable to the ~80% that were reported for days to anthesis in the maize NAM population (Peiffer et al. 2014). These values clearly exceed the prediction accuracies reported for many bi-parental populations (Nakaya and Isobe 2012). The NAM population design may cause this high predictive power, however also the increased population size, compared to classical bi-parental populations might explain this feature (VanRaden et al. 2009).

However, what can be stated without doubt is that the number of detectable QTL is massively increased in a multi-parental population, since there is a high number of alleles present in the set of founders. In HEB-25 initially eight major flowering time genes could be defined (Maurer et al. 2015). By applying numerous rounds of cross-validation and counting the number of significances (Valdar et al. 2009) for each

marker 35 additional minor QTL for this trait could be detected (Maurer et al. 2016a), reflecting the complex regulation of flowering time (Blümel et al. 2015). This is an outstanding example of the increased QTL detection power of HEB-25 compared to classical concepts where usually less QTL are detected. QTL detection could be further improved by combining several multi-parental populations. Li et al. (2016c) combined the US-NAM and the CN-NAM populations of maize, resulting in approximately 8,000 lines used for GWAS. They revealed an increased QTL detection rate as compared to analyzing them separately. Meng et al. (2016) report similar findings when combining different MAGIC populations in rice. In this regard it would be exciting to see how combinations of the barley MAGIC population (Sannemann et al. 2015), the AB-NAM population (Nice et al. 2016) and HEB-25 would perform.

A specific characteristic of HEB-25, compared to the original NAM population design of Yu et al. (2008), is the backcross of F_1 plants that was included in the population design. Therefore, each line carries a smaller amount of wild genome (25% vs. 50%), which might enable more accurate effect estimation since background noise is reduced. However, one has to keep in mind that at the same time the number of lines carrying a specific segment is reduced. This is a major constraint especially of AB-NAM, where a specific locus is only represented in 12.5% of lines in each family. Taken together with the low number of lines per family (30) some family-specific QTL that are only present in 3-4 lines might be hard to detect (Nice et al. 2016). BC-based NAM populations facilitate introgression of desirable wild alleles into elite material, since the first steps of backcrossing have already been performed. In addition, the estimated wild allele effect has already been evaluated in the genetic background of an elite parent. However, QTL × genetic background interaction should be considered when introgressing the desired QTL into cultivars other than the reference parent (Bernardo 2014). MAGIC populations overcome this limitation by enabling effect estimation across different genetic backgrounds, since several founders are intercrossed. Due to this specific mating design also a higher number of recombination can be expected, which positively affects mapping resolution (Gardner et al. 2016). Ladejobi et al. (2016) highlight the benefits of MAGIC over NAM with regard to creation of novel recombinants in order to increase diversity for trait mapping. However, identification of the founder

carrying a desirable allele is impeded in MAGIC populations, whereas in a NAM design a wild allele's effect can be directly traced back to the specific donor of a family. In MAGIC populations usually extensive haplotype-modelling approaches have to be performed to identify the most promising founder allele. The cumulation method, which has been developed within the frame of this thesis (Maurer et al. 2016b), represents a universal approach, enabling effect estimation for each founder QTL allele in both NAM and MAGIC populations.

Multi-parental populations are one of the key techniques to accelerate crop improvement in a changing climate (Batley and Edwards 2016). Different NAM and MAGIC population designs exist that show different benefits. They all have in common that they represent improvements over classical bi-parental or association mapping designs. It might be a promising approach to identify useful exotic alleles in multi-parental populations and incorporate them into breeding programmes via classical crossing or genome editing techniques like CRISPR-Cas9 (Doudna and Charpentier 2014), which has recently been successfully applied in barley (Lawrenson et al. 2015).

Summary

Future world crop production is expected to be severely impacted by prospected effects of climate change and human population growth. Therefore modern agriculture is confronted with a number of challenges to secure global demands. Plant breeding is regarded to play a key role in meeting this goal. However, one major constraint plant breeding has to deal with is a reduced adaptability of the current elite breeding germplasm to a changing environment. This resulted from a depletion of allelic diversity during domestication and modern elite breeding. One way to replenish the elite breeding pool with new favorable alleles is to introgress exotic alleles from wild ancestors of cultivated crops. Classically, exotic alleles of interest have been identified through mapping of quantitative trait loci (QTL) or genome-wide association studies (GWAS). However, in recent years the concept of multi-parental mapping populations evolved whereby mapping resolution and QTL detection power could be increased.

In the present thesis the development of the worldwide first wild barley nested association mapping (NAM) population, HEB-25, is reported and utilized to identify favorable exotic alleles via GWAS. HEB-25 is a multi-parental mapping population based on recurrent crosses of the German elite barley cultivar Barke with 25 highly divergent wild barley accessions, originating from the Fertile Crescent and Tibet. In total, the population consists of 1,420 individual BC_1S_3 lines, subdivided into 25 families of up to 75 genotypes. The genetic architecture of flowering time, a potentially relevant trait to escape abiotic stresses, was investigated via GWAS (Chapter 1, Maurer et al. (2015)). It was shown that the statistical models developed within the frame of this work are well-suited to analyze the genetic architectures of agronomically important quantitative traits in HEB-25. Eight major flowering time QTL were defined that could be assigned to seven key flowering time genes, known from *Arabidopsis* and barley, serving as a proof of concept. The remaining QTL represents a novel locus that has not been described before. Furthermore, it could be shown that there is a high degree of allelic diversity available in HEB-25 to fine-tune flowering time with up to 25 different exotic alleles at each QTL.

Since not only flowering time itself but also further developmental key stages are critical in determining yield, the next step was to evaluate also time of shooting and

maturity. To assess their impact on physiological processes, also plant height and thousand grain weight were investigated (Chapter 2, (Maurer et al. 2016a)). It turned out that the previously defined eight major QTL also had major impacts on other developmental stages as well as on plant height and thousand grain weight. However, there were slight differences in the effects single QTL exerted on different stages. For instance, exotic alleles at the *denso/sdw1* locus accelerated shooting more than flowering, resulting in an extended shoot elongation phase and increased plant height. Simultaneously, *denso/sdw1* increased thousand grain weight, presumably as a result of an extended ripening phase. A further outcome of this study was that in general maturity is negatively correlated with flowering time. Early flowering plants therefore have an extended ripening period. It seems that environmental signals prematurely terminated grain filling in Halle. Possibly, a warm and dry period between 80 and 90 days after sowing caused HEB plants to end their life cycle earlier than the genetic constitution would allow. Preponing flowering time might, thus, be an appropriate way to increase yields in a drought-prone environment.

At each QTL, theoretically up to 26 different alleles are present in HEB-25. Thus, the next task was to develop a model estimating donor-specific QTL allele effects (Chapter 3, (Maurer et al. 2016b)). Since each HEB family is comparably small, consisting of on average 60 BC_1S_3 lines, a direct calculation of QTL allele effects in single families did not result in reliable estimates. Therefore, a conceptually different method to overcome this limitation was developed. The so-called cumulation method produced SNP effect estimates across the full HEB-25 population and calculated a donor-specific QTL effect by summing up SNP effect estimates in a defined genetic region surrounding a QTL position. By applying the cumulation method, high QTL detection power and reliable donor-specific effect estimation were obtained. An extraordinary high variation of donor-specific QTL effects was observed for *Vrn-H1* and *Vrn-H3*, two major flowering time QTL.

To conclude, HEB-25 is a multi-parental exotic population, perfectly suitable to investigate agronomically important traits and to replenish the elite barley breeding pool with selected wild barley alleles. Since it exhibits high allelic diversity, HEB-25 may be exploited worldwide to fine-tune plant development and

to improve abiotic stress tolerance as well as yield and yield-stability by introgressing selected favorable QTL alleles.

Zusammenfassung

Die zukünftigen Getreideerträge werden maßgeblich von den prognostizierten Auswirkungen des Klimawandels und dem gleichzeitig stattfindenden globalen Bevölkerungswachstum beeinflusst. Die moderne Landwirtschaft muss eine Vielzahl von Herausforderungen bewältigen, um den steigenden Bedarf zu decken. Die Pflanzenzüchtung nimmt bei der Bewerkstelligung dieser Herausforderungen eine Schlüsselrolle ein. Eine Hürde stellt dabei jedoch die reduzierte Anpassungsfähigkeit moderner Elitesorten an die sich ändernden Umweltbedingungen dar. Dies ist auf den Verlust allelischer Vielfalt im Zuge der Domestikation und Züchtung zurückzuführen. Eine Möglichkeit die genetische Vielfalt in der modernen Pflanzenzüchtung wieder zu erhöhen stellt das Einkreuzen von exotischen Allelen der ursprünglichen Vorfahren kultivierter Arten dar. Klassischerweise wurden vorteilhafte exotische Allele mittels QTL (quantitative trait locus)-Kartierung oder mithilfe von genomweiten Assoziationsstudien (GWAS) identifiziert. In den vergangenen Jahren hat sich dafür jedoch das Konzept multiparentaler Populationen etabliert, wodurch die Fähigkeit QTL mit hoher Präzision zu bestimmen erhöht werden konnte.

In der vorliegenden Arbeit werden die Entwicklung der weltweit ersten nested association mapping (NAM)-Population in Wildgerste, HEB-25 genannt, sowie deren Nutzung zur Identifikation von vorteilhaften exotischen Allelen mittels GWAS, dargestellt. HEB-25 ist eine multiparentale Kartierungspopulation, die auf rekurrenten Kreuzungen der deutschen Sommergerstensorte Barke mit 25 hochdiversen Wildgerstenakzessionen aus dem Fruchtbaren Halbmond und Tibet beruht. Insgesamt besteht die Population aus 1420 unterschiedlichen BC_1S_3-Linien, die in 25 Familien mit bis zu 75 Genotypen untergliedert sind. Mittels GWAS wurden in dieser Population die genetischen Grundlagen, die den Zeitpunkt der Blüte bestimmen, untersucht (Kapitel 1, Maurer et al. (2015)). Der Blühzeitpunkt ist unter Umständen entscheidend, um abiotischen Stress zu vermeiden. Es konnte gezeigt werden, dass die statistischen Modelle, die im Laufe der genannten Arbeit entwickelt wurden, bestens dazu geeignet sind die genetischen Grundlagen agronomisch relevanter quantitativer Merkmale in HEB-25 zu erforschen. Acht Haupt-QTL für den Blühzeitpunkt konnten dabei identifiziert werden, die sieben

bereits bekannten Blühgenen aus *Arabidopsis* und Gerste entsprachen. Dies kann als Konzeptnachweis für das Funktionieren der Methode angesehen werden. Der verbleibende QTL stellte sich als neu heraus, da er bis dato noch nicht beschrieben wurde. Darüber hinaus konnte gezeigt werden, dass ein hoher Grad an allelischer Diversität in HEB-25 vorhanden ist, um den Blühzeitpunkt mit bis zu 25 verschiedenen Wildallelen pro QTL feinzujustieren.

Da nicht nur der Blühzeitpunkt an sich, sondern auch weitere Kardinalpunkte der Pflanzenentwicklung für die Ertragsbildung entscheidend sind, wurden im nächsten Schritt zusätzlich der Zeitpunkt des Schossens sowie der Reifezeitpunkt untersucht. Um deren Einfluss auf pflanzenphysiologische Prozesse zu beschreiben, wurden außerdem die Pflanzenhöhe und das Tausendkorngewicht analysiert (Kapitel 2, Maurer et al. (2016a)). Es zeigte sich, dass die vorher definierten acht Haupt-QTL für den Blühzeitpunkt auch die anderen Entwicklungsstadien sowie Pflanzenhöhe und Tausendkorngewicht maßgeblich beeinflussten. Dabei zeigten sich jedoch leichte Unterschiede wie stark die verschiedenen Entwicklungsstadien von einem bestimmten QTL beeinflusst wurden. Zum Beispiel beschleunigten exotische Allele am *denso/sdw1*-Locus das Schossen stärker als die Blüte, wodurch die Sprossstreckungsphase verlängert und die Pflanzenhöhe erhöht wurde. Gleichzeitig verursachte *denso/sdw1* eine Erhöhung des Tausendkorngewichtes, was vermutlich aus einer verlängerten Reifephase resultierte. Ein weiteres Ergebnis dieser Arbeit war, dass der Reifezeitpunkt generell negativ mit dem Blühzeitpunkt korreliert war. Früh blühende Pflanzen hatten demnach eine verlängerte Abreifeperiode. Es scheint, dass bestimmte Umweltfaktoren ein frühzeitiges Ende der Kornfüllungsphase in Halle herbeiführten. Möglicherweise war eine Periode von warmen Temperaturen in Kombination mit anhaltender Trockenheit 80 bis 90 Tage nach der Aussaat dafür ausschlaggebend, dass Pflanzen eher abreiften als es ihre genetischen Voraussetzungen erlaubt hätten. Eine Verfrühung der Blüte könnte demnach ein probates Mittel sein den Ertrag in trockenheitsanfälligen Gebieten zu sichern.

Da an jedem QTL theoretisch bis zu 26 unterschiedliche Allele in HEB-25 unterschieden werden können, bestand die nächste Aufgabe darin ein Modell zu entwickeln, um donorspezifische QTL-Allel-Effekte zu bestimmen (Kapitel 3, Maurer et al. (2016b)). Aufgrund der relativ kleinen Familiengröße von

durchschnittlich 60 BC_1S_3-Linien führte eine direkte Schätzung der Effekte in einzelnen Familien nicht zu verlässlichen Ergebnissen. Deshalb wurde in Form der sogenannten Kumulationsmethode ein anderes Konzept erarbeitet. Dabei werden SNP-Effekte über die gesamte HEB-25-Population geschätzt und anschließend die donorspezifischen Effekte errechnet, indem die SNP-Effekte einer definierten, den QTL umgebenden genomischen Region, aufsummiert werden. Durch dieses Vorgehen konnte eine hohe QTL-Detektionsrate bei gleichzeitiger zuverlässiger Schätzung donorspezifischer Effekte realisiert werden. Es zeigte sich, dass eine außerordentlich hohe Variation hinsichtlich donorspezifischer Effekte besonders bei *Vrn-H1* und *Vrn-H3*, zweier Haupt-QTL für den Blühzeitpunkt, existiert.

Alles in allem stellt HEB-25 eine multiparentale Population dar, die vorzüglich dazu geeignet ist agronomisch relevante Merkmale zu erforschen und die moderne Gerstenzüchtung mit ausgewählten exotischen Allelen zu bereichern. Aufgrund der reichhaltigen Alleldiversität könnte HEB-25 weltweit genutzt werden, um Pflanzenentwicklung, abiotische Stresstoleranz, sowie Ertrag und Ertragsstabilität durch die Introgression ausgewählter exotischer Allele zu optimieren.

References

Alonso-Peral MM, Oliver SN, Casao MC, Greenup AA, Trevaskis B (2011) The promoter of the cereal VERNALIZATION1 gene is sufficient for transcriptional induction by prolonged cold. PLoS One 6:e29456

Alqudah AM, Schnurbusch T (2014) Awn primordium to tipping is the most decisive developmental phase for spikelet survival in barley. Funct Plant Biol 41:424-436

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

Baik B-K, Newman CW, Newman RK (2011) Food Uses of Barley. In: Ullrich SE (ed) Barley: Production, Improvement, and Uses. Wiley-Blackwell, Chichester, pp 532-562

Bajgain P, Rouse MN, Tsilo TJ, Macharia GK, Bhavani S, Jin Y, Anderson JA (2016) Nested Association Mapping of Stem Rust Resistance in Wheat Using Genotyping by Sequencing. PLoS One 11:e0155760

Bandillo N, Raghavan C, Muyco PA, Sevilla MA, Lobina IT, Dilla-Ermita CJ, Tung CW, McCouch S, Thomson M, Mauleon R, Singh RK, Gregorio G, Redona E, Leung H (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6:11

Batley J, Edwards D (2016) The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. Curr Opin Plant Biol 30:78-81

Bernardo R (2010) Breeding for quantitative traits in plants, second edition. Stemma Press, Woodbury, Minnesota

Bernardo R (2014) Essentials of plant breeding. Stemma Press, Woodbury, Minnesota

Blümel M, Dally N, Jung C (2015) Flowering time regulation in crops—what did we learn from Arabidopsis? Curr Opin Biotech 32:121-129

Boden SA, Weiss D, Ross JJ, Davies NW, Trevaskis B, Chandler PM, Swain SM (2014) EARLY FLOWERING3 Regulates Flowering in Spring Barley by Mediating Gibberellin Production and FLOWERING LOCUS T Expression. Plant Cell 26:1557-1569

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331

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

Calixto CPG, Waugh R, Brown JWS (2015) Evolutionary Relationships Among Barley and Arabidopsis Core Circadian Clock and Clock-Associated Genes. J Mol Evol 80:108-119

Campoli C, Pankin A, Drosse B, Casao CM, Davis SJ, von Korff M (2013) HvLUX1 is a candidate gene underlying the early maturity 10 locus in barley: phylogeny, diversity, and interactions with the circadian clock and photoperiodic pathways. New Phytol 199:1045-1059

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

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

Chailakhyan MK (1937) Hormonal Theory of Plant Development. Akad. Nauk USSR, Moscow

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

Cockram J, Jones H, O'Sullivan DM (2011) Genetic variation at flowering time loci in wild and cultivated barley. Plant Genet Resour 9:264-267

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

Dai F, Nevo E, Wu D, Comadran J, Zhou M, Qiu L, Chen Z, Beiles A, Chen G, Zhang G (2012) Tibet is one of the centers of domestication of cultivated barley. Proc Natl Acad Sci USA 109:16969-16973

Dawson IK, Russell J, Powell W, Steffenson B, Thomas WTB, Waugh R (2015) Barley: a translational model for adaptation to climate change. New Phytol 206:913-931

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol 16:1-23

Distelfeld A, Avni R, Fischer AM (2014) Senescence, nutrient remobilization, and yield in wheat and barley. J Exp Bot 65:3783-3798

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096

Dunnett CW (1955) A multiple comparison procedure for comparing several treatments with a control. J Am Stat Assoc 50:1096-1121

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

Eshed Y, Abu-Abied M, Saranga Y, Zamir D (1992) Lycopersicon esculentum lines containing small overlapping introgressions from L. pennellii. Theor Appl Genet 83:1027-1034

FAO (2016) FAOSTAT. http://www.fao.org/faostat. Accessed 19 December 2016

Faure S, Turner AS, Gruszka D, Christodoulou V, Davis SJ, von Korff M, Laurie DA (2012) Mutation at the circadian clock gene EARLY MATURITY 8 adapts domesticated barley (Hordeum vulgare) to short growing seasons. Proc Natl Acad Sci USA 109:8328-8333

Flint-Garcia SA, Thornsberry JM, IV ESB (2003) Structure of Linkage Disequilibrium in Plants. Annual Review of Plant Biology 54:357-374

Gardner KA, Wittern LM, Mackay IJ (2016) A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. Plant Biotechnol J 14:1406-1417

Gooding MJ, Ellis RH, Shewry PR, Schofield JD (2003) Effects of Restricted Water Availability and Increased Temperature on the Grain Filling, Drying and Quality of Winter Wheat. J Cereal Sci 37:295-309

Guo B, Wang D, Guo Z, Beavis WD (2013) Family-based association mapping in crop species. Theor Appl Genet 126:1419-1430

Guo BH, Sleper DA, Beavis WD (2010) Nested Association Mapping for Identification of Functional Markers. Genetics 186:373-383

Gupta P, Roy J, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80:524-535

Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. PLoS Biol 2:1610-1615

Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12:186

Hill CB, Li C (2016) Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. Front Plant Sci 7:1906

Hoffmann A, Maurer A, Pillen K (2012) Detection of nitrogen deficiency QTL in juvenile wild barley introgression lines growing in a hydroponic system. BMC Genet 13:88

Honsdorf N, March TJ, Berger B, Tester M, Pillen K (2014a) High-Throughput Phenotyping to Detect Drought Tolerance QTL in Wild Barley Introgression Lines. PLoS One 9:13

Honsdorf N, March TJ, Hecht A, Eglinton J, Pillen K (2014b) Evaluation of juvenile drought stress tolerance and genotyping by sequencing with wild barley introgression lines. Mol Breed 34:1475-1495

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol J 10:826-839

Jia Q, Li C, Shang Y, Zhu J, Hua W, Wang J, Yang J, Zhang G (2015) Molecular characterization and functional analysis of barley semi-dwarf mutant Riso no. 9265. BMC Genomics 16:927

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

Jordan D, Mace E, Cruickshank A, Hunt C, Henzell R (2011) Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. Crop Sci 51:1444-1457

Jung C, Müller AE (2009) Flowering time control and applications in plant breeding. Trends Plant Sci 14:563-573

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

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

Ladejobi O, Elderfield J, Gardner KA, Gaynor RC, Hickey J, Hibberd JM, Mackay IJ, Bentley AR (2016) Maximizing the potential of multi-parental crop populations. Appl Transl Genomics 11:9-17

Landegren U, Nilsson M, Kwok P-Y (1998) Reading Bits of Genetic Information: Methods for Single-Nucleotide Polymorphism Analysis. Genome Res 8:769-776

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199

Lawrenson T, Shorinola O, Stacey N, Li C, Østergaard L, Patron N, Uauy C, Harwood W (2015) Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. Genome Biol 16:258

Li C, Li Y, Sun B, Peng B, Liu C, Liu Z, Yang Z, Li Q, Tan W, Zhang Y, Wang D, Shi Y, Song Y, Wang T, Li Y (2013) Quantitative trait loci mapping for yield components and kernel-related traits in multiple connected RIL populations in maize. Euphytica 193:303-316

Li D, Li Z, Hu J, Lin Z, Li X (2016a) Polymorphism analysis of multi-parent advanced generation inter-cross (MAGIC) populations of upland cotton developed in China. Genet Mol Res 15:gmr15048759

Li H, Singh S, Bhavani S, Singh RP, Sehgal D, Basnet BR, Vikram P, Burgueno-Ferreira J, Huerta-Espino J (2016b) Identification of Genomic Associations for Adult Plant Resistance in the Background of Popular South Asian Wheat Cultivar, PBW343. Front Plant Sci 7:1674

Li Y-x, Li C, Bradbury PJ, Liu X, Lu F, Romay CM, Glaubitz JC, Wu X, Peng B, Shi Y, Song Y, Zhang D, Buckler ES, Zhang Z, Li Y, Wang T (2016c) Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. Plant J 86:391-402

Liller CB, Walla A, Boer MP, Hedley P, Macaulay M, Effgen S, von Korff M, van Esse GW, Koornneef M (2017) Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. Theor Appl Genet 130:269-281

Lin SY, Sasaki T, Yano M (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, Oryza sativa L., using backcross inbred lines. Theor Appl Genet 96:997-1003

Lundqvist U, Franckowiak JD (2003) Diversity of barley mutants. In: von Bothmer R, van Hintum T, Knüpffer H, Sato K (eds) Diversity in Barley (Hordeum vulgare). Elsevier Science B.V., Amsterdam, pp 77-96

Ma X, Li C, Wang A, Duan R, Jiao G, Nevo E, Chen G (2012) Genetic diversity of wild barley (Hordeum vulgare ssp. spontaneum) and its utilization for barley improvement. Sciences in Cold and Arid Regions 4:453-461

Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, Greenland AJ, Horsnell R, Howells R, O'Sullivan DM (2014) An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3 (Bethesda) 4:1603-1610

Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics 16:290

Maurer A, Draba V, Pillen K (2016a) Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. J Exp Bot 67:2507-2518

Maurer A, Sannemann W, Leon J, Pillen K (2016b) Estimating parent-specific QTL effects through cumulating linked identity-by-state SNP effects in multiparental populations. Heredity, doi: 10.1038/hdy.2016.121

McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G (2013) Agriculture: Feeding the future. Nature 499:23-24

Meng L, Zhao X, Ponce K, Ye G, Leung H (2016) QTL mapping for agronomic traits using multi-parent advanced generation inter-cross (MAGIC) populations derived from diverse elite indica rice lines. Field Crop Res 189:19-42

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819-1829

Milner SG, Maccaferri M, Huang BE, Mantovani P, Massi A, Frascaroli E, Tuberosa R, Salvi S (2015) A multiparental cross population for mapping QTL for agronomic traits in durum wheat (Triticum turgidum ssp. durum). Plant Biotechnol J 14:735-748

Molina-Cano JL, Moralejo M, Igartua E, Romagosa I (1999) Further evidence supporting Morocco as a centre of origin of barley. Theor Appl Genet 98:913-918

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

Mrízová K, Jiskrová E, Vyroubalová Š, Novák O, Ohnoutková L, Pospíšilová H, Frébort I, Harwood WA, Galuszka P (2013) Overexpression of cytokinin dehydrogenase genes in barley (Hordeum vulgare cv. Golden Promise) fundamentally affects morphology and fertility. PLoS One 8:e79029

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59:651-681

Nakaya A, Isobe SN (2012) Will genomic selection be a practical method for plant breeding? Ann Bot 110:1303-1316

Naz AA, Arifuzzaman M, Muzammil S, Pillen K, Leon J (2014) Wild barley introgression lines revealed novel QTL alleles for root and related shoot traits in the cultivated barley (Hordeum vulgare L.). BMC Genet 15:12

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

Nevo E, Fu Y-B, Pavlicek T, Khalifa S, Tavasi M, Beiles A (2012) Evolution of wild cereals during 28 years of global warming in Israel. Proc Natl Acad Sci USA 109:3412-3415

Nice LM, Steffenson BJ, Brown-Guedira GL, Akhunov ED, Liu C, Kono TJ, Morrell PL, Blake TK, Horsley RD, Smith KP, Muehlbauer GJ (2016) Development and Genetic Characterization of an Advanced Backcross-Nested Association Mapping (AB-NAM) Population of Wild x Cultivated Barley. Genetics 203:1453-1467

Nishida H, Ishihara D, Ishii M, Kaneko T, Kawahigashi H, Akashi Y, Saisho D, Tanaka K, Handa H, Takeda K (2013) Phytochrome C Is A Key Factor Controlling Long-Day Flowering in Barley. Plant Physiol 163:804-814

Oliver SN, Deng W, Casao MC, Trevaskis B (2013) Low temperatures induce rapid changes in chromatin state and transcript levels of the cereal VERNALIZATION1 gene. J Exp Bot 64:2413-2422

Pankin A, Altmüller J, Becker C, von Korff M (2016) Targeted re-sequencing reveals the genetic signatures and the reticulate history of barley domestication. bioRxiv, doi: 10.1101/070078

Pankin A, Campoli C, Dong X, Kilian B, Sharma R, Himmelbach A, Saini R, Davis SJ, Stein N, Schneeberger K (2014) Mapping-by-sequencing identifies HvPHYTOCHROME C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. Genetics 198:383-396

Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, Bouchet JP, Le QH, Chauchard B, Verschave P, Causse M (2015) Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnol J 13:565-577

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CA, McMullen MD, Holland JB, Bradbury PJ (2014) The genetic architecture of maize height. Genetics 196:1337-1356

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

Pillen K, Zacharias A, Leon J (2004) Comparative AB-QTL analysis in barley using a single exotic donor of Hordeum vulgare ssp spontaneum. Theor Appl Genet 108:1591-1601

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

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

Ren X, Nevo E, Sun D, Sun G (2013) Tibet as a Potential Domestication Center of Cultivated Barley of China. PLoS One 8:e62700

Reuscher S, Kolter A, Hoffmann A, Pillen K, Kramer U (2016) Quantitative Trait Loci and Inter-Organ Partitioning for Essential Metal and Toxic Analogue Accumulation in Barley. PLoS One 11:22

Rötter RP, Tao F, Höhn JG, Palosuo T (2015) Use of crop simulation modelling to aid ideotype design of future cereal cultivars. J Exp Bot 66:3463-3476

Saade S, Maurer A, Shahid M, Oakey H, Schmöckel SM, Negrão S, Pillen K, Tester M (2016) Yield-related salinity tolerance traits identified in a nested association mapping (NAM) population of wild barley. Sci Rep 6:32586

Saal B, von Korff M, Leon J, Pillen K (2011) Advanced-backcross QTL analysis in spring barley: IV. Localization of QTL x nitrogen interaction effects for yield-related traits. Euphytica 177:223-239

Sakuma S, Salomon B, Komatsuda T (2011) The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. Plant Cell Physiol 52:738-749

Sallam A, Martsch R (2015) Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (Vicia faba L.). Genetica 143:501-514

Sannemann W, Huang BE, Mathew B, Léon J (2015) Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Mol Breed 35:1-16

Savin R, Nicolas M (1996) Effects of Short Periods of Drought and High Temperature on Grain Growth and Starch Accumulation of Two Malting Barley Cultivars. Funct Plant Biol 23:201-210

Schmalenbach I, Korber N, Pillen K (2008) Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. Theor Appl Genet 117:1093-1106

Schmalenbach I, Leon J, Pillen K (2009) Identification and verification of QTLs for agronomic traits using wild barley introgression lines. Theor Appl Genet 118:483-497

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

Schmalenbach I, Pillen K (2009) Detection and verification of malting quality QTLs using wild barley introgression lines. Theor Appl Genet 118:1411-1427

Schnaithmann F, Kopahnke D, Pillen K (2014) A first step toward the development of a barley NAM population and its utilization to detect QTLs conferring leaf rust seedling resistance. Theor Appl Genet 127:1513-1525

Schnaithmann F, Pillen K (2013) Detection of exotic QTLs controlling nitrogen stress tolerance among wild barley introgression lines. Euphytica 189:67-88

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44:825-830

Semenov MA, Stratonovitch P, Alghabari F, Gooding MJ (2014) Adapting wheat in Europe for climate change. J Cereal Sci 59:245-256

Soto-Cerda BJ, Cloutier S (2012) Association mapping in plant genomes. In: Caliskan M (ed) Genetic Diversity in Plants. InTech, pp 29-54

Sreenivasulu N, Schnurbusch T (2012) A genetic playground for enhancing grain number in cereals. Trends Plant Sci 17:91-101

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

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191-203

Thépot S, Restoux G, Goldringer I, Gouache D, Mackay I, Enjalbert J (2015) Efficiently tracking selection in a multiparental population: the case of earliness in wheat. Genetics 199:609-623

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159-162

Tuinstra MR, Ejeta G, Goldsbrough PB (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95:1005-1011

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

Valdar W, Holmes CC, Mott R, Flint J (2009) Mapping in Structured Populations by Resample Model Averaging. Genetics 182:1263-1277

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited Review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16-24

Vilhjálmsson BJ, Nordborg M (2013) The nature of confounding in genome-wide association studies. Nat Rev Genet 14:1-2

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

von Korff M, Wang H, Leon J, Pillen K (2006) AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (H-vulgare ssp spontaneum). Theor Appl Genet 112:1221-1231

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

von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, Casas A, Chen TH, Hayes PM, Skinner JS (2005) Molecular and structural characterization of barley vernalization genes. Plant Mol Biol 59:449-467

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407-4414

Wang G, Schmalenbach I, von Korff M, Leon J, Kilian B, Rode J, Pillen K (2010) Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC(2)DH population and a set of wild barley introgression lines. Theor Appl Genet 120:1559-1574

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388

Würschum T, Kraft T (2015) Evaluation of multi-locus models for genome-wide association studies: a case study in sugar beet. Heredity 114:281-290

Würschum T, Liu W, Gowda M, Maurer H, Fischer S, Schechert A, Reif J (2012) Comparison of biometrical models for joint linkage association mapping. Heredity 108:332-340

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proc Natl Acad Sci USA 103:19581-19586

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303:1640-1644

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci USA 100:6263-6268

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539-551

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

Zakhrabekova S, Gough SP, Braumann I, Muller AH, Lundqvist J, Ahmann K, Dockter C, Matyszczak I, Kurowska M, Druka A, Waugh R, Graner A, Stein N, Steuernagel B, Lundqvist U, Hansson M (2012) Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. Proc Natl Acad Sci USA 109:4326-4331

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983-989

Zamir D (2008) Plant breeders go back to nature. Nat Genet 40:269-270

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468

Zhang H, Mittal N, Leamy LJ, Barazani O, Song B-H (2017) Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. Evol Appl 10:5-24

Zohary D, Hopf M, Weiss E (2012) Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin, 4 edn. Oxford University Press, Oxford

Acknowledgements

First and foremost I am grateful to Klaus Pillen, who supervised my PhD project and gave me plenty of rope to realize my own ideas. I also want to thank him for the inspiring discussions and the critical review of my manuscripts.

Furthermore I would like to thank Jochen Reif and Yong Jiang, as well as Tobias Würschum, who encouraged me to apply multiple linear regression in the HEB-25 population and supported me in implementing the models.

I am also thankful to Jens Léon and Wiebke Sannemann, who agreed to share their data of the MAGIC barley population, which definitely valorized the third publication.

Many thanks also to all my colleagues in the plant breeding group, who made work fruitful and enjoyable.

Last but not least a big thank you to my family and friends for supporting me all the time during my PhD.

Eidesstattliche Erklärung / *Declaration under Oath*

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

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

Halle, 06.02.2017 _____

Unterschrift / *Signature*

Curriculum Vitae

Andreas Maurer

6 February 2017

Personal Details

Date of birth	04.04.1986
Nationality	German
Address	Walter-Hülse-Str. 10 06120 Halle (Saale) Germany
Telephone	+49 345 1714653
Mobile phone	+49 178 8893682
E-mail	maurerandi@gmx.de

Education

Since 11/2016	Martin Luther University, Halle Research associate at the chair of plant breeding (PhD student)
11/2014 – 10/2016	Martin Luther University, Halle Research associate at the chair of plant breeding (PhD student) ERA-CAPS project *“BARLEY-NAM – Locating exotic genes that control agronomic traits under stress in a wild barley nested association mapping (NAM) population”*
11/2011 – 10/2014	Martin Luther University, Halle Research associate at the chair of plant breeding (PhD student) DFG priority program SPP1530: „Flowering Time Control – from Natural Variation to Crop Improvement“; PP-9: *„Allele mining in wild barley: finding new exotic genes which control flowering time in the barley nested association mapping (NAM) population HEB-25”*
10/2009 – 10/2011	Martin Luther University, Halle M.Sc. Crop Plant Sciences Thesis title: *Detection of nitrogen deficiency QTL in juvenile wild barley introgression lines growing in a hydroponic system* Degree: 1.3

10/2006 – 10/2009	Martin Luther University, Halle B.Sc. Agricultural Sciences Thesis title: *Comparison of anther and microspore technique to produce doubled haploids in spring barley* Degree: 1.6
1996 – 2005	Helene-Lange-Gymnasium, Fürth (Abitur 2.0)

Publications

Maurer, Andreas, Wiebke Sannemann, Jens Léon, and Klaus Pillen (2016) "Estimating parent-specific QTL effects through cumulating linked identity-by-state SNP effects in multiparental populations." Heredity, doi: 10.1038/hdy.2016.121

Saade, Stephanie, **Andreas Maurer**, *et al.* (2016) "Yield-related salinity tolerance traits identified in a nested association mapping (NAM) population of wild barley." Scientific Reports 6:32586

Maurer, Andreas, Vera Draba, and Klaus Pillen (2016) "Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping." Journal of Experimental Botany 67:2507-2518

Maurer, Andreas, *et al.* (2015) "Modelling the genetic architecture of flowering time control in barley through nested association mapping." BMC Genomics 16:290

Hoffmann, Astrid, **Andreas Maurer**, and Klaus Pillen (2012) "Detection of nitrogen deficiency QTL in juvenile wild barley introgression lines growing in a hydroponic system." BMC Genetics 13:88

Conference contributions

06/2016, 12th International Barley Genetics Symposium in Minneapolis, USA. Poster: Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping.

03/2016, Main Conference of the Genome Research Working Group of the German Plant Breeding Association (GPZ) in Bonn, Germany. Poster: Identity-by-state versus identity-by-descent GWAS performance in the barley NAM population HEB-25.

09/2015, 18th Conference of the Genome Research Working Group of the German Plant Breeding Association (GPZ) in Düsseldorf, Germany. Talk: Genomic dissection of plant development and its impact on yield components in barley through nested association mapping.

09/2014, Main Conference of the Genome Research Working Group of the German Plant Breeding Association (GPZ) in Kiel, Germany. Poster: A step towards a causal model of flowering time in barley.

06/2014, Joint EUCARPIA Cereal Section & I•T•M•I Conference in Wernigerode, Germany. Talk: Allele mining in wild barley – finding new exotic genes which control flowering time in the barley nested association mapping (NAM) population HEB-25.

03/2014, International Symposium: Genetic Variation of Flowering Time Genes and Applications for Crop Improvement in Bielefeld, Germany. Talk: Allele mining in wild barley – finding new exotic genes which control flowering time in the barley nested association mapping (NAM) population HEB-25.

02/2014, 17th Conference of the Genome Research Working Group of the German Plant Breeding Association (GPZ) in Cologne, Germany. Poster: Mapping new flowering time alleles in the barley nested association mapping (NAM) population HEB-25.

02/2012, Main Conference of the Genome Research Working Group of the German Plant Breeding Association (GPZ) in Gießen, Germany. Poster: The genetic architecture of the barley nested association mapping (NAM) population HEB-25.

Teaching experience

2016–2017	Master course; Lecture and exercise on molecular markers in plant breeding
2015–2016	Master course; Exercise on quantitative genetics in plant breeding